

Hacia una infraestructura de datos abiertos enlazados para una gestión agroforestal sostenible

ALCARRIA, Ramón; MANSO, Miguel Ángel; ROBLES, Tomás; NAVARRO, Mariano; ESTRADA, Jesús

La provisión de soluciones de integración e interoperabilidad en la gestión de información geoespacial es una cuestión aún no resuelta. Los datos abiertos enlazados (Linked Open Data - LOD) se presentan como una solución adecuada para establecer relaciones entre modelos de datos geoespaciales que representan la realidad geográfica. En este trabajo se propone una infraestructura de datos abiertos enlazados que permita la integración e interoperabilidad semántica de grandes catálogos de datos, proporcionados por INSPIRE, GMES y GEOSS. La contribución principal de este trabajo es el diseño y desarrollo de la arquitectura de nuestra solución, que comprende una lista de funcionalidades soportadas por herramientas que están actualmente en desarrollo y un conjunto de fases para la adquisición y el tratamiento de los datos.

Esta infraestructura de datos abiertos enlazados se utilizará en una prueba piloto Hispano-Portuguesa en el ámbito del proyecto *SmartOpenData*, financiado por la Unión Europea a través del programa ICT-ENV-2013. La prueba piloto se encuentra actualmente en desarrollo y pretende generar y utilizar datos abiertos enlazados a partir de la información proporcionada por dos principales participantes, organismos públicos de ambos países (TRAGSA y Direção-Geral do Território), que disponen de información medioambiental y geoespacial que debe de ser reorganizada convenientemente. A través de esta prueba piloto y definiendo un caso de uso de gestión agroforestal validamos la arquitectura propuesta en este trabajo.

PALABRAS CLAVE

Datos abiertos enlazados, IDE, GeoSPARQL, Web Semántica, RDF.

INTRODUCCIÓN

La habilidad que ofrecen las tecnologías de la información para tratar datos geográficos ha mejorado en las últimas décadas y los sistemas de información geográfica (GIS) contemplan soluciones de integración y gestión de datos geográficos y medioambientales a nivel local, nacional e internacional. Sin embargo, estas soluciones de integración necesitan interpretar las relaciones entre modelos de datos geoespaciales que representan la realidad geográfica.

Los datos abiertos enlazados [1] permiten la integración de grandes catálogos de datos pertenecientes a distintos dominios, mediante la correcta estructuración de los datos y la identificación y asociación de recursos. Actualmente están apareciendo trabajos que establecen y consideran relaciones entre datos de naturaleza geográfica. La directiva INSPIRE proporciona conexiones con conjuntos de datos geoespaciales enlazados a través de la enmienda de los Anexos II+III [2], mientras que GMES [3] y GEOSS [4] ofrecen un conjunto reducido de funcionalidades relacionadas con el intercambio de datos a través de servicios conectados entre sí.

El proyecto SmartOpenData [5] - "Linked Open Data for Environmental Protection in Smart Regions" (SmOD), financiado por el programa FP7 ENV.2013.6.5-3, tiene como objetivo la creación de una infraestructura de datos enlazados (incluyendo conjuntos de datos y también las herramientas para su tratamiento) para la biodiversidad y la protección medioambiental de zonas protegidas en Europa, y que satisfagan los requisitos de los actores involucrados: entidades públicas, investigadores, grandes compañías, SMEs, y ciudadanos.

Un elemento principal en el proyecto es el despliegue de cinco pilotos en áreas relacionadas con la gestión agroforestal, investigación medioambiental y en biodiversidad, gestión y monitorización de

agua, sostenibilidad forestal, y reutilización de datos medioambientales. Estos pilotos requieren de funciones de armonización de metadatos, integración de datos de diversa naturaleza y también la visualización de información de acuerdo a los requisitos de los interesados.

Este trabajo está enmarcado en el proyecto SmOD y propone el diseño de una arquitectura que incorpora las funcionalidades de indexación semántica, administración y notificación y, por último, visualización cartográfica de los resultados. Se describe también el acceso a dos tipos de fuentes de datos: estructuradas (es decir que satisfacen algún estándar que permite una transformación directa a RDF) y no estructuradas (que requieren herramientas especiales e incluso adaptadas a un modelo de datos concreto).

Esta infraestructura de datos abiertos enlazados se utilizará en una prueba piloto Hispano-Portuguesa, que tiene los siguientes objetivos: (i) optimizar el uso agroforestal del suelo; (ii) diversificar la economía rural y mejorar la calidad de vida de las comunidades rurales; (iii) mejorar las actividades de gestión de bienes y suelos; (iv) preservar la biodiversidad y el entorno; (v) involucrar a los sectores público y privado en cuestiones relacionadas con la gestión forestal; (vi) mejorar la eficiencia de los procesos administrativos; (vii) mejorar la gestión de cuencas hidrográficas integrando información de uso del suelo; (viii) crear consenso acerca de los requisitos legales y normativos del uso del suelo.

El resto del documento se estructura del siguiente modo: a continuación se presenta la problemática que intenta resolver la infraestructura mediante la presentación de un ejemplo del ámbito agroforestal, posteriormente se presenta la arquitectura de la solución y se describen las tres funcionalidades soportadas. Finalmente, se describe el piloto Hispano-Portugués, como escenarios de aplicación de nuestra solución, seguido de unas conclusiones del trabajo.

MOTIVACIÓN

Los proveedores de datos y servicios se encuentran habitualmente con dos problemas o cuestiones que responder:

- ¿Cómo ofrecer contenido y funcionalidades en el marco de las obligaciones legislativas (INSPIRE), iniciativas (Open Government partnership) y expectativas de los usuarios?
- ¿Cómo enlazar contenidos y servicios desde sus portales, a través de interfaces estándar?

Estos problemas surgen por la falta de mecanismos que permitan consultar la información pública, y también por el bajo nivel de calidad de la información a menudo asociado a la misma.

En este apartado describimos un ejemplo de recomendación de semillas para el cultivo tanto agrícola como forestal en una determinada zona:

Un propietario o silvicultor desea adquirir unas semillas. Para ello necesita conocer la ubicación de los viveros de semillas o semilleros públicos más próximos. En el caso de los viveros de semillas, encontrar esta información es más difícil, ya que la reserva para visitas no es accesible por Internet. El procedimiento actual es contactar con un funcionario de la Administración Pública y rellenar un requerimiento formal para la reserva. El funcionario debe rellenar documentación asociada a los detalles de cada visitante. Además, un supervisor debe visitar el semillero público para asegurarse de que las semillas se recogen en la cantidad adecuada, y debe rellenar a su vez un informe con las fechas y lugares de recogida, compañía encargada de la recogida, etc. Esto se traduce en procesos claramente complejos y lentos. La localización de los semilleros públicos y su disponibilidad para la recogida de semillas podría estar accesible en Internet y la información introducida por el supervisor acerca de los detalles de las recogidas podría actualizarse en tiempo real. De esta forma, cualquier cambio en el estado o información sobre cualquiera de los semilleros públicos sería inmediatamente accesible y estaría disponible para cualquiera que lo necesitase.

Como problema adicional, este proceso administrativo se lleva a cabo a nivel local, no nacional. Con el tiempo, la administración española recibe toda esta información desde las Comunidades Autónomas, pero esto significa una duplicidad de esfuerzos debido a la ausencia de un conjunto de datos unificados y a la ausencia de mecanismos de intercambio de información entre las entidades nacionales y autonómicas.

Otro problema es la falta de información analítica en la selección de la especie a cultivar. Como se explicó anteriormente, los silvicultores y propietarios normalmente seleccionan las especies a sembrar en sus parcelas basándose en la experiencia o en su intuición. No tienen información o datos analíticos disponibles que definan cuál es la especie más recomendable. En este sentido, un mapa que muestre una recomendación sobre la distribución de las especies según algunas variables sería muy adecuado. Estas variables pueden ser la geología, el clima (temperaturas, precipitaciones), la vegetación, orientación, pendiente, etc.

Una característica interesante a implementar es ofrecer, de manera accesible, la disponibilidad de semillas en los semilleros en tiempo real. De esta forma, los propietarios y silvicultores elegirían el vivero más conveniente a sus intereses, lo que ahorraría tiempo y dinero.

Existen multitud de estudios en relación con la gestión de recursos genéticos forestales, que recomiendan la siembra de una determinada especie dependiendo de la región de origen y de la calidad de las semillas. Sin embargo, estos estudios son engorrosos y su búsqueda y consulta es lenta. Además son difíciles de interpretar por usuarios sin la debida formación y conocimientos.

Todos estos problemas impiden un proceso de reforestación correcto, al no considerar los factores relevantes: especie, región de origen, calidad de la semilla, planificación de la producción, tiempo óptimo de siembra, etc.

El resultado de esta falta de planificación es el peor uso de los recursos genéticos forestales y la disminución de la producción forestal. La solución viene de la mano de hacer disponibles los datos y servicios necesarios para las partes interesadas con el fin de hacer el proceso más fácil. La Figura 1 pretende esquematizar la necesidad de la consulta de información geográfica para el estudio de los factores que permiten una adecuada selección de la especie a cultivar.

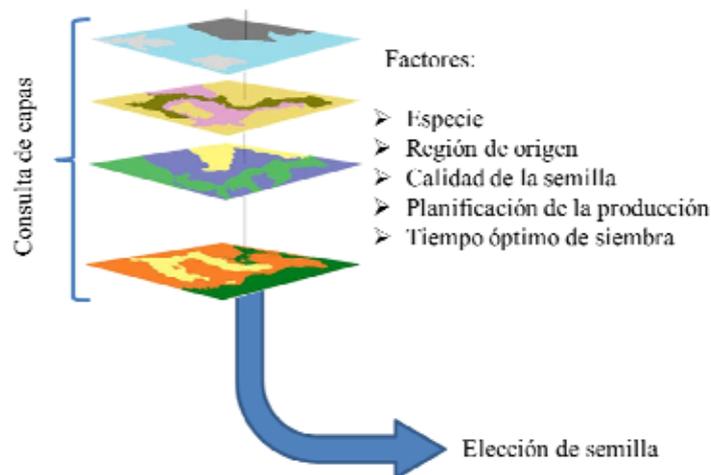


Figura 1. Consulta de información geográfica y generación de resultado de recomendación

Estos problemas introducen los siguientes retos:

- Descubrimiento: Un descubrimiento eficaz de los datos puede conseguirse mediante la implementación de grandes catálogos de metadatos.
- Federación: Proporcionar una publicación y mantenimiento de catálogos de datos más descentralizado y próximo a los actores y dueños de los datos.
- Interoperabilidad: Compatibilidad entre catálogos de datos y servicios, aunque hayan sido desarrollados por distintos proveedores. Armonización de metadatos geospaciales (basados en las normas ISO19115/19119) con los principios de la Web Semántica. En el contexto de SmOD los metadatos sirven como un componente de intercambio de información compatible con INSPIRE hacia/desde otros modelos de datos.
- Visualización de datos abiertos enlazados: Publicación de la información resultante del proceso de estructuración y enlazado de acuerdo con los requisitos de los actores implicados. Los metadatos servirán como punto de entrada para la provisión de la información necesaria para la transformación de datos espaciales a estructuras RDF (*Resource Description Framework*).

ARQUITECTURA DE LA SOLUCIÓN

Esta sección propone el diseño de una arquitectura de una infraestructura de datos abiertos enlazados que permita la indexación semántica de datos, la notificación de cambios en los datos obtenidos y, por último, la visualización cartográfica de los resultados.

La Figura 2 describe la arquitectura propuesta, que integra las funciones de indexación semántica, administración y notificación y visualización cartográfica de los resultados. Se describe también el acceso a dos tipos de fuentes de datos: estructuradas (satisfacen algún estándar que permite una transformación directa a RDF) y no estructuradas (que requieren herramientas especiales e incluso adaptadas a un modelo de datos concreto).

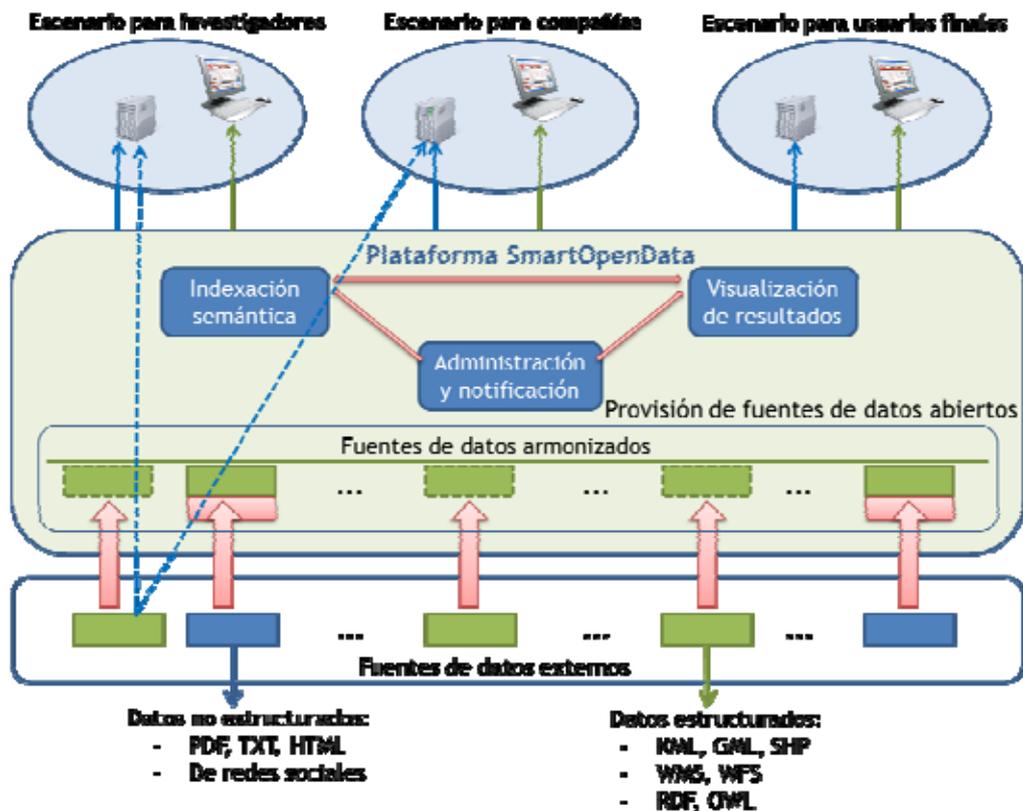


Figura 2. Arquitectura de la plataforma SmartOpenData

Las funciones definidas siguen un proceso que se describe a través de una serie de fases:

1. Definición de escenarios y pilotos
2. Análisis de los almacenes de datos actuales
3. Definición de nuevos modelos RDF compatibles con INSPIRE
4. Definición de modelos de transformación para la transformación semi-automática de dichos datos
5. Desarrollo de una interfaz para la consulta de datos utilizando SPARQL y GeoSPARQL
6. Visualización cartográfica y textual de los resultados de las consultas

En las siguientes secciones de este trabajo se describen las funciones de indexación semántica, administración y notificación y, por último, visualización cartográfica de los resultados.

INDEXACIÓN SEMÁNTICA

El modelo de datos de INSPIRE se define por el Centro Común de Investigación (Joint Research Centre - JRC) y se publica como un conjunto de modelos pertenecientes a distintas temáticas. INSPIRE considera un conjunto de 34 temas, agrupados en tres anexos. Para el proyecto SmOD se necesitarán los temas "Nombres Geográficos" y "Sitios protegidos" (esquema de aplicación simple), que se encuentran en el Anexo I.

El proyecto SmOD busca el uso de INSPIRE como base para las estructuras de datos de cada piloto. Sin embargo, las directrices INSPIRE se han definido como esquemas XML y GML, es decir, utilizando tecnologías no enlazadas. Existen algunos trabajos para la utilización de datos enlazados en el contexto de INSPIRE, en particular, en el proyecto GeoKnow [6]. El JRC se encuentra en proceso de

convertir los esquemas de INSPIRE a RDF/OWL pero este trabajo aún no está completo y la mayor parte de los que se ha completado aún no está disponible públicamente.

La indexación semántica requiere que los datos suministrados por los proveedores de datos se transformen a RDF para que puedan ser enlazados y utilizados por los pilotos de forma apropiada.

La Figura 3 muestra el proceso que deben seguir los datos que se encuentran en una base de datos relacional para que puedan ser recuperados por los usuarios. El primer paso es transformar los datos estructurados en tablas a datos estructurados mediante RDF. Además de ese proceso de transformación se necesita un proceso de enlazado para que se establezcan las relaciones entre entidades convenientemente. Estos dos procesos producen un grafo RDF que contiene las relaciones semánticas entre los conceptos definidos en la base de datos.

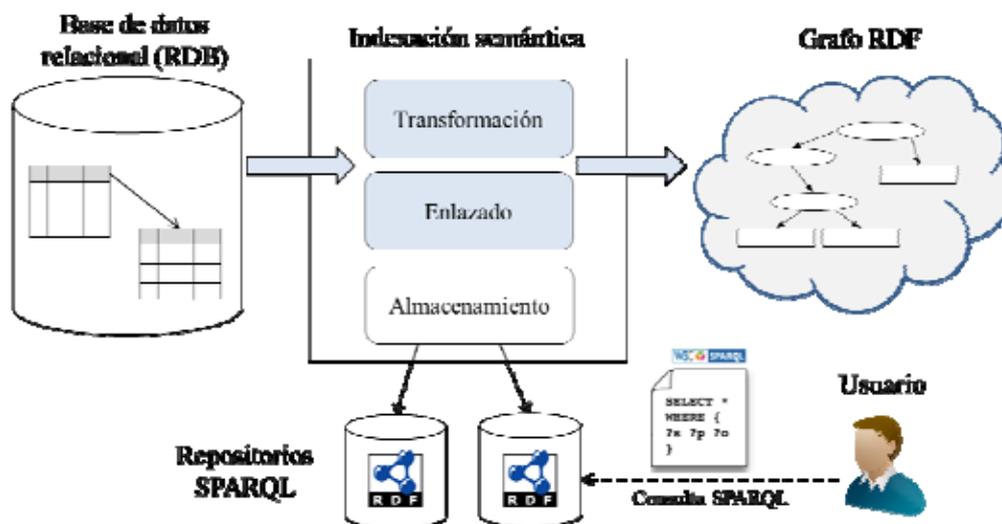


Figura 3. Proceso de transformación de los datos a estructuras RDF

Para el desarrollo de esta funcionalidad se han estudiado algunas herramientas:

Para las transformaciones de RDB a RDF se han analizado *Sparqlify*¹, que reescribe las peticiones SPARQL en SQL y permite definir vistas en RDF de bases de datos relacionales, para su posterior consulta en SPARQL. También se ha analizado *TripleGeo*², una herramienta *open source* para la extracción de características geoespaciales en forma de tripletas RDF.

Para establecer relaciones en el campo de los datos enlazados, se han analizado *GeoLift*³, que es un *framework* de mapeo espacial implementado en Java para el enriquecimiento de datos RDF con información geo-espacial, y *LIMES*⁴, utilizado para el descubrimiento de enlaces a gran escala de la Web de los Datos (Web of Data).

Por último, de entre las herramientas relacionadas con el almacenamiento de los datos, una vez transformados en RDF destacamos *Virtuoso*⁵, que es un gestor de base de datos que proporciona herramientas para la gestión de datos en RDF, además de soportar N3/N-Triples, serialización de datos y peticiones SPARQL; *Parliament*⁶, es un almacén de tripletas (Triplestore) compatible con los estándares RDF, RDFS, OWL, SPARQL y GeoSPARQL, y por último, *uSeekM*⁷, que es una extensión para Sesame para extender las bases de datos semánticas con indexación y capacidades de consulta.

El empleo de características de LOD depende del establecimiento de identificadores únicos (Uniform

¹ Sparqlify: <http://aksw.org/Projects/Sparqlify>

² TripleGeo: <https://github.com/GeoKnow/TripleGeo>

³ GeoLift: <http://aksw.org/Projects/GeoLift>

⁴ LIMES: <http://aksw.org/Projects/LIMES>

⁵ Virtuoso: <http://virtuoso.openlinksw.com/>

⁶ Parliament: <http://parliament.semwebcentral.org/>

⁷ uSeekM: <https://dev.opensahara.com/projects/useekm>

resource identifier - URI) para los recursos a enlazar, que corresponden con objetos reales, definiciones y documentos. Para la función de indexación semántica se ha trabajado en construir URIs para los objetos especiales que resulten lo más persistentes posibles. El patrón que hemos contemplado para objetos especiales es de la forma:

```
http://{domain}/{collection*}/{so}/{class}/{inspireLocalId}/{inspireVersionId}
```

donde [/so] significa que esta URI identifica a un objeto espacial (*spatial object*). Aunque es opcional, se recomienda encarecidamente su utilización, ya que diferencia de forma inmediata si el objeto es un objeto espacial [/so], un lugar del mundo real [/id], o un documento que describe un elemento del mundo real [/doc]. Esta entidad es de un tipo definido mediante la clase [{class}]. {inspireLocalId} es un parámetro opcional y responde al identificador del lugar según el modelo Inspire.

Para cada entorno de aplicación, es decir, cada piloto dentro del proyecto SmOD, esta estructura de datos tendrá que extenderse para incluir vocabularios específicos del dominio y requerirá la definición de nuevos conceptos y relaciones. Para hacer esto, se recomienda el uso del documento de buenas prácticas publicado por el W3C [7].

ADMINISTRACIÓN Y NOTIFICACIÓN

Esta funcionalidad tiene el objetivo de facilitar el intercambio de datos y la gestión de usuarios, procesos y datos. La funcionalidad de administración estará disponible para determinados usuarios, de forma que estos puedan habilitar el acceso a los datos y a los servicios que procesan esos datos a los proveedores y consumidores de los mismos.

La Figura 4 muestra la interfaz desarrollada para la administración de los usuarios. Se puede ver una lista de usuarios registrados, pertenecientes a distintos grupos (administradores, propietarios de un terreno, etc.). Cada uno de estos usuarios puede haber configurado unas notificaciones para que se les avise bajo ciertas condiciones; por ejemplo si un semillero público vuelve a tener stock de la semilla *Pinus sylvestris*, buscada por el usuario *propPinarMaceda*.



Figura 4. Panel de administración de usuarios y visualización de notificaciones asociadas

El servicio de notificación tiene la función de avisar a los usuarios cuando existen nuevos datos relacionados con las consultas efectuadas o de acuerdo a un conjunto de reglas.

El sistema de notificación funciona mediante consultas asíncronas siguiendo el modelo de publicación suscripción. La siguiente figura muestra el proceso de notificación que se ha definido y está actualmente desarrollándose.

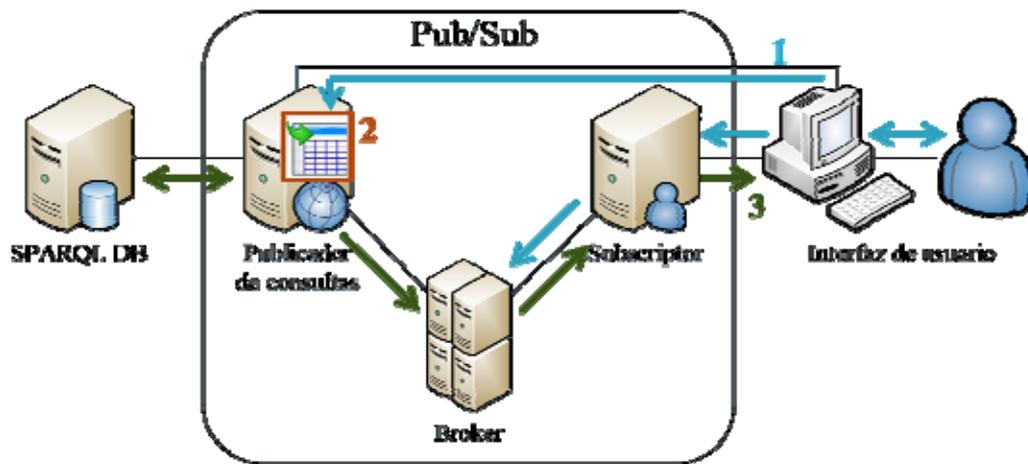


Figura 5. Proceso de notificación de cambios en las consultas efectuadas

El usuario realiza una petición de información geoespacial al servidor de suscripción. Este servidor almacena la petición y la asocia a un usuario (1). Además envía un mensaje de suscripción al Bróker, que es el elemento que se encarga de mantener la correspondencia entre las entidades publicadoras y subscriptoras. El publicador almacena la petición y el resultado de la operación (2) (que puede comunicar al cliente si este está interesado en el resultado). El publicador continúa ejecutando la consulta con una frecuencia determinada y comunica a través del Bróker si los resultados devueltos por la base de datos han cambiado respecto a la última petición. De esta forma el subscriptor reconoce los cambios y notifica al cliente la información solicitada, que pueden ser esos cambios o un aviso de que se han producido.

VISUALIZACIÓN CARTOGRÁFICA DE LOS RESULTADOS

La función de visualización de los resultados proporciona un visor web que permite realizar consultas SPARQL y GeoSPARQL y procesa los resultados de esas consultas representando información en un mapa y ofreciendo información adicional.

Hasta ahora no se ha desarrollado un cliente de visualización completo, pero sí varias pruebas de concepto utilizando información ya estructurada en RDF.

Por ejemplo en la Figura 6 se puede observar una consulta SPARQL sobre la población de un municipio de España.

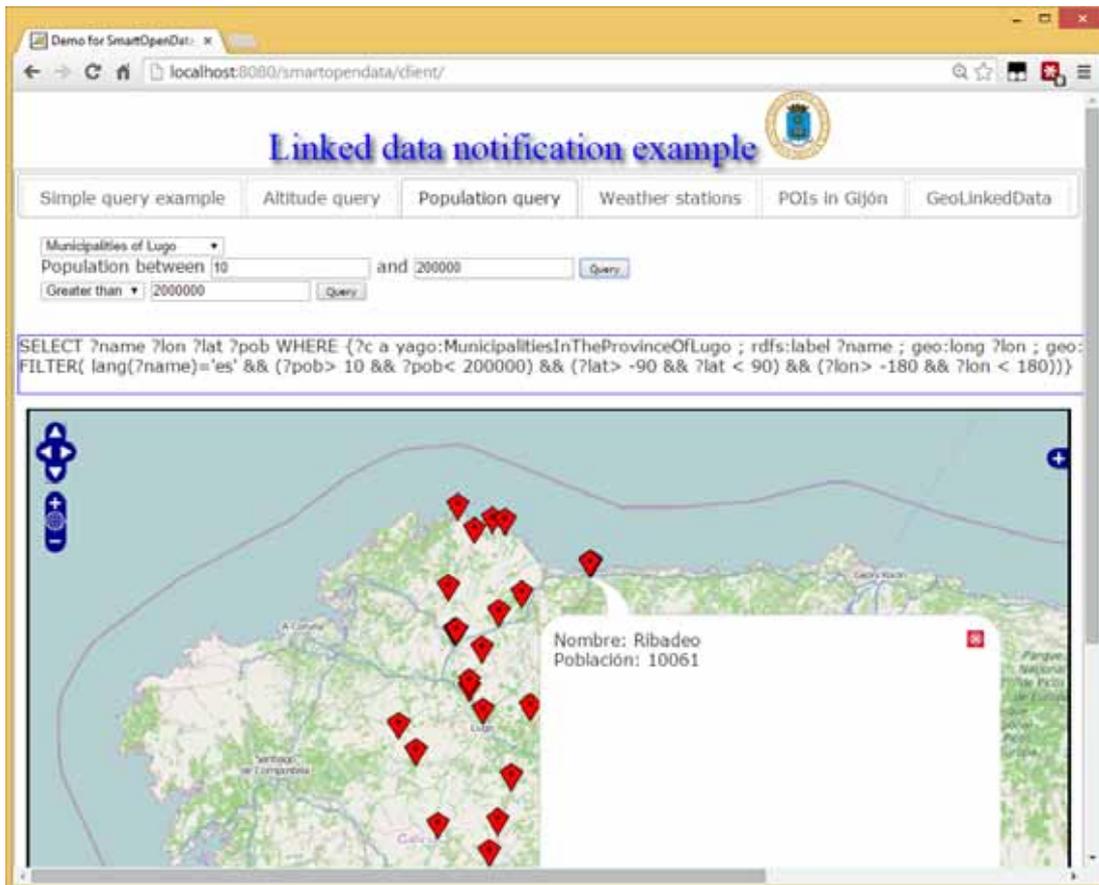


Figura 6. Resultado de consulta SPARQL visualizada sobre un mapa de OpenStreetMaps

La consulta realizada es la mostrada en la Figura 7. Como puede observarse, se establecen unas condiciones sobre los conceptos `yago:MunicipalitiesInTheProvinceOfLugo`, `dbpedia-owl:populationTotal`, `geo:lan`, y `geo:lon`.

```
SELECT ?name ?lon ?lat ?pob WHERE {
  ?c a yago:MunicipalitiesInTheProvinceOfLugo;
  rdfs:label ?name;
  geo:long ?lon;
  geo:lat ?lat;
  dbpedia-owl:populationTotal ?pob. FILTER(
    lang(?name)='es' &&
    (?pob > 10 && ?pob < 200000) &&
    (?lat > -90 && ?lat < 90) &&
    (?lon > -180 && ?lon < 180)
  )
}
```

Figura 7. Consulta SPARQL para la obtención de la población de los municipios de Lugo

ESCENARIO DE APLICACIÓN

La arquitectura definida en el proyecto SmOD se validará en un escenario real enmarcado en el caso de uso de recomendación en la selección de semillas, definido en la sección de motivación. El piloto será realizado como colaboración entre España y Portugal a través de empresas y organismos líderes

en el sector (TRAGSA y Direção-Geral do Território). El escenario se encuentra en los municipios de Allariz y Macera, en la región de Galicia, la noroeste de España. Esta área está principalmente utilizada para agricultura siendo el turismo y urbanismo en esta área actividades secundarias.

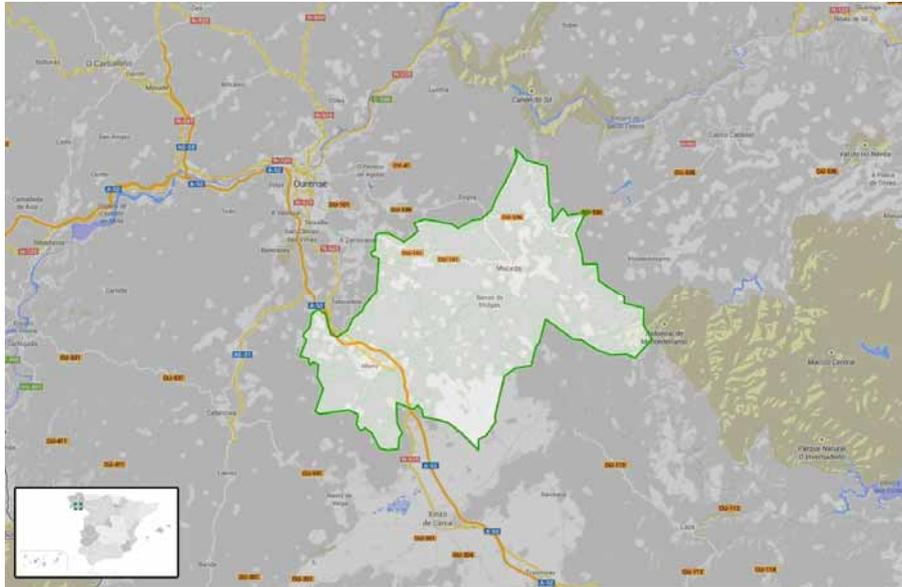


Figura 8: Situación del área de Macera y Allariz

Los participantes en esta prueba piloto son propietarios de bosques y parcelas destinadas a la agricultura, administraciones públicas, empresas de distribución y venta de semillas, y también viveros de plantas y semillas. Estos grupos de interés participan en una cadena que se inicia con la selección de las mejores especies para una parcela determinada y termina con la cosecha y obtención de las semillas una vez determinada la ubicación de la cosecha.

La principal preocupación para los propietarios y silvicultores es saber qué especie es la más adecuada para cada una de las parcelas. A continuación, tiene que descubrir dónde pueden obtener las semillas de esta especie particular. Obviamente puede recurrir a viveros de semillas, como el que existe en Macera y que además participa en esta prueba piloto. Sin embargo también pueden acudir a semilleros públicos. Estos semilleros, controlados por la administración no pueden ser visualizados de forma online, por lo que se debe contactar con un funcionario de la administración para obtener la información sobre cuáles están disponibles y su ubicación geográfica. Una vez hecho esto, se debe contactar con los semilleros para obtener información de las semillas disponibles, su cantidad y calidad.

Para conseguir que la información pública esté disponible para todas las partes interesadas y se pueda visualizar la localización de los semilleros públicos, el stock y la calidad de las semillas de las que disponen se debe realizar un modelo en el que se expongan de manera clara los servicios necesarios y las fuentes de datos a utilizar. Hemos realizado un esquema (Figura 9) que describe esta información.

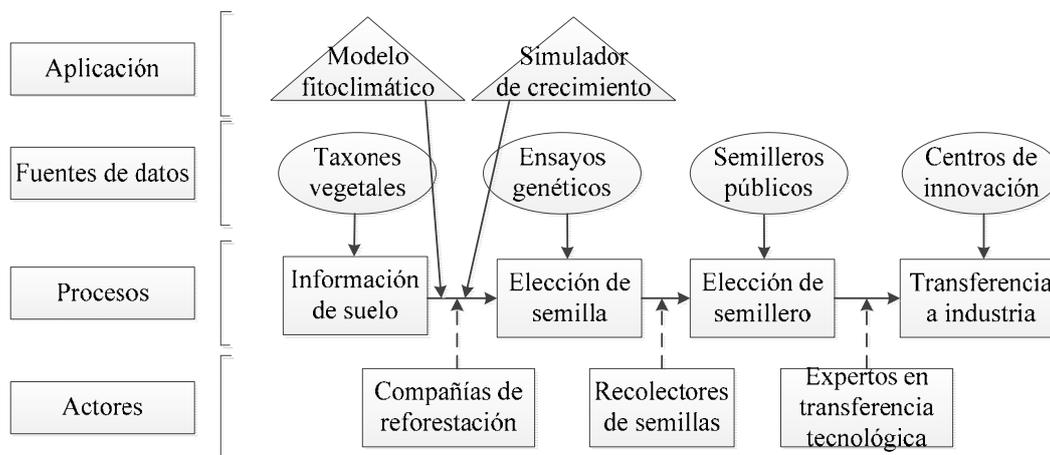


Figura 9: Elementos considerados en el escenario de aplicación

Como puede verse en la figura, para la fase de selección de la semilla contaremos con información del suelo en el área de Maceda y Allariz, procedente de las bases de datos de taxones vegetales, geobotánica, etc., y con bases de datos de ensayos genéticos como *Silvadat*. También contaremos con la utilización de aplicaciones de modelos fitoclimáticos y simuladores de crecimiento.

Las aplicaciones y servicios que soportan modelos fitoclimáticos, como *Climaforest*, permiten conocer el grado de idoneidad de una especie en un determinado territorio. Los simuladores de crecimiento, como *GESMO*[®], facilitan a los selvicultores y gestores forestales la utilización de modelos de crecimiento complejos, ocultando la complejidad matemática y de procesamiento de datos de las interfaces de usuario y proporcionando, por tanto, una herramienta sencilla de utilizar para expertos de dominio.

Para el proceso de transferencia a la industria, consideramos la industria maderera como la principal interesada en las regiones de Maceda y Allariz, por contener predominantemente bosques. La transferencia a la industria se realizará a través del Centro de Innovación y Servicios Tecnológicos de la Madera de Galicia, perteneciente a la *Consellería de Economía e Industria* de la Xunta de Galicia.

Los resultados esperados tras la puesta en funcionamiento de este piloto son, de forma general, que los procesos sean más automáticos, transparentes y permitan publicar de manera accesible la información relevante para los interesados en cada fase. Particularmente se considerará:

- Selección de las especies más adecuadas para cada parcela.
- Selección de un semillero público basado en las especies que administra y su disponibilidad.
- Selección de un vivero de semillas basado en las especies, calidades y proximidad.

Se considerarán como métricas para evaluar los beneficios de este piloto los siguientes:

- Grado de cooperación de las distintas entidades participantes en el piloto para desarrollar y adoptar la solución.
- Registros de peticiones para la selección de semillas y semilleros.
- Número de accesos y de ejecuciones de esos servicios.
- Comparación de utilización de estos servicios frente a la utilización de los servicios tradicionales de consulta de semilleros y semillas. Comparación con el ratio de consultas a los servicios tradicionales de los años anteriores.
- Utilización de los indicadores de calidad definidos en el standard ISO/IEC 25022, que contemplan satisfacción, adecuación, usabilidad, fiabilidad, seguridad en la información y mantenibilidad.

Esta información será analizada para mejorar el modelo desplegado en el piloto, desde el punto de vista de los indicadores de calidad y de las características ofrecidas. La información recolectada será procesada para identificar posibles problemas técnicos y deficiencias de usabilidad.

RESUMEN Y TRABAJOS FUTUROS

En este trabajo hemos definido el trabajo efectuado para el desarrollo de una infraestructura sostenible de LOD para promover la protección medioambiental y la compartición de datos entre entidades públicas en la Unión Europea. Los datos abiertos enlazados permiten la integración de datos estructurados en RDF para realizar consultas semánticas utilizando el lenguaje SPARQL.

Este trabajo, enmarcado dentro del proyecto SmartOpenData (programa FP7 ENV.2013.6.5-3), ha descrito el diseño de una arquitectura de la infraestructura de datos abiertos enlazados que permita llevar a cabo las funciones de indexación semántica de datos, la notificación de cambios en los datos obtenidos y, por último, la visualización cartográfica de los resultados. Se han descrito estas funciones con detalle, mostrando el estado actual de cada uno de los desarrollos.

Relacionado con la **indexación semántica**, consideramos fundamental el uso de la directiva INSPIRE para la estructuración de datos en los escenarios del proyecto. La estructuración en RDF requiere de la definición de un modelo de datos y de unas herramientas para su transformación, enlazado y almacenamiento.

La función de **administración y notificación** facilita el intercambio de datos y la gestión de usuarios. Se utiliza para ello el modelo de publicación suscripción para que los usuarios finales obtengan en tiempo real cambios en los datos en los que están interesados.

La función de **visualización cartográfica** proporciona un visor web que permite realizar consultas SPARQL y GeoSPARQL presentando la información en un mapa. El cliente desarrollado permite realizar algunas consultas SPARQL y reflejar la información como *mashup* de datos.

La arquitectura definida en este trabajo se validará en un escenario real de selección óptima de semillas para la producción de madera en las regiones de Maceda y Allariz, en Galicia, España. Se pretende demostrar el impacto de la compartición y explotación de datos provenientes de diferentes fuentes de información y su aplicación a un escenario en el que intervienen varios actores (compañías, administraciones públicas y propietarios de bosques y parcelas).

Como resultados preliminares del desarrollo de las herramientas y la estructuración de los datos podemos decir que la fiabilidad y exactitud de la visualización de los datos está siempre supeditada a la calidad de los datos y al correcto funcionamiento de los repositorios de datos, por lo que un correcto modelado, estructuración y enlazado de los datos es fundamental para asegurar el éxito de la prueba piloto.

Como trabajos futuros, pretendemos completar nuestro modelo de datos para el escenario descrito y así proceder a una transformación semiautomática de los mismos a datos en RDF. La automatización en este campo es muy necesaria, ya que cualquier cambio en los datos debe ser traducido a RDF y re-enlazado, para que las aplicaciones que consuman los datos desde puntos de acceso SPARQL puedan ser conscientes de estos cambios.

Se estudiarán también otros posibles enfoques como utilizar la ambiciosa propuesta de GeoJSON-LD, extensión del estándar JSON-LD [8], para añadir esta alternativa al conjunto de opciones a proporcionar a los futuros actores dispuestos a extender su información espacial hacia la dimensión semántica.

REFERENCIAS

- [1] Bizer, C., Heath, T., & Berners-Lee, T. Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS), 5(3), 1-22. doi:10.4018/jswis.2009081901 (2009).

- [2] Regulation (EU) No 1253/2013: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:331:0001:0267:EN:PDF>
- [3] Aschbacher, J. & Milagro-Pérez, M.P. The European Earth monitoring (GMES) programme: Status and perspectives, *Remote Sensing of Environment*, Volume 120, Pages 3-8, ISSN 0034-4257, doi:10.1016/j.rse.2011.08.028 (2012).
- [4] Christian. E. Planning for the Global Earth Observation System of Systems (GEOSS), *Space Policy*, Volume 21, Issue 2, Pages 105-109, ISSN 0265-9646, doi:10.1016/j.spacepol.2005.03.002 (2005).
- [5] Archer, P., Charvat, K., Navarro, M., Iglesias, C. A., O'Flaherty, J., Robles, T., & Roman, D. Linked Open Data for Environment Protection in Smart Regions-The SmartOpenData Project., ENVIP conference (2013).
- [6] Athanasiou, S., Hladky, D., Giannopoulos, G., García-Rojas, A., & Lehmann, J. GeoKnow: Making the Web an Exploratory Place for Geospatial Knowledge, *ERCIM News* 2014, No. 96 (2014).
- [7] Hyland, B., Atemezing, G., & Villazón-Terrazas, B. Best Practices for Publishing Linked Data, *W3C Note*, 9 January 2014. <http://www.w3.org/TR/Id-bp/>
- [8] A JSON-based Serialization for Linked Data. *W3C Recommendation* 16 January 2014. <http://www.w3.org/TR/json-ld/>

AUTORES

Ramón ALCARRIA
 ramon.alcarria@upm.es
 Universidad Politécnica de Madrid
 Departamento de Ingeniería Topográfica y Cartografía

Miguel Ángel MANSO
 m.manso@upm.es
 Universidad Politécnica de Madrid
 Departamento de Ingeniería Topográfica y Cartografía

Tomás ROBLES
 tomas.robles@upm.es
 Universidad Politécnica de Madrid
 Departamento de Ingeniería de Sistemas Telemáticos

Mariano NAVARRO
 mnc@tragsa.es
 Grupo Tragsa
 Unidad I+D+i

Jesús ESTRADA
 jmev@tragsa.es
 Grupo Tragsa
 Unidad I+D+i