



Memoria descriptiva CPP

Infraestructura semántica basada en el
paradigma de datos abiertos para la gestión
de investigación de las universidades
Españolas

Índice

1	Principales objetivos del proyecto.....	3
1.1	Web semántica	5
1.2	Sistema de gestión de las universidades	8
1.3	Cloud Computing.....	11
1.4	Administración Electrónica.....	12
1.5	Datos abiertos y enlazados.....	14
1.5.1	Iniciativas de Open Data y Linked Data.....	16
1.6	Relación entre Open Data y Big Data.....	20
1.7	Relación con líneas de investigación estratégicas internacionales	20
1.8	Escenarios de uso en gestión de la investigación.....	23
2	Estimación de tareas (costes) del proyecto.....	25
2.1.1	PT 1: Gestión y dirección	25
2.1.2	PT 2: Marco tecnológico, Estrategias Cloud y eAdmon	25
2.1.3	PT 3: Gestión de calidad y riesgos	27
2.1.4	PT 4: Desarrollo de un estudio para la selección de módulos del nuevo sistema de gestión de investigación universitaria.....	27
2.1.5	PT 5. Creación de la infraestructura semántica	29
2.1.6	PT 6. Casos de uso de explotación semántica.....	30
2.1.7	PT 7. Plan de pruebas.....	30
2.1.8	PT8. Implantación en la universidad licitante de la CPP.....	31
2.1.9	PT9. Gestión del Cambio.....	31
3	Bibliografía.....	33

1 Principales objetivos del proyecto

Actualmente existen en España 79 Universidades distribuidas por todo el territorio nacional:



Cada una de ellas cuenta con un sistema de gestión propio, desarrollado por la propia universidad en algunos casos. La colaboración interuniversitaria en materia de “Sistemas de Gestión” hasta la fecha ha sido muy pobre y puramente voluntarista, sin ninguna coordinación por parte del Gobierno central. Únicamente desde la CRUE, a través de su sectorial CRUE-TIC, se han venido, y desde hace pocos años, dando pasos simplemente orientativos y de buenas prácticas.

En la práctica, esto significa que cada Universidad tiene definido unos procedimientos de gestión particulares, con unos modelos y esquemas de datos definidos según sus criterios y cuya explotación combinada por parte de terceros presenta los problemas tradicionales de interoperabilidad de fuentes de información heterogénea, no basada en especificaciones semánticas y formales de los esquemas de datos y conocimiento.

Por mencionar algunos de ellos:

- Análisis sesgado de las necesidades de información
- No guardar ciertos datos de interés
- Aplicar clasificaciones incompatibles de los datos
- Representación de entidades a distinto nivel de granularidad
- Dificultad para discernir si datos procedentes de distintas fuentes tienen el mismo significado

Además de la docencia, las universidades participan activamente en los programas de investigación. La gestión de dicha participación es llevada a cabo por estos sistemas de gestión propios que son dispares y no están normalizados, y que dan soporte, entre otros, a los procesos de:

- Preparación y seguimiento de proyectos
- Gestión de resultados de proyectos de I+D+i

- Gestión económica y administrativa integrada. Contabilidad analítica.

Esta situación y en concreto lo que respecta a los sistemas de gestión de la investigación, provoca grandes ineficiencias e ineficacia en la gestión de la información y el conocimiento del sistema de investigación de las universidades españolas. Ello conlleva costes adicionales derivados de la realización de tareas de explotación con conjuntos parciales de los datos en cada universidad, que luego es necesario homogeneizar.

No es diferente la situación en el resto de áreas de gestión de las universidades, siendo las principales las de gestión académica, económica y de personal.

Por lo tanto, es necesario un nuevo sistema de gestión de investigación universitaria, que disponga de capacidades semánticas y que sea homogéneo para todas las universidades (o para un gran número de ellas) para poder conseguir los siguientes beneficios:

- Realizar explotación conjunta de información.
- Unificar los criterios para la obtención de información, ofreciendo mayores garantías de una adecuada interpretación de la información y, con ello, la exactitud de los indicadores obtenidos.
- Poder establecer sinergias y colaboraciones entre universidades y grupos de investigación
- Transparencia
- Facilidades para la transferencia tecnológica y para la colaboración universidad empresa
- Facilitar el conocimiento de la producción científica, para el resto de investigadores y para la sociedad en general.
- Facilitar la integración del currículum vitae normalizado. Mayor facilidad para la movilidad del PDI entre las universidades españolas.
- Se va a producir una externalidad positiva del hecho de que se homogeneicen los datos que manejamos las universidades. Esta es la de proporcionar con mayor facilidad al usuario, al contribuyente y a la sociedad datos estadísticos que puedan ser relevantes desde el punto de la transparencia en el ejercicio del servicio público, la recuperación del prestigio que las Universidades no han perdido pero tampoco pueden fácilmente demostrar a diario y volver a generar seducción entre los posibles alumnos que desean acertar eligiendo el mejor centro Universitario público acorde a sus expectativas
- Y, en definitiva, permitir la explotación conjunta de información de investigación de todas las universidades, permitiendo con ello una total transparencia en la gestión universitaria.

El desarrollo de este proyecto, convertiría a España en un referente internacional, lo que facilitaría, asimismo, la internalización de las empresas adjudicatarias.

Actualmente CRUE cuenta con 79 universidades. Gran parte de ellas han manifestado, por escrito y con la firma de su rector, su interés por el proyecto Hércules.

Para las empresas que participen en el desarrollo y la innovación, significa la apertura de un nuevo mercado para la implantación del mismo en el resto de universidades interesadas, como mínimo. Para el estado, un importante ahorro de costes mediante la optimización de la gestión de información.

Tras la creación de la infraestructura semántica y el sistema de gestión de investigación universitaria, será necesario realizar implantaciones del mismo en las diferentes universidades

que decidan utilizarlo. Esto supone tanto una migración de datos y estructuras de datos como la implantación de un nuevo sistema.

Es importante tener en cuenta que el ámbito del proyecto abarca únicamente la implantación del prototipo resultante en algunos centros de la Universidad licitante de la Compra Pública Precomercial (en adelante CPP), para constatar que el prototipo desarrollado funciona y cumple los objetivos propuestos. Quedará fuera del ámbito del mismo, la implantación y migración del resto de universidades interesadas.

1.1 Web semántica

La gestión de la información y el conocimiento del sistema de investigación de las universidades españolas requiere, por tanto, de un escenario tecnológico en el que prime la interoperabilidad de los datos. Esto quiere decir que los datos generados por los distintos sistemas sean combinables y que dicha combinación pueda ser realizada por las máquinas.

Dentro del catálogo tecnológico existente, las tecnologías de la web semántica son una propuesta cuyo objetivo fundamental es servir como espacio natural de integración de datos, posibilitando que los generadores de datos los puedan compartir de forma semántica, esto es, de forma que una máquina pueda ser capaz de entender el significado de estos datos.

El término Web Semántica fue acuñado en el año 2001 por Tim Berners-Lee [Berners-Lee et al, 2001] para designar una Web que propone reestructurar y enriquecer los documentos y componentes Web con información semántica explícita, independiente de la presentación al usuario, y susceptible de ser procesada de forma automática por un programa.

Un elemento fundamental en el paradigma de la Web Semántica es la ontología, que puede definirse como una descripción explícita y formal de una conceptualización [Gruber, 1993]. El motivo de su creciente popularidad es principalmente facilitar una comprensión compartida y común de un dominio que puede ser comunicado entre personas y aplicaciones informáticas. Como tales, el uso de las ontologías ofrece una oportunidad de mejorar las posibilidades de realizar cualquier tarea relacionada con la gestión de información y conocimiento. En los últimos años se ha avanzado sustancialmente en la estandarización de las tecnologías semánticas y ontológicas, donde el uso de lenguajes como RDF, OWL y SPARQL ha permitido la penetración de estas tecnologías en el ámbito industrial. Dichos lenguajes son recomendaciones del W3C, por lo que son estándares de facto para trabajar en entornos de Web Semántica, por lo que el proyecto tratará de utilizar los mismos u los que se determine en el desarrollo del mismo y que mejor se adapten a las necesidades de datos dinámicos. La comunidad es muy activa desarrollando estos y otros lenguajes semánticos, con diferentes características y potencia para abordar diferentes escenarios de uso, buscando facilitar la penetración de esta tecnología y la adopción efectiva de la Web 3.0.

Debido a sus propiedades, la tecnología ontológica ha sido propuesta como una solución a nivel de arquitectura de modernización de los sistemas de información empresariales, como los ERP [Wang et al, 2013]. Otras propuestas suponen arquitecturas de integración basadas en semántica como Smart-M3, que proponen modelos semánticos de compartición de información entre elementos hardware y software para conseguir la integración del mundo real y el mundo

digital en espacios Smart, a través del modelo ODD (ontology-driven development) [J Honkola et al, 2010].

El campo de aplicación donde existe un mayor número de ontologías es la biomedicina, pues en repositorios como Bioportal existen más de 300 ontologías y vocabularios controlados, siendo Gene Ontology [Ashburner et al, 2000] el caso de éxito más importante en cualquier ámbito. Gene Ontology se originó debido a la elevada dificultad que tenían los investigadores para comprender los resultados de investigación obtenidos por investigadores pertenecientes a otras comunidades de investigación biológica debido a la heterogeneidad terminológica existente. Esto les llevó a acordar catálogos de funciones moleculares, componentes celulares y procesos biológicos, y su impacto ha sido brutal en el avance científico por las posibilidades de interoperabilidad de las anotaciones de productos génicos. Vemos que es una problemática similar a la que presentamos en este proyecto y para la que el uso de tecnologías semánticas provocó una gran mejora en cuanto a la eficacia y la eficiencia.

Sin embargo, no es el único ámbito en el que se han creado ontologías, puesto que en el ámbito empresarial existen ontologías como BORO (Business Object Reference Ontology) [Partridge and Stefanova, 2001], que persiguen la interoperabilidad semántica de los sistemas operacionales de la empresa, o la Good Relations Ontology (<http://www.heppnetz.de/projects/goodrelations/>) que proporciona un vocabulario para transacciones electrónicas y que está siendo empleada actualmente por los buscadores más importantes como Google o Yahoo para la estructuración de información. Si bien en el ámbito de los Sistemas de Gestión no conocemos propuestas de estandarización semántica, existen algunos esfuerzos como OntoNaviERP [Hepp and Wechselberger, 2008] que persiguen un acceso semántico a la documentación de ayuda del software de gestión de SAP, y el uso de Web Semántica en Sistemas de Gestión ha sido recomendado como tecnología que podría conllevar grandes beneficios [Alper, 2006].

En el ámbito educativo en general, y universitario en particular, también existen ontologías y vocabularios que persiguen la interoperabilidad de datos académicos y que incluyen la representación de universidades como organizaciones (AIISO¹), publicaciones y comunidades científicas (BIBO², VIVO³), descripción de cursos (MLO⁴, XCRI-CAP⁵, TEACH⁶), recursos de investigación (LSC⁷, CERIF⁸) e incluso propuestas para la publicación semántica de información científica como las Nanopublications⁹. El uso de varios de los recursos anteriores está siendo promovido por iniciativas internacionales como Linked Universities, que persigue que las universidades europeas pueden interoperar entre sí, fundamentalmente a nivel de información académica.

¹ <http://vocab.org/aiiso>

² <http://bibliontology.com/>

³ <http://vivoweb.org/files/vivo-core-public-1.5.owl#>

⁴ <http://www.xcri.co.uk/library/127-metadata-for-learning-opportunities-advertising-mlo-ad-standard.html>

⁵ <http://www.xcri.org/>

⁶ <http://linkedscience.org/teach/ns/>

⁷ <http://linkedscience.org/lsc/ns/>

⁸ <http://eurocris.org/ontology/cerif/>

⁹ <http://nanopub.org/wordpress/>

Las buenas prácticas en construcción de ontologías sugieren la reutilización de ontologías existentes para obtener ontologías de mayor calidad, por lo que algunas de las comentadas anteriormente serán empleadas y combinadas en este proyecto para crear una infraestructura de ontologías que den soporte a las tareas de compartición, reutilización y explotación de datos. De esta forma, al estar alineados con iniciativas internacionales, **será mucho más fácil la integración de las universidades españolas en un sistema de gestión de la información y el conocimiento a nivel europeo**. Si bien se ha comentado antes la existencia de la iniciativa Linked Universities, ésta no deja de ser un esfuerzo de universidades concretas de distintos países europeos, por lo que este proyecto posicionaría a España en una posición de **liderazgo** y sería **pionera e innovadora** en este ámbito.

Una de las ventajas de las tecnologías semánticas frente a tecnologías tradicionales de gestión de información es que están pensadas para ser explotadas en Internet, esto es, en entornos altamente distribuidos. Esto facilita el trabajo de la explotación integrada de conjuntos de datos semánticos, puesto que permiten lanzar una consulta de datos que requiera recuperar datos de N recursos distribuidos, que compartan el modelo de datos (tal y como se obtendría con la realización de un sistema de gestión común para todas las universidades). Un escenario actual a nivel nacional es que todas las universidades vuelcan datos en el Sistema Integrado de Información Universitaria (SIIU), de forma que:

- Cada universidad tiene su repositorio de datos y tiene que enviar la información al Ministerio, por lo que tiene que alinear su esquema de datos con el que usa el Ministerio, con las posibles interpretaciones distintas por parte de cada entidad participante en el proceso.
- El Ministerio tiene que ocuparse del mantenimiento del repositorio centralizado así como de implementar, en un sistema cerrado, todas las explotaciones de los datos.

Es más, las Universidades españolas tienen que responder a solicitudes de información de distintas administraciones públicas, por lo que tienen que hacer esfuerzos específicos para atender cada una de las peticiones por la heterogeneidad de conjuntos y formatos de datos solicitados.

El escenario que plantea nuestro proyecto es que:

- Al tener el mismo sistema de gestión o EPR de investigación en las universidades, sus repositorios siguen el mismo esquema y ponen a disposición del sistema de investigación de la Universidades españolas su conjunto de datos, existiendo un directorio de dichos repositorios.
- Cualquier usuario del sistema de investigación (Ministerio, universidades, etc.) podría realizar consultas distribuidas que atacaran a todos los repositorios necesarios.
- Cada usuario consultor de información, sólo se tendría que preocupar de programar las consultas que le interesen para sus análisis de datos, las cuales podrían ser directamente compartibles al resto del sistema universitario.
- En un entorno de cooperación del sistema de investigación de las universidades españolas, se podrían desarrollar proyectos cooperativos para el desarrollo de funcionalidades adicionales a las que se pueden abarcar en esta propuesta.
- Los consultores de información estarían trabajando siempre con versiones actualizadas de los datos.

- La información estaría disponible y cada entidad interesada podría recuperarla sistemáticamente cuando fuera necesario, ganando en eficiencia y eficacia en publicador y consumidor.

Otra ventaja de las tecnologías semánticas es que permiten explotar estructuras jerárquicas de tipos de información. Las ontologías están compuestas fundamentalmente por conceptos, propiedades, individuos y axiomas formales. Algunos tipos de propiedades que relacionan conceptos generan una estructura jerárquica en la ontología. Un ejemplo de relación jerárquica sería que la Microbiología es una subárea de la Biología y ésta a su vez es subárea de Ciencias de la Vida. Las tecnologías semánticas, mediante las posibilidades de inferencia automática a través del uso de razonadores posibilitarían obtener directamente como resultado de una consulta del tipo “artículos publicados en Ciencias de la Vida” aquellos artículos del área de Microbiología. Este es sólo uno de los posibles usos de las jerarquías en ontologías, puesto que también se podrían utilizar con fines de encontrar resultados semánticamente similares. Por ejemplo, si Microbiología Clínica es un subárea de Microbiología y estamos buscando en nuestro sistema un experto en Microbiología Clínica y no disponemos de ninguno, el sistema podría recomendar un experto en Microbiología antes que uno en Genética por ejemplo, ya que son áreas más cercanas semánticamente. No es que este tipo de operaciones no se puedan hacer con tecnologías tradicionales, pero serían más costosas puesto que tendríamos que implementar cada una de estas consultas avanzadas, mientras que vienen de serie con el uso de tecnologías semánticas.

El desarrollo de la infraestructura semántica para la gestión del sistema de investigación sería un **resultado exportable** parcial o totalmente a otros países europeos con el objetivo de conseguir ese intercambio y explotación a nivel europeo. Por parcial nos referimos a que puede haber aspectos que debidos a la legislación de cada país pudieran requerir extensiones o tratamientos específicos. Además, no sólo vemos posibilidades de innovación de las tecnologías semánticas en el ámbito del proyecto presente, sino que creemos que será un punto de partida para un alcance mucho mayor a medio plazo.

1.2 Sistema de gestión de las universidades

Es necesario, para la consecución de los objetivos del proyecto, el desarrollo de un prototipo de un nuevo sistema de gestión de investigación universitaria que disponga de capacidades semánticas. De forma inicial y dentro del proyecto, se realizaría la implantación del prototipo obtenido en la universidad licitante de la Compra Pública Precomercial y quedaría disponible para el resto de las Universidades Españolas como software libre.

Por lo tanto, se propone un modelo de software común y libre para todas las universidades españolas que decidan implantarlo, como una forma de extender los beneficios del proyecto y la compartición de datos entre diversas universidades.

Este modelo de sistema de gestión único es similar al que siguen un gran número de países europeos como es el caso de Cineca en Italia, Amue en Francia, His en Alemania, etc...

Para la creación del “Sistema de Gestión de Investigación Universitaria”, no se partiría desde cero en los desarrollos sino que se reutilizarían módulos software ya existente en las universidades o empresas que estén dispuestas a cederlos. Para ello sería necesario realizar una primera etapa, en coordinación con la CRUE, para la selección de los módulos que servirán de partida de entre las distintas universidades/empresas que estuvieran en disposición de cederlos como software libre para este proyecto, valorando tanto aspectos funcionales como técnicos para su elección como modulo candidato.

Como marco tecnológico para la creación de este Sistema de Gestión Universitaria se propone partir de la base del estándar de la CRUE “Estándar D.O. CRUE-TIC” (más información en <http://wiki.rediris.es/crue-tic>) que está basado en tecnologías de código abierto y que fue aprobado por la CRUE-TIC en 2011 y actualizado en 2015, para su uso en los nuevos desarrollos para universidades, con el objetivo principal de establecer una arquitectura base común y unificada. El estándar, tiene en cuenta las tendencias más novedosas en el ámbito del Cloud Computing y la movilidad. Será conveniente, partiendo de dicho framework, actualizar sus componentes.

El uso del “Estándar D.O. CRUE-TIC” aporta, entre otros, los siguientes beneficios:

- Contar con una arquitectura tecnología para el desarrollo del proyecto y con una base sólida y actualizada de las tecnologías y metodologías de desarrollo.
 - Frameworks de desarrollo adecuados a la arquitectura definida.
 - Requerimientos a seguir durante el desarrollo para que las aplicaciones obtengan características Cloud como la escalabilidad, el auto aprovisionamiento, el aislamiento del dato o el uso compartido del mutitenant.
 - Aspectos claves en materia de seguridad.
- Disponer de las capas horizontales necesarias ensambladas, de forma que los desarrollos de los distintos módulos del Sistema de Gestión Universitaria sean rápidos y robustos, y que incluyan características propias del uso de tecnologías Cloud y relacionadas con la movilidad.
- Tener identificados los módulos transversales a desarrollar para que puedan ser reutilizables por los diferentes sistemas.

Con el desarrollo de este proyecto, no solo conseguiríamos la estandarización de los sistemas de investigación existentes en las universidades para mejorar la interoperabilidad tanto entre ellas como con el Ministerio y con otras instituciones Europeas, sino también, y casi tan importante como lo anterior, un segundo beneficio consistente en reducir costes al compartir tareas de mantenimiento y actualización de estos sistemas entre las universidades que se incorporen.

Por tanto, en una primera fase operativa, se propone incorporar las capacidades semánticas al sistema de gestión de investigación. El desarrollo del sistema y las integraciones que se deban llevar a cabo, deberán estar enfocadas en una estrategia orientada a la interoperabilidad y a una futura implementación de la administración electrónica.

Ampliable a otras áreas de gestión, con este sistema desarrollado se podría dar respuesta a las necesidades actuales de información de investigación del panorama universitario que se

requieren tanto desde los distintos ministerios como de forma interna en las universidades y entre ellas.

A continuación se describen las principales funcionalidades de los sistemas de gestión a integrar dentro de la plataforma.

Estas funcionalidades deseadas, serán una guía para el proceso de selección de los sistemas a reutilizar de las distintas universidades que lo cedan, con el criterio de seleccionar aquél que más funcionalidades de las descritas incluya.

Proyectos

Gestión de los proyectos de investigación, su definición, origen, propósito, gestión económica, anualidades y otra información básica asociada.

Grupos de investigación

Gestión de la solicitud, creación, mantenimiento, cancelación, incorporación de altas y bajas en grupos de investigación.

Convocatorias y ayudas

Gestión de distintas convocatorias para reparto de fondos, concesión de ayudas y becas. Gestión económica y trámites, concesión, denegación y sistema de baremo o evaluación.

Personal de proyectos

Gestión del personal asociado a proyectos y los pagos a personal por proyectos.

Producción científica

Gestión de la producción científica de los investigadores (artículos, tesis, congresos, publicaciones varias, etc.).

Currículum vitae

Gestión del currículum vitae de los investigadores en formato normal y también en formato CVN.

Contratos y patentes

Gestión de los contratos y las patentes, y de trabajo de investigación en los que colaboran universidades y empresas privadas.

Gestor de páginas de grupos

Gestión de la información asociada a los grupos de investigación para aumentar su visibilidad a través de páginas web propias.

Boletín de investigación

Gestión de un sistema de noticias para la difusión de información relevante para la comunidad investigadora.

Consortios y partners

Proporciona información sobre los consorcios y una valoración por parte de los grupos de investigación de los partners (clasificados por tipo, universidad, SME, Empresa, Centro Tecnológico,...) con los cuales ha participado, permitiendo, a la Universidad, conocer cuáles son los partners mejor valorados. Además permitiría a la Universidad seleccionar a los partners con los cuales formar consorcio en base a la experiencia previa.

1.3 Cloud Computing

El Cloud es una cuestión fundamental en la convergencia que este sistema de gestión de investigación universitaria quiere suscitar. Téngase en cuenta que las universidades han ido creando ecosistemas tecnológicos con especificidades en algunas partes hardware que condicionan con más frecuencia de la deseable decisiones de negocio. El Cloud convierte la base tecnológica en una commodity y permitirá tomar decisiones de negocio alienadas en todas las universidades una vez que se liberen de las diferencias.

El sistema de gestión de investigación ha de ser diseñado teniendo en cuenta las bondades del cloud entre las que destacamos:

1. Escalabilidad y elasticidad en el aprovisionamiento de recursos, de acuerdo con el desarrollo y la puesta en producción
2. Arquitectura en base a componentes homogéneos, de fácil contratación, con competencia de proveedores y con un know-how de gestión muy extendido.
3. Coste de propiedad tecnológica ajustado y siempre en relación con el grado de implantación
4. Reducción de costes energéticos y logísticos relacionados con abastecimiento y aseguramiento de la disponibilidad eléctrica, refrigeración, gestión de proveedores de hardware, averías, etc..

Se pueden proporcionar diferentes niveles de acceso al Cloud desde la visión de la propia infraestructura (IaaS) hasta la de software (Sistema de Gestión Universitaria) como servicio (SaaS). Tal vez esta cuestión permita la transición desde distintas sensibilidades en cuanto al control de resultado final hasta que todas las Universidades lleguen al deseado SaaS.

Los sistemas seleccionados ya deberán ser aptos en origen para su despliegue en la nube y por tanto, ya deberán estar compuestos por frameworks, librerías y componentes que permitan el despliegue en Cloud. Se realizará una consultoría de las adaptaciones necesarias en los módulos que no estén completamente preparados para el cloud, adaptando aquellos que requieran cambios menores.

La arquitectura de integración entre módulos que se diseñe deberá ser apta igualmente para su despliegue en Cloud.

1.4 Administración Electrónica

Otro aspecto fundamental de este sistema de gestión de investigación es el soporte de procesos de administración electrónica. La implantación de sistemas de e-Administración requiere del despliegue de infraestructuras de servicios y herramientas transversales, que posibiliten la tramitación electrónica de procedimientos.

Estas plataformas, diseñadas bajo el concepto SOA (arquitectura orientada a servicios), proveen de funcionalidades comunes para la firma de documentos, validación de certificados, archivado, generación de copias, notificaciones, registro de documentos, etc. El diseño y la configuración del sistema de gestión de investigación a desarrollar deberá ser tal que permita la integración natural de estos servicios a posteriori.

Las principales ventajas de estos sistemas son su escalabilidad y su gran capacidad de interoperabilidad con otros sistemas. Por el contrario, requieren un mayor esfuerzo en su despliegue y disponer de los recursos suficientes para su mantenimiento.

Se incluye, en el alcance del proyecto, la realización de una consultoría que sirva de base a la toma de decisiones sobre los servicios concretos que se podrán implantar dentro de esta infraestructura y cuáles externalizar, reutilizando aplicativos y/o sistemas de la Administración General del Estado u otras entidades colaboradoras e impulsando el concepto de “cloud computing”.

Adicionalmente, también resulta esencial que la consultoría a realizar incluya las recomendaciones para adecuar el sistema de gestión de investigación a los requisitos derivados de los Esquemas Nacionales de Seguridad (ENS) e Interoperabilidad (ENI) y las normas técnicas que los desarrollan.

Se pueden destacar como principales características deseables las siguientes:

- Múltiples interfaces de integración. Como mínimo, cualquier infraestructura de este tipo debe soportar una interfaz SOAP, aunque también puede ser importante considerar opciones adicionales, tanto por motivos de rendimiento, como por compatibilidad con tecnologías menos recientes.
- Comunicación con otras organizaciones. Estas infraestructuras deben estar capacitadas para la comunicación interoperable con otras entidades, incluso a través de Red SARA, consumiendo servicios de la Administración General del Estado.
- Archivo electrónico. Deben proporcionar soluciones de archivado y gestión de documentos electrónicos, bien disponibles desde la propia infraestructura o mediante la integración de soluciones de gestión documental de terceros.

Algunos ejemplos de pasos de tramitación electrónica para los que el sistema de gestión de investigación debería estar preparado son los siguientes.

- Autenticación de usuario basada en certificado electrónico.
- Firma electrónica de documentos.
- Emisión de recibos y justificantes electrónicos.

- Comprobación de la validez de un documento electrónico.
- Digitalización de documentos.
- Generación de notificaciones electrónicas.
- Archivado y gestión de documentos electrónicos.

Será también necesario disponer de herramientas horizontales concretas que apoyen la tramitación interna de ciertos procedimientos asociados al sistema de gestión de investigación, como las siguientes.

- Gestores de trámites. Son aquellas herramientas que permiten establecer los flujos de trabajo y circuitos de validación de cada procedimiento administrativo, poniendo a disposición de los tramitadores responsables de los mismos, los documentos, expedientes y datos necesarios en el orden adecuado, facilitando la visualización, gestión y firma electrónica de los mismos, así como la cumplimentación de las tareas que sean necesarias.
- Portafirmas electrónicos. Son herramientas específicas para la revisión y firma/rechazo de documentos electrónicos, posibilitando la opción de firmar un conjunto de documentos en un único proceso.

1.5 Datos abiertos y enlazados

En los últimos años, además de realizarse acciones a nivel de desarrollo de la Web Semántica investigando y proporcionando soluciones formales, se han ido desarrollando iniciativas para conseguir la publicación semántica, reutilizable e interoperable de datos. Por un lado, la iniciativa Open Data (datos abiertos) persigue que las distintas organizaciones publiquen sus datos de forma que terceros puedan acceder y usar esos datos. Esto ha dado lugar a portales a nivel nacional y europeo de Open Data, donde podemos encontrar distintos tipos de formatos, incluyendo los formatos de datos relacionados con la Web Semántica, como RDF. Según la Open Knowledge Definition, Datos Abiertos (Open Data) es una filosofía y práctica que persigue que determinados datos estén disponibles de forma libre a todo el mundo, sin restricciones de copyright, patentes u otros mecanismos de control.

En España, la Secretaría de Estado de Administraciones Públicas aprobó la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información (Resolución de 19 de febrero de 2013, BOE 4 de marzo de 2013), que se basa en la publicación de datos siguiendo la iniciativa de datos abiertos y siguiendo recomendaciones procedentes del mundo de la Web Semántica, incluyendo incluso un modelo de plantilla RDF para la definición de catálogos y registros.

En gran medida, estas recomendaciones vienen a adoptar principios de la iniciativa de Linked Open Data del W3C, que persigue crear vínculos entre los datos con tipo de contenido a partir de diferentes fuentes de datos. Técnicamente, Linked Data se refiere a la publicación de datos en la Web de tal manera que sean legibles por el ordenador, su significado se defina explícitamente, que esté relacionado con conjuntos de datos externos y que a su vez pueda ser vinculado desde otros de conjuntos de datos externos [Bizer et al., 2009]. Para la publicación de datos en la Web, Berners-Lee, [2006] describe un conjunto de "reglas" llamados "Principios de Linked Data" que proporciona una receta básica para la publicación y la conexión de datos utilizando la infraestructura de la Web de forma que todos los datos publicados se convierte en parte de un único espacio global de datos:

1. Usar URI para nombrar los recursos web
2. Usar URI HTTP para que se puedan buscar esas URI
3. Cuando se busca una URI, proporcionar información útil sobre la entidad de datos usando estándares como RDF o SPARQL.
4. Incluir enlaces a otras URI para que se pueda descubrir más información.

Cuando hablamos de datos enlazados que cumplen esta filosofía, estamos hablando de Datos Enlazados Abiertos o Linked Open Data (LOD). El objetivo de Linked Open Data es mejorar la información publicada para eliminar barreras en su consumo por los usuarios finales. Dependiendo del grado de compromiso (estrellas) será más sencillo automatizar tareas de utilización de la información publicada, eliminándose intermediarios que aportan poco valor a la información.

Existen cinco niveles, ordenados de menos a más abiertos, para hacer una apertura progresiva de los datos.

- Una estrella (*): Disponible en la web (en cualquier formato, por ejemplo PDF) con una licencia libre para ser considerados datos abiertos.
- Dos estrellas (**): Disponible en la web de manera estructurada y legible por máquinas (por ejemplo en formato Excel en lugar de una imagen escaneada de una tabla).
- Tres estrellas (***): Disponible en la web de manera estructurada y en formatos no propietario (por ejemplo CSV en lugar de Excel).
- Cuatro estrellas (****): Siguiendo todo lo anterior, pero dentro de los estándares establecidos por el W3C (RDF y SPARQL) para identificar “cosas”, de forma que la gente pueda apuntar a tu información desde el exterior.
- Cinco estrellas (*****): Siguiendo todo lo anterior, y además enlazar tus datos a los de otras personas para proporcionar un contexto. Por ejemplo, si estamos publicando datos de publicaciones científicas biomédicas que tienen un identificador PubMed y existe un repositorio de datos abiertos de PubMed, lo apropiado sería que la instancia de la publicación en nuestro repositorio haga referencia a la instancia existente en el repositorio de PubMed.

Berners-Lee expuso la siguiente explicación en cuanto a los costes y beneficios que, en teoría, implican para una organización el despliegue de datos abiertos, de acuerdo al nivel de compromiso de los mismos¹⁰:

Para el consumidor	
Datos *	<ul style="list-style-type: none"> • Puede consultarlos. • Puede imprimirlos. • Puede hacer copias locales. • Puede introducirlos en otro sistema. • Puede modificarlos como quiera. • Puede compartir los datos con quien quiera.
Datos **	<ul style="list-style-type: none"> • Todo lo de los datos *. • Puede procesarlos directamente con software propietario para realizar cálculos, visualizarlos, etc. • Puede exportarlos a otro formato (estructurado).
Datos ***	<ul style="list-style-type: none"> • Todo lo de los datos **. • Puede procesar los datos como quiera, sin las restricciones de ningún software concreto.
Datos ****	<ul style="list-style-type: none"> • Todo lo de los datos ***. • Puede enlazar a ellos desde cualquier otro sitio (de la web o localmente). • Puede marcarlos como favoritos. • Puede reutilizar parte de los datos. • Puede reutilizar herramientas y librerías existentes. Incluso si solo entiende parte del patrón utilizado por quien los publicó. • Entender la estructura de un “Grafo” de datos RDF puede ser más costoso que una tabla (Excel/CSV) o árbol (XML/JSON). • Puede combinar los datos con otros datos sin problemas. Las URIs son un esquema global, así que si dos cosas tienen la misma URI será

¹⁰ <http://5stardata.info/>

	intencionadamente, y si es así entonces irá por el buen camino hacia ser datos 5*.
Datos *****	<ul style="list-style-type: none"> • Todo lo de los datos *****. • Puede descubrir más datos (relacionados) mientras los consume. • Puede aprender el esquema de datos directamente. • Ahora tiene que manejar enlaces rotos de datos, como los errores 404 de las páginas web. • Exponer datos de un link arbitrario como verídico es tan arriesgado como dejar a la gente incluir contenido de cualquier web en tu sitio web. La precaución, confianza y el sentido común siguen siendo necesarios.

Para el publicador	
Datos *	<ul style="list-style-type: none"> • Publicar es simple. • No tiene que explicar repetidamente que pueden usar sus datos.
Datos **	<ul style="list-style-type: none"> • Publicar sigue siendo simple.
Datos ***	<ul style="list-style-type: none"> • Podría necesitar conversores o plugins para exportar los datos desde un formato propietario. • Sigue siendo bastante simple publicar.
Datos ****	<ul style="list-style-type: none"> • Tiene control de granularidad fina sobre los elementos de datos y puede optimizar su acceso (balanceo de carga, cacheo, etc.). • Otros publicadores pueden enlazar ahora a sus datos, promoviéndolos a 5*. • Normalmente invertirá tiempo en trocear y dividir sus datos. • Necesitará asignar URIs a sus elementos de datos y pensar en cómo representarlos. • Necesitará buscar patrones existentes que reutilizar o crear uno propio.
Datos *****	<ul style="list-style-type: none"> • Hace sus datos descubribles. • Incrementa el valor de sus datos. • Su organización obtendrá el mismo beneficio de los enlaces que los clientes. • Necesitará invertir recursos en enlazar sus datos con otros datos de la web. • Puede necesitar reparar enlaces rotos o incorrectos.

1.5.1 Iniciativas de Open Data y Linked Data

La irrupción de la corriente de Linked Open Data ha permitido llevar a la práctica una parte de la web semántica. Son muchas las instituciones, administraciones públicas, empresas privadas, universidades, hospitales, etc., que bajo esta corriente están publicando sus datos en el entorno web siguiendo las directrices marcadas por Tim Berners-Lee para la consecución de la Web de Datos. En la tabla siguiente se muestran algunas iniciativas en marcha relacionadas con datos abiertos y/o enlazados.

Fuentes de Datos	Ámbito	Web
------------------	--------	-----

Fuentes de Datos	Ámbito	Web
European Union Open Data Portal	Europa	https://open-data.europa.eu/
Italian Public Administration Open Data initiative	Italia	http://www.dati.gov.it
Linked Data from The Open University	Reino Unido	http://data.open.ac.uk
University of Southampton Open Data Service	Reino Unido	http://data.southampton.ac.uk
Open Data about the University of Oxford	Reino Unido	http://data.ox.ac.uk
Linked Open Data at University of Muenster	Alemania	http://data.uni-muenster.de
Data.overheid.nl: het open dataportaal van de Nederlandse overheid	Holanda	https://data.overheid.nl/
Basic Data Programme	Dinamarca	http://eng.gst.dk/
Catalog of Open Data (Government)	Dinamarca	http://digitaliser.dk/kataloger
Eesti avatud andmete kogukond (Open Data Estonia)	Estonia	http://www.opendata.ee/en/
Open Data Portal Spain	España	http://datos.gob.es/
International Catalogue of Open Data	Internacional	http://datahub.io/
Catalogue of Linked Data sets	Internacional	http://linkeddata.org/data-sets
Governmental data site	Rumanía	http://data.gov.ro/
EURES (European Employment Services) network via ANOFM platform	Europa	https://ec.europa.eu/eures/page/homepage?lang=en

A continuación describimos actividades relacionadas con datos abiertos y enlazados en el sector público, privado y educativo.

La publicación de datos abiertos es un valor en alza en las instituciones públicas para facilitar el acceso a información pública completa, fiable y de calidad.

La iniciativa de apertura de datos surge desde el ámbito europeo (Directiva europea

2003/98/CE) para, entre otros objetivos, concienciar a los empleados públicos del valor que posee la información del sector público con la que trabajan. Basada en tres pilares fundamentales: la transparencia, participación y colaboración. A nivel nacional se traduce en la Ley del 37/2007 de 16 de noviembre (RISP, Reutilización de la Información del Sector Público) y el Real Decreto 1495/2011, de 24 de octubre, que la desarrolla, así como la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información mencionada anteriormente.

En el caso de organizaciones de carácter público, según la normativa RISP, se recomienda que los datos estén enlazados desde la “sede electrónica” de la organización. El portal del Catálogo de Información Pública del Sector Público, gestiona y permite el acceso desde un único punto a recursos y sitios web del Sector Público que ofrecen información pública, incluyendo fichas-tipo para “publicitar” los datos publicados. Si bien en dicho catálogo se pueden encontrar datos abiertos en RDF, y por tanto de 4* y 5*, la mayoría no llegan a este nivel. Existen cada vez más secciones webs en entidades públicas dedicadas a la publicación de sus datos.

Los datos son el nuevo capital de la economía global y la búsqueda por parte de las organizaciones de crecimiento renovado, mayor rendimiento y un compromiso con el cliente más significativo, hace una inmensa presión hacia la explotación de datos [Deloitte, 2012]. Cada vez más empresas tienen una estrategia para explotar el rápido crecimiento de los datos abiertos. La amplitud de este tipo de datos actualmente disponibles hace que empresas de todas las industrias puedan beneficiarse de ellos. Por ejemplo, las aseguradoras podrían recopilar datos sobre mortalidad, salud, tráfico, transporte o crimen; los comercios pueden usar datos demográficos, riqueza de la población o bienestar.

Por otro lado, las empresas empiezan a abrir sus propios datos para revolucionar la forma en que compiten. Líderes como Google o Asos [Thomson, 2012] reconocen que nuevos modelos de negocio basado en el acceso gratuito a sus datos tiene el potencial para generar un buen beneficio de la inversión. Esto permite inspirar compromiso en los clientes, que ejercen una gran presión en las organizaciones para que sean más transparentes en cuanto al tráfico de negocio y las responsabilidades sociales y corporativas. Empresas como Nike¹¹ o Enel¹² son proactivas en este sentido y publican sus datos para demostrar compromiso de transparencia y sostenibilidad. Otros beneficios de publicar sus datos abiertos para las empresas pueden ser cumplir con obligaciones legales, vender servicios relacionados con los datos o establecer relaciones de colaboración con otras empresas o el sector público.

En el ámbito educativo existen diversas iniciativas de datos abiertos y enlazados, como LinkedUp¹³ y Linked Universities¹⁴. LinkedUp es un proyecto que fomenta la explotación y uso de datos abiertos de instituciones educativas europeas. LinkedUp proporciona algunas actividades y recursos para la persecución de dicho objetivo. Un ejemplo es el LinkedUp Challenge, una competición en la que se buscan herramientas interesantes e innovadoras que analizan o integran datos web abiertos para fines educativos.

¹¹ <http://www.nikeresponsibility.com/>

¹² <http://data.enel.com/>

¹³ <http://linkedup-project.eu/>

¹⁴ <http://linkeduniversities.org>

Linked Universities es una alianza de universidades europeas que tiene como objetivo publicar sus datos públicos como datos enlazados. Linked Universities gestiona un portal colaborativo en el que se publican vocabularios y herramientas que pueden ser reutilizados para publicar datos enlazados universitarios. Algunas universidades que forman parte de esta iniciativa son la University of Münster, en Alemania, la University of Southampton, en Inglaterra, o la Universitat Pompeu Fabra, en Barcelona. Todas ellas poseen un portal de datos abiertos enlazados.

A continuación se incluyen algunos datos de universidades europeas con iniciativas de datos abiertos:

País	Universidad	Website
Reino Unido	University of Leeds	http://data.leeds.ac.uk/
Reino Unido	University of Southampton	http://data.southampton.ac.uk
Reino Unido	University of Oxford	https://data.ox.ac.uk/
Reino Unido	University of Lincoln's	http://data.online.lincoln.ac.uk/
Reino Unido	The Open University	http://data.open.ac.uk/
Finlandia	Aalto University	http://data.aalto.fi/
Alemania	University of Münster	http://lodum.de/
Grecia	Aristotle University of	http://www.swu.auth.gr/en
España	Universidad Pablo de Olavide	http://www.upo.es/datos-abiertos/
España	Universidad de León	http://datosabiertos.unileon.es
España	Universitat Pompeu Fabra	http://data.upf.edu
España	Universidad de Zaragoza	http://zaguan.unizar.es/collection/OpenData?ln=es&as=1

En resumen, cada vez más organizaciones se incorporan al mundo de Open Data y/o Linked Open Data de manera natural, de forma que ser capaces de procesar y explotar información relevante para las organizaciones en estos formatos supone dar herramientas a las mismas para mejorar sus procesos de gestión de información y su productividad. Repositorios de datos abiertos internacionales como datahub.io incluyen actualmente más de 5000 conjuntos de datos abiertos, pero únicamente en torno al 10 % está disponible en formatos abiertos con al menos 4* como propone este proyecto.

Por tanto, disponer de un infraestructura que permitiera la publicación y explotación de la información y el conocimiento en formatos abiertos semánticos que garantizan que los datos sean al menos 4* supondría **situar al sistema universitario español como líder internacional.**

Como se ha comentado anteriormente, disponer de datos 4* no sólo supone un ventaja tecnológica, sino que supone un salto cualitativo en la visibilidad de los datos del sistema de gestión e investigación de las universidades españolas, con los consiguientes beneficios a nivel de acceso a los datos y reputación del sistema, que se debería traducir en que **entidades nacionales e internacionales pudieran encontrar expertos españoles con los que colaborar de manera más sencilla.**

Del mismo modo, desde el punto de vista de administración gestora de la actividad investigadora a nivel nacional, la tecnología facilitaría el desarrollo de cuadros de mando que permitirían gran flexibilidad en el análisis de los datos por la posibilidad de explotación de las estructuras jerárquicas de las tecnologías semánticas.

El hecho de utilizar un enfoque semántico basado en estándares como RDF para la publicación y explotación por parte de las máquinas de conjuntos de datos abiertos no significa generar una mayor dificultad para la publicación de los mismos en formatos comprensibles para los humanos, puesto que la representación RDF se puede considerar la canónica a partir de la cual generar versiones en PDF, XLS, CSV, etc. preservando el significado de cada dato. Además, actualmente existen herramientas como Pubby que permiten una publicación automática en formato HTML de los conjuntos de datos RDF, visualización que se puede ajustar mediante el diseño de plantillas específicas. Además, este tipo de herramientas permiten la sindicación de contenidos, que significa que a un usuario humano se le podría mostrar el contenido en HTML y a una máquina el contenido en RDF.

1.6 Relación entre Open Data y Big Data

En los últimos años están apareciendo diversos términos relacionados si bien no significan lo mismo, como Open Data y Big Data. Por Open Data nos referimos a publicar los datos de manera abierta, de forma que terceros puedan hacer uso de los mismos, mientras que Big Data se refiere a grandes volúmenes de datos. Esto quiere decir que algunos conjuntos de datos abiertos se pueden considerar Big Data. Se puede considerar que las Universidades españolas, teniendo en cuenta los conjuntos de datos históricos de que disponen, están en condiciones de generar Big Data. Si bien en este proyecto no se plantea como objetivo directo el desarrollo de técnicas específicas orientadas a Big Data, los avances científicos muestran que nuestra infraestructura semántica sería una buena base para desarrollos posteriores orientados a Big Data [Calvanese et al 2013; Bennett 2013; Giese et al 2013, Hitzler and Janowicz, 2013).

1.7 Relación con líneas de investigación estratégicas internacionales

Esta propuesta sitúa sus objetivos en el contexto del área identificada en las prioridades de investigación europeas “Content technologies and information management”. En esta área existen convocatorias de investigación ICT-15-2014 4 ICT-16-2015 relacionadas con investigación e innovación relacionadas con Open Data y Big Data. Estas acciones tienen como objetivo crear y explotar recursos semánticos e interoperables con el objetivo de situar a Europa en una posición dominante a nivel internacional.

Nuestra propuesta se ajusta a las prioridades y retos de Horizonte 2020, puesto que refleja las prioridades políticas de la estrategia europea 2020 y afronta una problemática de interés para toda la ciudadanía europea. Nos sentimos identificados con el reto “Europe in a changing world

- Inclusive, innovative and reflective societies“, para el que el uso de las TIC siempre se ha considerado un elemento fundamental.

Nuestra propuesta realizará contribuciones en el contexto de este reto europeo:

- Uso de tecnologías emergentes en los sectores público y privado, mostrando cómo se pueden usar de manera efectiva, orientada y útil.
- Modernización de la administración pública.
- Innovación de mercado, puesto que, como se ha considerado anteriormente, la explotación semántica basada en datos abiertos tiene el potencial de generar un mercado alrededor de la misma.

En la tabla siguiente se describe una selección de proyectos internacionales de investigación recientes o en curso relacionados con Open Data, Linked Data y Big Data.

Proyecto	Objetivos del proyecto
BIG- Big Data Public Private Forum ¹⁵ (318062 FP7- ICT-2011.4.4)	<ul style="list-style-type: none"> • Crear una comunidad industrial sostenible de Big Data en Europa • Promover la adopción de la tecnología de Big Data • Afrontar y resolver adecuadamente las barreras políticas y de regulación
OPTIQUE- Scalable End-user Access to Big Data ¹⁶ (318338 FP7- ICT-2011.4.4)	<ul style="list-style-type: none"> • Proporcionar una conexión semántica entre usuarios y fuentes de datos • Permitir a los usuarios formular consultas de forma intuitiva usando conceptualizaciones y vocabularios familiares • Integración de datos de múltiples fuentes distribuidas • Explotación de paralelismo masivo para lograr soluciones escalables
LinkedUp- Linking Web Data for Education Project & Open Challenge in Web-scale Data Integration ¹⁷ (317620 FP7- ICT-2013.4.4)	<ul style="list-style-type: none"> • Explotación y adopción de datos públicos y abiertos en la web, especialmente por parte de organizaciones e instituciones educativas • Repositorios y catálogos de recursos en forma de conjuntos de datos abiertos con fines educativos
EUCases- European and National CASE Law and Legislation Linked in Open Data Stack ¹⁸ (611760 FP7- ICT-2013.4.3)	<ul style="list-style-type: none"> • Desarrollo de una plataforma multilingüe europea de datos abiertos y enlazados sobre legislación.
DAPAAS-Data Publishing	<ul style="list-style-type: none"> • Desarrollo de infraestructura software que combina Data-as-a-Service (DaaS) y Platform-as-a-Service (PaaS) para optimizar la

¹⁵ <http://www.big-project.eu/>

¹⁶ <http://www.optique-project.eu/>

¹⁷ <http://linkedup-project.eu>

¹⁸ <http://www.eucases.eu>

Proyecto	Objetivos del proyecto
<p>through the Cloud: A Data- and Platform-as-a-Service Approach for Efficient Data Publication and Consumption¹⁹ (610988 FP7- ICT-2013.4.3)</p>	<p>publicación de datos abiertos y el desarrollo de aplicaciones basadas en los mismos.</p> <ul style="list-style-type: none"> • Desarrollar una plataforma intuitiva que simplifique la publicación y el consumo de datos abiertos
<p>MELODIES- Maximising the Exploitation of Linked Open Data In Enterprise and Science²⁰ (603525 FP7 ENVIRONMENT ENV.2013.6.5-3)</p>	<ul style="list-style-type: none"> • Desarrollar servicios innovadores y sostenibles basados en datos abiertos para investigadores, administraciones, industria y público en general

¹⁹ <http://project.dapaas.eu/>

²⁰ <http://www.melodiesproject.eu/>

1.8 Escenarios de uso en gestión de la investigación

El uso de la tecnología semántica permitirá una explotación avanzada de los datos relacionados con gestión de la investigación en las universidades españolas.

A continuación se describen cómo podría contribuir el desarrollo del proyecto a mejorar algunas actividades relacionadas con la gestión de la investigación nacional:

- **Mejorar la estructura de financiación de la investigación y contabilidad analítica:** La unificación de los módulos de gestión económica también permitiría conocer y analizar a distintos niveles los tipos de gasto realizados por los grupos nacionales en los proyectos de investigación financiados. Esto permitiría tomar mejores decisiones sobre las necesidades reales de financiación por conceptos de cada área y de cada tipo de proyectos, lo que permitiría tomar decisiones informadas sobre reparto de presupuestos en convocatorias de proyectos de investigación, becas, etc. para que la investigación se realice en las condiciones que aseguren sus mejores resultados. Este proceso también permitiría disponer de la información para realizar análisis fiables y flexibles de productividad de grupos, proyectos, áreas, etc.
- **Mapa del conocimiento nacional:** El uso de un esquema común de datos y de conocimiento permitiría la creación de un mapa del conocimiento nacional, posibilitando la identificación objetiva de polos de conocimiento en áreas temáticas, en línea con el objetivo EECTI 2013-20 de especialización regional inteligente, puesto que se podría identificar cuáles podrían ser las especializaciones estratégicas para cada región o grupo de regiones a partir de los datos existentes.
- **Cuadros de mando flexibles de gestión de la investigación:** La explotación semántica permitiría la puesta en marcha de cuadros de mando flexibles, en los que la estructura jerárquica de las ontologías permitirían modificar el nivel de profundidad de los análisis, permitiendo por ejemplo hacer estudios a nivel macro y a nivel micro en cuanto a especificidad de las áreas de investigación, tipología de proyectos, origen de fondos, etc. tanto unidimensionales como multidimensionales.
- **Búsqueda de socios a nivel nacional:** La cooperación de grupos de investigación es un aspecto fundamental para el avance científico. Habitualmente los grupos que trabajan en una misma área temática suelen tener contacto directo y conocerse, pero no siempre ocurre así. Cada vez es mayor la necesidad de realizar investigación multidisciplinar y suele ser difícil para este tipo de proyectos encontrar los socios adecuados. Es por ello que la explotación semántica de datos de investigación nacionales contribuirían a poner en marcha buscadores de grupos con perfiles concretos y disponibles para la participación en determinadas convocatorias de investigación, ya que se podría incluso filtrar aquellos que presentaran incompatibilidades o tuvieran un determinado nivel científico (aspecto subjetivo que se podría concretar a través de distintos tipos de filtros).
- **Selección de grupos:** Actualmente es muy difícil o laborioso para las administraciones públicas y entidades privadas seleccionar objetivamente aquellos grupos o expertos que puedan ser los más adecuados para realizar un proyecto por encargo, ser contratados para determinadas tareas relacionadas con investigación o la organización de programas formativos interuniversitarios de excelencia. La explotación semántica permitirá encontrar aquellos cuyo perfil académico-investigador se ajuste más e incluso poner en marcha método de diseño automático de consorcios óptimos a partir de los datos existentes. En este aspecto podemos incluir por ejemplo la puesta en marcha de institutos y redes nacionales de investigación, procurando que participen los grupos más adecuados.

- Mejorar las posibilidades de obtención de fondos europeos para investigación: Uno de los aspectos mejorables en nuestro sistema de investigación es la captación de fondos europeos. Si bien únicamente la puesta en marcha de este sistema no va a mejorar por sí sólo este aspecto, sí puede contribuir a que desde las entidades gestoras de la investigación puedan realizar su trabajo en mejores condiciones, lo que de forma indirecta debe generar mejores posibilidades de financiación. Siendo la excelencia un criterio de evaluación fundamental en H2020, disponer del mapa de conocimiento y de las posibilidades avanzadas de explotación de la información del sistema de gestión de investigación nacional, permitiría que desde las oficinas regionales y nacionales de proyectos europeos conocieran mejor el perfil de nuestros grupos de investigación y pudieran realizar una tarea activa más precisa de recomendación de convocatorias, promoción de grupos de investigación ante Bruselas, etc. Estamos hablando de que de manera objetiva se podría escoger a las personas que España propusiera para ser parte de comités, grupos de trabajo, etc. en los foros internacionales. Esta actividad tiene otra vertiente. La internacionalización de los conjuntos de datos permitirían ofrecer a los socios potenciales de grupos nacionales una puerta de entrada que facilitara la identificación de aquellos con los que desearan participar en propuestas internacionales.

Este proyecto desarrollará alguno de los anteriores como casos de estudio, quedando fuera del ámbito del proyecto, por su dimensión económica, la puesta en marcha de todos ellos. Al principio del proyecto se determinará cuáles se afrontarán mediante el correspondiente análisis de interés y viabilidad.

2 Estimación de tareas (costes) del proyecto

Se identifican dos fases principales para la realización del proyecto que son:

- 1.- Creación de una infraestructura semántica basada en el paradigma de datos abiertos, para la gestión de la información y el conocimiento del sistema de gestión de información universitaria.
- 2.- Aplicación de la infraestructura semántica al sistema de gestión de investigación universitario libre español que siga el “Estándar D.O. CRUE-TIC”, que esté preparado para el Cloud y con una estrategia de Open Data.

Para el desarrollo de estas tareas es necesario realizar, de forma previa una consultoría que abarque el análisis, estudio y selección de los diferentes módulos de los sistemas de gestión de la Investigación de las universidades/empresas que serán los que se implantarán en el nuevo sistema de gestión investigación semántico.

Las tareas de alto nivel a desarrollar en cada una de las fases son las descritas en los siguientes paquetes de trabajo:

2.1.1 PT 1: Gestión y dirección

Se realizarán todas las tareas necesarias para llevar a cabo el proyecto con éxito, es decir, en el plazo determinado, con el alcance definido y los recursos planificados. En concreto, se realizará el establecimiento de las metodologías para el inicio, planificación, ejecución, seguimiento, control y cierre del proyecto así como las actividades necesarias para realizar todas estas tareas.

1. Inicio: En esta etapa se hace la redacción de la propuesta (objeto, objetivos, alcance, calidad y riesgos del proyecto), y describe cómo se llevará a cabo. Incluye también estimaciones de coste y tiempo, y efectúa la Integración de todo lo anterior con lo que sigue.
2. Planificación: Se realiza la planificación de todas las actividades necesarias para llevar a cabo el proyecto, considerando las prioridades del proyecto, los recursos necesarios, los tiempos esperados para ejecutar cada una de las tareas.
3. Ejecución: Se refiere a la implementación o puesta en marcha del proyecto, consiste en poner en práctica la planificación llevada a cabo previamente. Durante la ejecución del proyecto, se debe poner énfasis en la comunicación para tomar decisiones lo más rápido posible en caso de que surjan problemas.
4. Seguimiento y Control: El fin de las actividades de control es asegurar que los objetivos sean alcanzados en el tiempo y calidad planificada, realizando una buena supervisión y medición del rendimiento de los resultados, con el objetivo de que se puedan tomar acciones correctivas.
5. Cierre: Cierre es la etapa final de un proyecto en la que éste es revisado, y se llevan a cabo las valoraciones pertinentes sobre lo planeado y lo ejecutado, así como sus resultados, en consideración al logro de los objetivos planteados. Se realizan las pruebas finales de corrección de la solución y la verificación.

2.1.2 PT 2: Marco tecnológico, Estrategias Cloud y eAdmon

En este paquete de trabajo se determinarán los estándares y las estrategias sobre las que se desarrollará el proyecto, siempre tomando como referencia las recomendaciones de la metodología D.O CRUE-TIC.

También se consideran las tareas necesarias para montar, tanto el entorno de desarrollo como el de producción, de acuerdo a las metodologías de desarrollo que se definan en este paquete de trabajo.

2.1.2.1 Cloud

Estudio de la conveniencia de la implantación de la solución en el Cloud y realización, en su caso, de las siguientes tareas:

- Identificar tipo y cantidad de componentes cloud o bloques que van a participar en la solución final:
 - Componentes de almacenamiento (disco raw y servidores de ficheros) y calidades requerida
 - Grupos de nodos de cómputo separados para cada rol (base de datos, servidor web, servidor de aplicaciones, otros, etc..)
 - Servicios de cortafuegos
 - Redes privadas virtuales
 - Balanceadores de carga
 - Monitores de estado y rendimiento del conjunto
 - Servicios de caché
 - etc..
- Realizar un diseño por capas de la solución final que mapee todas las necesidades de la arquitectura de la aplicación.
- Revisar que se respetan los requisitos de escalabilidad en todo el conjunto.
- Identificar proveedores de servicios cloud que cumplan con los requisitos
 - Técnicos para proporcionar los componentes cloud que hemos identificado previamente para la solución
 - Legales para el depósito y custodia de datos atendiendo a la legalidad a la que están sometidos los datos Universitarios
- Establecer la correspondencia de cada elemento de arquitectura de la aplicación propuesto al correspondiente objeto cloud de los principales proveedores válidos para el proyecto.

2.1.2.2 Administración electrónica

Las actividades a realizar dentro de este paquete de trabajo están orientadas a un trabajo de consultoría que nos permita sentar las bases sobre:

- Servicios y herramientas que constituirán la plataforma de e-Administración, identificando qué servicios sería necesario implementar, cuáles reutilizar y cuáles integrar a partir de soluciones cloud.
- Identificación de las fases procedimentales y tareas de tramitación que requieran integración con la plataforma de e-Administración.

- Diseño de la arquitectura de la plataforma y sus frameworks de desarrollo, alineados con el marco tecnológico escogido.

2.1.3 PT 3: Gestión de calidad y riesgos

Uno de los elementos clave a la hora de asegurar el éxito del proyecto, medido en términos de cumplimientos de plazos, costes, alcance funcional y calidad final de la solución, es la gestión de los riesgos.

Implantar la gestión de riesgos será un elemento decisivo para asegurar el proyecto, mediante la identificación y el análisis previo de los riesgos potenciales que pueden afectar al proyecto y la elaboración de los planes de contingencia adecuados para evitar que aparezcan o minimizar su impacto sobre el proyecto en caso de que finalmente el riesgo se verifique.

Las tareas necesarias para la gestión de riesgos se basarán en la correcta identificación, catalogación y proposición de acciones para evitarlos y mitigarlos, así como asignar responsables.

Además, es necesario garantizar la calidad del producto generado en términos del conjunto de propiedades y características que le confieren aptitud para satisfacer las necesidades del proyecto. La gestión de la calidad del proyecto comprende todos los procesos necesarios para asegurar que el proyecto satisface aquellas necesidades para las que fue concebido.

Las tareas necesarias para la gestión de la calidad son las siguientes:

1. Planificación de la calidad: Identificando que estándares de calidad son de relevancia para el proyecto y determinando como implantarlos.
2. Aseguramiento de la calidad: Evaluando de forma global los rendimientos del proyecto, de forma que se asegure el cumplimiento de los estándares seguidos.
3. Control de calidad: Controlando y supervisando los resultados del proyecto, comprobando que cumplen con los estándares y eliminando las causas de insatisfacción posibles.

2.1.4 PT 4: Desarrollo de un estudio para la selección de módulos del nuevo sistema de gestión de investigación universitaria.

En dicha fase se encuentra el grueso del proyecto. A continuación se especifican las dos fases de este plan de trabajo:

Fase Inicial:

En coordinación con la CRUE, se realizará un estudio comparativo sobre los diferentes sistemas vigentes en las diversas universidades/empresas que hayan cedido sus sistemas de forma gratuita al proyecto, para seleccionar los módulos que servirán de partida para la implantación del nuevo sistema de gestión integrado. En este fase, se realizará un análisis tanto técnico como funcional de los módulos que cumpliendo ya con una serie de requisitos iniciales prefijados, sean susceptibles de implantarse y servir como punto de partida para el desarrollo del sistema de gestión de investigación universitaria español.

Como base para realizar el estudio, se desarrollará un “Catálogo de Indicadores” que nos permitan comparar y por tanto seleccionar, los módulo más adecuados para el nuevo sistema. Este catálogo tocará diferentes ámbitos como:

- Tipología de Bases Datos y facilidad de migración de datos
- Capacidad de interoperabilidad e interfaces

- Tecnología de construcción utilizada
- Formatos de datos admitidos
- Capacidad de integración en una nueva arquitectura
- Funcionalidades que abarque cada uno de los módulos
- Cumplimiento de estándares
- Etc.

Para desarrollar el estudio se seleccionará un equipo de expertos capacitados para llevar a cabo el objetivo deseado y que estará compuesto por:

- Personal técnico y funcional de la empresa o empresas colaboradoras.
- Y por diversos grupos de expertos de las universidades que firmen el convenio CRUE de participación activa en el proyecto.

La información sobre la que se basará el estudio será la que hayan proporcionado las universidades participantes. Para la obtención de dicha información se planificarán grupos de trabajo de las diversas universidades que abarcarán los temas a tratar de manera estructurada y planificada.

Los módulos objeto de este estudio son:

1. Gestión de la investigación
2. Gestión económica y contabilidad analítica
3. Gestión Académica.

Se prevé que la recogida de documentación y el traspaso de información a los expertos que deban seleccionar los sistemas para implantar, se pueda realizar en un periodo de unos tres días para cada módulo objeto de estudio, teniendo en cuenta que podrían ser ampliables si se considerase necesario, eso sí, respetando siempre el límite de esta fase que se ha establecido en 3 meses.

Fase Operativa:

Los módulos a implantar anteriormente mencionados, formarán parte del desarrollo del nuevo sistema de gestión de investigación universitaria.

Para este desarrollo queda clara la importancia, no solo de conocer el tipo de soluciones seleccionadas (desarrollo propio o software comercial) si no también la tecnología con la que se ha desarrollado.

En esta fase se tendrán que realizar todos los trabajos necesarios para adecuar los sistemas cedidos a las especificaciones y requisitos que se hayan especificado anteriormente y de esta manera, poderlos integrar en el nuevo sistema.

Este paquete de trabajo está fuertemente integrado con el paquete de trabajo 5 y 6 y habrá actividades que se desarrollen en paralelo ya que algunos de los resultados del paquete de trabajo 5 influirán sobre el diseño de la arquitectura así como el desarrollo de la implantación del módulo de investigación.

Para poder desarrollar la implementación e integración se tienen en cuenta los siguientes puntos:

1. Análisis de requisitos generales y adaptaciones necesarias para el conjunto de aplicativos.
 - a. Establecimiento de grupos de expertos de las diferentes universidades para asesoramiento del proyecto.

- b. Reuniones de análisis con los expertos funcionales y técnicos.
 - c. Documentación de requisitos y adaptaciones.
 2. Diseño de la arquitectura global
 - a. Integración de requisitos con las necesidades definidas para la creación de la infraestructura semántica.
 - b. Diseño de la arquitectura del nuevo sistema
 - c. Diseño de las interfaces
 - d. Diseño del modelo de datos integrado
 3. Desarrollo de la implantación de los módulos indicados
 - a. Gestión de la investigación y las partes relacionadas o necesarias de los módulos de gestión académica y gestión económica.

2.1.5 PT 5. Creación de la infraestructura semántica

Se deberá abordar la creación de una infraestructura semántica basada en el paradigma de datos abiertos, para la gestión de la información y el conocimiento del sistema de gestión universitaria desarrollado.

Este paquete de trabajo pretende desarrollar la capa semántica que permitirá una explotación avanzada de los conjuntos de datos relacionados con el módulo de gestión de la investigación.

Las actividades en este paquete de trabajo se pueden clasificar en dos grandes grupos:

- 1) actividades de investigación y desarrollo para el diseño y desarrollo de la infraestructura para la explotación semántica del módulo de gestión de la investigación, y
- 2) actividades de innovación e implantación de la infraestructura que requerirá el dimensionamiento de los sistemas informáticos para la optimización de su rendimiento y su despliegue en entornos de alto rendimiento.

Si bien se buscará la obtención de resultados directamente ligados a los escenarios y casos de estudio planteados en el proyecto, el enfoque metodológico será general. Se pretende que la metodología y los artefactos producidos puedan ser fácilmente extendidos y combinados con futuros desarrollos que afecten a otros módulos del sistema.

Este paquete de trabajo hará uso de recomendaciones y estándares del W3C en materia de representación y explotación semántica de contenidos en Internet (RDF, OWL, SPARQL). Se aplicarán las buenas prácticas relacionadas con la construcción de artefactos semánticos y software promoviendo la reutilización y modularidad de los componentes. Se potenciará el uso de soluciones gratuitas y de código abierto desarrolladas a nivel nacional e internacional.

Este paquete de trabajo incluye la mayor parte de las actividades de I+D del proyecto:

1. Estudio de los últimos avances científicos y tecnológicos relacionados con los tópicos de este paquete de trabajo para seleccionar las tecnologías, herramientas y enfoques metodológicos más apropiados para el logro de los objetivos del proyecto.
2. Análisis y definición del modelo de datos del módulo de gestión de la investigación para identificar las entidades de datos implicadas, las relaciones entre las mismas y el rol que desempeñan en el módulo de gestión de la investigación.
3. Diseño de la arquitectura de la infraestructura semántica a emplear.
4. Desarrollo del conjunto de ontologías que formará parte de la infraestructura semántica a partir del análisis, estudio y reutilización de ontologías existentes, buscando alinear el

trabajo desarrollado con propuestas e iniciativas internacionales, lo que favorecerá futuros esfuerzos a nivel internacional. Algunas ontologías a desarrollar estarán relacionadas con la estructura básica del dominio del sistema y serán indispensables para cualquier caso de uso de explotación de datos y otras podrán ser específicas para los casos de uso implementados en el proyecto.

5. Definición de los interfaces de comunicación entre la infraestructura semántica y otros componentes del Sistema de Gestión Universitaria.
6. Formalización de las correspondencias entre el modelo de datos del módulo de gestión de la investigación y el conjunto de ontologías, para asegurar que los datos de cualquier universidad española usuaria del sistema son representados atendiendo al mismo esquema de conocimiento, facilitando de esta forma su compartición, comprensión y reutilización.
7. Diseño y puesta en marcha de una metodología de publicación y mantenimiento de conjuntos de datos abiertos para el módulo de gestión de la investigación.
8. Selección de la colección de datos que serán objeto de la puesta a disposición Link Open Data
9. Creación de modelos de explotación y consulta de datos para el módulo de gestión de la investigación, poniendo especial atención en los casos de uso definidos al inicio del proyecto.
10. Desarrollo de métricas de similitud semántica en los conjuntos de datos del módulo de gestión de la investigación con el fin de empleo en los casos de estudio que requieran selección o recomendación basada en criterios semánticos.
11. Evaluación técnica de la infraestructura semántica desarrollada, comprobando que se cumplen los requisitos definidos al inicio del proyecto y que es capaz de realizar las tareas previstas.

2.1.6 PT 6. Casos de uso de explotación semántica.

Este paquete de trabajo pondrá en marcha una serie de aplicaciones que explotan los conjuntos de datos relacionados con gestión de la investigación. Al principio del proyecto se determinarán cuáles se ejecutarán.

Para cada uno de ellos, el enfoque metodológico será el siguiente

1. Análisis de requisitos del caso de uso en cuanto al objetivo del caso de uso y los tipos de consultas a realizar.
2. Si fuera necesario, desarrollo de ontologías específicas que se integrarían en la infraestructura semántica desarrollada en el paquete de trabajo 5.
3. Diseño de los interfaces de consulta y explotación de datos a partir de los modelos desarrollados en el paquete de trabajo 5.
4. Implementación de los interfaces de consulta y explotación de datos
5. Evaluación técnica.

2.1.7 PT 7. Plan de pruebas

Se explicita el alcance, enfoque, recursos requeridos, calendario, responsables y manejo de riesgos del proceso de pruebas.

Se realizarán tantos planes de pruebas como el equipo de dirección del proyecto/oficina de proyecto considere necesarios para los distintos módulos e infraestructuras a desarrollar.

1. Análisis de requisitos
2. Diseño del plan de pruebas que incluye:

- a. identificación del plan y alcance del mismo, donde se incluyen los tipos de pruebas a realizar y las propiedades/elementos del software a ser probados.
 - b. ítems a probar, que indica la configuración a probar y las condiciones mínimas que deben cumplir para empezar a aplicarle el plan.
 - c. Estrategia: Describe la técnica, patrón y/o herramientas a utilizar en el diseño de los casos de pruebas.
 - d. Categorización de la configuración: Explicita las condiciones bajo las cuales el plan debe ser suspendido, repetido o culminado.
 - e. Tangibles: Especifica los documentos a entregar al finalizar el proceso descrito en el plan.
 - f. Calendario: describe los hitos del proceso de pruebas y el grafo de dependencia en el tiempo de las tareas a realizar.
 - g. Manejo de riesgos.
 - h. Recursos para el plan: Especifica las propiedades necesarias y las deseables para el plan de pruebas, tanto el hardware como el software. También se incluye una estimación de los recursos humanos necesarios.
3. Integración del personal y elementos necesarios para el desarrollo de las pruebas, en base al descrito en el plan que se ha diseñado.
 4. Realización de las pruebas. Documentación de los resultados obtenidos en el proceso en base a las especificaciones descritas en el plan de pruebas.
 5. Evaluación de los resultados y nueva planificación. Gestión de las incidencias.

2.1.8 PT8. Implantación en la universidad licitante de la CPP

Una vez finalizadas todas las tareas y logrados los objetivos del proyecto, se prevé la implantación del prototipo desarrollado en algún centro de la universidad licitante de la CPP para certificar el éxito del proyecto y cumplimiento de los objetivos.

Para ello será necesario dotar a la misma de la infraestructura (del tipo que se haya determinado a lo largo del proyecto) y licencias necesarias. En el ámbito de la implantación del prototipo se realizará una migración progresiva de los datos del módulo de gestión de la investigación, en base a la volumetría y complejidad de los mismos.

Se crearán sitios piloto (3 centros de la universidad licitante) para probar el trabajo en sus dimensiones técnicas y en términos de la organización y las respuestas de los usuarios. La experiencia de los sitios pilotos nos permitirá prepararnos para la puesta en funcionamiento, nos proporcionará un mejor entendimiento de los requisitos y la posibilidad de identificar las dificultades de forma temprana.

2.1.9 PT9. Gestión del Cambio

Para poder implantar con éxito el nuevo sistema habrá que tener en cuenta la “gestión del cambio” y poner en práctica las pautas definidas en los paquetes de trabajo previos para realizar con éxito esta gestión.

Se elaborará un plan de formación dotando a las personas que considere conjuntamente entre la universidad y la empresa proveedora, de conocimientos y habilidades necesarias para que puedan realizar sus funciones y tareas de la forma esperada y no se sientan inseguros en ese nuevo entorno. Dichas personas serán formadoras a un segundo nivel del conocimiento recibido.

Para esta formación la universidad dispondrá de las instalaciones que sean adecuadas para este traspaso de información.

Con el fin de conseguir un mejor traspaso de información se estiman realizar jornadas de formación que no superen las 500 horas totales. Los destinatarios de estos cursos serán tanto personal técnico como usuarios finales.

Se dejará por escrito toda la información necesaria para el equipo técnico de la Universidad de Extremadura y se realizarán, si fuera preciso, cursos de formación para garantizar el éxito del proyecto.

1. Análisis de equipamiento necesario
2. Instalación de los módulos de software
3. Migración de datos
4. Pruebas
5. Formación y Documentación

3 Bibliografía

- Alper, A. (2006) Disruptive technologies: Semantic web. Managing automation. http://www.managingautomation.com/maonline/magazine/read/view/Disruptive_Technologies__Semantic_Web_3571723 (último acceso: 24 de julio de 2014)
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25-29.
- Bennett, M. (2013). The financial industry business ontology: Best practice for big data. *Journal of Banking Regulation*, 14(3), 255-268.
- Berners-Lee, T. (2006). Linked Data - Design Issues. W3C. Available at: http://www.w3.org/DesignIssues/Linked_data.html.
- Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific American*.
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3), 1-22
- Calvanese, D., Giese, M., Haase, P., Horrocks, I., Hubauer, T., Ioannidis, Y., ... & Zheleznyakov, D. (2013). Optique: OBDA Solution for Big Data. In *The Semantic Web: ESWC 2013 Satellite Events* (pp. 293-295). Springer Berlin Heidelberg.
- Deloitte. Open data. Driving growth, ingenuity and innovation. 2012.
- Giese, M., Calvanese, D., Haase, P., Horrocks, I., Ioannidis, Y., Kllapi, H., ... & Waaler, A. (2013). Scalable End-user Access to Big Data. *Big Data Computing*, 205.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-220.
- Hepp, M. and Wechselberger, A. OntoNaviERP: Ontology-Supported Navigation in ERP Software Documentation. *International Semantic Web Conference 2008, Lecture Notes in Computer Science*, 2008, Volume 5318/2008, 764-776,
- Hitzler, P., & Janowicz, K. (2013). Linked data, big data, and the 4th paradigm. *Semantic Web*, 4(3), 233-235.
- Partridge, C. and Stefanova, M. (2001). A synthesis of estate of the art enterprise ontologies. In *Lessons Learned. 2001, The BORO Program, LADSEB CNR*.
- Thompson, Victoria. Asos to open data to external web developers. *Retail Week*. [En línea] 20 de diciembre de 2011. <http://www.retail-week.com/technology/asos-to-open-data-to-external-web-developers/5032186.article#>.
- Wang, G., Ren, B. Y., Che, Y., & Lv, M. (2013). Research on Model Transformation Method Oriented to Model Driven ERP System. *Advanced Materials Research*, 670, 208-215.
- Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 de noviembre de 2003, relativa a la reutilización de la información del sector público. *Diario Oficial n° L 345 de 31/12/2003* p. 0090 - 0096, Bruselas: Parlamento Europeo, 17 de noviembre de 2003. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:ES:HTML>.
- Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público. *BOE-A-2007-19814*, Madrid : Jefatura del Estado, 16 de noviembre de 2007. <http://www.boe.es/buscar/doc.php?id=BOE-A-2007-19814>.
- Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, para el ámbito del sector público estatal. *BOE-A-2011-17560*, Madrid : Ministerio de la Presidencia, 24 de octubre de 2011. http://www.boe.es/diario_boe/txt.php?id=BOE-A-2011-17560.

Resolución de 19 de febrero de 2013, de la Secretaría de Estado de Administraciones Públicas, por la que se aprueba la Norma Interoperabilidad de Reutilización de recursos de la información. BOE-A- 2013-2380, Madrid: Ministerio de Hacienda y Administraciones públicas, 19 de febrero de 2013. http://www.boe.es/diario_boe/txt.php?id=BOE-A-2013-2380.

PMBOK. PMI. PMP Dirección de proyectos pmi.org