

Como mejorar la visibilidad y el posicionamiento en los motores de búsqueda de un repositorio digital mediante el uso de Schema.org.

- Trabajo final de máster -

Nuria Nevado Chiné

Màster de Gestió i Direcció de Biblioteques i Serveis de Informació.

Facultat de Biblioteconomia i Documentació. Universitat de Barcelona

Tutor del TFM: Rubén Alcaraz

Septiembre 2019

SUMARIO

Resumen ejecutivo	3
1 Contexto, objetivos y metodología	5
1.1 Contexto y justificación	5
1.2 Objetivos	6
1.3 Metodología	7
2 Estado del arte	9
2.1 Introducción	9
2.2. Visibilidad de los repositorios en la web	9
2.3. Posicionamiento web	20
2.4 Schema.org	35
2.5 Software para la creación de repositorios	51
2.6 Análisis de casos y buenas prácticas	57
3 Metodología	61
3.1 Implementación de Schema.org en el repositori RODERIC	61
4 Resultados	71
5 Conclusiones	76
6 Discusión y trabajo futuro	79
Bibliografía	81
Glosario	90
Índice de imágenes	92
Índice de notas a pie de página	93
Anexo 1. Metadatos RODERIC	96
Anexo 2. Mapeo RODERIC - Schema.org	99
Anexo 3. Código JSON-LD para RODERIC	100
Anexo 4. Indicadores y estadísticas	104
Agradecimientos	108

Resumen ejecutivo

La Web 3.0 ha transformado la forma de acceder y compartir el conocimiento. La Web Semántica abre un nuevo espacio de relación entre la información que se publica en la Web y la comprensión que de ellas pueden extraer las máquinas para dar una mejor respuesta al usuario en sus búsquedas. En este contexto, la optimización de motores de búsqueda (SEO) se convierte en un factor crucial como método para mejorar la visibilidad de un sitio o página web en los resultados de búsqueda de un motor de búsqueda.

En este camino de mejora de la interoperabilidad semántica, Google, Yahoo y Bing presentan en 2011 Schema.org, un vocabulario creado para hacer que el contenido web sea comprensible a los rastreadores web y otras máquinas. Un vocabulario que permite describir la información que contienen las webs con una serie de propiedades que se insertan dentro del código HTML, *semantizando* sus contenidos y haciendo sus datos legibles e interpretables por aplicaciones informáticas.

Los repositorios digitales se enfrentan al desafío de que los usuarios encuentren su contenido en un entorno tan grande como Internet. Si consideramos que el comportamiento de los investigadores respecto a cómo descubren, leen y utilizan la literatura académica ha cambiado considerablemente con el avance de la Web, el planteamiento de trabajo por parte de los repositorios digitales en pro de la visibilidad e interoperabilidad con la web, necesita afrontar mejoras semánticas para no convertirse en repositorios invisibles, porque no pueden ser recuperados de forma apropiada por los motores de búsqueda en internet.

En este trabajo se realiza una introducción a los conceptos básicos y diferentes tecnologías, estándares y recomendaciones que conforman el entorno de la Web Semántica y se plantea el reto de implementar una herramienta como Schema.org en el repositorio institucional de la Universitat de València (RODERIC) como ingrediente de mejora en el entendimiento entre la web, la información que contiene y los buscadores.

Para ello se presenta un mapeo que relaciona los metadatos Dublin Core de RODERIC con las propiedades que establece Schema.org. Este mapeo se traslada a formato JSON-LD para poder implementarlo en las colecciones de Libros, Artículos académicos y Tesis del repositorio y se revisa con la "Herramienta de datos estructurados de Google" el código

implementado. Finalmente se solicita a los buscadores Google, Yahoo! y Bing la reindexación del repositorio.

Para evaluar el impacto de la Web semántica sobre el repositorio y su posicionamiento en el ranking de los resultados (SERP) y ver si se produce mayor indexación de documentos mejorando el tráfico hacia ellos, se definieron unos indicadores referidos a impresiones, posición media del repositorios en búsquedas, número de clics, tipo de fuentes de tráfico y sesiones y análisis de búsquedas específicas¹. El periodo de análisis se produjo entre el 12 de abril y el 31 de julio de 2019 y se utilizó Google Analytics, Google Search Console, Roderic Statistics y el conteo de las búsquedas específicas como herramientas de análisis.

Tras comparar el periodo de estudio de 2019 con los mismos meses de 2018, si se observa un ligero aumento de tráfico de visitas entre estos meses, en concreto un 4,63%, un aumento considerable del 111,66% de las impresiones que se producen en el buscador de Google por parte de RODERIC. Pero a penas hay diferencia en el posicionamiento de los resultados del estudio de búsquedas hechas y el CTR (porcentaje de clics) sufre un ligero descenso pasando de 3,89% a 2,06%. La posición media en el SERP si sufre un descenso más negativo pasando de la posición media de 23 en 2018 a la 33 en 2019.

El resultado del estudio es poco concluyente debido al escaso tiempo de análisis que requiere esta implementación. Es importante concluir que son muchos los factores que influyen en el posicionamiento orgánico en buscadores y el lenguaje semántico es solo uno de los ingredientes a considerar. Mejorar la visibilidad de los repositorios es un trabajo de fondo que debe gestionarse con un plan de trabajo a medio plazo. De ahí que una de las recomendaciones sea un estudio más amplio y un plan de evaluación que permita observar también las tendencias de los usuarios e ir acondicionando la implementación de Schema.org con otras mejoras cualitativas en el repositorio.

¹ Ver Anexo 4. Indicadores y estadísticas.

1 Contexto, objetivos y metodología

1.1 Contexto y justificación

En la actualidad las bibliotecas afrontan un nuevo reto, el de gestionar repositorios institucionales. Tarea que implica reunir, describir, dar acceso, divulgar y preservar la producción intelectual desarrollada en las propias instituciones como universidades, centros de investigación o de memoria.

Los repositorios digitales se han consolidado como la vía verde para implementar el acceso abierto. Una posición que invita a mejorar la accesibilidad y visibilidad de los datos y las publicaciones científicas, como uno de los pasos a dar en la eliminación de las barreras que inhiben el acceso a los resultados de la investigación científica, académica y cultural.

Los repositorios digitales deben abordar un nuevo contexto tecnológico en el que la Web Semántica juega un papel fundamental. Las plataformas de repositorios siguen utilizando tecnologías y protocolos diseñados hace casi veinte años, antes del auge de la web y el predominio de Google, las redes sociales, la web semántica y los dispositivos móviles (Rodriguez, 2017). Para que los repositorios digitales puedan contribuir y aprovechar la Web Semántica, la información que contienen debe ser legible por las máquinas mediante la aplicación de diferentes tecnologías, idiomas, estándares y recomendaciones. La información debe convertirse en información que la Web Semántica pueda leer y transformar y así poder dejar de ser invisibles para la web.

En este camino hacia la accesibilidad e interoperabilidad con la Web Semántica, Google, Yahoo, Bing y Yandex presentan en 2011 Schema.org. Un proyecto común para unificar la forma de etiquetar y enriquecer la información de los contenidos web para que los buscadores puedan indexarlos, clasificarlos y mostrarlos, con el fin de mejorar la interacción con los usuarios.

En este ecosistema juega un papel muy importante la visibilidad en Google, como uno de los principales buscadores que es. El denominado SEO (Search Engine Optimization) debe ser una de las prácticas fundamentales a trabajar en el camino a mejorar el impacto, la visibilidad, la difusión y el acceso al conocimiento depositado en los repositorios.

La realización de este trabajo tiene su origen en la carrera profesional de su autora. Hace unos 6 años, tuve la oportunidad de iniciar y desarrollar el repositorio del Departament de Salut de la Generalitat de Catalunya, Scientia Dipòsit d'Informació Digital del Departament de Salut². Un proyecto ambicioso liderado por la Biblioteca de Ciències de la Salut de Catalunya y que me permitió ver y trabajar de primera mano en el diseño, la implementación y gestión del repositorio. Una experiencia que quería reflejar en una propuesta que supusiera dar un paso adelante en la gestión de los repositorios. Este trabajo, surgido de una conversación con el profesor y tutor, Rubén Alcaraz, pretendía explorar nuevas posibilidades de mejora para los repositorios y ser un paso más en el camino hacia la visibilidad y el posicionamiento de los mismos.

1. 2 Objetivos

El propósito de este trabajo es realizar una introducción a los conceptos básicos y principales tecnologías que conforman la Web Semántica y presentar una visión específica de cómo un vocabulario como Schema.org puede maximizar el acceso al contenido alojado en repositorios institucionales.

El objetivo principal se centra en **mejorar la visibilidad y el posicionamiento en los motores de búsqueda de un repositorio digital a partir de la implementación del vocabulario Schema.org.**

Entre los objetivos específicos se busca:

- Definir, contextualizar y determinar los principales factores de posicionamiento en el SEO académico (ASEO).
- Hacer una revisión del vocabulario Schema.org y aplicabilidad a un repositorio digital.
- Analizar Schema.org como herramienta para aumentar la riqueza de los metadatos de un repositorio y herramienta con fines de posicionamiento web.
- Realizar un análisis de estas tecnologías y ver cómo pueden incrementar las visitas desde los buscadores (tanto en Google Scholar, como en Google Search).

² <https://scientiasalut.gencat.cat/>

1. 3 Metodología

La metodología utilizada para estudiar la implementación de Schema.org como herramienta de mejora en el posicionamiento en los motores de búsqueda y aumentar la visibilidad, el impacto y la interacción con el usuario, ha sido la siguiente:

1. Revisión del estado del arte

Se ha realizado una revisión sobre la situación actual de los repositorios digitales al respecto de su visibilidad en la web. Se han analizado de forma teórica los principales elementos que caracterizan el SEO y el SEO académico (ASEO). Se ha realizado una aproximación a Schema.org y se analiza su implementación como herramienta de marcado semántico propuesta por los principales motores de búsqueda. Se han revisado los principales formatos de implementación del (Microdatos y JSON-LD) y cual es la tendencia actual. Finalmente, se ha realizado un repaso por diferentes software para gestionar repositorios y las posibilidades estas aplicaciones que ofrecen frente al SEO semántico.

Para afrontar todo este recorrido propuesto, se realizaron búsquedas en diferentes recursos como el descubridor Cercabib de la Universitat de Barcelona (UB) o Trobes+ de la Universitat de València, bases de datos como (WOS, LISA, Dialnet), en el buscador académico Google Scholar, repositorios y recolectores (Digital CSIC, Core, Zenodo) en publicaciones especializadas (code4lib) y se ha realizado un recorrido por los artículos de interés propuestos por el tutor y la bibliografía reseñada de estos. Estas búsquedas bibliográficas han afrontado campos como el de autor, palabras clave y título, todos ellos en diferentes idiomas (Español, Inglés, Catalán) y dando prioridad a los artículos de los últimos 5 años. Entre las principales palabras clave utilizadas están: SEO; Academic SEO; Semantic SEO; Repository; Schema.org; Visibility; Semantic Web; Microdata; JSON-LD; Linked data; Indexing; Google Scholar; Interoperability.

Para abordar algunos aspectos más locales sobre el acercamiento de los investigadores a la hora de realizar una búsqueda y la interoperabilidad del repositorio RODERIC con Europeana, se realizaron dos consultas personales: una con el profesor de la UB Àngel Borrego y la otra con la responsable de RODERIC por parte de la Universitat de València Rosario García, quienes recomendaron algunos artículos para abordar esos aspectos.

2. Análisis de casos y buenas prácticas

Observar experiencias en este campo ayudó a ver casos reales, como el que más tarde se desarrollaría en este trabajo. Se ha intentado abordar este aspecto intentado ver casos a nivel internacional, europeo y nacional. Cabe destacar que en el ámbito nacional ha sido complicado encontrar casos de buenas prácticas aplicadas a repositorios, aunque sí hay experiencias de aplicación de Schema.org en bibliotecas virtuales (como la Biblioteca Virtual Miguel de Cervantes)³.

No obstante, algo que se ha podido constatar es que no son muchas las experiencias de bibliotecas, ni siquiera en el ámbito internacional, que hayan implementado Schema.org en los repositorios digitales. Algo que resulta sorprendente, si consideramos que las bibliotecas siempre buscan estar un paso adelante en acciones para mejorar el acceso abierto en el mundo académico y de investigación. También así lo refiere Macgregor (2019) al concluir que solo se ha documentado una cantidad limitada de trabajos anteriores sobre los enfoques de descubrimiento de repositorios y su evaluación.

3. Implementación y análisis de Schema.org en formato JSON-LD, en el repositorio institucional de la Universitat de València, RODERIC.

A principios de enero de este 2019, surgió la posibilidad de probar la implementación de Schema.org en el repositorio institucional de la Universitat de València, RODERIC. Una posibilidad única que nos permitía observar en un repositorio real y con un bagaje consolidado, la implementación de un vocabulario semántico, objeto de estudio de este trabajo.

Para ello se realizó un estudio previo de los metadatos Dublin Core, para poder completar un mapeo entre estos metadatos y las propiedades de Schema.org. Se definió un código en formato JSON-LD que se incorporó en el repositorio con el apoyo de José Ángel Navalón del Servei d'Informàtica de la Universitat de València. Finalmente, se ha estudiado el comportamiento del repositorio en la Web, a través de un conjunto de indicadores relacionados con el tráfico y el posicionamiento web.

³ <http://data.cervantesvirtual.com/about?lang=es>

2 Estado del arte

2.1 Introducción

Durante las dos últimas décadas el acceso a la información digital y el uso de recursos electrónicos ha sufrido una sustancial mejora, tanto cuantitativa como cualitativa, en lo que refiere al acceso y consulta de información científica.

La Web 3.0 ha transformando la manera de acceder y compartir el conocimiento, orientada hacia el entendimiento entre los motores informáticos y procesadores de información. La Web semántica busca definir el significado de un contenido Web más allá del propio significado textual, abriendo la puerta a un conjunto de elementos SEO (Search Engine Optimization) que influyen en el posicionamiento y la visibilidad.

La propuesta de Web Semántica elaborada por Berners-Lee et al. (2001) y la posibilidad de describir semánticamente la información en la web (Berners-Lee and Hendler, 2001) abre un mundo lleno de posibilidades para la publicación y recuperación de información.

Tanto en la Web como en los repositorios, hay una gran cantidad de documentación científica que en ocasiones resulta difícil recuperar. Lewandowski y Mayr (2006) lo definieron como la *Academic Invisible Web* aquella que “consta de todas las bases de datos y colecciones relevantes para el mundo académico pero que no pueden ser buscadas por los motores de búsqueda de Internet de uso general”. La indexación de esta parte de la web invisible es fundamental para los motores de búsqueda científicos.

La gran cantidad de información hace más complicada una recuperación precisa y relevante a una determinada búsqueda o consulta. De modo que dotar de significado semántico a la información científica disponible en la web es más necesario que nunca. La descripción o marcado de los recursos de información de manera semántica abre un nuevo espacio lleno de posibilidades a la hora de facilitar la recuperación de información de una manera más precisa y relevante. Además de favorecer la interoperabilidad y la reutilización de información por parte de las máquinas y/o motores de búsqueda.

2.2. Visibilidad de los repositorios en la web

El acceso abierto (OA) ha convertido a los repositorios en una pieza clave para la difusión y preservación del conocimiento científico. En este sentido, el número de repositorios, particularmente Repositorios Institucionales (RI), ha crecido constantemente en la última década. En agosto de 2019, OpenDOAR cifra en 4,242 los repositorios 'institucionales' en todo el mundo, una cifra similar a los 3,252 repositorios de “investigación institucional o departamental” registrados en el Registro de Repositorios de Acceso Abierto (ROAR)⁴. En España, de los datos de la evolución y crecimiento de los repositorios institucionales que ofrece el directorio “OpenDOAR” se observa que, en diciembre de 2005, aparecía registrado un solo repositorio institucional y en julio de 2017 el número asciende a 87 (Ferrerías, 2018).

El mandato⁵ del acceso abierto busca definir estrategias y metas para acelerar el proceso de dar acceso abierto a documentos científicos en el mundo, y es por eso que los objetivos de los repositorios pasan, entre otros por:

- Dar acceso a literatura científica.
- Favorecer la accesibilidad, la visibilidad y el impacto de la investigación científica.
- Recoger y preservar la producción científica.
- Transmitir y compartir conocimiento.
- Potenciar la investigación.

Los repositorios institucionales requieren una nueva visión que integre los aspectos que conlleva la visibilidad científica y las nuevas funcionalidades tecnológicas. La mayoría de repositorios están basados en estándares abiertos para garantizar que el contenido seleccionado sea de fácil acceso, se pueda buscar y se pueda recuperar para su uso posterior. Respaldar la visibilidad de los repositorios y sus contenidos, pasa por una revisión de los estándares de calidad, relevancia, uso y medidas de impacto.

2.2.1 Estándares de calidad

La unificación de criterios de calidad resulta ser de gran importancia para lograr el máximo aprovechamiento de la información que contienen. Por ello, desde 2007, la Fundación

⁴ http://v2.sherpa.ac.uk/view/repository_visualisations/1.html

⁵ El mandato de acceso abierto son una serie de políticas y medidas impulsadas por Instituciones y Gobiernos relacionadas con el acceso abierto.

Española para la Ciencia y la Tecnología (FECYT) y la Red de Bibliotecas Universitarias (REBIUN) de la CRUE colaboran en el proyecto RECOLECTA (o Recolector de Ciencia Abierta). Para alcanzar un estándar de calidad mayor se elaboró una Guía para la evaluación de repositorios institucionales de investigación que ofrece a toda la comunidad nacional un conjunto de directrices, basadas en criterios internacionales ya existentes, que garantizan la interoperabilidad de todos los recursos de acceso abierto y aseguran un acceso de calidad a sus contenidos (Barrueco; Cruz et al., 2017)

La *Guía para la evaluación de repositorios institucionales de investigación* (2017) tiene como finalidad principal el seguir siendo un instrumento de auditoría interna para mejorar la calidad de los repositorios. La guía presenta una selección de 25 criterios a valorar. Los aspectos a evaluar tienen que ver con:

- **Visibilidad:** se estudia si el repositorio está visible en la propia institución y presente en los directorios nacionales e internacionales. También se refiere a la proporción de documentos del repositorio en acceso abierto.
- **Políticas:** los indicadores de este apartado se refieren a si existe una definición de la misión y objetivos del repositorio, y si existe una política pública sobre depósito de documentos, preservación de contenidos y reutilización de metadatos.
- **Aspectos legales:** se comprueba si existe una autorización del autor de cada documento introducido en el caso de que el depósito sea mediado, y si está disponible información sobre derechos de autor en el repositorio.
- **Metadatos:** estos indicadores van dirigidos a conocer si los propios metadatos siguen una política de indización, y si se permite la exportación de metadatos en algún otro formato que el Dublin Core.

En la misma línea, Serrano-Vicente et al. (2014) hacen una primera revisión de los indicadores para la evaluación de repositorios institucionales de acceso abierto. En el estudio presentan ocho casos de buenas prácticas relacionadas con las siguientes categorías de evaluación: marketing, factores económicos y la financiación, interoperabilidad y cooperación, políticas del repositorio, contenidos y los servicios de valor añadido.

Los enfoques de evaluación de repositorios son muy variados. Sean unos u otros los indicadores a mirar, la evaluación debe ser una herramienta adecuada a los objetivos de la institución que lo ha creado y enfocada a mejorar el servicio que recibe el usuario final.

2.2.2 Impacto e indicadores en repositorios

Un ejemplo de indicador bibliométrico cuantitativo que ayuda a evaluar la calidad de un repositorio institucional es el Ranking Web de Repositorio Digitales⁶.

El Ranking Web de Repositorio Digitales es una iniciativa del Laboratorio de Cibermetría, que pertenece al Consejo Superior de Investigación Científica (CSIC) de España y tiene por objetivo apoyar las iniciativas de acceso abierto y gratuito a las publicaciones científicas. Mide la visibilidad y el impacto de los repositorios científicos, utilizando indicadores web (tamaño, visibilidad, ficheros ricos y Google Scholar).

Los indicadores web utilizados en este ranking miden la visibilidad e impacto de los repositorios científicos. El Ranking Web se basa en un indicador que combina datos de presencia web y de impacto web (visibilidad hipertextual) (Aguillo, et al., 2010). Los indicadores actualmente en uso y sus respectivos porcentajes son:

- Tamaño: número de páginas web de acuerdo al buscador Google (10%).
- Visibilidad: se tiene en cuenta la raíz cuadrada del número de enlaces externos multiplicada por el número de dominios web que originaron dichos enlaces. Las fuentes de información son: Majestic SEO⁷ y ahrefs⁸. Se utiliza el máximo de los citados productos normalizados entre ambos proveedores (50%).
- Ficheros ricos: la suma total del número de ficheros de los siguientes formatos (según Google): pdf, doc+docx, ppt+pptx y ps+eps (10%)
- Scholar: el número de artículos (no citas) recogidos en Google Scholar durante el periodo de 5 años comprendido entre 2008 y 2012 (30%).

El Ranking Web presenta una serie de recomendaciones y buenas prácticas para mejorar la posición en el ranking:

- Asignación de un URL estable con un dominio que no cambie.
- Creación de contenidos por la institución como guías, protocolos, productos bibliotecarios o equipos que creen contenido para el repositorio
- Conversión de contenidos no electrónicos a electrónicos para subirlos al repositorio
- Interoperabilidad e interconexión entre páginas web
- Accesibilidad en inglés.

⁶ <https://repositories.webometrics.info/es>

⁷ <https://es.majestic.com/>

⁸ <https://ahrefs.com/>

- Uso de ficheros enriquecidos como PDF o doc frente a otros como HTML que permite una mejor distribución y recolección.
- Diseños amigables sin demasiados menús y motores de búsqueda sencillos y dinámicos.
- Monitorización de la popularidad y estadísticas para vigilar su visibilidad en otros motores de búsqueda.
- Copia de seguridad del material viejo o desfasado para garantizar su preservación
- Estándares de metadatos como Dublin Core que permitan añadir significado a títulos e información complementaria.

Serrano-Vicente, et.al (2018) realizan un segundo estudio enfocado en identificar los indicadores existentes que se utilizan para evaluar los repositorios. La propuesta define un total de 48 indicadores divididos en cinco categorías (tecnología, procedimientos, contenido, marketing y personal). Una propuesta que recorre aspectos internos y externos, incluyendo aspectos relacionados con la web 2.0 y la difusión y que permite observar que aspectos necesitan una revisión y cuales pueden dar un paso adelante.

2.2.3 Fuentes de información y repositorios

Con el paso de las fuentes de información de papel a digital, el comportamiento de los investigadores respecto a cómo descubren, leen y utilizan la literatura académica para su trabajo ha cambiado considerablemente. El acceso a la información se multiplica si tenemos en cuenta la gran variedad de plataformas de acceso y publicación. Diversos estudios han analizado el comportamiento de los investigadores a la hora de buscar información poniendo como eje central qué fuentes utilizan para buscar información.

Brown y Swan (2007) presentan un estudio sobre el uso de las bibliotecas académicas y sus servicios, del que destacamos los siguientes datos:

- Un número considerable de consultas (más del 80%) provenían de buscadores web genéricos, principalmente Google.
- En general, el 6% de los investigadores informan que usan repositorios institucionales "con frecuencia" para encontrar y usar información, otro 11% dice que los usan "a veces" y el 11% de los investigadores los usan "de vez en cuando". En este momento, el 60% de los investigadores nunca han utilizado un repositorio institucional para encontrar información.

Figure 36: Researchers' use of Open Access discovery services and content

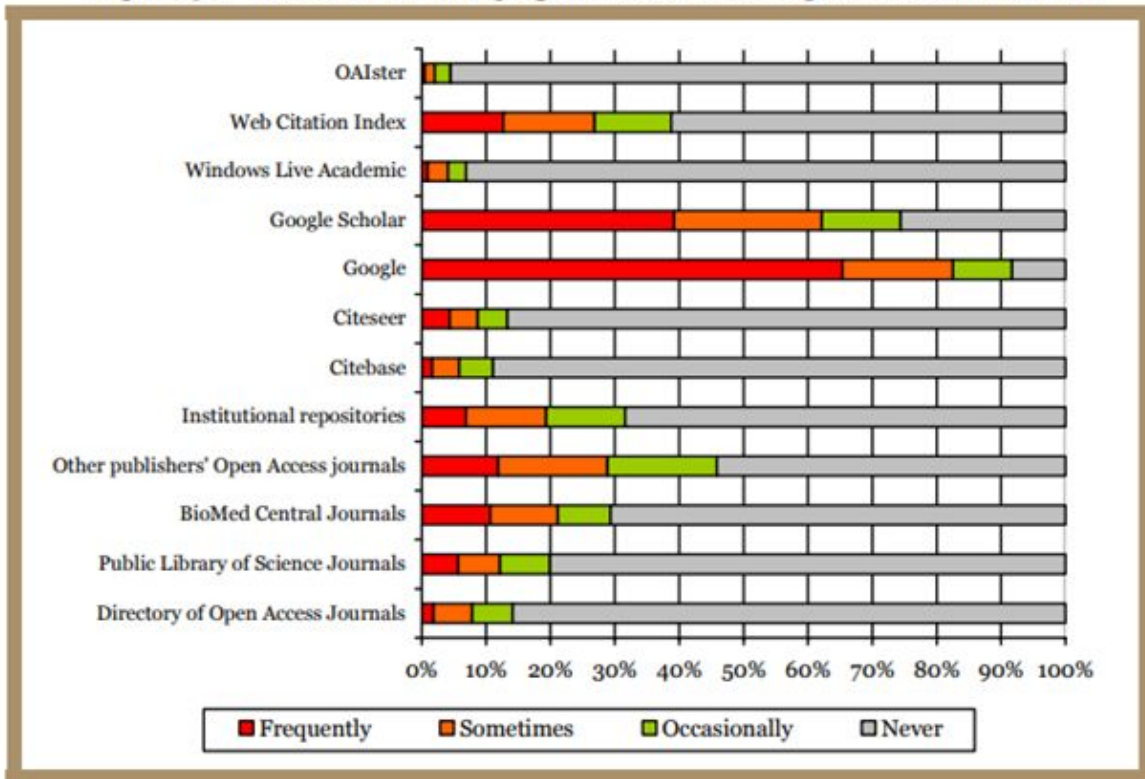


Imagen 1. Acceso a la información por parte de los investigadores (Brown y Swan (2007).

Uno de los últimos estudios en esta misma línea examina usos y costumbres de los investigadores (Tenpoir, 2019). En su encuesta realizada en 2018 y con un total de 963 respuestas, algunos de los datos que evidencia son:

- De las lecturas descubiertas mediante búsquedas, más de la mitad (56%,) se encontraron en motores de búsqueda generales como Google o Google Scholar, y los servicios de resumen e indexación representaron otro 22%.
- Los encuestados de 50 años o menos eran más propensos a usar un repositorio institucional o de temas y los encuestados de 30 años o menos tenían más probabilidades de obtener un artículo de un colega.

Al comprender la búsqueda de información, la lectura y el uso de los investigadores, las instituciones, bibliotecas y editores pueden satisfacer mejor las necesidades de sus usuarios.

En esta línea algunos ejemplos de buenas prácticas para ampliar el uso y visibilidad de los repositorios y así aumentar las visitas, consultas, descargas, citas..., son:

- La actualización constante de un Repositorio Institucional (RI) es una de las estrategias para incrementar la visibilidad de los recursos y la producción científica de forma abierta.
- Facilitar que los contenidos en abierto sean recolectados por herramientas como OpenAire, Recolecta, BASE o Europeana, ayuda a mejorar su visibilidad aumentando el número de consultas y descargas de la producción científica.
- La integración en buscadores como Google Scholar también está aumentando el número de consultas y descargas de la producción científica depositada en abierto.
- Los recolectores o herramientas de descubrimiento son uno de los medios empleados por las instituciones para ampliar la visibilidad de sus repositorios. Facilita las tareas de recolección de información, organización, elaboración y difusión de contenidos. Las tecnologías de descubrimiento permiten que los contenidos de los repositorios se ofrezcan de forma simultánea con los registros de los catálogos de biblioteca y con las colecciones de recursos electrónicos disponibles en la institución.

2.2.4. Indexación web

El incremento de la visibilidad de la producción científica y académica tiene como resultado una mayor audiencia, mayor transmisión del conocimiento y el fomento de la investigación. Para que un repositorio evolucione hacia la diseminación del conocimiento a través de Internet, se requiere que sean repositorios actualizados de acuerdo con las tendencias tecnológicas.

Para asegurar que los repositorios cumplen adecuadamente su misión es fundamental conocer su presencia e impacto en la Web, especialmente en Google (motor de búsqueda por excelencia). Conseguir que los repositorios estén correctamente indexados para hacer visible la producción científica, académica e institucional es una tarea fundamental para poder mejorar el objetivo de favorecer la accesibilidad, la visibilidad y el impacto de la investigación científica. Una correcta indexación ayudará a mejorar el posicionamiento en los primeros lugares en dichos sistemas.

El grado de indexación de los repositorios en Google y Google Scholar se convierte en un elemento fundamental. Prieto-Jimenez (2015) repasa los conceptos de indexación de los repositorios en motores de búsqueda. En el apunta los siguientes elementos de enlace a considerar:

- Software actualizado: las mejoras en cada versión.
- Sitemaps: un sitemap es un archivo XML que enumera las URL de un sitio junto con metadatos adicionales sobre cada una de ellas (cuándo se actualizó por última vez, con qué frecuencia suele cambiar y qué tan importante es, en relación con otras URL del sitio). Permite rastreos más inteligentes para que los motores de búsqueda.
- Robots.txt: un archivo robots.txt proporciona información a los rastreadores de buscadores sobre las páginas o los archivos que pueden solicitar o no del sitio web.
- Últimos registros añadidos: facilitar desde la página principal del repositorio un enlace a un listado de los últimos registros añadidos ya que facilita el rastreo de los registros nuevos.
- Redirecciones de la página de descarga: se desaconseja redireccionar la página de descarga a alguna página intermedia. Daña posicionamiento en índices de buscadores.
- Protocolo OAI: da acceso a los últimos registros añadidos o modificados y a los metadatos descriptivos de los recursos a indexar.
- Posicionamiento: estar enlazados con otras herramientas o repositorios. Disponer de URL estables, Handle.

Arlitsch y O'Brian (2012) apuntaron que las principales causas del mal posicionamiento de los repositorios son el esquema de metadatos utilizado y las características de la arquitectura de información y navegabilidad, que no ayudan al robot de búsqueda a realizar la indexación. La falta de visibilidad en la web de los repositorios está sustentada por los bajos índices de indexación de su contenido (tanto en Google como en Google Scholar), ya que una baja presencia en la web determina una baja visibilidad en la web. Un problema común con herramientas como los OPAC de bibliotecas, con poca capacidad para posicionarse en la web.

Como hemos apuntado anteriormente, si observamos los patrones de búsquedas, en general, las personas que están investigando no van directamente a un repositorio específico para descubrir lo que allí está depositado, sino que en la inmensa mayoría de los casos buscan a través de Google o Google Scholar. En otras palabras, la visibilidad de los

contenidos depende en manera importante de esos buscadores. Años atrás, se decía “si no estás en Internet no existes”, ahora podríamos decir, “si estás en Internet, pero no eres visible en Google, tampoco existes” (Espinak, 2014).

Estudios como el de Orduña-Malea; Delgado López-Cózar (2015) plantean casos específicos de esa falta de visibilidad en la web de los repositorios. El artículo destaca la insuficiente visibilidad que tiene la producción académica latinoamericana en la Web. Para una muestra de 127 repositorios, los resultados indicaron que el índice de indexación era bajo en Google y prácticamente no existía en Google Scholar. Resultados principalmente atribuibles a las limitaciones derivadas del uso de esquemas de descripción que son incompatibles con Google Scholar (diseño del repositorio) y la confiabilidad de los indicadores web (motores de búsqueda).

2.2.5. La interoperabilidad sintáctica y semántica

El acceso abierto se basa en la interoperabilidad, la capacidad de dos sistemas para comunicarse y compartir información de ida y vuelta en un formato utilizable (COAR, 2011). La interoperabilidad de los repositorios y la unificación de criterios de calidad resulta ser de gran importancia para lograr el máximo aprovechamiento de la información que contienen. “La finalidad de todo repositorio institucional es la difusión y aumento del impacto entre la comunidad científica de los resultados de investigación, evitando convertirse en un nuevo depósito infrautilizado y carente de sentido” (Sánchez y Melero, 2007).

Mientras el énfasis en la creación, diseño y difusión del producto, toma como centro de gravedad el concepto Open Access, generalmente se olvida que el producto es, ni más ni menos, un sitio web y, por tanto, se deben cuidar ciertos aspectos relacionados con la navegabilidad, la usabilidad, la visibilidad en los motores de búsqueda, etc. (Orduña-Malea, 2013).

Respetar los estándares de compatibilidad e interoperabilidad hace posible que el contenido de un repositorio sea más fácilmente recuperable y visible, no sólo para la propia institución sino para la comunidad científica a nivel global y la sociedad en general.

La interoperabilidad se puede administrar en distintas capas. Las directrices DRIVER apuntan a alcanzar una interoperabilidad en dos capas, la sintáctica (utilización de OAI-PMH y de OAI_DC) y semántica (utilización de vocabularios) (DRIVER Project, 2008).

En la interoperabilidad sintáctica, para lograr el incremento de la visibilidad e impacto, los repositorios trabajan con metadatos como elemento esencial para recuperar, identificar, localizar y reutilizar los contenidos alojados. Los repositorios institucionales basan su interoperabilidad sintáctica en los modelos de metadatos combinados con herramientas como el protocolo OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting), con el fin de lograr la interoperabilidad del repositorio y facilitar su integración a redes globales.

Los repositorios conviven en un ecosistema cada vez más complejo que incluye la interacción con otras infraestructuras locales y sistemas externos que se ocupan de publicaciones (por ejemplo sistemas de *e-learning*, plataformas de publicación, sistemas CRIS, etc.). Esto evidencia la necesidad de extender las actividades de interoperabilidad más allá de los esfuerzos de depósito a repositorio para incluir la interoperabilidad en esta diversidad de sistemas que existen en este ecosistema.

Simple Web-service Offering Repository Deposit - SWORD es un protocolo que permite a los autores depositar un artículo a través de una única interfaz y luego enruta el elemento a múltiples repositorios. El objetivo de SWORD es minimizar la barrera para depositar a fin de hacer el proceso tan simple y sencillo como sea posible para los investigadores. Con el fin de que cualquiera de estos protocolos se pueda trabajar, los sistemas deben estar configurados de manera consistente e interoperable.

Existen también otros protocolos que también intentan aprovechar el potencial de la interoperabilidad para facilitar el intercambio de contenidos (no solo registros de los metadatos) entre los sistemas. Un ejemplo es el protocolo iniciativa de Archivos Abiertos, reutilización e intercambio de materiales OAI-ORE, que define los estándares para la descripción y el intercambio de agregaciones a los recursos de la web, algunas veces llamados como objetos digitales compuestos, y que permite que los elementos ya depositados en un repositorios sean copiados en otras colecciones (COAR, 2011).

Hakimjavadi (2013) presenta un listado de estos protocolos en una comparativa a nivel de software de repositorios con diferentes protocolos de interoperabilidad sintáctica.

	OAI-PMH	OAI-ORE	ATOM	RSS	SWORD	AtomPub	Z39.50	SRU/W
CDS Invenio	✓	No	No	3rd	No	No	No	-
CONTENTdm	✓	No	No	No	No	No	✓	No
DigiTool	✓	Future	No	No	Future	No	✓	No
Digital Commons	✓	No	No	✓	No	No	No	✓
DSpace	✓	✓	✓	✓	✓	✓	No	✓
EPrints	✓	✓	✓	✓	✓	✓	No	No
ETD-db	✓	No	No	No	No	No	No	No
Fedora	Opt	Opt	No	Opt	✓	No	No	✓
Fedora-FEZ	✓	Opt	No	No	✓	No	No	✓
Greenstone	✓	No	No	No	No	No	✓	✓
MyCore	✓	NO	Opt	No	✓	✓	No	No
OPUS	✓	✓	No	No	-	No	No	-
Open Journal System	✓	✓	✓	✓	✓	✓	No	Future
VITAL	✓	Opt	Opt	No	✓	No	No	✓
Zentity	✓	✓	✓	✓	✓	✓	No	No

Notes: *Future:* Planned to be available in a future version; *Opt:* Optional – must be installed or enabled separately; *3rd:* 3rd Party – add-ins provided by third party suppliers; *[blank]* – Needs more investigations

Table IV.
Supported protocols in 15
common IR solutions

Imagen 2. Protocolos que admiten los diferentes softwares de repositorios (Hakimjavadi 2013).

Con la tecnología en constante evolución y las nuevas herramientas que se desarrollan en todo momento, los repositorios deben ser capaces de soportar más servicios complejos e innovadores que se basan en requisitos adicionales de interoperabilidad.

La interoperabilidad semántica busca garantizar que el significado preciso de la información intercambiada sea comprensible de manera consistente; permite el intercambio fiel de significado entre máquinas y personas. Introducir información semántica en el contenido de las páginas web, con vocabularios que si pueda leer la web semántica hará la información más comprensible para “robots” y buscadores. Los investigadores tendrán una forma más consistente de encontrar y recuperar elementos relevantes a través de búsquedas.

En esta línea, cabe destacar la iniciativa de Schema.org. Una propuesta presentada en 2011, por Google, Yahoo, Bing y Yandex. El propósito de Schema.org es el de incluir metadatos dentro de los propios contenidos web haciendo que una página web sea un objeto procesable para la extracción de información semántica. Estas propiedades se insertan dentro del código HTML, *semantizando* sus contenidos y permitiendo una mejora en el proceso de recuperación de la información. Schema.org provee de una colección de esquemas/vocabularios para marcar HTML y hacerlo comprensible para la mayoría de motores de búsqueda.

Estos vocabularios para la web semántica pueden presentarse en los siguientes formatos:

- Microformatos
- RDFa

- Microdata (HTML5)
- JSON-LD (JSON for Linking Data)

Mejorar la interoperabilidad semántica en el ámbito de los repositorios puede tener múltiples aplicaciones y beneficios. Desde, mejorar el entendimiento “buscador-interrogador”, mejorar el posicionamiento en buscadores y destacar en los SERPs⁹, atraer visitantes y fomentar un CTR¹⁰ más alto, aplicar implementaciones con fragmentos enriquecidos, ratings, obtener estadísticas de revisión, búsquedas localizadas (geolocalización), etiquetado de correos y vinculación con otras herramientas (calendarios, Google Now...)

Algunos ejemplos de implementaciones con Schema.org:

- En 2012, OCLC agregó Schema.org a sus registros bibliográficos de WorldCat .
WorldCat Linked Data Vocabulary
- Koha 3.14.0 se lanzó con soporte para microdatos de Schema.org en su OPAC de código abierto.
- Evergreen, otro SIGB de código abierto ahora está haciendo una implementación similar.
- Las bibliotecas de la Universidad de Indiana están utilizando Schema.org. en diferentes herramientas.
- *DigitalGeorgetown* (Repositorio Institucional de la Universidad de Georgetown)

2.3. Posicionamiento web

2.3.1. Search Engine Optimization (SEO)

Codina (2004) define el concepto de SEO (Search Engine Optimization) como el “conjunto de procedimientos que permiten posicionar un sitio o una página web en un lugar óptimo entre los resultados proporcionados por un motor de búsqueda”.

⁹ Ver glosario.

¹⁰ Ver glosario.

El posicionamiento web es el proceso para mejorar la visibilidad de un sitio web en los resultados orgánicos¹¹ de los diferentes buscadores. Los métodos SEO se centran en optimizar el ranking de sitios web y sus contenidos en los motores de búsqueda.

Entender cómo funcionan los buscadores es el primer paso para mejorar nuestro posicionamiento. El algoritmo RankBrain de Google actualiza sus algoritmos unas 500 o 600 veces por año, esto se traduce en que hay 2 o 3 actualizaciones por día. Muchos creen que su objetivo principal es medir cómo los usuarios interactúan con los resultados de búsqueda y clasificar los resultados en consecuencia.

Existen dos grupos de factores SEO a trabajar para mejorar la visibilidad de un sitio web.

- *SEO interno*: también conocido como *SEO On Site* o *SEO On Page*. El SEO interno va encaminado a todas aquellas acciones que realizamos en nuestra propia web para mejorar su optimización. La consultora de marketing digital Top Position (2012), resalta que se trata de “optimizar aquellos aspectos internos del diseño y estructura de la página web que se quiere posicionar”.

Google presenta una *Guía de optimización en buscadores (SEO) para principiantes*¹² con consejos y optimizaciones a llevar a cabo, propiciando un impacto notable en la experiencia de usuario y en el rendimiento de un sitio web en los resultados de búsqueda orgánica.

Algunos de los aspectos a considerar a la hora de optimizar el posicionamiento en los motores de búsqueda con este tipo de SEO son:

- El primer paso para conseguir que el sitio web se muestre en Google es mediante un sitemap, que es un archivo que se encuentra en el sitio web y que informa a los buscadores sobre las páginas nuevas o los cambios de la web.
- El dominio: se deberá elegir aquel que sea más acorde a la filosofía de la empresa y la marca, que haga referencia el producto o servicio, además de que incluya las principales palabras claves para las que queremos posicionarnos.

¹¹ Los resultados de una búsqueda pueden ser orgánicos o de pago (SEM). Los resultados orgánicos son aquellos resultados “naturales” en función de la relevancia que más se acerquen a la consulta de búsqueda del usuario. Los resultados inorgánicos son webs patrocinadas o de pago mostrados en formato de anuncios.

¹² <https://support.google.com/webmasters/answer/7451184?hl=es>

- URL semánticas: URLs “amigables”, sencillas y que sugieran el contenido de la página.
- Un archivo "robots.txt" que indica a los buscadores si pueden acceder a determinadas partes de un sitio web y, por tanto, rastrearlas.
- Títulos de las páginas y artículos: títulos únicos, no deben superar los 70 caracteres. Evitar utilizar un mismo título en todas las páginas de un sitio web o en un gran número de ellas.
- Descripciones: La metaetiqueta de descripción de una página ofrece a Google y a otros buscadores un resumen del contenido de la página. Es importante poner descripciones en todas las páginas y artículos sin repetir excesivamente contenido. Hay que evitar utilizar una única metaetiqueta de descripción en todas las páginas de un sitio web o en un gran número de ellas. Las metaetiquetas de descripción deben estar colocadas dentro del elemento <head> de un documento HTML. Estas descripciones influyen se muestran en los fragmentos de vista previa de una página determinada y son una herramienta indispensable de para lograr un aumento de “clics”.
- Encabezados: instrucciones dentro del código de programación que rastrearán los robots de los buscadores. Si queremos indicar a los buscadores cuál es el tema principal de nuestra página deberemos jerarquizar (a través de etiquetas, H1, H2, H3...) el contenido dentro de dicha página web.
- Añadir marcado de datos estructurados: Los datos estructurados son código que se puede añadir a las páginas de un sitio web para describir su contenido y que así los buscadores puedan comprender mejor lo que contienen. Además de utilizar el marcado de datos estructurados para mostrar resultados enriquecidos, es posible que se usen para servir resultados relevantes en otros formatos, como ubicación, horario, eventos, etc.
- Palabras clave: entre 1 y 3, elegir las que mejor se adapten al tema, así como otras relacionadas con ese mismo tema que puedan ser de interés. Google tiene en cuenta sinónimos y raíces. Es importante incluir las palabras clave en la descripción del sitio web y cuanto más arriba en la página mejor. La etiqueta HTML <description> es la encargada de identificar esta descripción. La descripción óptima debe estar entre 150-160 caracteres y intentando que sean únicas y no se repitan. Contenido único y de calidad, con un aspecto visual limpio, claro y conciso.
- Buen diseño: que atraiga a los visitantes, optimizada para tablets y smartphones, que se cargue rápidamente y facilite la navegación por ella.

Un resumen de estos elementos lo podemos observar en la siguiente imagen:



Imagen 3: Elementos del SEO On Page (Santamaría, 2017)

- *SEO externo*: también conocido como *SEO Off Site* o *SEO Off Page*. “Son todas aquellas acciones que realizamos para aumentar la presencia de nuestra web en la Red” (Blog Bitmarketing, 2013). El SEO externo busca aumentar los enlaces externos que apuntan a una página web con el fin de aumentar su popularidad y notoriedad y así mejorar su posicionamiento orgánico en las páginas de resultados de búsqueda de Google.

El PageRank es una métrica utilizada por Google para valorar entre otras cosas la relevancia de una web. Cuando una web es enlazada por otra, Google interpreta que esta es relevante, que debe confiar en ella. El valor del Page Rank de un sitio web viene determinado por la cantidad y la calidad de los enlaces que apuntan a la misma, la cantidad de dominios únicos o de referencia, la antigüedad de los mismos, su tráfico, etc.

El SEO off page también apunta como beneficios mayor visibilidad de la marca, publicidad, incremento del tráfico web o mejor posicionamiento para los términos de búsqueda. Entre los elementos que influyen vemos:



Imagen 4: Elementos del SEO Off Page (Santamaría, 2017)

Una propuesta integradora de los dos tipos de SEO es la diseñada por Search Engine Land que elaboró una *Tabla Periódica de Factores de Éxito de SEO*¹³. Una tabla que pondera cada uno de los elementos de SEO y su grado de afectación en las páginas de resultados de motores de búsqueda como Google.

En esta cuarta edición, propone como factores de éxito, aquellos factores relacionados con el móvil, las respuestas directas y el aumento de la velocidad. Los factores relacionados con el historial de búsqueda han disminuido su importancia.

¹³ Fuente: <https://searchengineland.com/seotable>

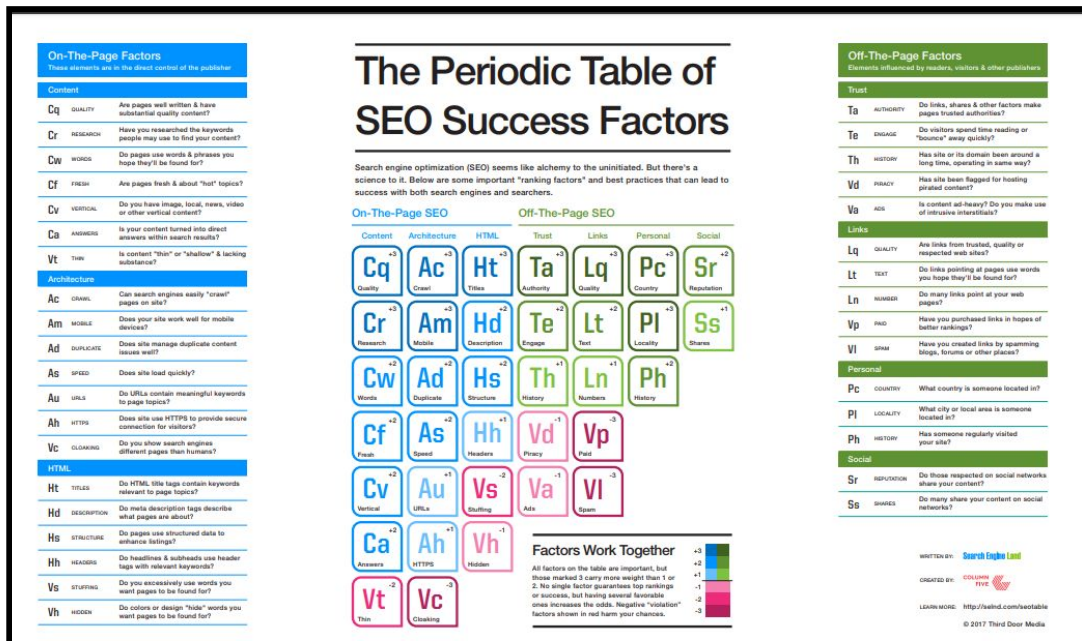


Imagen 5. Tabla Periódica de Factores de Éxito de SEO

2.3.2. SEO en repositorios

Los repositorios digitales de todo tipo enfrentan un desafío común: que los usuarios interesados encuentren su contenido en un mar lleno de información en Internet. Las bibliotecas invierten cada año en la creación, desarrollo y mantenimiento de sitios web y repositorios digitales, pero la optimización para los motores de búsqueda es a menudo una implementación tardía, que hace que la biblioteca digital se use una fracción de lo que podría ser. Muchos repositorios no han disfrutado de mantenimiento más allá del establecimiento del repositorio en sí y el desarrollo continuo de su colección académica.

El SEO merece una atención especial desde las direcciones de bibliotecas debido a su potencial para ayudar a estas a llegar a más usuarios y alinear el SEO con las prioridades institucionales e integrarlo en el plan estratégico.

Cualquier persona que haya analizado el tráfico de su repositorio (por ejemplo, utilizando Google Analytics o similar) notará que un número significativo de visitantes llega a través de un motor de búsqueda como Google o Yahoo!. Si queremos ofrecer la difusión más amplia posible y a la vez incrementar al máximo su visibilidad sería útil obtener una mejor comprensión de los factores que influyen en los motores de búsqueda y que contribuyen a apoyar la capacidad de descubrimiento del contenido alojado en los repositorios institucionales.

El trabajo SEO en repositorios va encaminado a visibilizar y mejorar el posicionamiento de estos en los buscadores. Trabajos que deben ir en la línea de garantizar que los rastreadores de los motores de búsqueda puedan navegar a través de sitios web y repositorios sin encontrarse con barreras de diseño, metadatos y sistemas que tengan un impacto negativo en los índices o clasificaciones en las SERP. Los sitios web lentos, el uso excesivo de imágenes y otros elementos multimedia que aumentan el tiempo de carga, enlaces rotos y contenido de mala calidad o metadatos redundantes son factores que pueden contribuir a la decisión de un motor de búsqueda de enviar a sus usuarios a otros sitios web (Arlitsch, 2013).

Un estudio sobre la invisibilidad de los repositorios institucionales (Arlitsch, 2012) enumeraba como problemas técnicos más frecuentes encontrados por los rastreadores los siguientes elementos:

- Conflictos entre sitemaps y archivos robots.txt.
- Tiempo de respuesta lento del servidor.
- Enlaces rotos o falta de redirecciones apropiadas.
- Laberintos creados por el software de repositorio, incluidos los conjuntos de marcos y JavaScripts mal implementados, así como múltiples URL para el mismo objeto.
- Aplicación deficiente de metadatos, incluida la reutilización de los mismos términos de metadatos para múltiples objetos.
- Esquemas de metadatos considerados inaceptables por el motor de búsqueda específico.

Sin duda, la visibilidad en Google desempeña un papel importante en la maximización del acceso al contenido alojado en repositorios institucionales. Por lo tanto, existe una imperiosa necesidad de obtener una mejor comprensión de las características SEO de los repositorios actuales para identificar ejemplos de mejores prácticas y enfoques defectuosos (Kelly, 2013).

En 2016, la Universidad de Strathclyde proyectó una serie de mejoras técnicas en materia de SEO en el repositorio Strathprints y supervisó su impacto durante 2016/2017 (Macgregor, 2017). Estas mejoras se concretaron en las siguientes actuaciones:

- Interfaz de usuario actualizada. Muchos repositorios continúan con bajos niveles de usabilidad. La evaluación heurística¹⁴ de la interfaz de usuario de Strathprints se realizó a principios de 2016 con la finalidad de mejorar la facilidad de uso y la participación del usuario.
- Sistema responsivo y adaptado a móviles. Que permita una lectura sea agradable, aportando una buena experiencia al usuario. El uso de Internet en el móvil está en alza, y cada vez Google concede mayor importancia a las SERPS móviles.
- Mejoras en la compatibilidad con Core Recommender y AltMetric API, herramientas que remiten a los usuarios a contenido adicional de Strathprints y promueven las interacciones de los usuarios.
- Implementación de una nomenclatura estándar para los archivos, fáciles de rastrear y mejoras en la "amigabilidad" del motor de búsqueda. Si se facilita que los nombres de archivos sean de contenido más semántico, el motor de búsqueda tendrá más facilidad para mostrarlo en su lista de resultados.
- Limpieza gradual de enlaces rotos dentro de Strathprints, lo que mejora la "salud del contenido" del sitio.
- Mejoras en la integración con herramientas sociales, alcanzando un crecimiento en las interacciones sociales resultado de los Tweets sobre nuevos contenido de Strathprints.
- Mejorar la descripción de los objetos digitales del repositorio con metadatos. Se implementó una configuración de "conector-lite" para lograr que Strathprints se convierta en un destino de texto completo para usuarios y máquinas por igual.

La Web semántica representa el nexo del desarrollo tecnológico y los datos estructurados, ofreciendo una dimensión adicional al contenido web para mejorar la capacidad de búsqueda. Su objetivo es reducir la ambigüedad del lenguaje y proporciona la estructura para entregar información más rica, precisa y relevante al usuario en los resultados de búsqueda (Arlitsch, 2014).

Todos aquellos elementos que pueden integrarse con la web semántica son elementos a tener en cuenta a la hora de hacer implementaciones en pro del posicionamiento pero sobre todo de la visibilidad. Un ejemplo claro es la estructura de las URL. La estructura de direcciones URL de un sitio debe ser lo más sencilla posible. Es importante organizar el contenido de modo que las URL se construyan de forma lógica e inteligible para los

¹⁴ Ver glosario

usuarios (Google. Cómo simplificar la estructura de direcciones URL). Es aconsejable utilizar palabras fáciles de leer en lugar de largos números de identificación. Por ejemplo, si se busca información sobre viajes, una URL como

<https://www.skyscanner.es/noticias/20-destinos-a-los-que-viajar-recomendados-por-los-usuarios-de-skyscanner>

te ayuda a decidir hacer clic en el enlace. En cambio, una URL como

http://www.example.com/index.php?id_sezione=360&sid=3a5ebc944f41daa6f89f730f1

resulta mucho menos interesante para los usuarios.

Otro elemento que influye en las páginas de resultados son los *rich snippets* o *resultados enriquecidos*. Estos fragmentos permiten mostrar más datos al usuario, con diferentes elementos visuales y enriquecer las páginas de resultados de búsqueda. Algunas aplicaciones de estos “fragmentos enriquecidos” son *Rich snippets de opiniones, de noticias, de eventos, de recetas, de ubicación con mapa...* Estos fragmentos son más atractivos en cuanto a CTR, pero su aplicabilidad en repositorios es prácticamente nula al no existir formatos específicos por el momento.

Kelly (2012) compartió una encuesta sobre “*la proporción de tráfico que llega a los repositorios institucionales de Google*”. El objetivo era identificar patrones de uso interesantes que puedan ayudar a mejorar la efectividad de los repositorios institucionales. Un elemento a tener en cuenta es conocer el modo en que los usuarios buscan información:

- No solo a nivel de herramienta (catálogos, bases de datos, etc) -como comentamos en el capítulo anterior- sino también qué tipos de dispositivos usan. Sin duda el tráfico de búsqueda móvil ha superado el tráfico de escritorio en todo el mundo (Sullivan, 2018).
- Con el crecimiento de los asistentes digitales activados por voz, más personas están haciendo consultas de voz. El futuro pasa por la eliminación progresiva de las teclas a la hora de realizar las búsquedas. Las búsquedas se hacen de una forma diferente, las keywords son más específicas, y el tono de las mismas más coloquial.
- Un diseño responsivo permitirá que el contenido se vea bien en los dispositivos móviles y que a su vez ayude a los motores de búsqueda a decidir qué tipo de

consultas debe clasificar. Gracias a un diseño web responsive el contenido se visualizará de forma correcta y sin problemas, con independencia del tipo de pantalla que se utiliza.

Algunas soluciones SEO son de naturaleza técnica, pero todos los aspectos de SEO requieren comunicación, administración y coordinación para garantizar que las personas trabajen juntas para lograr objetivos comunes. Incluir metas y objetivos de SEO en el plan estratégico de una biblioteca eleva su importancia y crea una mejor posibilidad de que las técnicas de SEO se apliquen de manera efectiva y sus resultados se midan con precisión. Conducir el SEO de manera estratégica ayudará a garantizar su éxito y establecerá metas y objetivos que permitirán a los administradores recopilar estadísticas de uso precisas para que los informes se conviertan en una parte integral del proceso de evaluación.

2.3.3. Google Scholar

En 2004, nace un buscador de Google especializado en buscar e identificar material bibliográfico publicado de carácter científico y académico: Google Scholar. Google Scholar (GS) o también Google Académico fue concebido principalmente como una herramienta de descubrimiento para la información académica. En él podemos encontrar artículos, tesis, libros, resúmenes y opiniones judiciales, precedentes de editores académicos, sociedades profesionales, repositorios en línea, universidades y otros sitios web. Este tipo de información la podemos encontrar en formato de:

1. Referencias bibliográficas con resúmenes
2. Textos completos en formato PDF, HTML, DOC...
3. Recuento de citas a los trabajos científicos

Google Scholar se alimenta de información procedente de diversas fuentes: editoriales universitarias, asociaciones profesionales, repositorios. Utiliza un software automatizado, conocido como "robots" o "rastreadores", para obtener sus archivos con el objetivo de incluirlos en los resultados de búsqueda. En este sentido, funciona de manera similar a la búsqueda regular de Google.

Google Académico ordena los resultados de búsqueda en función de su relevancia, teniendo en cuenta el contenido del texto completo de cada artículo, el autor, dónde fue publicado y con qué asiduidad ha sido citado en otras fuentes especializadas. Aunque se

desconocen los componentes exactos de este algoritmo y el peso de cada variable, por razones de propiedad industrial, spam y posibles malas praxis, varios estudios empíricos han demostrado que la cantidad de citas recibidas por un documento es uno de los factores clave de clasificación (Beel y Gipp, 2009; Martín-Martín et al., 2017). Gracias a los filtros los resultados pueden aparecer ordenados por fecha, por relevancia, en el idioma que se seleccione, que busque citas y/o patentes.

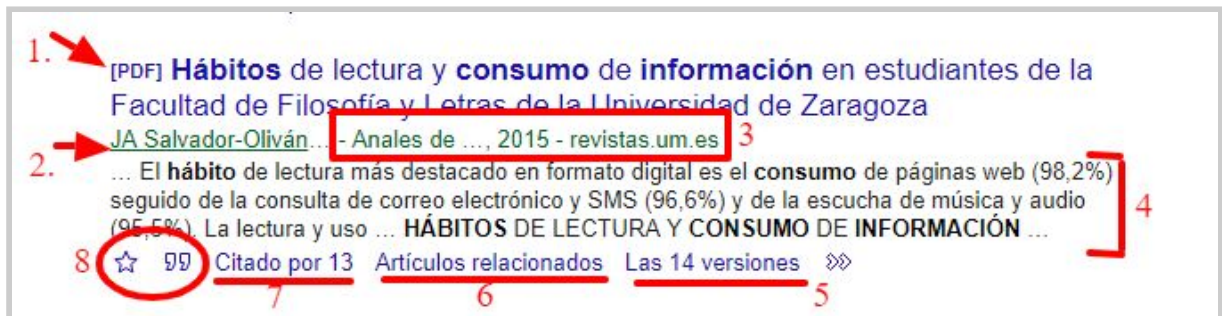


Imagen 6. Composición de un snippet de Google Scholar

1. Título del documento y tipo de registro o archivo.
2. Autor/es.
3. Fuente del documento.
4. Parte del texto (relacionado con nuestra búsqueda).
5. Número de versiones de las que se dispone.
6. Artículos relacionados con el documento.
7. Número de citas recibidas.
8. Posibilidad de guardar el registro u obtener la referencia bibliográfica.

Google Scholar entre otras funcionalidades y posibilidades permite:

- Crear alertas a partir de una búsqueda.
- Enlazar a la referencia existente en la Web o al documento íntegro, siempre que sea posible, además de presentar un listado de varias versiones por fuente cuando sea posible.
- Acceder a textos completos de pago depende de los derechos que se tengan en el lugar en que se consulta. Muchas bibliotecas universitarias permiten mediante el acceso remoto al campus que Google Scholar indique qué textos están disponibles para sus usuarios.

- Facilitar información de las citas recibidas por los documentos y tiene dos herramientas de análisis métrico: Mis Citas (para autores) que permite la creación de perfiles personales, privados o públicos con las citas recibidas por los documentos propios) y Estadísticas (rankings de revistas y repositorios por idiomas).
- Realizar una recopilación personal de referencias desde el espacio “Mi biblioteca”

Entre las críticas que recibe el buscador están las diferencias en la cobertura, calidad y capacidad para reemplazar las bases de datos controladas como fuente de literatura científica confiable, sobre todo, porque no admite descargas de datos y, por lo tanto, es difícil de usar como única fuente bibliométrica. Google Scholar carece de control de calidad y pautas claras de indexación y aunque ha ampliado significativamente su cobertura a través de los años, sin embargo la calidad de los recursos indexados y la política general aún no se conocen. En cuanto a las citas y métricas, se debe tener precaución cuando se confía en GS porque estas se pueden manipular fácilmente y su calidad de indexación sigue siendo un desafío (Halevi, 2017).

El alcance de Google Scholar es un misterio y una fuente de especulación. Uno de los últimos estudios sobre Google Scholar, apuntaba que este tenía un tamaño estimado de unos 389 millones de registros (Gusenbauer, 2019) posicionándolo como el motor de búsqueda académica más completo. Este motor de búsqueda, a diferencia de las bases de datos tradicionales, indexa automáticamente la información de la web académica. Su facilidad de uso, junto con su amplia cobertura y rápida velocidad de indexación, la han convertido en la primera herramienta a la que recurren investigadores de diferentes disciplinas cuando necesitan realizar una búsqueda bibliográfica.

Un estudio en Reino Unido (Researchers of tomorrow, 2012) preguntaba sobre las principales fuentes utilizadas para encontrar recursos y destacaba que, del total de la muestra de la encuesta, el 30% utilizó Google o Google Scholar como su fuente principal para encontrar la información de investigación que buscaban.

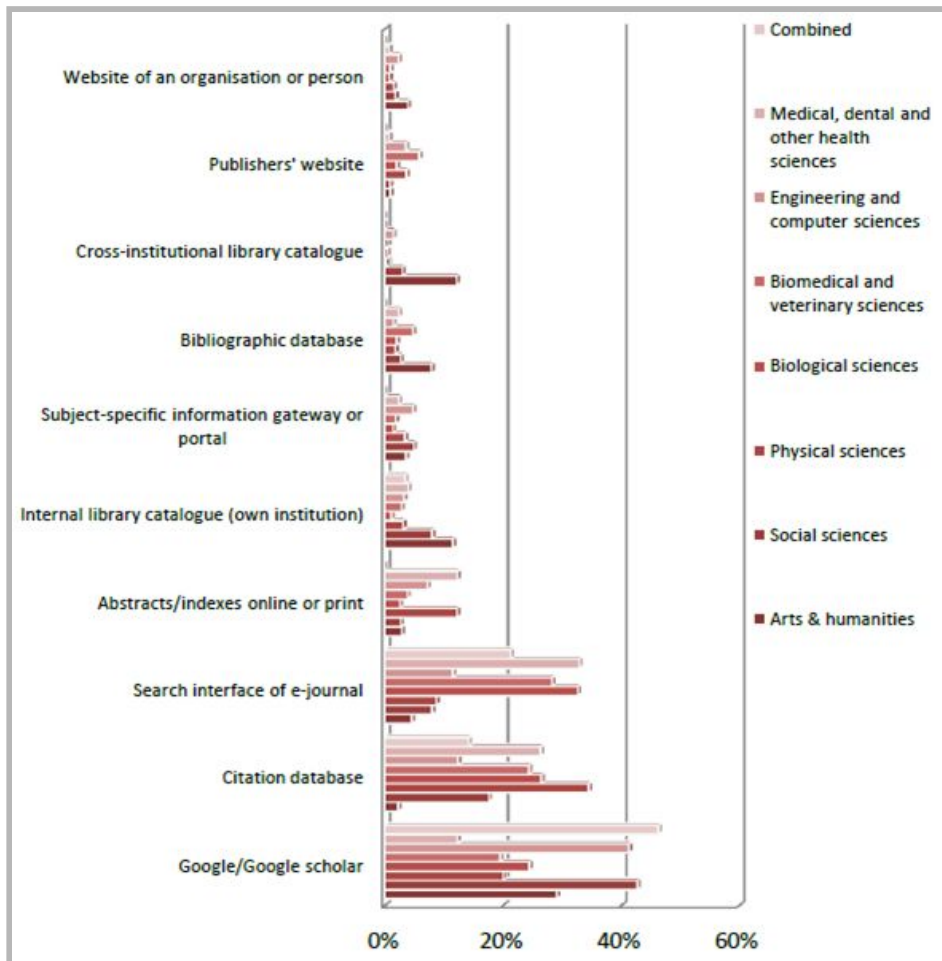


Imagen 7: Principales fuentes utilizadas para encontrar recursos por disciplina de la asignatura.

En un ámbito más cercano, la tendencia es la misma. Los académicos complementan sus bases de datos bibliográficas con motores de búsqueda en Internet, como Google y Google Scholar (Ollé y Borrego, 2010). Un estudio posterior sobre el “Comportamiento informativo del profesorado de las universidades catalanas” (Borrego, 2014; Borrego y Anglada, 2016) planteaba dos preguntas sobre las estrategias empleadas por el profesorado universitario para buscar bibliografía científica. En primer lugar, se pedía a los informados que eligieran la fuente más habitual a la hora de iniciar una búsqueda bibliográfica. De las cuatro opciones presentadas, la mitad de los encuestados (50%) se dirigía a una base de datos especializada. La segunda opción eran los motor de búsqueda en Internet (30%) y, la tercera el catálogo de la biblioteca (17%). La opción de dirigirse a la biblioteca físicamente era totalmente residual (3%).

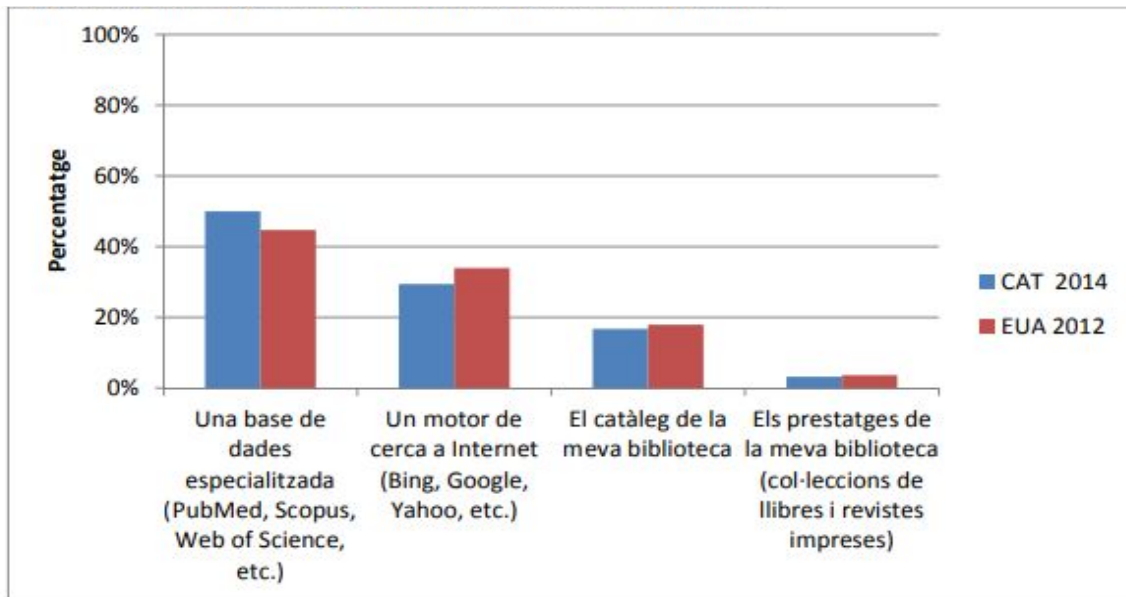


Imagen 8. Fuentes empleadas para iniciar una búsqueda bibliográfica (Borrego, 2014)

Google Scholar se ha posicionado en los últimos tiempos como el mejor motor de búsqueda de facto disponible para el contenido de RI (Arlitsch y O'Brien, 2012). Según el último informe Ithaka S+R US Faculty Survey publicado en 2019 el incremento de uso de este buscador se ha incrementado pasando del 23 % en 2015 al 33 % en 2018 por parte de científicos y pasando del 33% al 47% en el caso de investigadores sociales (Blankstein, 2019).

Google y Google Scholar utilizan diferentes formas de indexar su información. GS tiene sus propios rastreadores que visitan repositorios y sitios de editores, entre otros, para recopilar contenido apropiado para su índice. Google Scholar es capaz de recolectar información de un repositorio siempre y cuando este cumpla los siguientes requisitos¹⁵:

- Si es un repositorio de universidad o institucional, se recomienda utilizar la última versión de Eprints (eprints.org), Digital Commons (digitalcommons.bepress.com) o DSpace (dspace.org).
- Configurar el repositorio para exportar datos bibliográficos en etiquetas HTML "<meta>".
- Google Scholar no admite OAI-PMH, y sugiere utilizar las etiquetas Dublin Core como último recurso. Las etiquetas Dublin Core (por ejemplo, DC.title) funcionan mal

¹⁵ Fuente: <https://scholar.google.com/intl/en/scholar/inclusion.html#overview>

para los artículos de revistas porque Dublin Core no tiene campos inequívocos para el título de la revista, el volumen, el número y los números de página.

- Formato de archivo: los archivos deben estar en formato HTML o PDF. Los archivos en PDF deben permitir la búsqueda de texto y poder buscar y encontrar palabras en el documento utilizando Adobe Acrobat Reader. Cada archivo no debe exceder los 5MB de tamaño.
- Permitir la indexación por los robots de búsqueda de Google.
- Los rastreadores automáticos deben poder descubrir y recuperar las URL de todos sus artículos. La interfaz de navegación es necesaria para que los robots de búsqueda descubran las URL de sus artículos. Se recomienda que se pueda acceder a la URL de cada artículo desde la página de inicio siguiendo como máximo diez enlaces HTML simples.
- El sitio web no debe exigir a los usuarios (o robots de búsqueda) que inicien sesión, instalen software especial, acepten exenciones de responsabilidad, rechacen publicidades emergentes o intersticiales, hagan clic en los enlaces o botones, o se desplacen por la página antes de que puedan leer el resumen completo del documento.
- Dado que Google remite a los usuarios a su sitio web para leer los documentos, sus páginas web deben estar disponibles para usuarios y rastreadores en todo momento. Los robots de búsqueda visitarán la web periódicamente para recoger las actualizaciones, para asegurarse de que sus URL aún estén disponibles.
- En el caso de que GS no indexe automáticamente el repositorio es posible solicitar el alta manual del repositorio.

Google Scholar ha ampliado significativamente su cobertura a través de los años, sin embargo, son muchos los repositorios con problemas de indexación. Arlitsch y O'Brien (2012) examinaron la integración de varios repositorios en GS. El estudio puso de manifiesto que la indexación de estos repositorios en GS no era buena. Encontraron que las principales causas apuntaban al esquema de metadatos utilizado y las características de la arquitectura de información y navegabilidad, que no ayudaban al robot de búsqueda a realizar la indexación. Un estudio similar examinó la cobertura de las revistas de Acceso Abierto indexadas en los repositorios de América Latina, GS y Google mostrando bajas tasas de indexación (Orduña-Malea y López-Cózar, 2015).

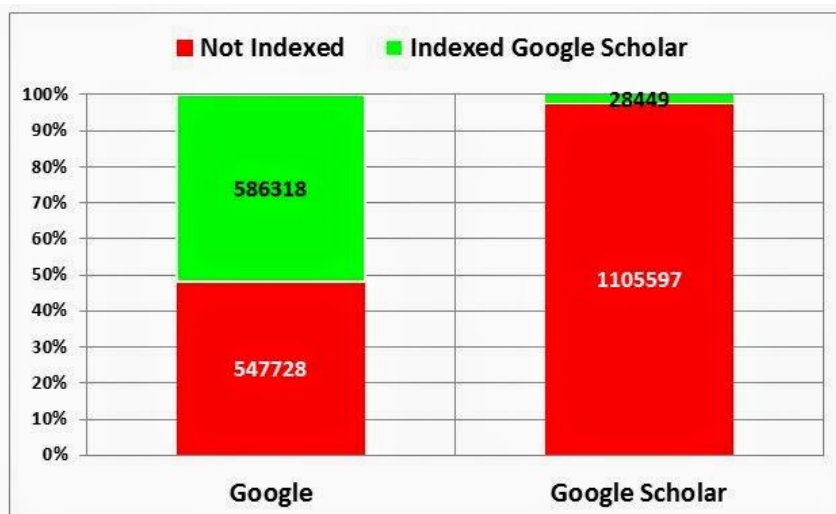


Imagen 9. Porcentaje de documentos de los 127 repositorios latinoamericanos indexados en Google y Google Scholar) (Orduña-Malea y Delgado López-Cózar, 2014)

Los bajos índices de indexación de los repositorios en Google Scholar influyen mucho en la visibilidad e impacto de la investigación depositada, limitando la capacidad de los Repositorios Institucionales para desempeñar un papel más importante en análisis bibliométricos como las tasas de citas de Google Scholar o la evaluación académica. De ahí la importancia de mejorar la indexación de los repositorios en esta herramienta. Los repositorios que utilizan los esquemas de metadatos recomendados por GS y los expresan en metaetiquetas HTML experimentan índices de indexación significativamente más altos (Arlitsch y O'Brien, 2012).

2.4 Schema.org

2.4.1 Definición

Internet ha cambiado nuestra forma de comunicarnos y la web, como una de sus máximas expresiones, sigue evolucionando para hacer que la información que presenta sea más fácilmente legible para las aplicaciones informáticas y por ende poder aproximar la información con mayor precisión por parte de los buscadores. Esta transformación en la forma de describir la información es el paso adelante que nos ofrece la *Web Semántica*.

Junto con el crecimiento de la Web Semántica, aparecieron un conjunto de estándares, tecnologías y herramientas necesarias para la publicación de datos legibles e interpretables por aplicaciones informáticas.

A continuación vemos un ejemplo de cómo ven los sistemas informáticos la información (izquierda) y como la ven las personas (derecha).

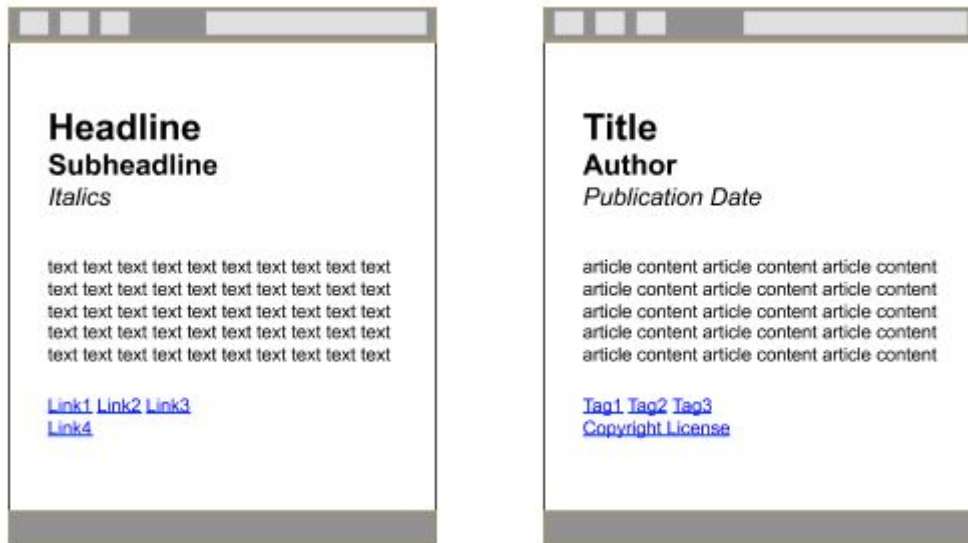


Imagen 10. Visualización de la información por parte de máquinas y de personas ¹⁶

En 2011, Google, Yahoo, Bing y Yandex presentan Schema.org, un proyecto basado en la adición de información de carácter semántico al contenido para mejorar las búsquedas y el posicionamiento en la web. El propósito de Schema.org es el de incluir metadatos dentro de los propios contenidos web haciendo que una página web sea un objeto procesable para la extracción de información semántica.

Schema.org es un conjunto de vocabularios organizados jerárquicamente e interrelacionados que definen una serie de propiedades con las que describir el contenido de diferentes tipos de sitios web (Sulé, 2015). Estas propiedades se insertan dentro del código HTML, habitualmente en forma de microdatos, RDFa o JSON-LD, *semantizando* sus contenidos y permitiendo una mejora en el proceso de recuperación de la información. La flexibilidad y adaptación de estos vocabularios se basa en la multiplicidad de vocabularios específicos.

Algunos de los vocabularios disponibles son:

- Creative works: Article, Book, Digitaldocument, MusicRecording, photography...
- Embedded non-text objects: AudioObject, ImageObject, VideoObject

¹⁶ Fuente: <https://www.w3.org/TR/rdfa-primer/>

- Event
- Organization
- Person
- Place, LocalBusiness, Restaurant...
- Product, Offer, AggregateOffer
- Review, AggregateRating

HTML5 permite describir con más expresividad las páginas y documentos de la Web semántica. Con una sintaxis de Microdatos, los editores web pueden colocar microdatos enriquecidos directamente en su web, facilitando la comprensión a los rastreadores web.

La implementación de un código de datos estructurados como el de Schema.org puede hacer que determinados elementos del contenido web sean más interpretables por Google y otros buscadores, y aparezcan de una manera destacada y más efectiva en sus resultados de búsqueda.

Schema.org propone una serie de clases, atributos y propiedades para crear los datos estructurados de los objetos que queremos definir. El enriquecimiento que propone Schema.org va más allá ofreciendo no solo etiquetas sino un vocabulario estructurado y definido con propiedades y tipologías. Esta forma de estructurar la información facilita a los motores de búsqueda entender y clasificar la información haciendo que los resultados de las búsquedas sean cada vez más precisos y acertados.

Las clases o tipos de objeto son la primera definición a asignar al objeto a describir. Pueden ser: personas, lugares, productos, trabajos creativos, eventos, acciones y cosas intangibles como ofertas o servicios. Cada tipo se puede describir con “propiedades” asociadas a él. Schema.org nos indica el tipo o tipos de datos que puede tener cada propiedad. Los tipos de datos que puede contener una propiedad pueden ser un valor, una enumeración o una entidad. El vocabulario básico de Schema.org consta actualmente de 598 tipos, 862 propiedades y 114 valores de enumeración¹⁷.

Al tipo de documento que se describe, Schema.org lo llama “type” o “itemtype”. Es la definición de lo que es. Y cada “type” a su vez tiene unas propiedades que proporcionan una información específica sobre un aspecto predefinido del elemento que se describe.

¹⁷ Fuente: <https://schema.org/docs/schemas.html>

Extracted structured data

Item	
type:	http://schema.org/article
property:	
name:	All About Schema.org

Imagen 11. Ejemplo tipos y propiedades de Schema.org¹⁸

Cada “type” tiene unas propiedades determinadas y definidas asociadas a él. Debido a que la estructura de schema.org es jerárquica, cada subtipo hereda las propiedades de su tipo principal. Estas propiedades pueden ser de tipo: texto, URL, acción, imagenObjeto, valor...

A continuación, se muestra un ejemplo de etiquetado para un libro, a partir del vocabulario Book de Schema.org:

```
<div itemprop="mainEntity" itemscope
  itemtype="http://schema.org/Book">
  
  <span itemprop="name">The Catcher in the Rye</span> -
  <link itemprop="bookFormat" href="http://schema.org/Paperback">Mass
  Market Paperback
  by <a itemprop="author" href="/author/jd_salinger.html">J.D.
  Salinger</a>
  <link itemprop="availability" href="http://schema.org/InStock">In
  Stock
</div>
Product details
<span itemprop="numberOfPages">224</span> pages
Publisher: <span itemprop="publisher">Little, Brown, and
Company</span> -
<meta itemprop="datePublished" content="1991-05-01">May 1, 1991
Language: <span itemprop="inLanguage">English</span>
```

¹⁸ Fuente: <http://www.seoskeptic.com/basic-vocabulary-for-schema-org-and-structured-data/>

```
ISBN-10: <span itemprop="isbn">0316769487</span>
</div>
```

Este ejemplo aplica el estándar de metadatos, e incluye entre otras propiedades:

- `itemtype="http://schema.org/Book"`
- `itemprop="name"`: Título del libro.
- `itemprop="author"`: Autor.
- `itemprop="numberOfPages"`: número de páginas.
- `itemprop="datePublished"`: fecha de publicación.
- `itemprop="availability"`: disponibilidad del libro.
- `itemprop="inLanguage"`: idioma del libro.
- `itemprop="image"`: imagen de la portada del libro.

García-Marco (2013) plantea que “los esquemas propuestos en Schema.org son microformatos o microplantillas de catalogación para diversos tipos de información que se expresan en RDF Schema”. Para García-Marco “La Web es otro ejemplo de redescubrimiento de los principios clásicos del tratamiento de la información”. Para él, Schema.org es “un paso más hacia la “cataloguización” (de los grandes motores de búsqueda) forzados por la pérdida de relevancia y exhaustividad que el crecimiento exponencial de sus bases de datos ha provocado”.

Aunque la lista de “@type” cada vez es más amplia, a continuación se describen algunos de estos atributos de interés para el ámbito del trabajo.

2.4.1.1. Schema: Thing

Es uno de los elementos más genéricos de Schema.org <http://schema.org/Thing>

Dentro de este tipo genérico encontramos otros tipos más específicos como:

- Action
- CreativeWork
- Event
- Organization
- Person
- Place
- Product

Tanto este “@type” como sus más específicos heredan las características o propiedades de este “@type”. Entre las propiedades que pueden ayudar a definirlo encontramos:

- “name” : nombre del objeto.
- “image” : imagen del objeto (puede ser una URL).
- “identifier” : identificador referido al documento. Puede ser un ISBN, una URL o incluso la imagen del objeto.

Un ejemplo de la aplicación de este atributo:

```
<script type="application/ld+json">
{
  "@context": "http://schema.org/",
  "@type": "Thing",
  "name": "Schema.org Ontology",
  "subjectOf": {
    "@type": "Book",
    "name": "The Complete History of Schema.org"
  }
}
</script>
```

2.4.1.2. Schema: Person

Referido a una persona (viva, muerta, ficticia...). Es un tipo dependiente jerárquicamente del atributo Thing (visto anteriormente) <https://schema.org/Person>

Tiene propiedades propias y otras heredadas del superior jerárquico “Thing”. Entre ellas destacamos:

- “name” : nombre de la persona.
- “additionalname” : otros nombres asociados a la persona.
- “affiliation”: Entidad u organización a la cual la persona está afiliada o depende. Puede ser una universidad, una asociación, un departamento, etc.
- “birthdate” or “deathdate” : fechas asociadas a nacimiento y muerte de la persona.

Un ejemplo de codificación:

```
<script type="application/ld+json">
{
  "@context": "http://schema.org/",
  "@type": "Person",
  "name": "Albert Einstein"
  "additionalname":
  "birthname": "1879"
  "deathname": "1955"
}
</script>
```

2.4.1.3. Schema: Organization

Referido a una entidad u organización, Universidad, ONG, etc.

<https://schema.org/Organization>

Este tipo puede describirse con mucha precisión. Tiene más de 60 propiedades, además de otros muchos atributos con los que se puede relacionar. Algunas de las propiedades que se puede describir son:

- "name" : nombre de la entidad.
- "additionalname" : otros nombres asociados a la entidad.
- "address: dirección
- "URL": dirección electrónica de la entidad
- "Departament": Departamento de la entidad relacionada
- "email": correo electrónico

Un ejemplo de este tipo y sus propiedades:

```
<script type="application/ld+json">
{ "@context": "https://schema.org",
  "@type": "Organization",
  "name": "Universitat de València",
  "additionalname" : "UV",
```

```
"url": "https://www.uv.es/",
"address": "Av. Blasco Ibáñez, 13. 46010 València. Espanya".
}
</script>
```

2.4.1.4. Schema: CreativeWork

Creativework define el contenido asociado es un trabajo creativo en general. <https://schema.org/CreativeWork> Este tipo puede asociarse a libros, películas, fotografías, programas de software, etc. También se enmarcan otros “Type” más concretos como Book, ScholarlyArticle y Thesis, que son algunos de los “Types” que objeto de estudio de este trabajo.

Entre las propiedades que pueden definir este atributo están:

- Audience : audiencia prevista, un grupo para el que se creó algo.
- Audio: un objeto de audio incrustado.
- Author: autor del contenido, sea persona u organización. Este atributo puede definirse específicamente mediante su etiqueta propia, aunque se puede usar indistintamente esta también.
- Citation: cita o referencia a otro trabajo.
- DateCreated: fecha en la que se creó el trabajo que se describe
- Keywords: palabras clave o etiquetas utilizadas para describir el contenido. Las entradas múltiples en una lista de palabras clave están generalmente delimitadas por comas.
- License: licencia que se aplica a este contenido, generalmente indicado por URL.

A continuación, se muestra un ejemplo de este tipo y sus propiedades:

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "CreativeWork",
  "author": "Crystal Dynamics",
  "datepublished": "5, march, 2013",
  "image": "videogame.jpg",
  "name": "Tomb Raider"
```

```
}  
</script>
```

2.4.1.5. Schema: Book

Este tipo define las características específicas para los libros. <https://schema.org/Book> Es un tipo específico del genérico "CreativeWork". Entre las propiedades que permite definir están todas las que comparte con su superior jerárquico y algunas específicas, como:

- Author: autor del contenido, sea persona u organización. Este atributo puede definirse específicamente mediante su etiqueta propia, aunque se puede usar indistintamente esta también.
- bookEdition: edición del libro que describimos.
- illustrator: persona que realiza la ilustración del libro que describimos.
- bookFormat: formato del libro que describimos.
- isbn: número ISBN del documento que se describe.
- DateCreated: fecha en la que se creó el trabajo que se describe.
- Keywords: palabras clave o etiquetas utilizadas para describir el contenido. Las entradas múltiples en una lista de palabras clave están generalmente delimitadas por comas.
- numberOfPages: número de páginas del documento.

A continuación se muestra un ejemplo de libro definido:

```
{  
  "@type": "Book",  
  "author": "Belloch, Amparo",  
  "bookFormat": "EBook",  
  "datePublished": "2018-05-01",  
  "inLanguage": "English",  
  "name": "Psychopathology exercise book",  
  "numberOfPages": "1234"  
  "publisher": "McGraw-Hill Interamericana de España"  
}
```

2.4.1.6 Schema: ScholarlyArticle

ScholarlyArticle, es la clase o tipo que nos ayuda a definir los artículos académicos. <https://schema.org/ScholarlyArticle> . Se enmarca dentro del type “Article” que es una clase más genérica que incluye otros tipos de artículos no necesariamente científicos o académicos como artículo de noticias, prensa o informe de investigación. En nuestro objeto de estudio hemos escogido este @type “ScholarlyArticle” por ser más próximo a nuestros tipos de documentos.

Entre las propiedades que pueden definirlo encontramos de propias y de heredadas del superior jerárquico CreativeWork y también de Thing:

- name: nombre-título del artículo.
- pagination: descripción de las páginas del artículo.
- pageEnd: página en la que finaliza el artículo.
- pageStart: página en la que comienza el artículo.
- Author: autor del contenido, sea persona u organización. Este atributo puede definirse específicamente mediante su etiqueta propia, aunque se puede usar indistintamente esta también.
- contributor: persona o entidad colaboradora del artículo.
- ISSN: número ISSN de la publicación del artículo.
- isPartOf: es parte de. Propiedad inversamente proporcional a “hasPart”
- URL: url del documento

Ejemplo de aplicación de esta clase:

```
{
  <script type="application/ld+json">{"@context":"https://schema.org"
  ,"@type":"ScholarlyArticle"
  ,"name":"Computer Animation to teach interpolation"
  ,"author":[{"@type":"Person","name":"Pla i Castells,
  Marta"}, {"@type":"Person","name":"García Fernández,
  Ignacio"}, {"@type":"Person","name":"Gimeno Sancho,
  Jesús"}, {"@type":"Person","name":"Ferrando Palomares, Irene"}]
  ,"keywords":["Matemàtica per a enginyers"
  "Informàtica"]
}
```

```

,"description":"In this work we present a learning experience based
on computer animation and using the perspective of mathematical
modelling. Our goal is to provide the students with a context that
motivates the study of function interpolation. We present a problem
statement that is intended to be solved by means of the Modeling
Cycle. The development of the activity and the strategies identi ed
during the process are presented and discussed."
,"publisher":""
,"datePublished":"2019"
}</script>

```

2.4.1.7 Schema: Thesis

Usamos este tipo para definir principalmente tesis académicas, aunque también lo utilizamos para otros tipos de trabajos de final de estudios como trabajos de final de grado o máster. Este tipo depende jerárquicamente de “CreativeWork” y “Thing” y hereda sus propiedades. <https://schema.org/Thesis>

Algunas propiedades destacables son:

- author: autor de la tesis.
- name: nombre-título de la tesis
- inSupportOf: calificación, candidatura, grado, solicitud que respalda la tesis.
- sourceOrganization: organización a la que pertenece el autor de la tesis.
- keywords: palabras clave o materias utilizadas para describir el contenido.
- URL: url asociada al documento

Un ejemplo de codificación de tesis:

```

{
  "@type": "schema:Thesis",
  "schema:author": "Konieczny, Piotr ",
  "schema:inSupportOf": "Departament de Genètica",
  "schema:name": "A Spinal Muscular Atrophy Reporter System for in
vivo",
  "schema:sourceOrganization": "Universitat de València"
}

```

```
</script>
```

2.4.2 Implementación

Una vez elegido qué clase o @type se van a utilizar y qué propiedades lo van definir, hay que convertir este vocabulario a uno de los formatos de datos estructurados o microdatos que entiendan Google y los demás buscadores.

Schema.org se puede utilizar con muchas codificaciones diferentes, incluyendo RDFa, Microdatos y JSON-LD. Estos vocabularios estructurados relacionan elementos con propiedades, cubren autorías, entidades, relaciones entre entidades y acciones.

Los microdatos se insertan dentro del texto del cuerpo HTML. Para ello es necesario convertir en código el vocabulario de schema elegido así como sus propiedades. Para ellos podemos usar 2 lenguajes de marcado como Microdatos o RDFa, o JSON-LD, este último es el formato actualmente sugerido por Google como mejor opción.

2.4.2.1 Microdatos y RDFa

Tanto los Microdatos como RDFa son una serie de atributos HTML que asocian el contenido de la página con el vocabulario de Schema.org.

RDFa es una sintaxis de marcado de atributos para expresar datos estructurados como Schema.org u otros lenguajes de marcado. RDFa provee un conjunto de atributos que pueden ser usados para el marcado de datos. La sintaxis RDFa proporciona una serie de características básicas y avanzadas que permiten a los autores expresar datos estructurados bastante complejos, como las relaciones entre personas, lugares y eventos en un documento HTML o XML (W3C, 2015).

Los microdatos utilizan los siguientes atributos HTML para hacer el marcado de datos: itemscope, itemtype, itemprop, itemid, itemref.

Un ejemplo de etiquetado con microdatos:

```
{  
<div itemscope itemtype="http://schema.org/Person">  
<h1><span itemprop="name">Edurne Zabala</span></h1>
```

```

<span itemprop="description"> Enfermera </span>
<div itemprop="address" itemscope
itemtype="http://schema.org/PostalAddress">
<span itemprop="streetAddress"> Sants 335</span>
<span itemprop="addressLocality">Barcelona</span>
</div> Teléfono: <span itemprop="telephone">938359030</span>
</div>
}

```

Al usar microdatos, se puede agregar información mejorada adicional a la que ya de por sí permite HTML5, siendo solo visible para los motores de búsqueda. De forma que esta información, una vez procesada, amplía las posibilidades de los buscadores y terceras aplicación en relación con la visualización de los datos.

2.4.2.2 JSON for Linking Data

JSON-LD (JSON for Linking Data) es un método de codificación basado en JSON (JavaScript Object Notation) para serializar datos vinculados. La sintaxis está diseñada para integrarse fácilmente en los sistemas implementados que ya utilizan JSON¹⁹. Por lo general no está incrustado en el diseño HTML de la página en sí, sino como datos sin procesar en el encabezado de una página.

JSON-LD, utiliza sus propias clases y datos que van separadas del código HTML, por lo que no necesita que los datos a marcar estén en el código HTML del contenido de la página web ya que se incluyen directamente en este formato. A diferencia de los Microdatos y RDFa, no añade nuevas etiquetas al contenido HTML, sino que únicamente se necesita crear un script al inicio de la página definiendo en ese bloque los elementos que tendrán carácter semántico, favoreciendo una implementación más sencilla.

Aquí tenemos un ejemplo²⁰ de descripción usando JSON para describir un libro:

```

<script type="application/ld+json">
{
  "@context": "http://schema.org/",

```

¹⁹ Fuente: <https://json-ld.org/spec/>

²⁰ Fuente: <https://schema.org/accessibilitySummary>


```

    "@type": "Book",
    "name": "Some graphic novel",
    "accessMode": ["textual", "visual"],
    "accessModeSufficient": ["textual", "visual"],
    "accessibilitySummary": "Visual elements are not described."
  }
</script>

```

Actualmente Google recomienda usar JSON-LD para implementar los datos estructurados siempre que sea posible (*Comprende cómo funcionan los datos estructurados*).

2.4.3 Aplicabilidad

Hoy en día los datos estructurados son cada vez más útiles y necesarios. La principal función de estos es la de facilitar a los buscadores el entendimiento del sitio web y refinar en sus resultados, sin embargo, si se combina estos datos estructurados con Schema.org, se tiene la posibilidad de tener fragmentos enriquecidos para la web que ayudarán a aumentar la probabilidad de que las personas encuentren la página web y mejorarán la experiencia de usuario de estos, traduciendo esto en una mejora en el posicionamiento web (Sancho-Revilla, 2018).

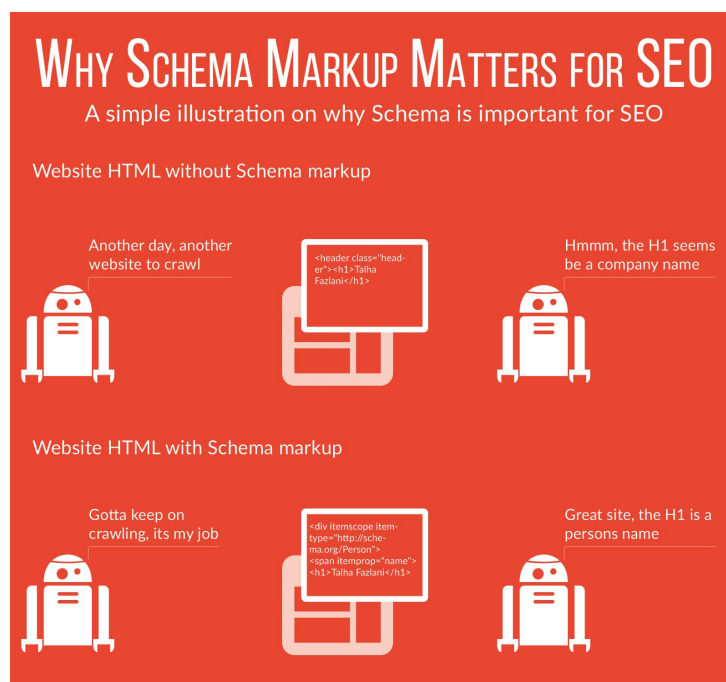


Imagen 12. Porque el marcado de Schema es importante para el SEO²¹.

²¹ Fuente: seopressor.com

Autores como Pastor-Sánchez (2012) apuntan que, “Las implicaciones de este tipo de solución en el funcionamiento de los motores de búsqueda son considerables: puesto que estos funcionan mejor con información estructurada, las búsquedas serían más precisas, las posibilidades de filtrado aumentarían la versatilidad del proceso de recuperación de información, los resultados podrían visualizarse de forma más personalizada a los gustos de usuario, etc.”

El éxito de Schema.org radica en haber conseguido estandarizar y simplificar el proceso de marcar el contenido web y proporcionar beneficios por ello, como con los fragmentos enriquecidos. Los fragmentos enriquecidos, más conocidos como **rich snippets**, son los datos de marcado que aparecen en los resultados de búsqueda.

Una aplicación directa de este marcado la observamos en la Búsqueda de Google, que también usa datos estructurados para habilitar funciones especiales y mejoras de los resultados de la búsqueda.

Alguno ejemplos de resultados de búsquedas con fragmentos enriquecidos:



Imagen 13. Receta rich snippet.



Imagen 14. Libro rich snippet.

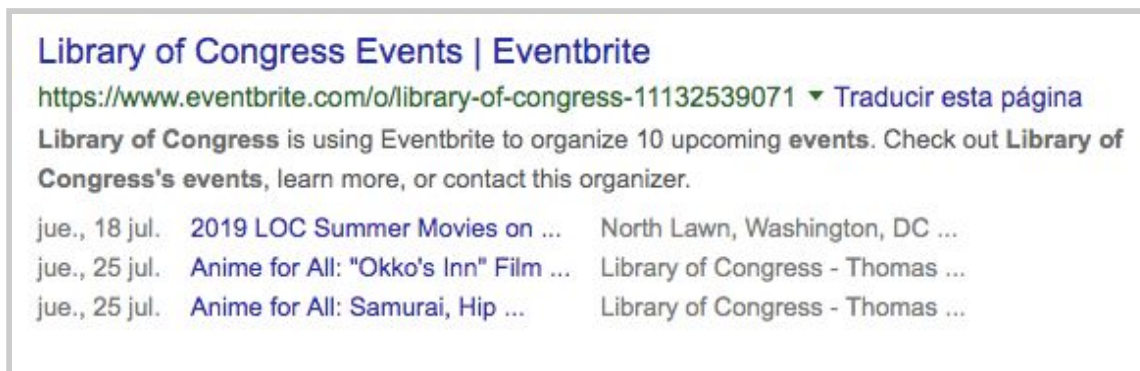


Imagen 15. Evento rich snippet.

Estos fragmentos enriquecidos hacen que el resultado muestre destacados (como las estrellas y valoraciones) en los resultados de búsquedas (Search Engine Optimization) ayudando a captar la atención de los usuarios y aumentando así el número de clics (CTR) y accesos.

Otros ejemplos de implementación las podemos observar en los correos electrónicos que confirman las reservas (restaurante, hotel, aerolínea, etc.), recibos de compra, etc., que han incorporado el marcado Schema.org con detalles de la transacción.



Imagen 16. Reserva vuelo rich snippet²²

Más de 10 millones de sitios utilizan Schema.org así como muchas aplicaciones de Google, Microsoft, Pinterest, IOS 9 y otras, que ya utilizan estos vocabularios para alimentar experiencias ricas y con múltiples posibilidades (Guha, Brickley y Dan Macbeth, 2016).

En el ámbito de las bibliotecas y en concreto repositorios, sin duda los desafíos en la difusión de contenidos abre una oportunidad muy interesante para poder mejorar el acceso e impacto de sus colecciones a través de la web. Uno de los primeros proyectos a nivel bibliográfico que apostó por Schema.org para mejorar su visibilidad fué el de OCLC, que

²² Fuente: <https://support.google.com/webmasters/answer/3069489?hl=es>

desde 2012 ha implementado en las páginas web de los registros bibliográficos de WordCat los datos descriptivos codificados con Schema.org (Sulé, 2015).

2.5 Software para la creación de repositorios

En la última década se han desarrollado una gran cantidad de repositorios. OpenDoar²³ cifra, este agosto de 2019, en 4.242 los Repositorios a nivel mundial. Los softwares para su gestión son varios y de distinta índole, ya que unos potencian más unas funciones que otras. En la siguiente gráfica se pueden observar la variedad de sistemas disponibles para gestionar repositorios y el porcentaje de uso según OpenDoar.

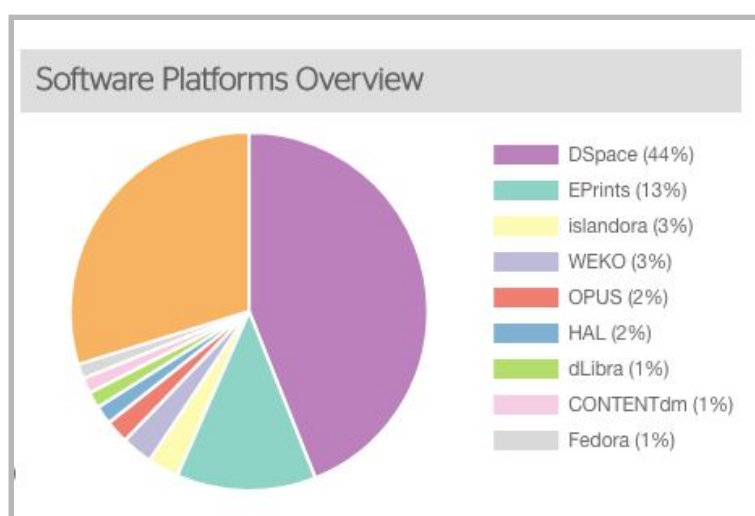


Imagen 17. Porcentaje de softwares, Agosto 2019 (OpenDoar)

Son bastantes los trabajos que abordan comparativas entre ellos y estudio de las diferentes posibilidades y sistemas de trabajo (Castagné, 2013; Masrek, 2012; Bankier, 2014; Pyrounakis, 2014). En un momento como el actual, en que que la Web semántica tiene un peso específico importante a la hora de relacionarnos con Internet, es interesante observar cuál es el grado de aproximación de estos programas a la Web Semántica. Para ello se han seleccionado 4 softwares (Dspace, Eprints, CONTENTdm e Invenio), en base a sus estadísticas de uso (OpenDoar) y a las experiencias que han realizado algunos repositorios con Schema.org.

²³ http://v2.sherpa.ac.uk/view/repository_visualisations/1.html

Los repositorios deben ser capaces de ofrecer la interoperabilidad adecuada para proporcionar acceso a través de los motores de búsqueda. De ahí que los aspectos que se quieren observar en su aproximación a la Web Semántica son los siguientes:

- Versión actual de la plataforma: número de versión de la plataforma relacionada con los datos que se quieren revisar.
- Tipo de software: Si es un sistema de acceso libre o es un software comercial propiedad de alguna empresa privada.
- Búsqueda y descubrimiento: observar si permite indexación de texto completo, navegación gráfica, opciones de navegación personalizables, búsqueda por voz, geolocalización u optimización en los principales motores de búsqueda.
- Interoperabilidad: Protocolos de intercambio de datos que permite utilizar, como por ejemplo OAI PMH, OAI_DC, OAI-ORE, SWORD, SRU / SRW.
- Integration con plataformas de descubrimiento.
- Integración con páginas de perfiles de investigación.
- Integración o descripción con vocabularios semánticos como Schema.org.
- Formato de salida de datos (JSON, XML, Metadatos)
- Search Engine Optimization (SEO): posibilidades del software para mejorar el SEO. Ver si permite generar archivos como robots.txt, Sitemap, si proporciona links atractivos semánticamente.
- Indexación en Google Scholar: optimización para motores de búsqueda especializados como Google Scholar.
- Diseño optimizado para dispositivos móviles: páginas optimizadas para dispositivos móviles que ayuden a garantizar una navegación óptima.
- WCAG (Web Content Accessibility Guidelines): Pautas de accesibilidad y recomendaciones para hacer que el contenido web sea más accesible.
- Informes de uso: informes de metadatos de publicación, uso y descargas, integración de Google Analytics.
- Instituciones de referencia que utilizan el software
- Experiencias de Repositorios con Schema.org

La revisión que presentamos busca observar con especial atención aquellos aspectos relacionados con el SEO y la visibilidad de los repositorios.

	Dspace	Invenio	Eprints	CONTENTdm
Versión actual de la plataforma	6.3	V 3.1.1.	EPrints v3.4.1	6.x CONTENTdm
Tipo de Software	Software libre	Software libre	Software libre	Software comercial
Búsqueda y navegación	Búsqueda de texto libre; búsqueda combinada por campos ; Navegación por índices y comunidades; Indexación de texto completo	Invenio utiliza Elasticsearch como su motor de búsqueda subyacente, Elasticsearch proporciona capacidades de búsqueda y agregación muy potentes y permite realizar por ejemplo consultas geoespaciales.	Búsqueda de texto libre; búsqueda combinada por campos ; Navegación ; Indexación de texto completo	Búsqueda avanzada, flexible y móvil
Interoperabilidad	OAI PMH, OAI_DC, OAI-ORE, SWORD, SRU / SRW.	OAI-PMH. Invenio exporta registros en formato MARCXML y BibTeX y admite fuentes RSS	OAI-PMH; EPrints exporta objetos de datos en muchos formatos, entre ellos METS, MPEG-21 Digital Item Declaration Language y BibTeX. También es compatible con SWORD, Atom y RSS.	OAI-PMH, XML, EAD, Z39.50
Integration con plataformas de descubrimiento	Si	Se desconoce	Si	Si
Integración con páginas de perfiles de investigación.	--	ORCID	ORCID	---
Integración con redes sociales	Permite la inclusión de una barra que	Permite la inclusión de una barra que	Permite la inclusión de una barra que	Permite la inclusión de una barra que

	conecta con redes sociales, por ejemplo "addthis"	conecta con redes sociales, por ejemplo "addthis"	conecta con redes sociales, por ejemplo "addthis"	conecta con redes sociales, por ejemplo "addthis"
Integración o descripción con vocabularios semánticos como Schema.org	No	No	No	No
Personalización de la meta-descripción	No se especifica	No se especifica	Editor de texto enriquecido , que permite añadir texto enriquecido y representarlo como HTML.	No se especifica
Formato de salida de datos (JSON, RDF, Metadatos, XML)	JSON o XML	JSON	JSON o XML	JSON o XML
Search Engine Optimization (SEO)	Sitemap; robots.txt	Sitemaps XML ; robots.txt	Sitemap; robots.txt	Sitemap; robots.txt
URL Semánticas (sin códigos ni números)	No	No	No	No. Aunque todas las páginas de visualización contienen una URL canónica con etiquetas recomendada por los principales motores de búsqueda;
Indexación en Google Scholar	Si	No recomendado	Si	No recomendado
Diseño optimizado para dispositivos móviles	Diseño responsive de las páginas	Diseño responsive de las páginas	Diseño responsive de las páginas	Si, (Mobile-friendly considerations)
WCAG (Web	---	---	---	Si

Content Accessibility Guidelines)				
Informes de uso	SOLR; Google Analytics	COUNTER-compliant statistics.	IRStats; Google Analytics	Resumen incorporado de uso; Google Analytics
Instituciones de referencia que utilizan el software	Dipòsit Digital de la Universitat de Barcelona	UAB (DDD, Dipòsit Digital de Documents de la UAB) ; Zenodo	E-Prints Complutense (UCM)	Memoria digital de Canarias (mdC)
Experiencias de Repositorios con Schema.org	DigitalGeorgetown (Universidad de Georgetown)	Zenodo	Strathprints (Universidad de Strathlyde)	---

De esta comparativa observamos que:

- Los softwares utilizan diferentes motores de búsqueda pero todos permiten la búsqueda por texto libre, búsqueda combinada por campos o facetas, navegación por índices y colecciones y permiten una indexación de texto completo. Destacan dos opciones que van más allá e integran elementos relacionados con la web como poder realizar consultas geoespaciales (Invenio) o dar un peso específico a las consultas desde dispositivos móviles (CONTENTdm) algo muy importante, este último elemento, para buscadores como Google en su última actualización.
- Todos los softwares incluyen el Protocolo OAI-PMH para la recolección de metadatos. Un protocolo que facilita la interoperabilidad de contenidos y permite que recolectores como RECOLECTA, EUROPEANA, CORE, puedan incluir en sus sistemas los documentos de esos repositorios. Un protocolo que a nivel web no tiene tanto significado ya que Google en 2008 retiró su apoyo en pro la importación mediante XML.
- Otro estándar de interoperabilidad utilizado cada vez más por los repositorios digitales es SWORD, un protocolo que permite la transferencia de contenido y que en su versión 3.0 utiliza JSON-LD para la alineación con Linked Data. Los metadatos

descriptivos que indican la información del recurso deben optimizarse utilizando diseños y estrategias en pro de datos abiertos vinculados.

- Elementos de indexación web como el Sitemap, o el archivo robots.txt son fundamentales para que el contenido de los repositorios sea buscable por los buscadores. Los cuatro softwares permiten o disponen de la integración de estos elementos pero solo CONTENTdm va más allá en el trabajo SEO, ofreciendo una URL canónica con etiquetas recomendadas por los principales motores de búsqueda. No obstante ninguno de ellos ofrece URL semánticas ni la posibilidad de personalizar las meta-descripciones.
- Todos ofrecen ya una interfície responsiva que se adapta a diferentes pantallas y dispositivos. CONTENTdm tiene en cuenta además las recomendaciones “mobile-friendly” con el fin de mejorar el ranking en el motor de búsqueda para móviles, reducir los clics hacia el contenido y acelerar el acceso al contenido por parte de los usuarios.
- CONTENTdm propone un acercamiento a las pautas de de accesibilidad WCAG mejorando el acceso a sus colecciones. Un elemento que todavía no incorporan otros sistemas.
- La integración con plataformas de descubrimiento ya es un elemento común para los 4 softwares, no así con Google Scholar, donde la indexación todavía es irregular. Solo Dspace y E-prints aparecen como softwares recomendados por el buscador especializado ya que presentan menos inconvenientes de interoperabilidad con los metadatos y su descripción. También queda pendiente la integración con páginas de perfiles de investigación. Solo Invenio y CONTENTdm empiezan a integrarse con plataformas como ORCID.
- Los 4 sistemas ofrecen informes de uso y la posibilidad de integrar herramientas como Google Analytics para revisar las tendencias de uso de los usuarios desde la web. Invenio propone además el uso de una herramienta diseñada para bibliotecarios y editores (entre otros), con el fin de facilitar el intercambio de estadísticas de uso en relación a estándares y protocolos internacionales.
- Pero sin duda el trabajo todavía por desarrollar es la incorporación de etiquetas semánticas para mejorar la interoperabilidad con los buscadores web. Ninguno de los softwares revisados añade vocabulario semántico a sus descripciones.

Esta comparativa no busca posicionarse por un software en particular, sino observar cual es el grado de aproximación de estos sistemas con la Web y la Web Semántica. Sin duda la

gran cuestión pendiente es cómo mejorar la integración de los repositorios en la web, posicionando mejor y haciendo más accesible su contenido. Una cuestión todavía pendiente si observamos las pocas implementaciones que tienen incorporadas los softwares para repositorios y aunque se trabaja para que estos sean interoperables con otras herramientas, todavía la integración e interoperabilidad con la web tiene un largo camino por hacer.

2.6 Análisis de casos y buenas prácticas

Una década después de la implementación de la Web semántica, su aplicación en bibliotecas digitales y repositorios es escasa, a pesar de los beneficios que ofrecen en el descubrimiento y recuperación de información.

Aun así, encontramos algunas bibliotecas académicas que han explorado el uso de Schema.org en sus repositorios institucionales. La biblioteca de la Universidad de Georgetown, implementó en su repositorio Digital Georgetown²⁴ diferentes vocabularios de Schema.org como estrategia para aumentar su impacto (Pekala, 2018). En el repositorio gestionado con Dspace, se implementó primero Schema.org en una de sus comunidades, extendiendo después su aplicación a todo el repositorio.

El mismo camino recorrieron la Columbia University y su repositorio Academic Commons²⁵ al añadir con formato de microdatos Schema.org (Hilliker, Wacker y Nurnberger, 2013). Se aplicaron, al repositorio, basado en Fedora y Blacklight, los atributos apropiados de schema.org en el código HTML.

En una línea similar, DataMed²⁶, portal que busca indexar y buscar de manera eficiente diversos tipos de conjuntos de datos biomédicos en los repositorios, lanzó en agosto de 2015 una primera versión de la especificación DATS 8, con el lenguaje JSON y en septiembre de 2016, presentaron una segunda versión de los elementos de metadatos adicionales producidos por los Metadatos de Accesibilidad WG7, un conjunto de esquemas JSON y un archivo de contexto con descripción Schema.org (Sansone, 2017).

Fenner (2018) nos presenta un listado con algunos de los repositorios de datos que han implementado o están implementado Schema.org en su arquitectura. Además de la lista de

²⁴ <https://www.library.georgetown.edu/digitalgeorgetown>

²⁵ <https://academiccommons.columbia.edu/>

²⁶ <https://datamed.org/>

repositorios de datos, proporciona información de las propiedades de Schema.org compatibles con estos repositorios de datos, centrándose en las propiedades requeridas y recomendadas en "A data citation roadmap for scholarly data repositories" de Fenner, 2019.

Fenner (2019) proporciona una hoja de ruta para los repositorios de datos académicos. Entre sus recomendaciones, se apunta a que todas las *landing page* de conjuntos de datos deben proporcionar metadatos legibles por las máquinas utilizando el marcado Schema.org en formato JSON-LD y se hace una recomendación expresa a utilizar JSON-LD como formato para representar los metadatos de schema.org. Un dato importante, para nuestro objeto de estudio es el grado de implementación de Schema.org, detectando un total de 32 repositorios en esta línea.

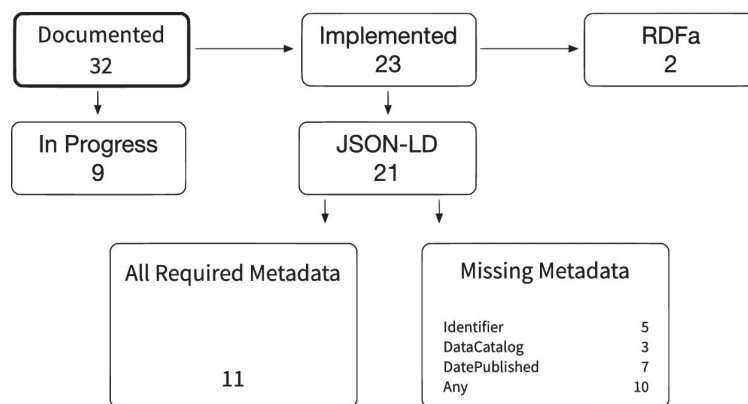


Imagen 18. Estado de implementación de los metadatos de Schema.org en las páginas de destino del repositorio (Fenner, 2019)

Fenner (2019) apunta que los elementos de metadatos que faltan con mayor frecuencia son el identificador y sorprendentemente, la fecha de publicación. También ofrece un mapeo de metadatos en diferentes lenguajes, que considera básicos en repositorios de datos.

Discovery Metadata	Dublin Core	Schema.org	DataCite	DATS
Description	description	description	description	dataType dimension Material..*
Keywords	subject	keywords	subject	keywords
License	license	license	rights	license
Related Dataset**	isPartOf isVersionOf references	isPartOf citation	relatedIdentifier	isPartOf
Related Publication***	bibliographicCitation	citation	relatedIdentifier	publication

Imagen 19. Metadatos de descubrimiento importantes para los repositorios de datos (Fenner, 2019)

Sin duda en la línea de los repositorios de datos, los metadatos semánticos son un elemento que se tiene presente. Proyectos como Bioschemas²⁷ cuyo objetivo es mejorar la capacidad de búsqueda de datos en las ciencias de la vida. Para ello, alienta a las personas que trabajan en el área de “Ciencias de la vida” a usar el marcado Schema.org en sus sitios web para que puedan ser indexados por los motores de búsqueda y otros servicios.

Bioschemas está haciendo dos contribuciones principales:

1. Proponer nuevos tipos y propiedades a Schema.org para permitir la descripción de los recursos de ciencias de la vida.
2. Perfiles sobre los tipos Schema.org que identifican las propiedades esenciales para usar en la descripción de un recurso.

Pero si hay otro proyecto importante, tanto por su cobertura como por su cantidad de registros descriptivos, que ha trabajado en la línea de implementar Schema.org es el portal Europea²⁸. Europea proporciona acceso a más de 54 millones de objetos de patrimonio cultural a través de su portal Europea Collections. Aunque su portal es el principal punto de entrada, la mayoría de los usuarios de Europea llegan a él a través de los motores de búsqueda, especialmente Google. Bajo esta premisa, Europea se propuso definir sus recursos utilizando el vocabulario de Schema.org, con el objetivo de permitir que las organizaciones externas en general, y los motores de búsqueda en particular, entiendan mejor sus datos en sus procesos de rastreo web (Wallis, 2017). El formato de implementación de Schema.org en sus registros fue JSON-LD. Este formato proporciona una carga útil más pequeña, en comparación con RDFa y Microdata; además de que JSON-LD es compatible con muchas herramientas y ya se utilizaba en algunos servicios de API de Europea.

Respecto a la implementación de Schema.org, Europea presenta como @type de sus documentos “CreativeWork” aunque recomienda que siempre que sea posible se debe proporcionar un tipo más preciso. Podemos observar el mapeo planteado por europea (Wallis, 2017) desde el siguiente link

<https://journal.code4lib.org/media/issue36/wallis/MappingFromEDMtoSchema.pdf>

²⁷ <https://bioschemas.org/>

²⁸ <https://www.europeana.eu/portal/es>

Aquí vemos un ejemplo de la descripción propuesta por Europeana:

```
"@graph":[
  {
"@id":
"http://data.europeana.eu/item/2021618/internetserver_Details_kunst_25027",
  "@type": ["CreativeWork", "VisualArtwork"],
  "name": "Mona Lisa | Leonardo da Vinci",
  "creator": ["_p:0", "_p:1"],
  "about": ["Mona Lisa", "_:p1"],
  "artform": "grafiek",
  ...
  }
]
```

Imagen 20. Registro de Europeana con Schema.org (Wallis, 2017)

En España se desconocen los datos de bibliotecas o instituciones que haya implementado Schema.org en sus repositorios. Es cierto que algunas bibliotecas como la Biblioteca Virtual Miguel de Cervantes han implementado Schema.org (Candela et al., 2015) y que otras muchas colaboran con el proyecto Europeana aportando sus registros y buscando una mejora en la interoperabilidad y en las descripciones (Ibai Sistemas, 2012). Pero a nivel de repositorios no hay datos de implementaciones como esta en repositorios.

3 Metodología

3.1 Implementación de Schema.org en el repositori RODERIC

El objeto de estudio de este trabajo es el de observar cómo la Web semántica y, en concreto, la implementación de Schema.org, puede ayudar a mejorar el posicionamiento en los motores de búsqueda de los repositorios.

El proyecto de estudio de este trabajo se propuso implementar en el repositorio de la Universitat de València, RODERIC²⁹. El repositorio, cumplía a finales de 2018, 10 años de su creación. En estos 10 años, RODERIC ha crecido hasta posicionarse entre los diez

²⁹ <http://roderic.uv.es/>

repositorios con más contenidos entre las universidades españolas, según datos de REBIUN. Si nos fijamos exclusivamente en contenidos de investigación, RODERIC se sitúa en cuarto lugar, siendo el único de los cuatro que ofrece contenidos exclusivamente a texto completo y 100% de acceso abierto³⁰.

El proyecto de estudio que se ha implementado en RODERIC, busca dar un paso adelante en favor de:

- La interoperabilidad con los buscadores.
- Aumentar la visibilidad del repositorio, en las búsquedas realizadas desde los motores de búsqueda.
- Facilitar la localización de recursos depositados en el repositorio y ampliar la visibilidad de la colección.

El proyecto se ha organizado en diferentes fases:

1. Estudio introductorio de RODERIC, software, estructura de colecciones, de los metadatos del repositorio roderic.
2. Mapeo y comparativa de los metadatos DC a propiedades Schema.
3. Selección de tipos de documentos y vocabularios a implementar.
4. Formato a implementar e implementación del código fuente.
5. Definición de indicadores.
6. Implementación del proyecto.
7. Análisis del impacto y rendimiento de la implementación de Schema.org.

3.1.1 RODERIC, software, estructura de colecciones y metadatos

RODERIC (Repositori d'Objectes Digitals per a l'Ensenyament la Recerca i la Cultura) es el repositorio institucional de la Universitat de València. RODERIC responde al compromiso de la Universitat de València con el movimiento de acceso abierto al conocimiento adquirido con su adhesión a la Declaración de Berlín (30 Septiembre de 2008).

RODERIC está destinado a recoger y difundir la producción digital generada por los miembros de la comunidad universitaria en materia de cultura, docencia, colecciones

³⁰ Fuente:

<https://www.uv.es/uvweb/servicio-bibliotecas-documentacion/es/novedades/repositorio-institucional-roderic-cumple-10-anos-1285923456427/Novetat.html?id=1286052422540>

digitalizadas, e investigación. Los contenidos abarcan todas las materias científicas que se imparten en la Universidad y recoge todo tipo de materiales digitales, tanto preprints como postprints, comunicaciones a congresos, documentos de trabajo, materiales docentes y objetos de aprendizaje, revistas editadas por la Universitat de València, así como documentos y materiales resultantes de la actividad institucional realizada por sus centros, unidades y servicios³¹.

RODERIC está gestionado con la plataforma Dspace y actualmente su arquitectura responde a 5 grandes comunidades:

- Cultura: material digital relacionado con la función de difusión cultural de la Universitat.
- Docencia: materiales de uso docente en las titulaciones impartidas en la Universitat
- Institucional: información sobre la Universitat de Valencia, eventos, actos y reuniones
- Investigación: producción científica del personal investigador de la Universitat de València
- Somni: fondo antiguo digitalizado procedente de la Biblioteca de la Universitat.

El Repositorio tiene desplegadas un total de 89 colecciones dependientes de las 5 comunidades mencionadas. Sus registros se basan en la descripción de metadatos Dublin Core (DC) y actualmente tiene implementadas 113 etiquetas o campos³².

En cuanto a su interoperabilidad, RODERIC utiliza el protocolo OAI-PMH que permite que los registros almacenados en una base de datos puedan ser agregados por los servicios de recolección más importantes del mundo. Gracias a él, RODERIC se interconecta con repositorios y recolectores como el de OpenAire, Recolecta, Core, Hispana, o Google Scholar, entre otros. En el caso de Google Scholar, RODERIC implementó metaetiquetas específicas para mejorar el posicionamiento en el recolector académico.

```
<meta name="DC.creator" content="Abad García, María Francisca" />
<meta name="DC.subject" content="producción científica" xml:lang="es_ES" />
<meta name="DC.title" content="Visibilidad internacional de la investigación de
la Universitat de València (Web of Science 2000-2009)" xml:lang="es_ES" />
<meta name="DC.type" content="info:eu-repo/semantics/book" xml:lang="es_ES" />
```

³¹ Fuente:

<https://www.uv.es/uvweb/servicio-informatica/es/repositorio-institucional-roderic-1285904072947.html>

³² Anexo 1. Metadatos RODERIC

```
<meta name="DC.subject" content="UNESCO::LINGÜÍSTICA::Lingüística
aplicada::Documentación" xml:lang="es_ES" />
<meta name="DC.identifier" content="057368" xml:lang="es_ES" />
```

3.1.2 Mapeo y comparativa de los metadatos DC a propiedades Schema

Actualmente RODERIC tiene implementadas 113 etiquetas o campos Dublin Core (DC). Para nuestro proyecto realizamos una comparativa entre ambos vocabularios en el que quisimos observar el grado de expresividad entre los términos, valorando del 0 al 5 esta apreciación. Donde 0 es una relación inexistente y 5 es una expresividad total. Esta comparativa de equivalencias puede verse en el Anexo 1. Metadatos RODERIC.

De las 113 etiquetas DC encontramos una alta relación (4-5) en 60 campos (un 53%). También observamos que algunas etiquetas no tienen una representación tan igual en Schema.org o son demasiado específicas para algunas descripciones. Otras tienen un valor relativo de 3, porque aunque permiten expresar términos y valores, la relación no es tan próxima y la correspondencia entre términos pierde detalles que sí permite especificar DC ya que ofrece un nivel de especificidad en el ámbito bibliográfico que otros lenguajes no tienen tan desarrollado. Por ejemplo, se puede indicar el campo es una materia o palabra clave (Keyword) pero no que tipo de materia (LCC, UNESCO o mesh). De ahí que en un mapeo inicial como este se pierdan matices que sí ofrece DC y todavía no Schema.org. De ahí que sea preciso revisar relaciones y especificar la etiqueta DC en función del tipo de documento.

Las bibliotecas de Hesburgh elaboraron un mapeo³³ de metadatos Dublin Core a Schema.org muy interesante que se ha tenido en cuenta a la hora de realizar el propio, teniendo en cuenta que la propuesta que ellos hacen la basan en codificación RDF y no es nuestro caso.

³³ http://ndlib.github.io/metadata_application_profile/elements/

Core Elements: Condensed View						
Name	Required?	Repeatable?	Autogenerated?	DC Terms	Schema.org	RDF Gem
Title	Yes	No	No	dc:title	name	map.title(in: RDF::DC)
Alternate Title	No	Yes	No	dc:alternative	alternateName	map.alternate_title(to: 'alternative', in: RDF::DC)
Creator	No	Yes	No	dc:creator dc:creator.artist dc:creator.editor dc:creator.illustrator dc:creator.organization dc:creator.photographer	author	map.creator(in: RDF::DC)
Author	No	Yes	No	dc:creator.author	author	map.author(to: 'creator#author', in: RDF::QualifiedDC)
Contributor	No	Yes	No	dc:contributor dc:contributor.advisor dc:contributor.artist dc:contributor.author dc:contributor.editor dc:contributor.illustrator dc:contributor.organization dc:contributor.photographer	contributor	map.contributor(in: RDF::DC)

Imagen 21. Digital Metadata Standards at the Hesburgh Libraries³⁴

La Dublin Core Metadata Initiative (DCMI) creó en 2012 un *DCMI Schema.org alignment task group*, cuyo objetivo era precisamente elaborar mapeos entre ambos estándares³⁵. Otra propuesta que trabajaba en la línea de mejorar las descripciones bibliográficas de Schema.org era la de Schema Bib Extend Community Group³⁶, con propuestas de descripción específicas para diversos tipos documentales. En el transcurso de preparación de nuestro proyecto, estos vocabularios fueron revisados, pero hemos de apuntar que en mayo de 2019, el vocabulario se ha considerado obsoleto y desaconsejan su propuesta³⁷.

Para Google incluir todas las propiedades requeridas propicia que un objeto sea apto para aparecer en la Búsqueda de Google con un aspecto mejorado. Sin embargo, es más importante agregar menos propiedades recomendadas, pero que sean detalladas y precisas, que tratar de incluir todas las propiedades recomendadas posibles con datos incompletos, inexactos o con formato incorrecto (Comprende cómo funcionan los datos estructurados).

3.1.3 Selección de tipos de documentos y metaetiquetas a implementar

Debido a la gran cantidad de tipos documentales que tiene RODERIC, (31 tipos distintos) se decidió hacer el proyecto piloto en aquellos tipos documentales de más volumen

³⁴ http://ndlib.github.io/metadata_application_profile/elements/

³⁵ http://www.dublincore.org/specifications/lrmi/lrmi_1/

³⁶ <https://bibliograph.github.io/BibloGraph-Frozen/bibliograph.net/schemas.html>

³⁷ <https://bibliograph.github.io/BibloGraph-Frozen/bibliograph.net/vocabularydeprecated.html>

bibliográfico y a su vez mayor relevancia para el ámbito académico, investigador y documental.

Los tipos documentales escogidos fueron:

- Artículos (ScholarlyArticle): actualmente RODERIC tiene 27836 registros.
- Tesis (Thesis): actualmente RODERIC tiene 5721 registros.
- Libros (Book): actualmente RODERIC tiene 283 registros

A cada @type (ScholarlyArticle, Book, Thesis) le asociamos unas propiedades para su descripción. Se puede observar toda la relación en el Anexo 2. Mapeo RODERIC - Schema.org.

En el proyecto que hemos implementado esta equivalencia se utilizaron condicionales que permiten evaluar el tipo de registro y aplicar uno u otro tipo según este mediante transformaciones XSLT:

```
<xsl:when
test="contains(dim:field[@element='type'],'info:eu-repo/semantics/doctoralThesis
')">

    <xsl:text>thesis</xsl:text>

</xsl:when>

<xsl:when
test="contains(dim:field[@element='type'],'info:eu-repo/semantics/masterThesis')
">

    <xsl:text>schema:Thesis</xsl:text>

</xsl:when>

<xsl:when
test="contains(dim:field[@element='type'],'info:eu-repo/semantics/book')">

    <xsl:text>book</xsl:text>

</xsl:when>

<xsl:when
test="contains(dim:field[@element='type'],'info:eu-repo/semantics/article')">

    <xsl:text>ScholarlyArticle</xsl:text>

</xsl:when>
```

3.1.4 Formato a implementar y código base

El formato de codificación a implementar escogido ha sido JSON-LD. Hemos dado algunos apuntes de las preferencias y recomendaciones de uso sobre este formato por parte de muchas aplicaciones o proyectos desarrollados. Sin duda uno de los principales argumentos a favor de este formato es la recomendación que hace Google sobre su uso. También es interesante añadir que JSON-LD es un formato más sencillo de implementar y automatizar, tal y como apuntan proyectos como el de Europeana (Wallis, 2017) o DataMed (Fenner, 2019). Otro apunte más a favor del uso de JSON-LD es el tipo de marcado. Al insertarse en una etiqueta de script html al inicio de la página no "ensuciamos el código fuente", al no tener que añadir las propiedades directamente sobre los elementos HTML. Facilitando así también futuras migraciones a otras versiones del estándar.

Para el desarrollo del código en formato JSON-LD tuvimos el apoyo del "Servei d'Informàtica de la Unviersitat de València". En un primer entorno de pruebas, se hizo una primera definición (a fecha de 1 de abril) que más tarde se perfiló en dos ocasiones, ya en el repositorio en vivo. El código que finalmente se implementó se puede observar en el Anexo 3. Código JSON-LD para RODERIC.

Adjuntamos aquí un pequeño ejemplo, relativo a la descripción de autores o contribuidores de la publicación:

```
<script type="application/ld+json">{"@context": "https://schema.org",
"@id": "<xsl:value-of select="$id_schema_org"/>",
"@type": "<xsl:value-of select="$type_modif_schema_org"/>"
"author": [ <xsl:for-each select="dim:field[@element='contributor' and
@qualifier='author']"> {"@type": "Person", "name": "
<xsl:value-of select="./node()"/>} <xsl:if
test="count(following-sibling::dim:field[@element='contributor' and
@qualifier='author'])>0">,
</xsl:if></xsl:for-each>]
<xsl:for-each select="dim:field[@element='creator']">,
"author": [ {"@type": "Person", "name": "<xsl:value-of select="./node()"/>"} ]
</xsl:for-each>
```

```
<xsl:for-each select="dim:field[@element='contributor' and
@qualifier='advisor']">,
"contributor": [{"@type": "Person", "name": "<xsl:value-of select="."/node()"/>"}]
```

3.1.5. Definición de indicadores y estudio estadístico

En el ámbito de los repositorios, son pocos los estudios, que apuntan a que datos e indicadores se han revisado para observar la evolución de una implementación de estas características. La implementación de Schema.org todavía es minoritaria en repositorios.

Wallis (2017) en referencia al proyecto de Europeana, invita a observar o identificar cambios en los números y / o patrones de visitas de los usuarios a los sitios web de Europeana. En el caso del repositorio Digital Georgetown, utilizaron Google Analytics para observar su evolución y comparar el número total de referencias de Google (Pekala, 2018).

En el caso de RODERIC, los indicadores que se consideraron observar tienen como punto de referencia dos cuestiones: el posicionamiento de RODERIC en los buscadores y el rendimiento de los resultados impresos en los resultados de búsqueda. Para recoger los datos de estudio de los indicadores se utilizaron dos herramientas: Google Analytics y Roderic Statistics³⁸.

En base a la primera cuestión, el posicionamiento de RODERIC en los buscadores se observó en relación a las consultas realizadas a través del buscador de Google:

- Impresiones: número de veces que una URL de su sitio ha aparecido en los resultados de búsqueda de un usuario; no incluye las impresiones de búsqueda de pago de Google Ads. Este indicador se ha observado desde Google Analytics, sección, Adquisiciones - Search Console - Consultas.
- Clasificación media de las URL de su sitio web con respecto a una o varias consultas. Por ejemplo, si una URL de su sitio obtiene la tercera posición con respecto a una consulta, y la séptima con respecto a otra consulta, la posición media de dicha URL sería la quinta $(3+7/2)$. Este indicador se ha observado desde Google Analytics, sección, Adquisiciones - Search Console - Consultas.

³⁸ <http://roderic.uv.es/statistics>

- Búsquedas específicas en cuatro buscadores (Google, Yahoo, Bing y Google Scholar) para observar si hay mejora en el posicionamiento.

Respecto a la segunda cuestión, el rendimiento en las búsquedas, observamos los siguientes datos:

- Tipos de fuentes de tráfico utilizadas por los usuarios y cantidad de accesos/sesiones por cada fuente. Este indicador se ha observado desde Google Analytics, sección, Adquisición - Todo el tráfico - canales - URL de referencia.
- Número de clics que han recibido las URL de su sitio web desde una página de resultados de búsqueda de Google; no incluye los clics de los resultados de búsqueda pagados de Google Ads. Este indicador se ha observado desde Google Analytics, sección, Adquisiciones - Search Console - Consultas.
- CTR: Porcentaje de clics= $\text{clics}/\text{impresiones} \times 100$. Este indicador se ha observado desde Google Analytics, sección, Adquisiciones - Search Console - Consultas.

En el Anexo 4 “Indicadores y estadística” se puede observar al completo tanto los indicadores como el estudio estadístico en su conjunto.

3.1.6 Implementación del proyecto

Los trabajos de implementación se realizaron con el soporte del “Servei d’Informàtica de la Universitat de València”, en el campus de Burjassot. A fecha de 12 de abril de 2019, nos reunimos para definir una primera versión completa del código e insertarlo en el sistema. Tras unas primeras observaciones, a fecha de 30 de mayo de 2019, se implementó una segunda versión, modificada con algunas correcciones. La herramienta para la revisión de código fue la “Herramienta de pruebas de datos estructurados” de Google³⁹.

³⁹ <https://search.google.com/structured-data/testing-tool/u/0/?hl=es>

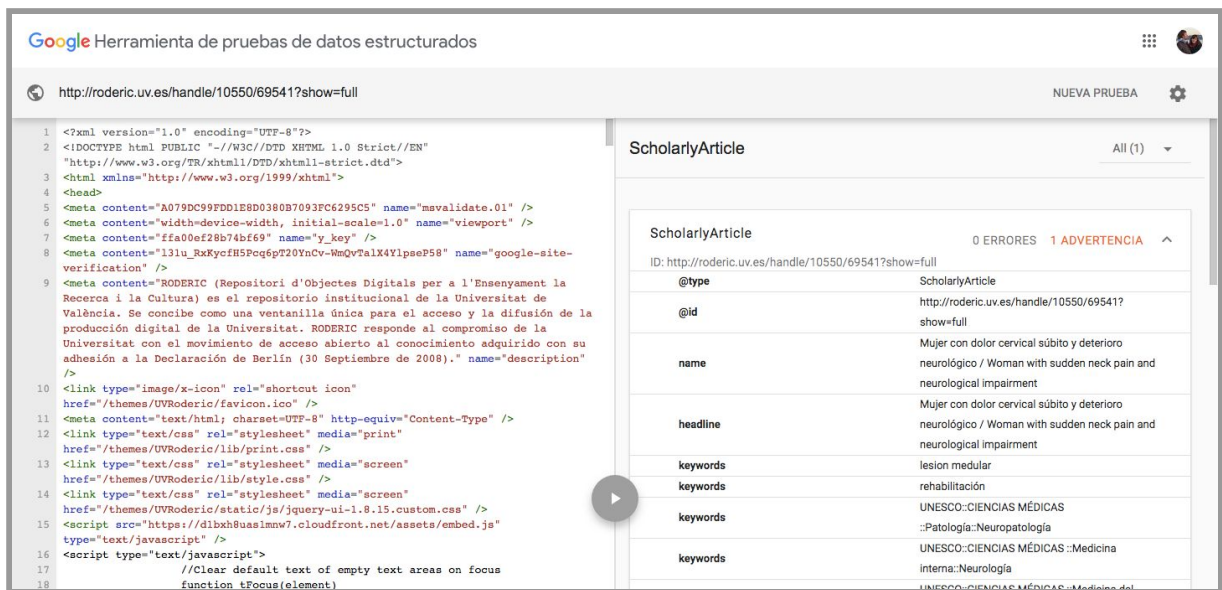


Imagen 22. Revisión de datos estructurados

Finalmente se solicitó a Google, Yahoo y Bing la reindexación del repositorio, enviando el “Mapa del sitio” en formato XML. Es importante que los mapas de sitio se actualicen periódicamente para indicar páginas nuevas y actualizadas, ya que las páginas pueden no haber cambiado visualmente, pero si los datos que contienen.

3.1.7 Análisis del impacto y rendimiento de la implementación de Schema.org

Analizar el impacto de una implementación como la de Schema.org requiere observar diversos factores. No existe un único dato a mirar y es importante apuntar que los datos que se observen deben mirarse en un contexto y considerando los límites (en este caso) de la muestra.

En el proyecto que estamos estudiando una de las principales herramientas usadas para observar el comportamiento del sitio web de RODERIC ha sido Google Analytics. Esta herramienta nos proporciona datos e información sobre el uso del sitio web y el comportamiento del usuario. Su implementación es relativamente sencilla, si se tiene acceso al módulo de administración de Dspace y en el caso de RODERIC ya estaba implementada.

Es cierto que Dspace como muchos de los principales sistemas de repositorios digitales vienen con sus propios informes de uso internos. Los informes que nos ofrece Dspace tienen que ver más con la interacción del usuario en el propio repositorio (consultas, visitas, descargas, etc.). En cambio los datos que queremos observar en nuestra implementación

tienen una relación más directa con el posicionamiento web que tiene el repositorio al respecto de los buscadores.

La información observada en Google Analytics y que se puede ver al completo en el Anexo 4 “Indicadores y estadística”, ha tenido relación con las fuentes de tráfico utilizadas por los usuarios y cantidad de accesos/sesiones por cada fuente, así como el rendimiento y el tráfico de búsqueda del sitio web.

Fuente	Adquisición			Comportamiento			Conversiones		
	Usuarios	Usuarios nuevos	Sesiones	Porcentaje de rebote	Páginas/sesión	Duración media de la sesión	Tasa de conversión del objetivo	Objetivos cumplidos	Valor del obje
	36.584 % del total: 87,67 % (41.727)	33.798 % del total: 87,08 % (38.814)	42.735 % del total: 85,54 % (49.957)	73,09 % Media de la vista: 71,53 % (2,17 %)	2,06 Media de la vista: 2,27 (-9,24 %)	00:01:14 Media de la vista: 00:01:24 (-11,43 %)	0,00 % Media de la vista: 0,00 % (0,00 %)	0 % del total: 0,00 % (0)	0,00 U % del total: 0 (0,00)
1. google.com	21.721 (58,94 %)	20.302 (60,07 %)	24.778 (57,98 %)	75,61 %	1,92	00:01:02	0,00 %	0 (0,00 %)	0,00 US\$ (0)
2. roderic.uv.es	3.505 (9,51 %)	3.084 (9,12 %)	4.045 (9,47 %)	86,35 %	1,25	00:00:54	0,00 %	0 (0,00 %)	0,00 US\$ (0)
3. dialnet.unirioja.es	2.704 (7,34 %)	2.306 (6,82 %)	3.180 (7,44 %)	54,50 %	2,40	00:01:44	0,00 %	0 (0,00 %)	0,00 US\$ (0)
4. google.es	2.179 (5,91 %)	1.942 (5,75 %)	2.552 (5,97 %)	75,90 %	2,06	00:01:02	0,00 %	0 (0,00 %)	0,00 US\$ (0)
5. scholar.google.es	1.231 (3,34 %)	1.098 (3,25 %)	1.382 (3,23 %)	61,79 %	1,95	00:01:17	0,00 %	0 (0,00 %)	0,00 US\$ (0)
6. google.com.mx	563 (1,53 %)	551 (1,63 %)	608 (1,42 %)	84,87 %	1,25	00:00:34	0,00 %	0 (0,00 %)	0,00 US\$ (0)
7. mobiroider.uv.es	434 (1,18 %)	381 (1,13 %)	569 (1,33 %)	61,34 %	2,35	00:01:19	0,00 %	0 (0,00 %)	0,00 US\$ (0)
8. scholar.google.com.pe	369 (1,00 %)	354 (1,05 %)	407 (0,95 %)	62,41 %	1,84	00:01:31	0,00 %	0 (0,00 %)	0,00 US\$ (0)
9. scholar.google.com.mx	353 (0,96 %)	342 (1,01 %)	392 (0,92 %)	70,41 %	1,59	00:01:09	0,00 %	0 (0,00 %)	0,00 US\$ (0)

Imagen 23. Seguimiento con Google Analytics del portal RODERIC

La otra herramienta complementaria que Google proporciona y permite integrarse con Google Analytics es Google Search Console. Esta herramienta ayuda a proporcionar métricas más precisas para revisar el SEO de un repositorio y su posicionamiento en un buscador tan importante como Google. Gracias a esta herramienta podemos observar datos como el de las impresiones, clics o posición media. Repositorios como el de Strathprints, de la Universidad de Strathclyde (Glasgow), ya apuntan en esta dirección de datos a observar (Macgregor, 2019).

Además de la herramienta de Google, otra pauta considerada para observar el posicionamiento han sido diferentes búsquedas específicas donde se observaba el ranking web que alcanzaba RODERIC en los diferentes buscadores (Google, Yahoo, Bing, Google Scholar). El periodo de observación se ha realizado entre los meses de abril y agosto. Se realizaron un total de 200 búsquedas, tomando en consideración, autores del repositorio, títulos de publicaciones existentes y palabras clave relacionadas con los documentos seleccionados por tipos (Book, ScholarlyArticle y Thesis) y poder recopilar datos y obtener una muestra para observar.

4 Resultados

El impacto de la implementación de Schema.org en el repositorio se supervisó y midió utilizando una variedad de métricas, tomando como referencia los datos de tráfico de búsqueda de la Consola de búsqueda de Google, los datos de uso del módulo de Dspace (Roderic/Statistics), los datos de seguimiento de Google Analytics y estadísticas de búsquedas específicas.

Los datos estadísticos que presentamos⁴⁰ se ciñen (en la medida de lo posible) a una comparativa temporal del mismo periodo (1 de abril a 31 de Julio) para los años 2018 y 2019. No es un periodo de observación muy amplio. Para poder observar si hay una perspectiva de mejora o no, hemos comparado el mismo periodo observado con el del año anterior. Y la comparativa es con el año anterior porque el curso académico y sus exigencias académicas son diferentes según el mes en que nos encontremos.

Tráfico

Respecto a los datos de tráfico que hemos mirado desde Google Analytics, observamos:

El tráfico total de visitas a páginas de RODERIC en los meses de abril a julio de 2018 fue de 416.133 frente a los 435.398 de 2019, presentando un ligero ascenso. Al igual que el número de visitas que ha tenido el repositorio, que en 2018 fueron 164.442 y en 2019, durante esos meses, 184.230.

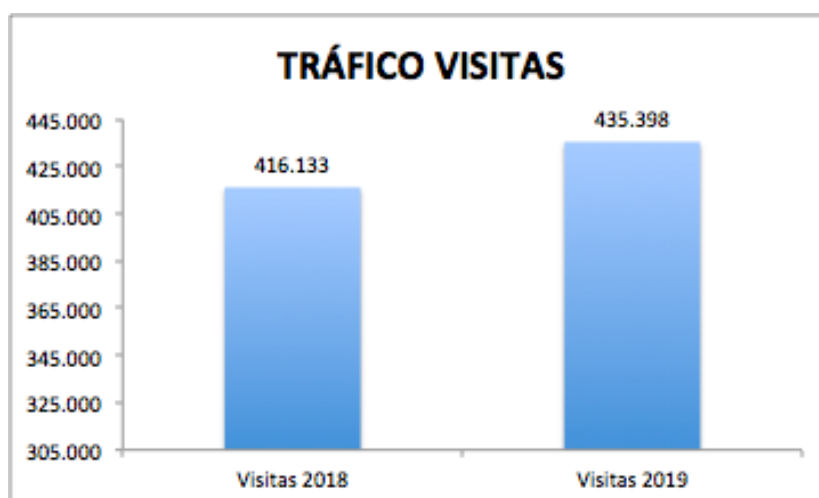


Imagen 24. Comparativa de tráfico de visitas RODERIC

⁴⁰ Ver Anexo 4. indicadores y estadísticas

Si comparamos el tráfico de referencia de Google y Google Scholar, para el mismo periodo, la cifra de tráfico presenta un ligero descenso pasando de 131.994 visitas referenciadas en 2018 a 119.277 visitas en este 2019.

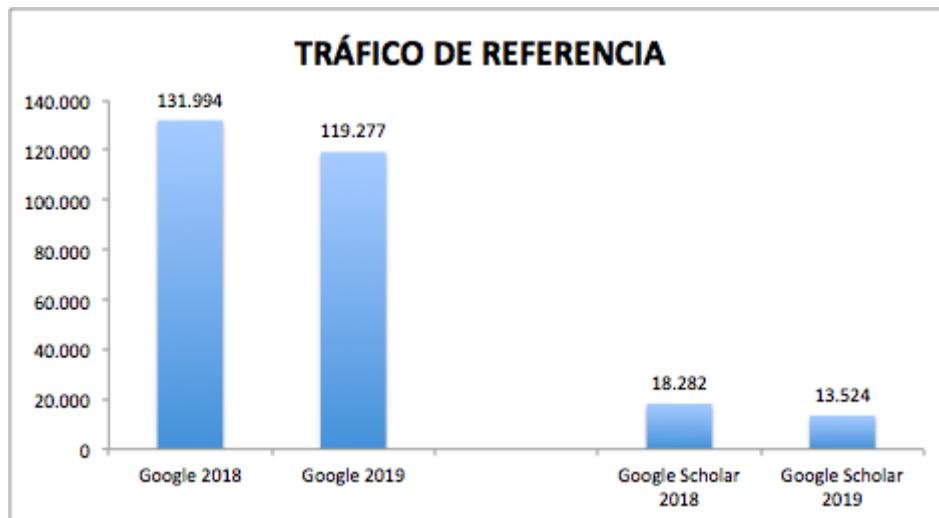


Imagen 25. Comparativa tráfico de referencia para Google y Google Scholar RODERIC

Google Search Console

Los datos de las consultas en Google se han extraído de Google Search Console en Google Analytics. Se han mirado 4 métricas: Impresiones, clics, CTR, posición en ranking web.

Para las impresiones observadas en el periodo de abril a julio de 2019, los datos en comparación a 2018 han crecido en su conjunto, pasando de 787.087 impresiones en el año 2018 a 1.665.928 en el 2019. Esto supone un crecimiento del 111,66 % respecto al mismo periodo del año anterior.

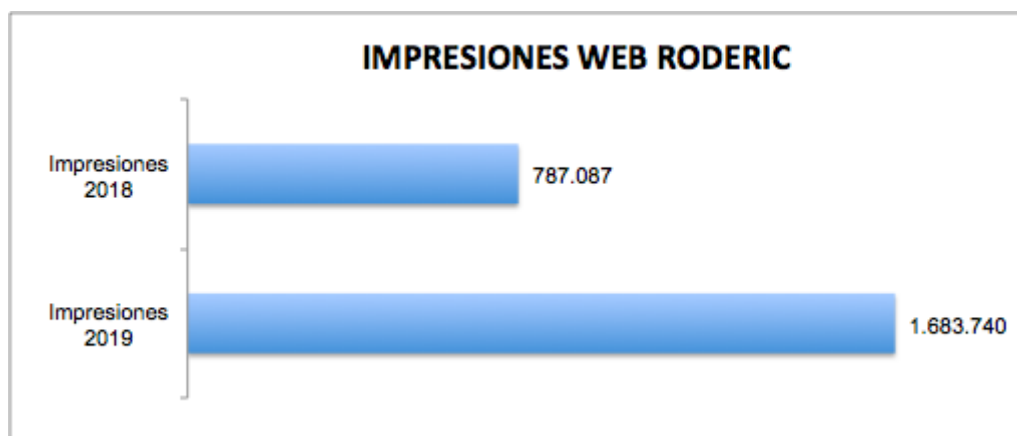


Imagen 26. Comparativa de impresiones web para RODERIC

Respecto a los clics, el pasado año entre los meses de abril y julio (2018) se registraron un total de 30.611 clics, frente a los 23.590 de este 2019. Un descenso del 23%. El CTR (porcentaje de clics=clics/impresiones*100) pasa de un 3,89% en 2018 a un 2,06 % en 2019 y una posición media de 23 en 2018 frente a la 33 en 2019.

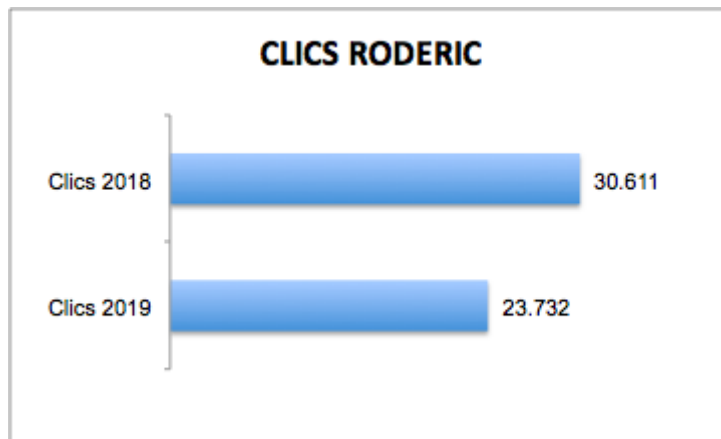


Imagen 27. Comparativa sobre clics web de RODERIC

Unos datos que no reflejan la reindexación que se solicitó a Google a finales de julio y que esperamos ayude a mejorar los datos de posicionamiento futuro. Las visitas frecuentes de los rastreadores y las reindexaciones de los repositorios son acciones necesarias para recopilar contenido nuevo. En nuestro caso, deberíamos haberlo realizado inmediatamente después de la versión final revisada.

Resultados de la búsqueda Rankings

Todas las búsquedas se realizaron en los buscadores Google, Yahoo! y Bing. Siempre desconectado de cualquier cuenta, en modo incógnito y mirando un máximo de 5 páginas de resultados. Hay que tener en cuenta que los resultados de búsqueda que un usuario ve se personalizan según el historial de búsqueda, la ubicación, el tipo de dispositivo y lo que Google sabe acerca de las preferencias del usuario (Comprende cómo funcionan los datos estructurados, 2019).

Para observar si los registros de los documentos obtuvieron una clasificación más alta después de la implementación, tomamos una muestra de veinte artículos seleccionados al azar de colecciones representativas (Artículos, Libros y Tesis) en las que se ha implementado Schema.org. Para ello se realizaron diferentes tipos de búsquedas (título,

autores, palabras clave del título) en el que se anotaba si el documento aparecía en los resultados de la búsqueda y en qué posición.

Una de las primeras cosas que observamos es que RODERIC no estaba indexado ni en Yahoo! ni en Bing de modo que ampliamos las búsquedas a Google Scholar para observar someramente y por comparativa el posicionamiento del repositorio. Se solicitó en 2 ocasiones la indexación a Yahoo! y Bing pero todavía sin resultados.

Tras un total de 200 búsquedas, observamos que la clasificación de los documentos de RODERIC no variaba demasiado. Por un lado encontramos documentos que aunque eran de años anteriores no estaban indexados y tras estos meses de estudio siguen sin indexarse. El resto de documentos que sí posicionaba sigue mostrando un ranking similar tras la implementación y observación de resultados. En general, los resultados de nuestra evaluación muestran un impacto todavía pequeño en los resultados de búsqueda.

Respecto de Google Scholar (GS), se observan varias cuestiones. La primera de ellas es respecto a los autores. Cuando estos tienen definido un buen perfil de autor, su posicionamiento y visibilidad es mejor respecto a los que no lo tienen. La segunda cuestión es respecto a la visibilidad de RODERIC en GS y su duplicidad de indexación entre mobilroderic.uv.es y roderic.uv.es.

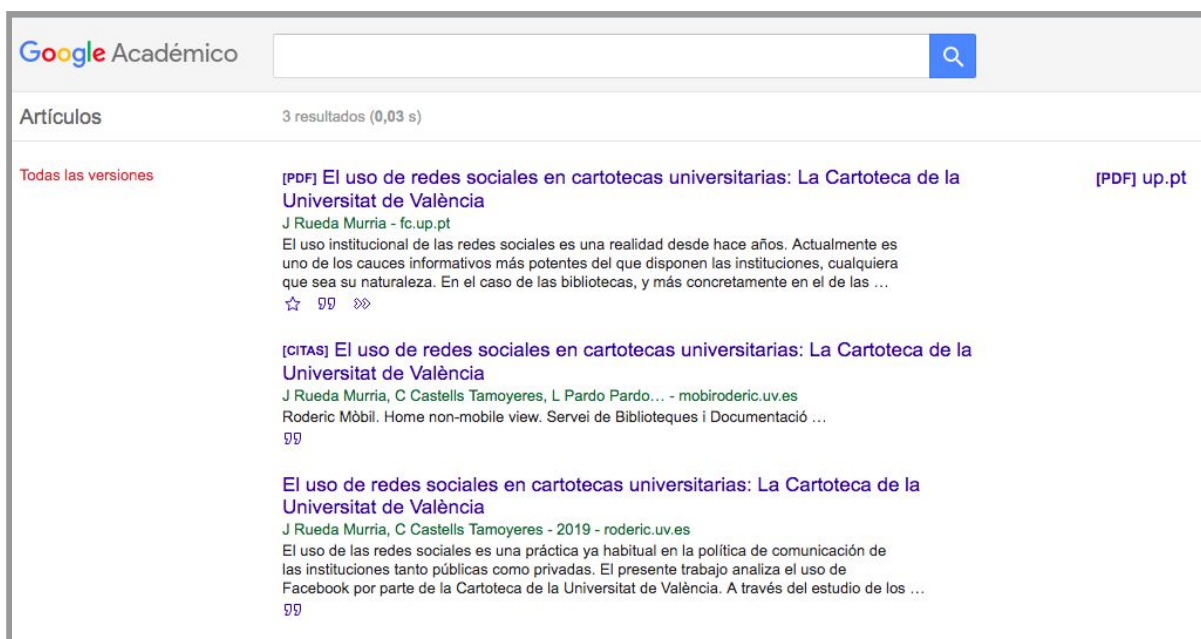


Imagen 28. Duplicidad de versiones web en GS de RODERIC

Faltaría definir bien la propiedad o identidad de los documentos. Ya sea con la marca de la universidad o con la del propio repositorio. Un ejemplo lo observamos en la imagen anterior donde ninguno de los dos registros que pertenecen a RODERIC tiene indicada la referencia, como sí sucede en el primer documento.

Quizás hay que tener en consideración al respecto del posicionamiento, el hecho de que los competidores también buscan posicionarse mejor y grandes portales como Pubmed, Dialnet, Researchgate o editoriales como Ovid o Elsevier, están trabajando mucho también para que su información aparezca en los primeros puestos.

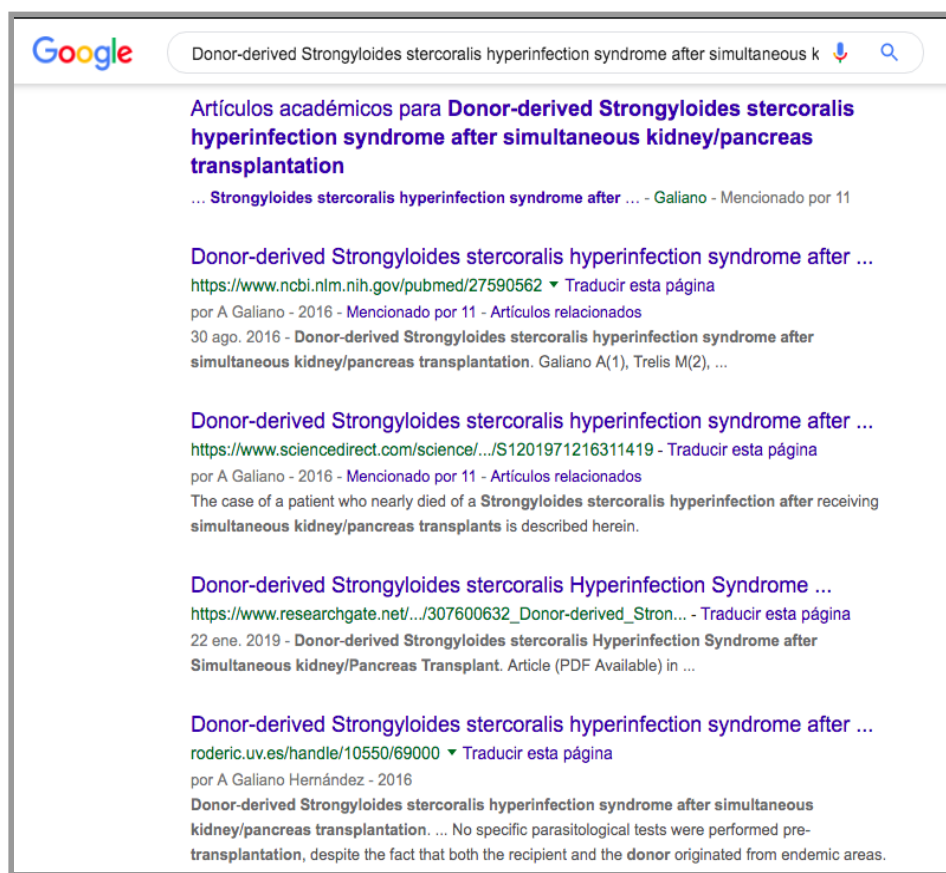


Imagen 29. Ranking web RODERIC

Y en esta línea también comentar que por ejemplo la página web de la Universitat de València posiciona bastante bien a los investigadores con página personal o de grupos de trabajo con marca de la universidad y en muchos casos mejor que el posicionamiento del propio RODERIC.

5 Conclusiones

Cuando se inició este trabajo los objetivos que se marcaron fueron:

- Definir, contextualizar y determinar los principales factores de posicionamiento en el SEO académico (ASEO).
- Hacer una revisión del vocabulario Schema.org y aplicabilidad a un repositorio digital.
- Análizar Schema.org como herramienta para aumentar la riqueza de los metadatos de un repositorio y herramienta con fines de posicionamiento web.
- Realizar un análisis de estas tecnologías y ver cómo pueden incrementar las visitas desde los buscadores (tanto en Google Scholar, como en Google Search).

Tras el proceso de estudio que hemos realizado observamos que:

La optimización en buscadores (SEO) y la visibilidad de una página Web se encuentran directamente relacionadas con el proceso de mejora de la posición en los resultados devueltos por motores de búsqueda como Google. El posicionamiento académico (ASEO) pasa por un trabajo a largo plazo que va más allá de implementar determinados cambios en repositorios o producciones académicas. En un sector como el SEO existen muchos factores que influyen en él y que deben ser valorados en el contexto de cada organización. Uno elemento que se debe considerar en la planificación del ASEO son las estrategias de búsqueda de los usuarios. Descubrir cómo las personas buscan nos ayudará a poner en orden las opciones a mejorar. Buscadores web como Google o Google Scholar, están ganando posiciones en las preferencias de búsquedas de los investigadores. Y es ahí donde los repositorios debe ser más visibles. No podemos olvidar que una de las misiones de los repositorios es la de dar acceso y visibilizar la producción científica, educativa y cultural que contienen. Y para ello se necesita planificar programas, objetivos y acciones vinculados a estrategias compartidas basadas en mejorar y ampliar el SEO académico.

En este nuevo contexto tecnológico, los repositorios digitales deben abordar el papel fundamental que juega la Web Semántica y herramientas como Schema.org para mejorar el entendimiento entre ambos. La Web 3.0 presenta un entorno semántico que las bibliotecas y los repositorios, como buenos conocedores de la catalogación, la descripción y la recuperación de información deben aprovechar. Las bibliotecas han dado tímidos pasos en esta dirección al empezar implementar en sus repositorios elementos semánticos que amplíen su relación con la web.

El potencial que precede a una herramienta como Schema.org en páginas web merecía el planteamiento de implementarla en repositorios como uno de los elementos que pueden marcar la diferencia a la hora de “visibilizar los repositorios”. Aunque no hay una gran cantidad de ejemplos sobre la aplicación de esta herramienta en repositorios, aquellos que lo han abordado si están en el camino de una mejora consolidada.

Dos son los grandes retos que plantea este proyecto. El primero es el de ajustar las descripciones realizadas con Dublin Core a las propiedades de Schema.org, ya que la comunidad bibliotecaria y la comunidad de la Web Semántica utilizan una terminología diferente para conceptos similares sobre metadatos.

En un repositorio como RODERIC, con 10 años de trayectoria y una gran cantidad de agentes implicados, una de las primeras cosas que apreciamos es la evolución en las descripciones de sus registros. Por un lado observamos descripciones muy cuidadas y trabajadas y por otro, otras más sencillas donde algunos datos importantes se aglutinan por ejemplo en la citación bibliográfica. Todo un reto que afrontar a la hora de trasladar en un mapeo entre las propiedades de Schema.org y los campos Dublin Core. Sin duda una descripción más precisa y homogénea ayudaría en el mapeo de metadatos y en la recuperación de la información en los buscadores. Cuanta más información podamos transformar para que el buscador pueda leer correctamente, más más relevantes serán nuestros documentos, nuestros datos y nuestra información.

Implementar una herramienta semántica como Schema.org no es un proceso cerrado, es necesario ir revisando términos y aplicando de nuevos. Schema.org es un vocabulario en crecimiento y tiene a su lado importantes grupos de trabajo que revisan propiedades y proponen nuevas, en función del área que se trabaje. De ahí el dinamismo y la necesaria actualización.

Sin duda el segundo de los retos que presenta este proyecto ha sido el de cómo medir el impacto de la implementación de Schema.org en los buscadores. Algunos autores como Pekala (2018) o Wallis (2017) apuntan que, si bien la presencia de microdatos en una página web se puede verificar mediante la herramienta de prueba de datos estructurados de Google⁴¹, no existe una forma única de medir su efecto en los resultados de búsqueda.

Cómo medir entonces el impacto de Schema.org en la visibilidad de los repositorios. Un primer apunte a considerar es que en la configuración del ranking o posicionamiento web, tanto de Google como de otros buscadores, influyen muchos aspectos que se desconocen. No en vano, Google actualiza su algoritmo más de seiscientos veces al año.

⁴¹ <https://search.google.com/structured-data/testing-tool/u/0/?hl=es>

Los repositorios que han implementado Schema.org en sus sistemas, han observado una variedad de métricas que van desde el seguimiento de resultados de búsquedas, a datos compilados de Google Analytics y Google Search Console (GSC) y proporcionar así una imagen más completa del impacto de Schema.org en los resultados de búsqueda de Google.

Wallis (2017) propone además ajustar el análisis de tráfico actual para identificar las visitas a dichas páginas y las rutas que toman los usuarios para llegar a estas páginas y poder identificar tendencias en la implementación de Schema.org.

Una buena práctica debe ser la reindexación del sitio del repositorio, tanto a Google como a otros buscadores. Ellos mismos la recomiendan (Comprende cómo funcionan los datos estructurados, 2019). Una acción que en nuestro caso se realizó prácticamente al final del proyecto y que sin duda tendríamos que haber realizado antes, para poder observar quizás cambios más significativos.

Mención a parte y en la que no entraremos en esta discusión, son en las métricas y el análisis para los buscadores de Yahoo y Bing. Un análisis con poco recorrido en la comunidad bibliotecaria, ya que aproximadamente el 80 por ciento de las bibliotecas académicas utiliza Google y Google Analytics para realizar sus análisis (O'Brien, 2017). En lo referente a nuestro proyecto, el análisis en estos buscadores ha sido un campo considerado, pero poco explotado debido a la falta de datos y a que el repositorio no ha sido indexado correctamente en estos dos buscadores.

No cabe duda de que un proyecto como este, que busca mejorar el posicionamiento de los repositorios en la web, necesita ser observado en un espectro temporal amplio. Pekala (2018) apunta que los cambios en los resultados de búsqueda no son inmediatamente visibles o consistentes; por lo tanto, la evaluación de múltiples métricas para obtener una imagen más amplia del impacto general es clave.

Esta es una carrera de fondo. Lograr un SEO óptimo conlleva tiempo, de ahí la importancia de mirar con perspectiva esta implementación para poder avalarla con números.

Afrontar un proyecto como el que se ha llevado a cabo en este trabajo de máster, ha sido en lo personal un proceso estimulante en el que sin duda se ha aprendido mucho. Presentaba el reto de salir de una zona de confort para estudiar nuevas posibilidades que ofrezcan a los repositorios dar un paso más en pro de su visibilidad. Aquellas bibliotecas y repositorios que se planteen su implementación, comentar que no es un proceso complicado, pero sí intenso en cuanto a estudio. Los dos retos que presenta este proyecto, así lo requieren. Un buen mapeo de datos requiere concreción en la selección de tipos y propiedades a usar.

Schema.org presenta muchas opciones y necesariamente para cada repositorio las opciones serán únicas, si se quiere ganar en precisión. Otro elemento que me gustaría destacar es la colaboración fundamental de los servicios de informática, tanto para la configuración del formato de Schema.org en JSON-LD, como para la inserción del código en el sistema. Sin duda son un apoyo fundamental para complementar una mirada más centrada en el que se quiere hacer que en el cómo se implementa. Y ellos la mayoría de las veces tienen la clave.

Dicen que lo complicado en la vida no es llegar, sino mantenerse. Un buen resumen del elemento clave para avanzar en un proyecto como este. Afrontar no solo la evaluación del impacto, sino el conseguir una dinámica de estudio para ir reconduciendo el proceso en la dirección más adecuada, se convierte en fundamental en un medio tan cambiante como el entorno Web. Entender al usuario en sus procesos de búsqueda puede tener la clave para que nos encuentren cuando nos necesitan.

6 Discusión y trabajo futuro

Las bibliotecas y sus instituciones invierten mucho esfuerzo y recursos en mejorar la producción científica y su impacto. Un trabajo que, debe estar alineado con el entorno tecnológico que tenemos y sus posibilidades y que tiene en los repositorios una pieza clave para el éxito.

Si hace 10 años los repositorios nacían con la misión de recoger la producción científica con el fin de dar acceso, divulgar y preservar la producción intelectual de una institución. El objetivo actual de los repositorios debe pasar por ganar interoperabilidad e integración con la web con el fin dar mayor visibilidad a toda esa producción que gestionan.

Queda patente, tras este estudio, que ese trabajo debe ir de la mano de una estrategia bien definida y una planificación a medio plazo que contemple el SEO como un ingrediente fundamental. También se considera fundamental desarrollar una dinámica de trabajo que nos permita observar y evaluar el trabajo implementado con Schema.org y poder desarrollar nuevas vías que nos ayuden a dar pasos adelante en pro de la visibilidad de los repositorios pero sobre todo de la investigación y trabajo que contienen. Y es que sin duda cuantos más estudios e implementaciones se presenten en esta dirección, más serán las mejoras que se podrán aplicar a los repositorios.

Como trabajo futuro y entendiendo cuál es el presente y a que apunta el futuro, las líneas de trabajo a desarrollar deberían pasar por 3 aspectos fundamentales:

1. Revisión y actualización de las descripciones y del mapeo a Schema.org

Realizar un procedimiento de trabajo que incorpore revisar y mejorar las descripciones y potenciar relaciones semánticas más precisas, revisar índices especialmente el de autores (para revisar duplicidades y errores tipográficos). Ampliar y mejorar la descripción implementada de Schema.org con nuevas propiedades y tipos y expandirla a otras colecciones documentales.

2. Plan de evaluación

Crear un plan de evaluación que permita observar en todo su conjunto métricas y tendencias. Con informes que permitan un estudio más amplio del SEO de RODERIC que incluya un estudio de las tendencias del usuario y análisis de búsquedas exitosas que permitan integrar la intención del usuario y los componentes del sistema para mejorar las interacciones persona-máquina. Ampliar la mirada con herramientas como Google Tag manager, o Google Search Console entre otras. Se están desarrollando varios proyectos en la comunidad bibliotecaria europea para establecer estándares y desarrollar herramientas para la presentación de informes de estadísticas para repositorios (Obrien, 2017).

3. Mejorar la estrategia móvil

Cada vez es más fundamental mejorar nuestra estrategia para dispositivos móviles porque cada vez son más las búsquedas que realizan desde ellos. Con la última actualización de la indexación de Google se rastrearán e indexarán las páginas con el agente para smartphones principalmente (Prepárate para la indexación "dispositivos móviles primero", 2019). Todos aquellos sitios web que no tengan en cuenta el contenido de su sitio móvil desde el punto de vista de la usabilidad, corren el riesgo de no posicionarse ni siquiera para las consultas desde el *escritorio*. Es lo que Google denomina *Mobile first indexing*.

Bibliografía

Comprende cómo funcionan los datos estructurados. [en línea], [sin fecha]. [Consulta: mayo 2019]. Disponible en:

<https://developers.google.com/search/docs/guides/intro-structured-data>

Prepárate para la indexación «dispositivos móviles primero». [en línea], 2019. [Consulta: julio 2019]. Disponible en:

<https://developers.google.com/search/mobile-sites/mobile-first-ind>

Web Content Accessibility Guidelines (WCAG) Overview. [en línea], [sin fecha]. [Consulta: agosto 2019]. Disponible en: <https://www.w3.org/WAI/standards-guidelines/wcag/>

DSpace 6.x Documentation - DuraSpace Wiki. [en línea], [sin fecha]. [Consulta: agosto 2019]. Disponible en: <https://wiki.duraspace.org/display/DSDOC6x>

Invenio 3.1.1 documentation. [en línea], [sin fecha]. [Consulta: agosto 2019]. Disponible en: <https://invenio.readthedocs.io/en/latest/index.html>.

Researchers of tomorrow. The research behaviour of Generation Y doctoral students. [en línea], 2012. [Consulta: agosto 2019]. Disponible en:

<https://www.jisc.ac.uk/reports/researchers-of-tomorrow>.

Schema.org. [en línea], [sin fecha]. [Consulta: mayo 2019]. Disponible en:

<https://schema.org/>.

ABAD GARCÍA, M.F., 2009. RODERIC, el Repositorio de Objetos Digitales para la Enseñanza y la Investigación y la Cultura de la Universitat de Valencia. En: *@tic. revista d'innovació educativa* [en línea], no. 3, pp. 39-46. [Consulta: mayo 2019]. Disponible en: <http://ojs.uv.es/index.php/attic/article/view/166>.

ADAMICH, T., 2019. OA/Open Data Designs and Digital Repository Strategies. En: *Computers in Libraries* [en línea], vol. 39, no. 4, pp. 4-8. [Consulta: agosto 2019]. Disponible en:

<http://www.infotoday.com/cilmag/may19/Adamich--OA-Open-Data-Designs-and-Digital-Repository-Strategies.shtml>.

AGUILLO, I.F., ORTEGA, J.L., FERNÁNDEZ, M. y UTRILLA, A.M., 2010. Indicators for a webometric ranking of open access repositories. En: *Scientometrics* [en línea], vol. 82, no. 3, pp. 477-486. [Consulta: julio 2019]. Disponible en:

<http://link.springer.com/10.1007/s11192-010-0183-y>.

ALTAMIRANO, M., LAYEDRA, N.R. y GONZALEZ, N., 2018. Impacto de la web semántica sobre los repositorios digitales en las universidades ecuatorianas. En: *Revista ESPACIOS* [en línea], vol. 39, no. 52. [Consulta: julio 2019]. Disponible en:

<http://www.revistaespacios.com/a18v39n52/18395211.html>.

- ANDRÉS ENCINAS, E., 2013. *Criterios de Calidad en los repositorios institucionales abiertos* [en línea]. [Consulta: agosto 2019]. Disponible en: <https://gredos.usal.es/handle/10366/120544>.
- ARLITSCH, K. y O'BRIEN, P.S., 2012. Invisible institutional repositories. En: *Library Hi Tech* [en línea], vol. 30, no. 1, pp. 60-81. [Consulta: julio 2019]. Disponible en: <https://www.emeraldinsight.com/doi/10.1108/07378831211213210>.
- ARLITSCH, K., OBRIEN, P., CLARK, J.A., YOUNG, S.W.H. y ROSSMANN, D., 2014. Demonstrating Library Value at Network Scale: Leveraging the Semantic Web With New Knowledge Work. En: *Journal of Library Administration* [en línea], vol. 54, no. 5, pp. 413-425. [Consulta: agosto 2019]. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/01930826.2014.946778>.
- ASSANTE, M., CANDELA, L., CASTELLI, D., MANGHI, P. y PAGANO, P., 2015. Science 2.0 Repositories: Time for a Change in Scholarly Communication. En: *D-Lib Magazine* [en línea], vol. 21, no. 1/2. [Consulta: julio 2019]. Disponible en: <http://www.dlib.org/dlib/january15/assante/01assante.html>.
- BANKIER, J.G. y GLEASON, K., 2014. *Institutional Repository Software Comparison* [en línea]. París: UNESCO. [Consulta: agosto 2019]. Disponible en: <http://dide.minedu.gob.pe/handle/123456789/2899?show=full>.
- BARRUECO CRUZ, J.M. y AZORÍN MILLARUELO, C., 2014. *Guía para la evaluación de repositorios institucionales de investigación* [en línea]. Madrid: FECYT. [Consulta: agosto 2019]. Disponible en: <https://rebiun.baratz.es/rebiun/record/Rebiun08546548>
- BEEL, J. y GIPP, B., 2009. Google Scholar's ranking algorithm: The impact of citation counts (An empirical study). En: *2009 Third International Conference on Research Challenges in Information Science* [en línea], pp. 439-446. [Consulta: agosto 2019]. Disponible en: <http://ieeexplore.ieee.org/document/5089308/>.
- BERNARDIS, S., 2014. Estudio del impacto del SEO. semántico en los motores de búsqueda: aplicando microdatos y RDFa Lite 1.1 en el ámbito de Schema.org. En: *Cuadernos de Gestión de Información* [en línea], vol. 3, pp. 85-104. [Consulta: julio 2019]. Disponible en: <https://revistas.um.es/gesinfo/article/view/207771>.
- BERNERS-LEE, T., HENDLER, J. y LASSILA, O., 2001. The semantic web. En: *Scientific American* [en línea], vol. 284, no. 5, pp. 29-37. [Consulta: agosto 2019]. Disponible en: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF2>.
- BLANKSTEIN, M. y WOLFF-EISENBERG, C., 2019. *Ithaka S+R US Faculty Survey 2018*. [en línea]. [Consulta: agosto 2019]. Disponible en: <http://sr.ithaka.org/?p=311199>.
- BORREGO, Á., 2017. Institutional repositories versus ResearchGate: The depositing habits of Spanish researchers. En: *Learned Publishing* [en línea], vol. 30, no. 3, pp. 185-192.

- [Consulta: mayo 2019]. Disponible en:
<https://onlinelibrary.wiley.com/doi/full/10.1002/leap.1099>.
- BORREGO, À., 2014. *Comportament informatiu del professorat de les universitats catalanes*. [en línea]. [Consulta: agosto 2019]. Disponible en:
<https://recercat.cat/handle/2072/242106>.
- BORREGO, À. y ANGLADA, L., 2016. Faculty information behaviour in the electronic environment. En: *New Library World* [en línea], vol. 117, no. 3/4, pp. 173-185. [Consulta: mayo 2019]. Disponible en:
<http://www.emeraldinsight.com/doi/10.1108/NLW-11-2015-0089>.
- BORREGO, A. y URBANO, C., 2007. Analysis of the behaviour of the users of a package of electronic journals in the field of chemistry. En: *Journal of Documentation* [en línea], vol. 63, no. 2, pp. 243-258. [Consulta: agosto 2019]. Disponible en:
<https://www.emeraldinsight.com/doi/10.1108/00220410710737204>.
- BRADLEY, A., 2013. Basic Vocabulary for schema.org and Structured Data. En: *SEO Skeptic* [en línea]. [Consulta: agosto 2019]. Disponible en:
<http://www.seoskeptic.com/basic-vocabulary-for-schema-org-and-structured-data/>.
- BROWN, S. y SWAN, A., 2007. *Researchers' use of academic libraries and their services: a report commissioned by the Research Information Network and the Consortium of Research Libraries*. [en línea]. [Consulta: agosto 2019]. Disponible en:
<https://eprints.soton.ac.uk/263868/>.
- CANDELA, G., ESCOBAR, P., CARRASCO, R. y MARCO, M., 2015. Migration of a library catalogue into RDA linked open data. En: *Semantic Web* [en línea], vol. 0, pp. 1-12. [Consulta: agosto 2019]. Disponible en:
<http://www.semantic-web-journal.net/content/migration-library-catalogue-rda-linked-open-data-0>.
- CASTAGNÉ, M., 2013. *Institutional repository software comparison: DSpace, EPrints, Digital Commons, Islandora and Hydra*. [en línea], [Consulta: agosto 2019]. Disponible en:
<https://open.library.ubc.ca/cIRcle/collections/graduateresearch/42591/items/1.0075768>.
- CHEN, H. y ZHANG, Y., 2014. Functionality Analysis of an Open Source Repository System: Current Practices and Implications. En: *The Journal of Academic Librarianship* [en línea], vol. 40, no. 6, pp. 558-564. [Consulta: agosto 2019]. Disponible en:
<https://linkinghub.elsevier.com/retrieve/pii/S0099133314001773>.
- COAR, 2011. *El caso de Interoperabilidad para Repositorios de Acceso Abierto*. [en línea]. [Consulta: agosto 2019]. Disponible en:
https://www.coar-repositories.org/files/de_la_investigación-a-través-de-redes-globales-de-Repositorios-de-Acceso-Abierto-final-version.pdf.

- CODINA, L., 2004. Posicionamiento Web: Conceptos y Ciclo de Vida. En: *Hipertext.net: Revista Académica sobre Documentación Digital y Comunicación Interactiva* [en línea], vol. 0, no. 2. [Consulta: mayo 2019]. Disponible en: <https://www.raco.cat/index.php/Hipertext/article/view/57793>.
- DRIVER PROJECT, 2008. *Directrices DRIVER 2.0 Directrices para proveedores de contenido - Exposición de recursos textuales con el protocolo OAI-PMH*. [en línea]. S.I.: [Consulta: 21 abril 2019]. Disponible en: https://www.recolecta.fecyt.es/sites/default/files/contenido/documentos/DRIVER_2_1_Guidelines_Spanish.pdf.
- FENNER, M., CROSAS, M., DURAND, G., WIMALARATNE, S., GRÄF, F., HALLETT, R., BERNAL LLINARES, M., SCHINDLER, U. y CLARK, T., 2018. *Listing of data repositories that embed schema.org metadata in dataset landing pages*. [en línea]. [Consulta: julio 2019]. Disponible en: https://zenodo.org/record/1263942#.XTV_zZP7TBI.
- FENNER, M., CROSAS, M., GRETHE, J.S., KENNEDY, D., HERMJAKOB, H., ROCCA-SERRA, P., DURAND, G., BERJON, R., KARCHER, S., MARTONE, M. y CLARK, T., 2019. A data citation roadmap for scholarly data repositories. En: *Scientific Data* [en línea], vol. 6, no. 1, pp. 28. [Consulta: mayo 2019]. Disponible en: <http://www.nature.com/articles/s41597-019-0031-8>.
- FERRERAS FERNÁNDEZ, T., 2018. Los repositorios institucionales: evolución y situación actual en España. En: MERLO VEGA, JA. *Ecosistemas del Acceso Abierto*, [en línea]. Salamanca: Ediciones Universidad de Salamanca, 2018, pp. 39-84. [Consulta: 22 mayo 2019]. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=6746331>.
- FONS, T., PENKA, J. y WALLIS, R., 2012. OCLC's Linked Data Initiative: Using Schema.org to Make Library Data Relevant on the Web. *Information Standards Quarterly* [en línea], vol. 24, no. 2/3, pp. 29. [Consulta: 25 julio 2019]. Disponible en: <http://www.niso.org/publications/isq/2012/v24no2-3/fons/>.
- GARCÍA GARCÍA, A., 2015. *Integración de contenidos semánticos en un portal web de científicos y humanistas valencianos: vestigium* [en línea]. S.I.: Universidad Católica de Valencia San Vicente Mártir. [Consulta: julio 2019]. Disponible en: <https://www.educacion.gob.es/teseo/imprimirFicheroTesis.do?idFichero=3GRWICacf4Q%3D>.
- GARCÍA-MARCO, F.J., 2013. Schema.org: la catalogación revisitada. En: *Anuario ThinkEPI* [en línea], vol. 7, no. 0, pp. 169-172. [Consulta: agosto 2019]. Disponible en: <https://recyt.fecyt.es/index.php/ThinkEPI/article/view/30355>.
- GOOGLE, [sin fecha]. *Guía de optimización en buscadores (SEO) para principiantes*. [en línea]. [Consulta: 21 agosto 2019]. Disponible en: https://support.google.com/webmasters/answer/7451184?hl=es&ref_topic=4564315.

- GOOGLE, [sin fecha]. *Cómo simplificar la estructura de direcciones URL*. [en línea]. [Consulta: agosto 2019]. Disponible en: <https://support.google.com/webmasters/answer/76329?hl=es>.
- GUHA, R. V., BRICKLEY, D. y MACBETH, S., 2016. Schema.org: evolution of structured data on the web. En: *Communications of the ACM* [en línea], vol. 59, no. 2, pp. 44-51. [Consulta: agosto 2019]. Disponible en: <http://dl.acm.org/citation.cfm?doid=2886013.2844544>.
- GUSENBAUER, M., 2019. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. En: *Scientometrics* [en línea], vol. 118, no. 1, pp. 177-214. [Consulta: agosto 2019]. Disponible en: <http://link.springer.com/10.1007/s11192-018-2958-5>.
- HAKIMJAVADI, H. y NOORMAN MASREK, M., 2013. Evaluation of interoperability protocols in repositories of electronic theses and dissertations. En: *Program* [en línea], vol. 47, no. 1, pp. 34-59. [Consulta: agosto 2019]. Disponible en: <https://www.emeraldinsight.com/doi/10.1108/00330331211296303>.
- HALEVI, G., MOED, H. y BAR-ILAN, J., 2017. Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the Literature. En: *Journal of Informetrics* [en línea], vol. 11, no. 3, pp. 823-834. [Consulta: agosto 2019]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S1751157717300676>
- HILLIKER, R.J., WACKER, M. y NURNBERGER, A.L., 2013. Improving Discovery of and Access to Digital Repository Contents Using Semantic Web Standards: Columbia University's Academic Commons. En: *Journal of Library Metadata* [en línea], vol. 13, no. 2-3, pp. 80-94. [Consulta: junio 2019]. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/19386389.2013.826036>.
- IBAI SISTEMAS, 2012. *Informe de adaptación de DSpace a Europea Fase Danubio Europea Data Model (EDM)*. [en línea]. [Consulta: mayo 2019]. Disponible en: <http://travesia.mcu.es>.
- KELLY, B., 2012. MajesticSEO Analysis of Russell Group University Repositories. En: *UK Web Focus* [en línea]. [Consulta: mayo 2019]. Disponible en: <https://ukwebfocus.wordpress.com/2012/08/29/majesticseo-analysis-of-russell-group-university-repositories/>.
- KELLY, B. y NIXON, W., 2013. *SEO analysis of institutional repositories: What's the back story?* [en línea]. [Consulta: mayo 2019]. Disponible en: <https://researchportal.bath.ac.uk/en/publications/seo-analysis-of-institutional-repositories-whats-the-back-story>.
- LEWANDOWSKI, D. y MAYR, P., 2006. *Exploring the academic invisible web*. [en línea], [Consulta: 1 agosto 2019]. Disponible en: <http://eprints.rclis.org/9203/>.

- LÓPEZ-CÓZAR, E.D., ORDUNA-MALEA, E. y MARTÍN-MARTÍN, A., 2018. *Google Scholar as a data source for research assessment*. [en línea]. [Consulta: mayo 2019]. Disponible en: <http://arxiv.org/abs/1806.04435>.
- LOPEZOSA, C., CODINA, L. y CALDERA-SERRANO, J., 2018. SEO semántico: Framework ISS para la optimización de sitios intensivos en contenidos. En: *Cuadernos de Documentación Multimedia* [en línea], vol. 29, no. 1, pp. 97-123. [Consulta: mayo 2019]. Disponible en: <http://revistas.ucm.es/index.php/CDMU/article/view/60607>.
- LORENZO GIL, E., BRAÑA FERREIRO, E. y NIETO CARAMÉS, S., 2015. Estudio de la integración de repositorios en el sistema científico-investigador: alternativas y estado actual. En: *XIV Workshop Rebiun de Poyectos Digitales / VI Jornadas OS-Repositorios (Córdoba, 11-13 de marzo de 2015)* [en línea]. [Consulta: mayo 2019]. Disponible en: <https://helvia.uco.es/xmlui/handle/10396/12631>.
- MACGREGOR, G., 2019. Improving the discoverability and web impact of open repositories: techniques and evaluation. En: *Code4Lib Journal* [en línea], no. 43. [Consulta: 19 julio 2019]. Disponible en: <https://journal.code4lib.org/articles/14180>.
- MACGREGOR, G., 2017. Reviewing repository discoverability : approaches to improving repository visibility and web impact - Strathprints. En: *Repository Fringe 2017, John McIntyre Conference Centre, University of Edinburgh* [en línea]. [Consulta: mayo 2019]. Disponible en: <https://strathprints.strath.ac.uk/61333/>.
- MARTIN-MARTIN, A., ORDUNA-MALEA, E., HARZING, A.-W. y DELGADO LÓPEZ-CÓZAR, E., 2017. Can we use Google Scholar to identify highly-cited documents?. En: *Journal of Informetrics* [en línea], vol. 11, no. 1, pp. 152-163. [Consulta: mayo 2019]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S175115771630298X?via%3Dihub>.
- MASEDA-SECO, D., BUENO-DE-LA-FUENTE, G. y MÉNDEZ, E., 2017. Análisis y categorización de los datos abiertos de las bibliotecas municipales españolas: metadatos, interoperabilidad y propuesta para la apertura y reutilización. En: *El Profesional de la Información* [en línea], vol. 26, no. 3, pp. 392. [Consulta: junio 2019]. Disponible en: <https://recyt.fecyt.es/index.php/EPI/article/view/epi.2017.may.05>.
- OBRIEN, P., ARLITSCH, K., MIXTER, J., WHEELER, J. y STERMAN, L.B., 2017. RAMP – the Repository Analytics and Metrics Portal. En: *Library Hi Tech* [en línea], vol. 35, no. 1, pp. 144-158. [Consulta: 27 junio 2019]. Disponible en: <http://www.emeraldinsight.com/doi/10.1108/LHT-11-2016-0122>.
- OBRIEN, P., ARLITSCH, K., STERMAN, L., MIXTER, J., WHEELER, J. y BORDA, S., 2016. Undercounting File Downloads from Institutional Repositories. En: *Journal of Library Administration* [en línea], vol. 56, no. 7, pp. 854-874. [Consulta: junio 2019]. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/01930826.2016.1216224>.

- OHNO-MACHADO y LUCILA, 2017. Finding useful data across multiple biomedical data repositories using DataMed. En: *Nature Genetics* [en línea], vol. 49, no. 6, pp. 816-819. [Consulta: julio 2019]. Disponible en: <https://www.nature.com/articles/ng.3864>.
- OLLÉ, C. y BORREGO, Á., 2010. A qualitative study of the impact of electronic journals on scholarly information behavior. En: *Library & Information Science Research* [en línea], vol. 32, no. 3, pp. 221-228. [Consulta: 21 agosto 2019]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0740818810000319>.
- ORDUÑA-MALEA, E., 2013. *Impacto de los repositorios a través de técnicas cibernéticas: el caso general de Latinoamérica y especial de Costa Rica*. [en línea], [Consulta: mayo 2019]. Disponible en: <http://digibug.ugr.es/handle/10481/34352>.
- ORDUÑA-MALEA, E. y DELGADO LÓPEZ-CÓZAR, E., 2015. The dark side of open access in Google and Google Scholar: the case of Latin-American repositories. En: *Scientometrics* [en línea], vol. 102, no. 1, pp. 829-846. [Consulta: 21 agosto 2019]. Disponible en: <http://link.springer.com/10.1007/s11192-014-1369-5>.
- ORDUÑA-MALEA, E., MARTÍN-MARTÍN, A., AYLLÓN MILLÁN, J.M. y DELGADO LÓPEZ-CÓZAR, E., 2014. *Are Latin-American repositories invisible on Google and Google Scholar?* [en línea], [Consulta: 21 agosto 2019]. Disponible en: <http://digibug.ugr.es/handle/10481/32363>.
- PARK, M., 2018. SEO for an open access scholarly information system to improve user experience. *Information Discovery and Delivery* [en línea], vol. 46, no. 2, pp. 77-82. [Consulta: mayo 2019]. Disponible en: <https://www.emeraldinsight.com/doi/10.1108/IDD-08-2017-0060>.
- PASTOR SÁNCHEZ, J.A., 2011. *Tecnologías de la web semántica*. Barcelona : UOC. ISBN 9788497884747.
- PASTOR-SÁNCHEZ, J.-A., 2012. Prospectiva de la web semántica: divergencia tecnológica y creación de mercados linked data. En: *Anuario ThinkEPI* [en línea], vol. 6, pp. 269-275. [Consulta: mayo 2019]. Disponible en: <https://recyt.fecyt.es/index.php/ThinkEPI/article/view/30436>.
- PASTOR-SÁNCHEZ, J.-A. y SAORÍN, T., 2017. Web semántica. Informe de situación 2014. En: *Anuario ThinkEPI* [en línea], vol. 1, no. 0, pp. 177. [Consulta: agosto 2019]. Disponible en: <https://recyt.fecyt.es/index.php/ThinkEPI/article/view/info.2015.12>.
- PEKALA, S., 2018. Microdata in the IR: A Low-Barrier Approach to Enhancing Discovery of Institutional Repository Materials in Google. En: *Code4Lib Journal* [en línea], no. 39. [Consulta: julio 2019]. Disponible en: <https://journal.code4lib.org/articles/13191>.
- PRIETO JIMÉNEZ, A.J., 2015. *Indexación y posicionamiento de los repositorios en motores de búsqueda*. [en línea]. [Consulta: mayo 2019]. Disponible en: <https://helvia.uco.es/handle/10396/12652>.

- PYROUNAKIS, G., NIKOLAIDOU, M. y HATZOPOULOS, M., 2014. Building Digital Collections Using Open Source Digital Repository Software. En: *International Journal of Digital Library Systems* [en línea], vol. 4, no. 1, pp. 10-24. [Consulta: mayo 2019]. Disponible en: <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/ijdls.2014010102>.
- RÍOS-HILARIO, A., MARTÍN-CAMPO, D. y FERRERAS-FERNÁNDEZ, T., 2012. Linked Data y Linked Open Data: su implantación en una biblioteca digital. El caso de Europea. En: *El Profesional de la Información* [en línea], vol. 21, no. 3, pp. 292-297. [Consulta: mayo 2019]. Disponible en: <https://recyt.fecyt.es/index.php/EPI/article/view/epi.2012.may.10>.
- RODRIGUES, E., et al., 2017. *Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group*. [en línea], [Consulta: agosto 2019]. Disponible en: <https://zenodo.org/record/1215014#.XUIYgZP7TBI>.
- RONALLO, J., 2012. HTML5 Microdata and Schema.org. En: *Code4Lib Journal* [en línea], no. 16. [Consulta: julio 2019]. Disponible en: <http://journal.code4lib.org/articles/6400>.
- RUBIO LUCAS, M.E., 2015. Propuesta de un vocabulario de microdatos basado en Schema.org para la descripción semántica de publicaciones científicas y la mejora de su visibilidad en entornos abiertos del conocimiento. En: *Cuadernos de gestión de la información* [en línea], vol. 5, no. 1, pp. 32. [Consulta: mayo 2019]. Disponible en: <https://digitum.um.es/digitum/handle/10201/51047>.
- SÁNCHEZ, S. y MELERO, R., 2007. *La denominación y el contenido de los Repositorios Institucionales en Acceso Abierto : base teórica para la "Ruta Verde"*. [en línea], [Consulta: abril 2019]. Disponible en: <https://digital.csic.es/handle/10261/1487>.
- SANCHO REVILLA, J., 2018. *Publicación de datos estructurados en la web usando WordPress* [en línea]. [Consulta: agosto 2019]. Disponible en: <http://uvadoc.uva.es/handle/10324/33161>.
- SANSONE, S.A., et al., 2017. DATS, the data tag suite to enable discoverability of datasets. En: *Scientific Data* [en línea], vol. 4, no. 1, pp. 170059. [Consulta: julio 2019]. Disponible en: <http://www.nature.com/articles/sdata201759>.
- SANTAMARÍA MORILLAS, R., 2017. *Análisis SEO. El sector de la moda online femenina* [en línea]. [Consulta: agosto 2019]. Disponible en: <http://repositorio.upct.es/handle/10317/6024>.
- SERRANO VICENTE, R., MELERO MELERO, R. y ABADAL, E., 2014. Indicadores para la evaluación de repositorios institucionales de acceso abierto. En: *Anales de Documentación* [en línea], vol. 17, no. 2. [Consulta: agosto 2019]. Disponible en: <https://revistas.um.es/analesdoc/article/view/analesdoc.17.2.190821>.

- SERRANO-VICENTE, R., MELERO, R. y ABADAL, E., 2018. Evaluation of Spanish institutional repositories based on criteria related to technology, procedures, content, marketing and personnel. En: *Data Technologies and Applications* [en línea], vol. 52, no. 3, pp. 384-404. [Consulta: agosto 2019]. Disponible en: <https://www.emeraldinsight.com/doi/10.1108/DTA-10-2017-0074>.
- SPINAK, E., 2014. Repositorios en América Latina tienen poca visibilidad en Google Scholar. En: *SciELO en Perspectiva*. [en línea]. [Consulta: agosto 2019]. Disponible en: <https://blog.scielo.org/es/2014/09/18/repositorios-en-america-latina-tienen-poca-visibilidad-en-google-scholar/#.XV0JvJP7TBJ>.
- SULÉ DUESA, A., 2015. Schema.org, la mejora de la visualización de los resultados en los buscadores y mucho más. En: *BiD: Textos universitarios de biblioteconomía i documentació* [en línea], no. 34. [Consulta: mayo 2019]. Disponible en: <http://bid.ub.edu/es/34/sule.htm>.
- SULLIVAN, D., 2016. FAQ: All about the Google RankBrain algorithm. En: *Search Engine Land* [en línea]. [Consulta: agosto 2019]. Disponible en: <https://searchengineland.com/faq-all-about-the-new-google-rankbrain-algorithm-234440>.
- SULLIVAN, D., 2018. A reintroduction to Google's featured snippets. [en línea]. [Consulta: mayo 2019]. Disponible en: <https://www.blog.google/products/search/reintroduction-googles-featured-snippets/>.
- TMAVA, A.M. y ALEMNEH, D.G., 2013. Enhancing content visibility in institutional repositories: Overview of factors that affect digital resources discoverability. En: *iConference 2013 Proceedings* [en línea]. S.l.: iSchools, pp. 855-859. [Consulta: agosto 2019]. Disponible en: <https://www.ideals.illinois.edu/handle/2142/42117>.
- TORRES, D.M.B., ROMERO, A.C. y SANABRIA, J.S.G., 2017. Web semántica, más de una década de su aparición. En: *Puente* [en línea], vol. 8, no. 1, pp. 61-69. [Consulta: julio 2019]. Disponible en: <https://revistas.upb.edu.co/index.php/puente/article/view/7177>.
- W3C, 2015. *RDFa Lite 1.1*. [en línea]. [Consulta: agosto 2019]. Disponible en: <https://www.w3.org/TR/rdfa-lite/>.
- WALLIS, R., 2017. Recommendations for the application of Schema.org to aggregated Cultural Heritage metadata to increase relevance and visibility to search engines: the case of Europeana. En: *Code4Lib Journal* [en línea], no. 36. [Consulta: junio 2019]. Disponible en: <https://journal.code4lib.org/articles/12330>.
- WOLFF, C., SCHONFELD, R. y ROD, A., 2016. *UK Survey of Academics 2015*. [en línea]. [Consulta: agosto 2019]. Disponible en: <http://sr.ithaka.org/?p=282736>.

Glosario

- **Acceso Abierto:** movimiento que promueve el acceso libre y gratuito a la literatura científica, fomentando su libre disponibilidad en Internet y permitiendo a cualquier usuario su lectura, descarga, copia, impresión, distribución o cualquier otro uso legal de la misma, sin ninguna barrera financiera, técnica o de cualquier tipo⁴².
- **ASEO:** (*Academic Search Engine Optimization*) conjunto de prácticas que promueven la visibilidad y la facilidad de acceso de las producciones académicas, típicamente, artículos de revistas científicas, pero también monografías, informes y cualquier otra clase de producto resultante de una investigación⁴³.
- **CTR (Click Through Ratio):** Proporción de clics que obtiene una web respecto de los resultados de búsqueda (impresiones). Se calcula $CTR = clics / impresiones$
- **Datos estructurados:** elementos etiquetados en una web que describen el contenido que se encuentra en una página web, *semantizando* el contenido de estas y que facilitan la comprensión para los motores de búsqueda. Su implementación puede hacerse mediante lenguajes estructurados como Schema.org.
- **Dublin Core:** estándar de metadatos respaldado por la DCMI (Dublin Core Metadata Initiative) para la descripción de documentos con el fin de transmitir un significado semántico a los elementos que describen.
- **Evaluación Heurística (EH):** método de inspección de la usabilidad sin usuarios. Consiste en examinar la calidad de uso de una interfaz por parte de varios evaluadores expertos, a partir del cumplimiento de unos principios reconocidos de usabilidad.
- **HTML5:** (*HyperText Markup Language*, versión 5) Lenguaje HTML que presenta en su versión quinta nuevos elementos y atributos permitiendo a los sitios Web y a las aplicaciones ser más diversas y de más alcance. Permite describir con mayor precisión cuál es el contenido de la web.
- **Impresiones:** Número de veces que una URL de un sitio ha aparecido en los resultados de búsqueda de un usuario.
- **Indexar:** acción de recuperar, leer y añadir una página al índice de un buscador.
- **Interoperabilidad:** capacidad de los sistemas de información para compartir datos y posibilitar el intercambio de información y conocimiento entre ellos de la manera más óptima posible.

⁴² <https://bibliotecas.usal.es/acceso-abierto>

⁴³ <https://www.lluiscodina.com/seo-academico-guia/>

- **JSON-LD:** (JSON for Linking Data) formato de codificación ligero de datos vinculados para entornos web, que facilita la transferencia de información en entornos de Web Semántica.
- **Linked data:** un método de publicación de datos estructurados para que puedan ser entrelazados y se vuelvan más útiles por medio de consultas semánticas⁴⁴.
- **Metadatos:** conjunto de elementos –autor, title, description, keywords, etc– que sirven para describir los contenidos de las distintas páginas web de cara a los buscadores.
- **Mobile First Indexing:** nuevo sistema de rastreo de Google que afecta a todas las páginas web y a la forma en la que se posicionan. Google dará prioridad en su ranking a la versión móvil de los sitios web frente a su versión escritorio, como se hacía hasta ahora.
- **OAI-PMH:** *Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)* – protocolo por el cual los repositorios permiten que sus metadatos sean cosechados y recogidos por otros sistemas.
- **RankBrain:** algoritmo de clasificación del buscador de Google para indexar las webs en el índice de Google.
- **Rastreador:** software automatizado que rastrea (recupera) páginas de la Web y las indexa.
- **Rastrear:** proceso de buscar páginas web nuevas o actualizadas por parte de buscadores. Google descubre URLs por ejemplo siguiendo enlaces o leyendo sitemaps.
- **Resultados de pago:** Los resultados inorgánicos son webs patrocinadas o de pago mostrados en formato de anuncios.
- **Resultados orgánicos:** Los resultados orgánicos son aquellos resultados “naturales” en función de la relevancia que más se acerquen a la consulta de búsqueda del usuario.
- **Rich snippets:** o fragmentos enriquecidos que
- **Robots.txt:** fichero de texto que dice a los buscadores qué directorios, archivos o páginas de una web no deben ser indexados.
- **Schema.org:** vocabulario desarrollado por Google, Yahoo! y Bing que permite marcar y enriquecer los contenidos web, con el fin de mejorar el entendimiento de los motores de búsqueda.
- **SEM:** (*Search Engine Marketing*) es la estrategia de marketing online para posicionar páginas web en los buscadores mediante publicidad online.

⁴⁴ <http://blog.sedici.unlp.edu.ar/2015/03/13/linked-data-recomendaciones-de-la-w3c/>

- **SEO:** (*Search Engine Optimization*) optimización para motores de búsqueda. Se trata de un conjunto de técnicas para mejorar el posicionamiento orgánico con el fin de aumentar la posición en los resultados de las búsquedas de los buscadores (SERP).
- **SERP:** hace referencia a los resultados que muestra una página de buscador como puede ser el de Google, Yahoo! y Bing. Generalmente los buscadores muestran unos diez enlaces/sitios web por página de resultados.
- **Sitemap:** archivo que engloba el listado de páginas, vídeos e imágenes de una web de manera jerarquizada, para ayudar a los buscadores a indexar una determinada web.
- **Sword:** protocolo estándar de interoperabilidad utilizado por los repositorios digitales para que permite la transferencia de contenido desde múltiples fuentes.
- **Web Semántica:** extensión de la Web actual dotada de significado, donde la información que contiene tendría un significado bien definido y etiquetado, de manera que pueda ser comprendida por las máquinas y mejor interpretada por los buscadores.

Índice de imágenes

1. Imagen 1. Acceso a la información por parte de los investigadores (Brown y Swan, 2007). Pág.14
2. Imagen 2. Protocolos que admiten los diferentes softwares de repositorios (Hakimjavadi 2013). Pág. 19
3. Imagen 3: Elementos del SEO On Page (Santamaría, 2017). Pág 23
4. Imagen 4: Elementos del SEO Off Page (Santamaría, 2017). Pág. 24
5. Imagen 5. Tabla Periódica de Factores de Éxito de SEO. Pág. 25
6. Imagen 6. Composición de un snippet de Google Scholar. Pág. 30
7. Imagen 7: Principales fuentes utilizadas para encontrar recursos por disciplina de la asignatura. Pág. 32
8. Imagen 8. Fuentes empleadas para iniciar una búsqueda bibliográfica (Borrego, 2014). Pág. 33
9. Imagen 9. Porcentaje de documentos de los 127 repositorios latinoamericanos indexados en Google y Google Scholar) (Orduña-Malea y Delgado López-Cózar, 2014). Pág.35
10. Imagen 10. Visualización de la información por parte de máquinas y de personas. Pág. 36
11. Imagen 11. Ejemplo tipos y propiedades de Schema.org. Pág. 38

12. Imagen 12. Porque el marcado de Schema es importante para el SEO. Pág. 48
13. Imagen 13. Receta rich snippet. Pág. 49
14. Imagen 14. Libro rich snippet. Pág. 49
15. Imagen 15. Evento rich snippet. Pág. 50
16. Imagen 16. Reserva vuelo rich snippet. Pág.50
17. Imagen 17. Porcentaje de softwares, Agosto 2019 (OpenDoar). Pág. 51
18. Imagen 18. Estado de implementación de los metadatos de Schema.org en las páginas de destino del repositorio (Fenner, 2019). Pág. 58
19. Imagen 19. Metadatos de descubrimiento importantes para los repositorios de datos (Fenner, 2019). Pág. 58
20. Imagen 20. Registro de Europeana con Schema.org (Wallis, 2017). Pág. 60
21. Imagen 21. Digital Metadata Standards at the Hesburgh Libraries. Pág. 64
22. Imagen 22. Revisión de datos estructurados. Pág. 69
23. Imagen 23. Seguimiento con Google Analytics del portal RODERIC. Pág. 70
24. Imagen 24. Comparativa de tráfico de visitas RODERIC. Pág. 71
25. Imagen 25. Comparativa tráfico de referencia para Google y Google Scholar RODERIC. Pág. 72
26. Imagen 26. Comparativa de impresiones web para RODERIC. Pág. 72
27. Imagen 27. Comparativa sobre clics web de RODERIC. Pág. 73
28. Imagen 28. Duplicidad de versiones web en GS de RODERIC. Pág. 74
29. Imagen 29. Ranking web RODERIC. Pág. 75

Índice de notas a pie de página

1. Ver Anexo 4. Indicadores y estadísticas. Pág. 4
2. <https://scientiasalut.gencat.cat/> Pág. 6
3. <http://data.cervantesvirtual.com/about?lang=es> Pág. 8
4. http://v2.sherpa.ac.uk/view/repository_visualisations/1.html Pág. 10
5. El mandato de acceso abierto son una serie de políticas y medidas impulsadas por Instituciones y Gobiernos relacionadas con el acceso abierto. Pág. 10
6. <https://repositories.webometrics.info/es> Pág. 12
7. <https://es.majestic.com/> Pág. 12
8. <https://ahrefs.com/> Pág. 12
9. Ver glosario. Pág. 20

10. Ver glosario. Pág. 20
11. Los resultados de una búsqueda pueden ser orgánicos o de pago (SEM). Los resultados orgánicos son aquellos resultados “naturales” en función de la relevancia que más se acerquen a la consulta de búsqueda del usuario. Los resultados inorgánicos son webs patrocinadas o de pago mostrados en formato de anuncios.
Pág. 21
12. <https://support.google.com/webmasters/answer/7451184?hl=es> Pág. 21
13. Fuente: <https://searchengineland.com/seotable> Pág. 24
14. Ver glosario. Pág. 27
15. Fuente: <https://scholar.google.com/intl/en/scholar/inclusion.html#overview> Pág. 33
16. Fuente: <https://www.w3.org/TR/rdfa-primer/> Pág.36
17. Fuente: <https://schema.org/docs/schemas.html> Pág. 37
18. Fuente:
<http://www.seoskeptic.com/basic-vocabulary-for-schema-org-and-structured-data/>
Pág. 38
19. Fuente: <https://json-ld.org/spec/> Pág. 47
20. Fuente: <https://schema.org/accessibilitySummary> Pág. 47
21. Fuente: seopressor.com Pág. 48
22. Fuente: <https://support.google.com/webmasters/answer/3069489?hl=es> Pág. 50
23. http://v2.sherpa.ac.uk/view/repository_visualisations/1.html Pág. 51
24. <https://www.library.georgetown.edu/digitalgeorgetown> Pág. 57
25. <https://academiccommons.columbia.edu/> Pág. 57
26. <https://datamed.org/> Pág. 57
27. <https://bioschemas.org/> Pág. 59
28. <https://www.europeana.eu/portal/es> Pág. 59
29. <http://roderic.uv.es/statistics> Pág. 60
30. Fuente:
<https://www.uv.es/uvweb/servicio-bibliotecas-documentacion/es/novedades/repositorio-institucional-roderic-cumple-10-anos-1285923456427/Novetat.html?id=1286052422540> Pág. 61
31. Fuente:
<https://www.uv.es/uvweb/servicio-informatica/es/repositorio-institucional-roderic-1285904072947.html> Pág. 62
32. Ver Anexo 1. Metadatos RODERIC. Pág. 62
33. http://ndlib.github.io/metadata_application_profile/elements/ Pág. 63

34. http://ndlib.github.io/metadata_application_profile/elements/ Pág. 64
35. http://www.dublincore.org/specifications/lrmi/lrmi_1/ Pág. 64
36. <https://bibliograph.github.io/BibloGraph-Frozen/bibliograph.net/schemas.html> Pág. 64
37. <https://bibliograph.github.io/BibloGraph-Frozen/bibliograph.net/vocabularydeprecated.html> Pág. 64
38. <http://roderic.uv.es/statistics> Pág. 67
39. <https://search.google.com/structured-data/testing-tool/u/0/?hl=es> Pág. 68
40. Ver Anexo 4. indicadores y estadísticas. Pág. 71
41. <https://search.google.com/structured-data/testing-tool/u/0/?hl=es>. Pág. 77
42. <https://bibliotecas.usal.es/acceso-abierto> Pág. 90
43. <https://www.lluiscodina.com/seo-academico-guia/> Pág. 90
44. <http://blog.sedici.unlp.edu.ar/2015/03/13/linked-data-recomendaciones-de-la-w3c/>
Pág. 91

Anexo 1. Metadatos RODERIC

DUBLIN CORE RODERIC	Definición DC	Schema.org Property	Schema.org Type	Definición Schema.org	Aproximación/Expresividad de términos
dc.contributor.advisor	Asesor de tesis	contributor	schema:Person	Colaborador secundario de CreativeWork o Event.	4
dc.contributor.author	Entidades o personas que tienen la responsabilidad del contenido intelectual de la obra.	author	schema:Person or schema:Organization	Autor de este contenido. Tenga en cuenta que el autor es especial porque HTML 5 proporciona un mecanismo especial para indicar la autoría a través de la etiqueta rel. Eso es equivalente a esto y puede usarse indistintamente.	5
dc.contributor.editor	Entidad responsable de la publicación de la obra; se trata del editor (comercial o no comercial) del recurso	editor	schema:Person	Especifica la persona que editó el trabajo creativo.	5
dc.contributor.illustrator	Ilustrador del libro.	illustrator	schema:Book	The illustrator of the book.	5
dc.contributor.other	Otros colaboradores	contributor	schema:Person or schema:Organization	Colaborador secundario de CreativeWork o Event.	5
dc.contributor	Una persona, organización o servicio responsable del contenido del recurso. Campo para contribuyentes no especificados.	contributor	schema:Person or schema:Organization	Colaborador secundario de CreativeWork o Event.	5
dc.coverage.spatial	Características espaciales del contenido.	contentLocation	schema:Place	Ubicación del contenido.	4
dc.coverage.temporal	Características temporales del contenido.	temporalCoverage	CreativeWork	cobertura temporal de CreativeWork indica el período al que se aplica el contenido	5
dc.creator	No utilizar; solo para metadatos cosechados.	author	schema:Person or schema:Organization	El creador / autor de este CreativeWork o UserComments. Esto es lo mismo que la propiedad de Autor para CreativeWork.	5
dc.date.accessioned	Fecha en que DSpace adquiere el artículo.	uploadDate	schema:Date	Fecha en que se cargó este objeto multimedia en este sitio.	5
dc.date.available	La fecha o fechas en que pasan a estar disponibles para el público.	availabilityStarts	schema:Date	Comienzo de la disponibilidad del producto o servicio	5
dc.date.created	Fecha de creación o fabricación de contenido intelectual si es diferente de la fecha de ...	dateCreated	CreativeWork	La fecha en que se creó CreativeWork.	5
dc.date.copyright	Fecha de copyright.	copyrightYear	CreativeWork		5
dc.date.issued	Fecha de publicación o distribución.	datePublished	CreativeWork	Fecha de la primera emisión / publicación.	5
dc.date.submitted	Recomendado para tesis / disertaciones.	uploadDate	schema:Date	Fecha en que se cargó este objeto multimedia en este sitio.	3
dc.date.updated	La última vez que se actualizó el elemento a través de Interfaz SWORD	dateModified	CreativeWork	La fecha en que CreativeWork se modificó más recientemente o cuando se modificó la entrada del elemento dentro de un DataFeed.	5
dc.date	Usar un formulario calificado si es posible.	startDate	DateTime	La fecha y hora de inicio del elemento (en formato de fecha ISO 8601).	5
dc.description	Usar para cualquier descripción no definida por otros campos.	description	schema:ScholarlyArticle or schema:Thesis schema:Book	Una breve descripción del artículo.	5
dc.description.abstract	Resumen o resumen en el idioma del documento.	description	schema:ScholarlyArticle or schema:Thesis schema:Book	Una breve descripción del artículo.	4
dc.description.abstractenglish	Resumen o resumen en inglés cuando el documento está escrito en un idioma diferente al inglés.	description	schema:ScholarlyArticle or schema:Thesis schema:Book	Una breve descripción del artículo.	4
dc.description.bindingdesc		description	CreativeWork	Una breve descripción del artículo.	2
dc.description.custevent		description	CreativeWork	Una breve descripción del artículo.	2
dc.description.decodesc	Descripción específica para libro antiguo	description	CreativeWork	Una breve descripción del artículo.	2
dc.description.epiteto	Descripción, epiteto. Usado solo para colecciones de historia natural	description	CreativeWork	Una breve descripción del artículo.	2
dc.description.explicit	Descripción	description	CreativeWork	Una breve descripción del artículo.	2
dc.description.extent	Descripción. Extensión	pagination / numberOfPages	ScholarlyArticle / Book	El número de páginas en el libro.	4
dc.description.form	Usado solo para colecciones de historia natural	description	CreativeWork	Una breve descripción del artículo.	
dc.description.gender	Usado solo para colecciones de historia natural	description	CreativeWork	Una breve descripción del artículo.	2
dc.description.handdesc	Utilizado para manuscritos. etiqueta 546 marc	description	CreativeWork	Una breve descripción del artículo.	2
dc.description.incipt	Incipit son las primeras palabras de un texto o documento,	description	CreativeWork	Una breve descripción del artículo.	2
dc.description.location		identifier	schema:ScholarlyArticle or schema:Thesis schema:Book	La propiedad del identificador representa cualquier tipo de identificador para cualquier tipo de cosa	4
dc.description.phase	Usado solo para colecciones de historia natural	description	CreativeWork	Una breve descripción del artículo.	3
dc.description.plumage	Usado solo para colecciones de historia natural	description	CreativeWork	Una breve descripción del artículo.	3
dc.description.private	Notas destinadas a la gestión interna del ítem, que deben quedar ocultas en la visualización de los datos.	description	CreativeWork	Una breve descripción del artículo.	3
dc.description.provenance	El historial de custodia del artículo desde su creación, incluidos los cambios que los custodios sucesivos le hayan hecho.	description	CreativeWork	Una breve descripción del artículo.	3
dc.description.statementofresponsibility	Para preservar la declaración de responsabilidad de los registros MARC.				
dc.description.tableofcontents	Tabla de contenido	description	CreativeWork	Una breve descripción del artículo.	3
dc.description.uri	Identificador uniforme de recursos que apunta a la descripción de este artículo.	URL	schema:ScholarlyArticle or schema:Thesis schema:Book	URL del artículo	4
dc.description.variedadadforma	Forma. Usado solo para colecciones de historia n	description	CreativeWork	Una breve descripción del artículo.	
dc.description.version	Estado de la revisión por pares de un artículo				
dc.embargo.liftdate	Fin del embargo	availabilityEnds	Date	Fecha final	4
dc.embargo.terms	Términos de embargo	description	CreativeWork	Una breve descripción del artículo.	1
dc.format	Campo para cualquier información de formato no definida por calificadores.	description	CreativeWork	Una breve descripción del artículo.	2
dc.format.extent	Tamaño o duración.	pagination / numberOfPages	ScholarlyArticle / Book		4
dc.format.height	Medida de la altura de un video / imagen	description	CreativeWork	Una breve descripción del artículo.	2
dc.format.medium	Physical medium.	description	CreativeWork	Una breve descripción del artículo.	2
dc.format.mimetype	Registered MIME type identifiers	description	CreativeWork	Una breve descripción del artículo.	2
dc.format.width	Medida del ancho de un video / imagen	description	CreativeWork	Una breve descripción del artículo.	2
dc.identifier	Campo para identificadores inequívocos no definidos por forma calificada; use identifier. other para un identificador conocido común a una colección local en lugar de una forma no calificada.	identifier	schema:ScholarlyArticle or schema:Thesis schema:Book	identificador	5
dc.identifier.citation	Cita bibliográfica estándar	CITATION	schema:ScholarlyArticle or schema:Thesis schema:Book	Una cita o referencia a otro trabajo creativo, como otra publicación, página web, artículo académico, etc.	5
dc.identifier.doi	Identificador DOI	sameAs	ScholarlyArticle		
dc.identifier.govdoc	Identificador número de documento del gobierno	identifier	schema:ScholarlyArticle or schema:Thesis schema:Book	Identificador	4
dc.identifier.idgrec	Identificador del registro en GREC	identifier	schema:ScholarlyArticle or schema:Thesis schema:Book	Identificador	4
dc.identifier.isbn	Número de ISBN	ISBN	schema:Book	ISBN del documento	5

Anexo 1. Metadatos RODERIC

DUBLIN CORE RODERIC	Definición DC	Schema.org Property	Schema.org Type	Definición Schema.org	Aproximación/Expresividad de términos
dc.identifier.ismn	International Standard Music Number	identifier	schema:ScholarlyArticle or schema:Thesis schema:Book	Identificador	
dc.identifier.issn	International Standard Serial Number	ISSN	ScholarlyArticle	ISSN de la publicación	5
dc.identifier.issue	Número de revista o publicación	identifier	schema:ScholarlyArticle or schema:Thesis schema:Book	Identificador	3
dc.identifier.local	Numero ID de Millenium o del Catálogo	identifier	schema:ScholarlyArticle or schema:Thesis schema:Book	Identificador	3
dc.identifier.other	Identificador de la colección local.	identifier	schema:ScholarlyArticle or schema:Thesis schema:Book	Identificador	3
dc.identifier.sici	Elemento de serie e identificador de contribución	identifier	schema:ScholarlyArticle or schema:Thesis schema:Book	Identificador	3
dc.identifier.slug	uri del encabezado o uri sugerida para el artículo	URL	schema:ScholarlyArticle or schema:Thesis schema:Book	URL del artículo.	
dc.identifier.thumbnail	Thumbnail	thumbnailUrl	schema:ScholarlyArticle or schema:Thesis schema:Book	Miniatura relevante de una cosa	5
dc.identifier.uri	Uniform Resource Identifier (URI)	URL	schema:ScholarlyArticle or schema:Thesis schema:Book	URL del artículo.	4
dc.identifier.urlrecursoscream	URL para materiales del CREAM	URL	schema:ScholarlyArticle or schema:Thesis schema:Book	URL del artículo.	4
dc.identifier.urivideoeng	Uri del video en inglés	URL	schema:ScholarlyArticle or schema:Thesis schema:Book	URL del artículo.	4
dc.identifier.urivideoval	Uri del video en valenciano	URL	schema:ScholarlyArticle or schema:Thesis schema:Book	URL del artículo.	4
dc.identifier.videoflvmp4	Nombre del archivo de video flv o mp4	fileFormat	CreativeWork	Tipo de medio o formato	3
dc.identifier.videomms	Nombre del archivo de video mms				
dc.identifier.volume	Volumen de revista	volumeNumber	PublicationVolume	Identifica el volumen de publicación o trabajo de varias partes	5
dc.language	Campo para formas no ISO del lenguaje del artículo	inLanguage	CreativeWork	Lenguaje del contenido o rendimiento o utilizado en una acción	4
dc.language.iso	ISO actual para el idioma del contenido intelectual, incluidos los códigos de país (por ejemplo, "en_US")	inLanguage	CreativeWork	Lenguaje del contenido o rendimiento o utilizado en una acción	4
dc.language.iso639-2	Lengua ISO	inLanguage	CreativeWork	Lenguaje del contenido o rendimiento o utilizado en una acción	
dc.language.rfc3066	the rfc3066 form of the language for the item				
dc.provenance	El historial de custodia del artículo desde su creación, incluidos los cambios que los custodios sucesivos le hayan hecho.				
dc.publisher	Entidad responsable de la publicación, distribución o impresión.	publisher	schema:ScholarlyArticle or schema:Thesis schema:Book	El editor del trabajo creativo.	5
dc.relation.haspart	Referencias de elementos contenidos física o lógicamente.	hasPart	CreativeWork	Indica un elemento o CreativeWork que forma parte de este elemento, o CreativeWork (en cierto sentido).	5
dc.relation.hasversion	Referencias versión posterior				
dc.relation.isbasedon	Fuente de referencias.	isBasedOn	CreativeWork	Un recurso del cual se deriva este trabajo o del cual es una modificación o adaptación.	5
dc.relation.isbasisfor	El material es la base el uno para el otro.				
dc.relation.isformatof	Referencias forma física adicional				
dc.relation.ispartof	Referencias que contienen elementos físicos o lógicos.	isPartOf or issueNumber	ScholarlyArticle or PublicationIssue	Identifica el número de la publicación	5
dc.relation.ispartofseries	Nombre y número de serie dentro de esa serie, si está disponible.	isPartOf or issueNumber	ScholarlyArticle or PublicationIssue	Identifica el número de la publicación	5
dc.relation.isreferencedby	Señalado por recurso referenciado.	claimReviewed	thing		3
dc.relation.isreplacedby	Referencias elemento siguiente	Version	CreativeWork	La versión de CreativeWork incorporada por un recurso específico.	3
dc.relation.isversionof	Referencias versión anterior	Version	CreativeWork	La versión de CreativeWork incorporada por un recurso específico.	4
dc.relation.references	El material hace referencia a otros	mentions	CreativeWork	Contiene una referencia a, pero no necesariamente se trata de un concepto.	4
dc.relation.replaces	Referencias artículo anterior				
dc.relation.requires	Se requiere un recurso referenciado para soportar la función, entrega o coherencia del artículo.				
dc.relation.uri	Referencias Uniform Resource Identifier for related item.	URL	schema:ScholarlyArticle or schema:Thesis schema:Book	URL del artículo.	3
dc.relation	Campo para referencias a otros artículos relacionados.	mentions	CreativeWork	Contiene una referencia a, pero no necesariamente se trata de un concepto.	4
dc.rights holder	El dueño de los derechos de autor				
dc.rights.uri	Referencia URI a los términos que rigen el uso y la reproducción.	license	schema:ScholarlyArticle or schema:Thesis schema:Book	licencia que se aplica a este contenido, normalmente indicado por URL	4
dc.rights	Derechos de uso y la reproducción	license	schema:ScholarlyArticle or schema:Thesis schema:Book	licencia que se aplica a este contenido, normalmente indicado por URL	4
dc.source	Do not use; only for harvested metadata.				
dc.source.uri	Do not use; only for harvested metadata.	URL	schema:ScholarlyArticle or schema:Thesis schema:Book	URL del artículo.	3
dc.subject	Materia	Keywords	schema:ScholarlyArticle or schema:Thesis schema:Book	Palabras clave o etiquetas utilizadas para describir este contenido. Las entradas múltiples en una lista de palabras clave generalmente están delimitadas por comas.	4
dc.subject.classification	Campo para el valor del sistema de clasificación local; los sistemas de clasificación global recibirán un calificador específico	Keywords	schema:ScholarlyArticle or schema:Thesis schema:Book	Palabras clave o etiquetas utilizadas para describir este contenido. Las entradas múltiples en una lista de palabras clave generalmente están delimitadas por comas.	3
dc.subject.ddc	Dewey Decimal Classification Number	indentifier	CreativeWork	Identificador	3
dc.subject.family	Usado solo para colecciones de historia natural	Keywords	CreativeWork	Palabras clave o etiquetas utilizadas para describir este contenido. Las entradas múltiples en una lista de palabras clave generalmente están delimitadas por comas.	4
dc.subject.group	Usado solo para colecciones de historia natural	identifier	CreativeWork	Identificador	3

Anexo 1. Metadatos RODERIC

DUBLIN CORE RODERIC	Definición DC	Schema.org Property	Schema.org Type	Definición Schema.org	Aproximación/Expresividad de términos
dc.subject.lcc	Library of Congress Classification Number	identifier	CreativeWork	Identificador	3
dc.subject.lcsh	Library of Congress Subject Headings	Keywords	schema:ScholarlyArticle or schema:Thesis schema:Book	Palabras clave o etiquetas utilizadas para describir este contenido. Las entradas múltiples en una lista de palabras clave generalmente están delimitadas por comas.	4
dc.subject.mesh	Medical Subject Headings	Keywords	schema:ScholarlyArticle or schema:Thesis schema:Book	Palabras clave o etiquetas utilizadas para describir este contenido. Las entradas múltiples en una lista de palabras clave generalmente están delimitadas por comas.	4
dc.subject.other	Vocabulario controlado localmente; los vocabularios globales recibirán un calificador específico.	Keywords	schema:ScholarlyArticle or schema:Thesis schema:Book	Palabras clave o etiquetas utilizadas para describir este contenido. Las entradas múltiples en una lista de palabras clave generalmente están delimitadas por comas.	4
dc.subject.species	Usado solo para colecciones de historia natural	Keywords	schema:ScholarlyArticle or schema:Thesis schema:Book	Palabras clave o etiquetas utilizadas para describir este contenido. Las entradas múltiples en una lista de palabras clave generalmente están delimitadas por comas.	4
dc.subject.subspecies	Usado solo para colecciones de historia natural	Keywords	schema:ScholarlyArticle or schema:Thesis schema:Book	Palabras clave o etiquetas utilizadas para describir este contenido. Las entradas múltiples en una lista de palabras clave generalmente están delimitadas por comas.	4
dc.subject.unesco	Campo subject que toma valores de un vocabulario controlado según la clasificación unesco con 2 dígitos	Keywords	schema:ScholarlyArticle or schema:Thesis schema:Book	Palabras clave o etiquetas utilizadas para describir este contenido. Las entradas múltiples en una lista de palabras clave generalmente están delimitadas por comas.	4
dc.title	Título	name	schema:ScholarlyArticle or schema:Thesis schema:Book	Nombre del documento o ítem	5
dc.title.alternative	Varía (o sustituye) la forma del título que aparece correctamente en el elemento, p. abreviatura o traducción	alternateName	schema:ScholarlyArticle or schema:Thesis schema:Book	Un alias para el artículo.	5
dc.type	Naturaleza o género de contenido.	additionalType	schema:ScholarlyArticle or schema:Thesis schema:Book	Tipo adicional para el elemento	4
dc.type.interactivitytype	Usado por LOM. Nivel de interacción soportado por el material.				
dc.type.learningresourcetype	Usado por LOM. Tipo de material de aprendizaje.	learningResourceType	CreativeWork	El tipo predominante o tipo que caracteriza el recurso de aprendizaje. Por ejemplo, 'presentación', 'folleto'.	3

Anexo 2. Mapeo RODERIC - Schema.org

Tipo documental	DUBLIN CORE RODERIC	Schema.org Property	Schema.org Type	Tipo documental	DUBLIN CORE RODERIC	Schema.org Property	Schema.org Type	Tipo documental	DUBLIN CORE RODERIC	Schema.org Property	Schema.org Type
Articulos	dc.contributor.author	author	schema:Person or schema:Organization	Tesis	dc.contributor.advisor	contributor	schema:Person	Libros / Monografias	dc.contributor.author	author	schema:Person or schema:Organization
Articulos	dc.contributor.editor	editor	schema:Person	Tesis	dc.contributor.author	author	schema:Person or schema:Organization	Libros / Monografias	dc.contributor.editor	editor	schema:Person
Articulos	dc.contributor.other	contributor	schema:Person or schema:Organization	Tesis	dc.contributor.editor	editor	schema:Person	Libros / Monografias	dc.contributor.illustrator	illustrator	schema:Book
Articulos	dc.contributor	contributor	schema:Person or schema:Organization	Tesis	dc.contributor.other	contributor	schema:Person or schema:Organization	Libros / Monografias	dc.contributor.other	contributor	schema:Person or schema:Organization
Articulos	dc.creator	author	schema:Person or schema:Organization	Tesis	dc.contributor	contributor	schema:Person or schema:Organization	Libros / Monografias	dc.contributor	contributor	schema:Person or schema:Organization
Articulos	dc.date.created	dateCreated	schema:ScholarlyArticle	Tesis	dc.creator	author	schema:Person or schema:Organization	Libros / Monografias	dc.creator	author	schema:Person or schema:Organization
Articulos	dc.date.issued	datePublished	schema:ScholarlyArticle	Tesis	dc.date.created	dateCreated	schema:Thesis	Libros / Monografias	dc.date.created	dateCreated	schema:Book
Articulos	dc.date.updated	dateModified	schema:ScholarlyArticle	Tesis	dc.date.issued	datePublished	schema:Thesis	Libros / Monografias	dc.date.copyright	copyrightYear	schema:Book
Articulos	dc.date	date	schema:Date	Tesis	dc.date.submitted	uploadDate	schema:Date	Libros / Monografias	dc.date.issued	datePublished	schema:Book
Articulos	dc.description	description	schema:ScholarlyArticle	Tesis	dc.date.updated	dateModified	schema:Thesis	Libros / Monografias	dc.date.updated	dateModified	schema:Book
Articulos	dc.description.abstract	description	schema:ScholarlyArticle	Tesis	dc.date	date	schema:Date	Libros / Monografias	dc.date	date	schema:Date
Articulos	dc.description.english	description	schema:ScholarlyArticle	Tesis	dc.description	description	schema:Thesis	Libros / Monografias	dc.description	description	schema:Book
Articulos	dc.description.extent	pagination	schema:ScholarlyArticle	Tesis	dc.description.abstract	description	schema:Thesis	Libros / Monografias	dc.description.abstract	description	schema:Book
Articulos	dc.description.version	pagination	schema:ScholarlyArticle	Tesis	dc.description.english	description	schema:Thesis	Libros / Monografias	dc.description.abstract.english	description	schema:Book
Articulos	dc.description.liftdate	pagination	schema:ScholarlyArticle	Tesis	dc.description.terms	description	schema:Thesis	Libros / Monografias	dc.description.terms	description	schema:Book
Articulos	dc.description.terms	pagination	schema:ScholarlyArticle	Tesis	dc.description.extent	description	schema:Thesis	Libros / Monografias	dc.description.extent	description	schema:Book
Articulos	dc.format.extent	pagination	schema:ScholarlyArticle	Tesis	dc.description.terms	pagination	schema:Thesis	Libros / Monografias	dc.description.extent	pagination	schema:Book
Articulos	dc.identifier	identifier	schema:ScholarlyArticle	Tesis	dc.format.extent	pagination	schema:Thesis	Libros / Monografias	dc.identifier	identifier	schema:Book
Articulos	dc.identifier.citation	CITATION	schema:ScholarlyArticle	Tesis	dc.identifier	identifier	schema:Thesis	Libros / Monografias	dc.identifier.citation	CITATION	schema:Book
Articulos	dc.identifier.doi	sameAs	schema:ScholarlyArticle	Tesis	dc.identifier.citation	Citation	schema:Thesis	Libros / Monografias	dc.identifier.isbn	ISBN	schema:Book
Articulos	dc.identifier.isbn	ISSN	schema:ScholarlyArticle	Tesis	dc.identifier.other	identifier	schema:Thesis	Libros / Monografias	dc.identifier.other	identifier	schema:Book
Articulos	dc.identifier.issue	issueNumber	schema:ScholarlyArticle	Tesis	dc.identifier.thumbnaill	thumbnailUrl	schema:Thesis	Libros / Monografias	dc.identifier.thumbnaill	thumbnailUrl	schema:Book
Articulos	dc.identifier.other	identifier	schema:ScholarlyArticle	Tesis	dc.identifier.uri	URL	schema:Thesis	Libros / Monografias	dc.identifier.uri	URL	schema:Book
Articulos	dc.identifier.thumbnaill	thumbnailUrl	schema:ScholarlyArticle	Tesis	dc.language	inLanguage	schema:Thesis	Libros / Monografias	dc.language	inLanguage	schema:Book
Articulos	dc.identifier.uri	URL	schema:ScholarlyArticle	Tesis	dc.language.iso	inLanguage	schema:Thesis	Libros / Monografias	dc.language.iso	inLanguage	schema:Book
Articulos	dc.identifier.volume	volumeNumber	schema:ScholarlyArticle	Tesis	dc.publisher	publisher	schema:Thesis	Libros / Monografias	dc.publisher	publisher	schema:Book
Articulos	dc.language	inLanguage	schema:ScholarlyArticle	Tesis	dc.relation.uri	URL	schema:Thesis	Libros / Monografias	dc.relation.uri	URL	schema:Book
Articulos	dc.language.iso	inLanguage	schema:ScholarlyArticle	Tesis	dc.relation.uri	URL	schema:Thesis	Libros / Monografias	dc.relation.uri	URL	schema:Book
Articulos	dc.publisher	publisher	schema:ScholarlyArticle	Tesis	dc.rights.uri	license	schema:Thesis	Libros / Monografias	dc.rights.uri	license	schema:Book
Articulos	dc.relation.haspart	hasPart	schema:ScholarlyArticle	Tesis	dc.subject	Keywords	schema:Thesis	Libros / Monografias	dc.subject	Keywords	schema:Book
Articulos	dc.relation.ispartof	isPartOf or isPartOfSeries	schema:ScholarlyArticle	Tesis	dc.subject.classification	Keywords	schema:Thesis	Libros / Monografias	dc.subject.classification	Keywords	schema:Book
Articulos	dc.relation.ispartof	isPartOf or issueNumber	schema:ScholarlyArticle or PublicationIssue	Tesis	dc.subject.unesco	Keywords	schema:Thesis	Libros / Monografias	dc.subject.unesco	Keywords	schema:Book
Articulos	dc.relation.uri	URL	schema:ScholarlyArticle or PublicationIssue	Tesis	dc.title	name	schema:Thesis	Libros / Monografias	dc.title	name	schema:Book
Articulos	dc.rights.uri	license	schema:ScholarlyArticle	Tesis	dc.title.alternative	alternateName	schema:Thesis	Libros / Monografias	dc.title.alternative	alternateName	schema:Book
Articulos	dc.subject	Keywords	schema:ScholarlyArticle	Tesis	dc.type	dc.type	schema:Thesis	Libros / Monografias	dc.type	dc.type	schema:Book
Articulos	dc.subject.classification	Keywords	schema:ScholarlyArticle	Tesis							
Articulos	dc.subject.mesh	Keywords	schema:ScholarlyArticle	Tesis							
Articulos	dc.subject.unesco	Keywords	schema:ScholarlyArticle	Tesis							
Articulos	dc.title	name	schema:ScholarlyArticle	Tesis							
Articulos	dc.title.alternative	alternateName	schema:ScholarlyArticle	Tesis							
Articulos	dc.type	additionalType	schema:ScholarlyArticle	Tesis							

Anexo 3. Código JSON-LD para RODERIC

```
<!-- MODIF 30-05-2019: JSON-LD PARA METADATOS schema.org-->
<!-- captura de variables -->
<xsl:variable name="type_modif_schema_org">
<xsl:choose>

<xsl:when test="contains(dim:field[@element='type'],'info:eu-repo/semantics/doctoralThesis')">
<xsl:text>thesis</xsl:text>
</xsl:when>
<xsl:when test="contains(dim:field[@element='type'],'info:eu-repo/semantics/masterThesis')">
<xsl:text>schema:Thesis</xsl:text>
</xsl:when>
<xsl:when test="contains(dim:field[@element='type'],'info:eu-repo/semantics/book')">
<xsl:text>book</xsl:text>
</xsl:when>
<xsl:when test="contains(dim:field[@element='type'],'info:eu-repo/semantics/article')">
<xsl:text>ScholarlyArticle</xsl:text>
</xsl:when>
<xsl:otherwise>
<xsl:value-of select="dim:field[@element='type']"/>
</xsl:otherwise>
</xsl:choose>
</xsl:variable>

<!--<xsl:variable name="additionaltype_schema_org" select="dim:field[@element='type']" />-->
<xsl:variable name="additionaltype_schema_org">
<xsl:choose>
<xsl:when test="contains(dim:field[@element='type'],'info:eu-repo/semantics/article')">
<xsl:text></xsl:text>
</xsl:when>
<xsl:otherwise>
<xsl:value-of select="dim:field[@element='type']"/>
</xsl:otherwise>
</xsl:choose>
</xsl:variable>

<!-- metadatos comunes a los tres tipos-->
```

```
<xsl:variable name="name_schema_org" select="dim:field[@element='title']" />
<xsl:variable name="alternate_name_org" select="dim:field[@element='title' and
@qualifier='alternative']" />
```

```
<xsl:variable name="id_schema_org">
<xsl:for-each select="dim:field[@element='identifier' and @qualifier='uri']">
<xsl:if test="contains(/.node(), 'https://hdl.handle.net/10550')">
<xsl:value-of select="/.node()"/>
</xsl:if>
</xsl:for-each>
</xsl:variable>
```

```
<xsl:variable name="contributor_schema_org" select="dim:field[@element='contributor']" />
<xsl:variable name="contrib_other_schema_org"
select="dim:field[@element='contributor'][@qualifier='other']" />
```

```
<xsl:variable name="about_schema_org"
select="/dri:document/dri:meta/dri:pageMeta/dri:metadata[@element='subject']" />
```

```
<xsl:variable name="description_schema_org">
<xsl:choose>
<xsl:when test="string-length(dim:field[@element='description' and @qualifier='abstract']) > 0">
<xsl:value-of select="dim:field[@element='description' and @qualifier='abstract']" />
</xsl:when>
<xsl:when test="not(string-length(dim:field[@element='description' and @qualifier='abstract']) >
0)">
<xsl:if test="string-length(dim:field[@element='description' and @qualifier='abstractenglish']) > 0">
<xsl:value-of select="dim:field[@element='description' and @qualifier='abstractenglish']"/>
</xsl:if>
</xsl:when>
<xsl:otherwise>
<xsl:text></xsl:text>
</xsl:otherwise>
</xsl:choose>
</xsl:variable>
```

```

<xsl:variable name="pagination_schema_org" select="dim:field[@element='description' and
@qualifier='extent']" />
<xsl:variable name="format_extent_schema_org" select="dim:field[@element='format' and
@qualifier='extent']" />

<xsl:variable name="publisher_schema_org" select="dim:field[@element='publisher']" />

<xsl:variable name="ispartof_schema_org" select="dim:field[@element='relation' and
@qualifier='ispartof']" />
<xsl:variable name="ispartof_series_schema_org" select="dim:field[@element='relation' and
@qualifier='ispartofseries']" />

<xsl:variable name="keywords_schema_org" select="dim:field[@element='subject' and
@qualifier='unesco']" />

<xsl:variable name="thumbnail_schema_org" select="dim:field[@element='identifier' and
@qualifier='thumbnail']" />

<!-- tesis -->
<xsl:variable name="advisor_schema_org"
select="/dri:document/dri:meta/dri:pageMeta/dri:metadata[@element='contributor'][@qualifier='advisor
']/node()" />

<!-- books -->
<xsl:variable name="contrib_ilustrator_schema_org"
select="dim:field[@element='contributor'][@qualifier='illustrator']" />

<xsl:variable name="datepublished_schema_org" select="dim:field[@element='date' and
@qualifier='issued']" />

<!-- representacion en json-ld -->
<script type="application/ld+json">{"@context":"https://schema.org"
,"@id":"<xsl:value-of select="$id_schema_org"/>"
,"@type":"<xsl:value-of select="$type_modif_schema_org"/>"
,"name":"<xsl:value-of select="$name_schema_org"/>"
,"headline":"<xsl:value-of select="$name_schema_org"/>"
<xsl:if test="string-length($thumbnail_schema_org) > 0">,"thumbnailUrl":"<xsl:value-of
select="$thumbnail_schema_org"/>"</xsl:if>

```

```

,"alternatename": "<xsl:value-of select=\"$alternate_name_org\"/>"
,"author": [ <xsl:for-each select="dim:field[@element='contributor' and
@qualifier='author']"> {"@type": "Person", "name": "
<xsl:value-of select=" ./node()"/>" <xsl:if test="count(following-sibling::dim:field[@element='contributor'
and @qualifier='author'])>0">,
</xsl:if></xsl:for-each>]
<xsl:for-each select="dim:field[@element='creator']">
,"author": [{"@type": "Person", "name": "<xsl:value-of select=" ./node()"/>"}]
</xsl:for-each>
<xsl:for-each select="dim:field[@element='contributor' and @qualifier='advisor']">
,"contributor": [{"@type": "Person", "name": "<xsl:value-of select=" ./node()"/>"}]
</xsl:for-each>
,"keywords": [ <xsl:for-each select="dim:field[@element='subject' and not(@qualifier)]"> <xsl:value-of
select=" ./node()"/> <xsl:if test="count(following-sibling::dim:field[@element='subject' and
@qualifier='unesco'])>0">, </xsl:if></xsl:for-each> <xsl:for-each select="dim:field[@element='subject'
and @qualifier='unesco']"> <xsl:value-of select=" ./node()"/> "

<xsl:if test="count(following-sibling::dim:field[@element='subject' and
@qualifier='unesco'])>0">, </xsl:if></xsl:for-each>]
<xsl:if test="string-length($description_schema_org) > 0">
,"description": "<xsl:value-of select="$description_schema_org\"/>"
</xsl:if>
<xsl:if test="string-length($pagination_schema_org) > 0">
,"description": "<xsl:value-of select="$pagination_schema_org\"/>"
</xsl:if>
<xsl:if test="contains($type_modif_schema_org, 'ScholarlyArticle')">
<xsl:if test="string-length($format_extent_schema_org) > 0">
,"pagination": "<xsl:value-of select="$format_extent_schema_org\"/>"
</xsl:if>

</xsl:if>
,"publisher": "<xsl:value-of select="$publisher_schema_org\"/>"
,"datePublished": "<xsl:value-of select="$datepublished_schema_org\"/>"
,"citation": {"@id": "<xsl:value-of select="$id_schema_org\"/>"} }</script>

```

[!- fin.schema.org ->](http://fin.schema.org)

Anexo 4. Indicadores y estadísticas

Agradecimientos

Sin duda este ha sido un largo viaje... Empezó en Barcelona y terminó en Valencia. Toda una travesía que me ha enseñado a volver a empezar y cuánto queda por descubrir. Por eso quiero agradecer en primer lugar a la Universitat de Barcelona, a mis profesores y compañeros del máster, por esta experiencia. Ha sido una gran aventura en la que he descubierto gente maravillosa, que me llevo conmigo y que me ha aportado muchísimo. En segundo lugar me gustaría dar las gracias a mi tutor, Rubén Alcaraz, que me sacó de mi zona de confort y que me llevó más allá. Que me invitó a investigar y que me propuso un proyecto como este, del que he aprendido tanto. Le agradezco todas esas horas por Skype, el estar siempre ahí, a pesar de los vaivenes. Gracias por tu compromiso, dedicación y por enseñarme tanto.

Quiero agradecer también, de una forma muy especial, a los compañeros de la Universitat de València, por tanto... Gracias!. Gracias por abrirme las puertas de vuestra casa y hacerme sentir un cariño sin igual, por apoyarme con un proyecto como este y por estar dispuestos a avanzar sin complejos ni miedos en un camino como el de la visibilidad y difusión de la información.

Dejo para el final a a 3 personas, que me han acompañado en este y en otros tantos viajes...y sin los que hoy no estaría aquí. A mi madre, que me inculcó siempre estudiar y aprender, que puso todo su esfuerzo en que estudiara sin importar los límites. A Néstor, mi Cruz del Mapa... que cree en mí más que yo misma, que siempre insiste en que haga más y que me ayuda a ser mejor. Y a mi sonrisa, Hugo, mi niño, mi luz... a quien he robado tantas horas de juegos, pero que ha aprendido que esforzarse tiene recompensa.

A todos.... Gracias!