

Hubs virtuales para facilitar el acceso a Datos Abiertos

El proyecto europeo ENERGIC OD desplegará distribuidores virtuales de datos para armonizar el acceso a datos y servicios geográficos heterogéneos

LATRE, Miguel Ángel; LOPEZ-PELLICER, Francisco J.; KAMALI, Nargess; BAUER, Michael; PREVITALI, Mattia; BRUMANA, Raffaella; BRAUMANN, Stefan; KUECHLY, Helga; MAZZETTI, Paolo; NATIVI, Stefano

El proyecto europeo ENERGIC OD (*European Network for Redistributing Geospatial Information to user Communities - Open Data*), financiado por la Unión Europea dentro del Programa Marco para la Competitividad y la Innovación (CIP), comenzó su andadura de tres años en octubre de 2014. El objetivo general del proyecto es el de facilitar el uso de datos geográficos abiertos procedentes de fuentes diversas a través de la creación de distribuidores virtuales de datos (*hubs virtuales*) y la creación de servicios y aplicaciones innovadores que los utilicen.

En ENERGIC OD, los *hubs* virtuales se conciben como sistemas de información capaces dar soporte al ciclo de vida completo de los datos abiertos: publicación, descubrimiento y acceso. Facilitarán el uso de datos abiertos eliminando las barreras que dificultan la utilización de información geográfica por parte de usuarios finales y desarrolladores de *software*.

Uno de los principales obstáculos para la utilización y reutilización de datos abiertos es la heterogeneidad existente tanto en datos y formatos como en servicios, que requiere a usuarios y desarrolladores grandes esfuerzos en términos de acceso a los datos y de armonización de los mismos para poder ser utilizados. La adopción de estándares que especifiquen modelos de datos y metadatos e interfaces de servicios, aunque la reduce, no consigue evitar completamente esta heterogeneidad, ya que distintas infraestructuras tienen que adoptar distintos estándares para responder a problemas específicos y para resolver casos de uso también específicos. Así pues, la heterogeneidad es inevitable, especialmente en contextos interdisciplinares. Los *hubs* virtuales de ENERGIC OD hacen frente al problema de la heterogeneidad adoptando un enfoque basado en la mediación: componentes especiales (intermediarios o *brokers*) tienen como misión armonizar interfaces de servicio, modelos de datos y metadatos, permitiendo descubrir y acceder a datos e infraestructuras heterogéneas.

Como proyecto de innovación, ENERGIC OD está integrando varias tecnologías ya existentes para implementar sus *hubs* virtuales como puntos únicos de acceso a información geográfica que es proporcionada por plataformas e infraestructuras ya existentes o en creación, incluyendo entre ellas sistemas conformes con INSPIRE y servicios de Copernicus. ENERGIC OD desplegará un total de cinco *hubs* virtuales a nivel nacional en Francia, Alemania, Italia, Polonia y España y otro más a nivel europeo.

El mayor impacto que se espera del establecimiento de los *hubs* virtuales de ENERGIC OD es la creación de nuevas oportunidades de mercado, abriendo todavía más el acceso a la información del sector público y a datos de investigación. Por lo tanto, ENERGIC OD se dirige no solo a usuarios finales, que tendrán la posibilidad de acceder a los *hubs* a través de un portal geográfico, sino también a desarrolladores de *software*, que podrán acceder a la funcionalidad de los *hubs* virtuales a través una interfaz de programación de aplicaciones (API) sencilla.

El consorcio responsable de ENERGIC OD está comenzando el desarrollo de diez aplicaciones que serán construidas sobre los *hubs* virtuales una vez que estén desplegados. Estas aplicaciones pretenden demostrar cómo los *hubs* virtuales facilitarán el desarrollo de nuevas aplicaciones multidisciplinares basadas en la explotación de datos geográficos abiertos y, por tanto, estimulando nuevas actividades de innovación y oportunidades de negocio.

PALABRAS CLAVE

Datos abiertos, Información geográfica, datos geográficos abiertos, *hub* de datos, mediación, *brokering*, armonización.

INTRODUCCIÓN

Los datos abiertos son un escenario tecnológico de gran actualidad y normalmente promocionado como una de las bases principales de la sociedad de la información en el futuro más inmediato. De hecho, el concepto de datos abiertos abarca no solo aspectos tecnológicos, sino también socio-económicos, previendo el nacimiento y crecimiento de nuevos modelos de negocio, mercados y la creación de valor social [1]. No parece existir ningún dominio científico o área de aplicación que escape a la tendencia actual a la apertura de datos [2]. Iniciativas europeas y mundiales de gran relevancia han iniciado también acciones de ayuda y guía dirigidas al movimiento de los datos abiertos, como la Research Data Alliance (RDA)¹ y la Open Knowledge Foundation (OKF)². Incluso organismos públicos que destinan financiación a determinados proyectos, como la Comisión Europea, explícitamente promueven la distribución de datos abiertos (y no solo datos «a secas»). En particular, la Comisión Europea reconoce la importancia de los datos abiertos en su Agenda Digital para Europa, centrándose en generar valor añadido a través de la reutilización de un tipo específico de datos abiertos: la información del sector público [3].

Sin embargo, existen diversas barreras que pueden actuar como un obstáculo para alcanzar una distribución y explotación efectiva de los datos abiertos. Aunque su promoción y publicitación adecuadas pueden eliminar algunas de ellas, como las legales, las técnicas y las relativas a la inercia en el mantenimiento de procesos de trabajo, otras, sin embargo, permanecerán. De hecho, incluso asumiendo que se consiguiera una apertura total de datos, esta apertura por sí sola no sería suficiente para garantizar los beneficios que se esperan de ella. No solo se necesita una mayor *apertura* de datos, sino también una mayor *usabilidad* de los mismos. La enorme cantidad de fuentes de datos abiertos y de portales que los ofrecen y la gran variedad de interfaces de búsqueda, formatos de metadatos, interfaces de acceso y formatos de los datos, tanto propietarios como no propietarios, son todos ellos ejemplos de obstáculos para la utilización práctica de la información que se publica de forma abierta.

Por lo tanto, es necesario el diseño, desarrollo e implementación de infraestructuras de datos avanzadas que proporcionen servicios que mejoren el grado de apertura de los datos que se vienen publicando de forma que estos sean realmente utilizables, y que aseguren a los usuarios la posibilidad de explotarlos completamente, liberando su valor intrínseco.

En el año 2013, la Comisión Europea comenzó la financiación del proyecto de innovación ENERGIC OD³ (*European Network for Redistributing Geospatial Information to user Communities - Open Data*), en respuesta a una convocatoria de proyectos que «deberían centrarse en el desarrollo de hubs virtuales que faciliten el uso de información geográfica abierta (libremente accesible) procedente de fuentes diversas para la creación de aplicaciones y servicios innovadores» [4]. El proyecto, que comenzó en octubre de 2014 y terminará en septiembre de 2017, está coordinado por el Consiglio Nazionale delle Ricerche (CNR) de Italia y en él participan 15 socios de seis países distintos de la Unión Europea (España, Italia, Francia, Alemania, Polonia y Reino Unido), entre los que hay ocho empresas privadas, dos administraciones públicas y cinco centros de investigación y universidades, entre los que se encuentra la Universidad de Zaragoza como socio español.

Este artículo describe la solución propuesta y desarrollada en el proyecto ENERGIC OD para facilitar el uso de información geográfica abierta, comentando los logros ya conseguidos y sus beneficios potenciales. En particular, se presentan los objetivos del proyecto, la metodología que se pretende seguir, el estado de desarrollo actual y los beneficios esperados.

OBJETIVOS

La necesidad de propiciar la usabilidad de los datos más allá de su apertura fue identificada desde los inicios del movimiento de datos abiertos por Tim Berners-Lee, inventor de la web, que sugirió la utilización de una escala basada en cinco estrellas para caracterizar el grado de apertura de los datos [5].

★	publica tus datos en la Web (con cualquier formato) y bajo una licencia abierta
★★	publicalos como datos estructurados (ej: Excel en vez de una imagen de una tabla escaneada)
★★★	usa formatos no propietarios (ej: CSV en vez de Excel)

¹ <https://rd-alliance.org/>

² <https://okfn.org/>

³ <http://www.energic-od.eu/>

★★★★	usa URIs para denotar cosas, así la gente puede apuntar a estas
★★★★★	enlaza tus datos a otros datos para proveer contexto

Figura 1. Clasificación de estrellas del grado de apertura de los datos de Tim Berners-Lee⁴

Según esta clasificación, cuantas más estrellas tiene un conjunto de datos, más usable es (Figura 1). Los proveedores de datos abiertos deberían, por tanto, tratar de mejorar el nivel en la clasificación de los datos que publican. Sin embargo, debido a la cada vez mayor cantidad de datos que se generan y se hacen disponibles, esta mejora no es posible de forma manual. En el caso de la información geográfica, este problema es evidente cuando se tiene en cuenta información procedente de sistemas de observación de la Tierra, como sensores automáticos e imágenes de satélite, que generan una cantidad de datos enorme que requieren técnicas automáticas para su gestión, descubrimiento, acceso y uso.

Esto plantea un gran reto para el mundo de la distribución de datos geográficos: ¿cómo es posible diseñar e implementar sistemas avanzados capaces de incrementar el grado de apertura y usabilidad de estos nuevos conjuntos de datos? El proyecto ENERGIC OD propone una aproximación innovadora para el diseño y desarrollo de estos sistemas (el concepto de *hub* virtual de ENERGIC OD) que aliviará, y posiblemente eliminará, las barreras que impiden la explotación completa de información geográfica disponible de fuentes de datos dispares (distribuidas y heterogéneas). Los *hubs* virtuales están diseñados en torno al concepto de las arquitecturas mediadas (que utilizan *brokers* o intermediarios) e implementados aprovechando el esfuerzo realizado en los últimos años en el desarrollo de Infraestructuras de Datos Espaciales, estándares y herramientas.

METODOLOGÍA

IDENTIFICACIÓN DE LOS USUARIOS

La distribución de datos abiertos es un proceso complejo en sí mismo, por lo que la complejidad solo puede reducirse hasta cierto punto. Sin embargo, la complejidad puede trasladarse de un actor a otro. Por ejemplo, las herramientas semánticas pueden ser de ayuda para los usuarios en sus búsquedas, pero requieren un esfuerzo no despreciable de diseño, implementación y mantenimiento.

El objetivo principal de ENERGIC OD es «la creación de aplicaciones y servicios innovadores», por lo que se considera como usuarios principales a los desarrolladores de aplicaciones web y aplicaciones para móviles sin experiencia específica en el desarrollo de aplicaciones en el dominio geoespacial. Los *hubs* virtuales les proporcionan servicios avanzados que les permiten la creación de aplicaciones basadas en datos geográficos abiertos de acuerdo con sus habilidades y a través de los formatos e interfaces en los que sean más expertos o se sientan más cómodos, sin necesidad de convertirse en expertos en tecnologías relativas a información geográfica

IDENTIFICACIÓN DE BARRERAS PARA COMPARTIR INFORMACIÓN GEOGRÁFICA

ENERGIC OD parte de la base de que la apertura de datos geográficos es tanto una necesidad como un beneficio en sí misma. Aunque esta afirmación no pueda generalizarse, ENERGIC OD se apoya en la gran cantidad de acciones en curso que tienen por objetivo conseguir un consenso en cuanto a la necesidad y utilidad de la distribución de datos abiertos en el mundo de la información geográfica. Por ejemplo, las actividades de la RDA y la definición de los *Principios para el intercambio de datos* del Global Earth Observation System of Systems (GEOSS). ENERGIC OD se basa así mismo en iniciativas que están contribuyendo a distribuir conjuntos de datos geográficos con licencias de uso abiertas, como el Portal de datos abiertos de la Unión Europea⁵, el programa Copernicus⁶ y el GEOSS Data Collection of Open Resources for Everyone (GEOSS Data-CORE)⁷.

Pero declarar unos datos como abiertos no quiere decir que efectivamente sean usables, debido a obstáculos como: ¿se pueden encontrar fácilmente esos datos abiertos? ¿Se puede acceder a ellos con facilidad? ¿Se puede leer el formato en el que están codificados de forma adecuada y cómoda?

⁴ Traducción obtenida de <http://5stardata.info/es/>

⁵ <http://open-data.europa.eu>

⁶ <http://www.copernicus.eu/main/data-access>

⁷ https://www.earthobservations.org/geoss_dsp.shtml

ENERGIC OD considera dos categorías de barreras tecnológicas

- Interoperabilidad, necesaria para salvar la heterogeneidad sintáctica y semántica de los datos y servicios geoespaciales.
- Usabilidad, necesaria para una gestión adecuada de los datos por parte de los usuarios.

CONCEPTO DE HUB VIRTUAL

ENERGIC OD pretende facilitar el uso de la información geográfica a través de *hubs* virtuales que proporcionan la funcionalidad necesaria para garantizar la interoperabilidad y la usabilidad de los datos.

La convocatoria a la que ENERGIC OD se presentó no especificaba exactamente qué es un *hub* virtual, aunque utilizaba explícitamente este término, haciendo referencia a ellos de forma general como instrumentos para facilitar el uso de datos abiertos. Sin embargo, el término *hub virtual* se utiliza habitualmente en el paradigma de distribución *hub-and-spoke*, donde significa: «sistema de distribución de mercancías, pasajeros o datos, en los que los elementos que se distribuyen son guiados desde la entrada a la salida de una ubicación central» [6]. Una arquitectura *software* que sigue el patrón de intermediación de mensajes suele denominarse arquitectura *hub-and-spoke* [7].

Siguiendo estas indicaciones, un *hub* virtual en ENERGIC OD se define como «un nodo virtual donde los usuarios pueden acceder a una cantidad potencialmente ilimitada de datos a través de intermediarios (*brokers*) de fuentes heterogéneas de datos geoespaciales abiertos»

HABILITANDO INTEROPERABILIDAD: SISTEMAS DE SISTEMAS MEDIADOS

En ENERGIC OD, un *hub* virtual es un punto de acceso único a datos geográficos abiertos. Es *virtual* en el sentido de que no almacena físicamente los datos, que permanecen en sus ubicaciones originales, según el paradigma de los sistemas de sistemas.

En los últimos años un gran número de sistemas con características bastante diferentes se han desarrollado para permitir el intercambio de datos geoespaciales en distintos ámbitos, como INSPIRE⁸ en el caso de las Administraciones Públicas europeas, los servicios de Copernicus para aplicaciones medioambientales y de seguridad a partir de datos europeos de observación de la Tierra, GEOSS para observaciones globales de la Tierra, y de las infraestructuras específicas en dominios concretos, como la Infraestructura Mundial de Información en Biodiversidad (Global Biodiversity Information Facility⁹, GBIF). Estos sistemas difieren tanto en soluciones arquitectónicas como tecnológicas, tienen sus propias misiones, mandatos y formas de gobierno y gestión y la necesidad de evolucionar de forma autónoma. La noción de *sistema de sistemas* (SoS por sus siglas en inglés, *System of Systems*) y su ingeniería surgen en muchos campos de aplicación para abordar el problema de la integración de numerosos sistemas independientes y autónomos. Los sistemas de sistemas pueden describirse como *sistemas heterogéneos integrados a gran escala, que se componen de subsistemas mantenidos y operados de forma independientemente, pero están conectados entre sí para la consecución de un objetivo común* [8].

Desde un punto de vista técnico, hay dos enfoques generales para la construcción de un sistema de sistemas: el *federado* y el *intermediado*.

En el enfoque *federado*, los participantes deben ponerse de acuerdo en un conjunto común de especificaciones que conforman su *modelo federado*. La federación puede variar desde una aproximación ligera en la que solo se necesite la adopción de un conjunto de interfaces, metadatos y modelos de datos, a aproximaciones muy estrictas en las que se impone la adopción de las mismas herramientas de *software* en todo el sistema federado. En cualquier caso, los participantes, para cumplir con el modelo federado (especificaciones o herramientas) tienen que hacer algún cambio, por mínimo que sea, en sus propios sistemas. Por lo tanto, este enfoque es factible cuando:

- Existe una fuerza legal lo suficientemente ponderosa como para imponer la adopción del modelo federado a todos los participantes y hacerla cumplir posteriormente, o los participantes tienen un fuerte interés y compromiso para participar en el sistema de sistemas.
- Las organizaciones participantes tienen la experiencia y las habilidades necesarias para la aplicación de la

⁸ <http://inspire.ec.europa.eu/>

⁹ <http://www.gbif.org/>

reingeniería para hacer que sus sistemas cumplan con el modelo federado.

Por ejemplo, el enfoque federado se ajusta bien en dominios como comercio electrónico, banca electrónica y e-gobierno. En el mundo de la información geográfica, el Open Geospatial Consortium (OGC) ha sido muy activo históricamente en el desarrollo de especificaciones y estándares [9], y la experiencia de la implantación de INSPIRE es un ejemplo en el que una autoridad central, la Unión Europea, a través de una Directiva, impuso una serie de principios de intercambio, junto con normas de ejecución, para el establecimiento de la Infraestructura de Información Espacial Europea [10].

En el mundo de la información geográfica, la adopción de un enfoque federado basado en herramientas de *software* comunes no es posible más allá de pequeñas comunidades altamente cohesionadas. De hecho, la definición de un estándar común ajustándose a todos los posibles sistemas de información geográfica es inherentemente difícil y plantea numerosos problemas de tipo teórico. En primer lugar, aun suponiendo que existiera un único conjunto de estándares comunes, su adaptación a las necesidades de dominio muy específicos, pero relevantes, sería muy costosa en términos de eficiencia, resultando en que la mayoría de los sistemas que lo adoptaran tendrían una baja tasa de rendimiento. En segundo lugar, un único estándar potencial que se ajustara a los múltiples requisitos planteados en entornos multidisciplinarios, sería extremadamente complejo: muchos desarrolladores de aplicaciones no tendrían capacidad técnica suficiente como para adoptarlo y aplicarlo. Por último, la tecnología evoluciona constantemente y la evolución de los sistemas pronto produciría un desajuste entre tecnologías y estándar.

En el enfoque *intermediado* [11], estas dificultades se pueden evitar, ya que no se define ni exige ningún modelo común, y los sistemas participantes pueden mantener sus modelos de interfaces, metadatos y datos. Componentes específicos (los *brokers*) son los encargados de acceder a los sistemas participantes y proporcionan las funciones de intermediación y armonización requeridas. No se requiere remodelar los sistemas existentes (o esta remodelación es mínima), reduciendo así la barrera de entrada para proveedores de datos. Este enfoque encaja bien en situaciones donde no existe un mandato legal para el establecimiento de un sistema de sistemas, y donde las organizaciones participantes no tienen fuertes intereses en ser parte del sistema de sistemas. En el mundo de la información geográfica, GEOSS es el ejemplo principal de una iniciativa global que un tercero, el Grupo de Observación de la Tierra (Group on Earth Observation, GEO), tiene un interés específico en la construcción de un sistema de sistemas que recolecte datos existentes de sistemas establecidos con sus propios mandatos. El enfoque intermediado también es útil cuando las organizaciones participantes no tienen la experiencia suficiente como para cumplir con especificaciones complejas. Una ventaja adicional de la intermediación es que la introducción de un nivel intermedio entre los clientes y los servidores, los *brokers*, puede añadir servicios avanzados de valor añadido (por ejemplo, semánticos) que complementan las capacidades del servidor. Los principales inconvenientes del enfoque de intermediado estriban en la complejidad de los mediadores y en el establecimiento de los mecanismos de gobierno o gobernanza de la infraestructura de intermediación que surge [12].

En ENERGIC OD, la elección de arquitecturas intermediadas para los *hubs* virtuales está plenamente justificada por tres razones principales:

- Datos de interés para el proyecto se proporcionan por distintas fuentes a través de protocolos heterogéneos;
- ENERGIC-OD no tiene ni el mandato ni la capacidad de imponer y hacer cumplir nuevos estándares o modelos;
- Las herramientas de intermediación avanzadas que se necesitan ya están disponibles para su reutilización.

RESOLVIENDO LA USABILIDAD: INTERFACES Y API

En el contexto de ENERGIC OD, la usabilidad se relaciona específicamente con el requisito general de facilitar el uso de datos geográficos abiertos. Es previsible que los usuarios de los *hubs* virtuales (es decir, los desarrolladores de aplicaciones) tengan competencias y experiencias heterogéneas en el campo de las tecnologías de información geográfica. Por lo tanto, ENERGIC OD tiene como objetivo proporcionar múltiples formas de interacción con los *hubs* virtuales, que van desde los basados en especificaciones geoespaciales potentes y generalizadas (por ejemplo, las del OGC e ISO), a especificaciones más ligeras (como OpenSearch), y, finalmente, a las interfaces simples de programación de aplicaciones (API, por sus siglas en inglés¹⁰) para un desarrollo rápido implementadas en JavaScript.

La Figura 2 representa la interacción de los usuarios con un *hub* virtual de ENERGIC OD. Un portal maneja la interacción

¹⁰ Application Programming Interface

de persona a máquina; la interacción máquina a máquina explota las interfaces disponibles y las API web. Dentro del *hub* virtual, un conjunto de intermediadores facilita el uso de datos abiertos al proporcionar el descubrimiento y acceso a funcionalidades armonizadas, enriquecidas con capacidades semánticas y de transformación.

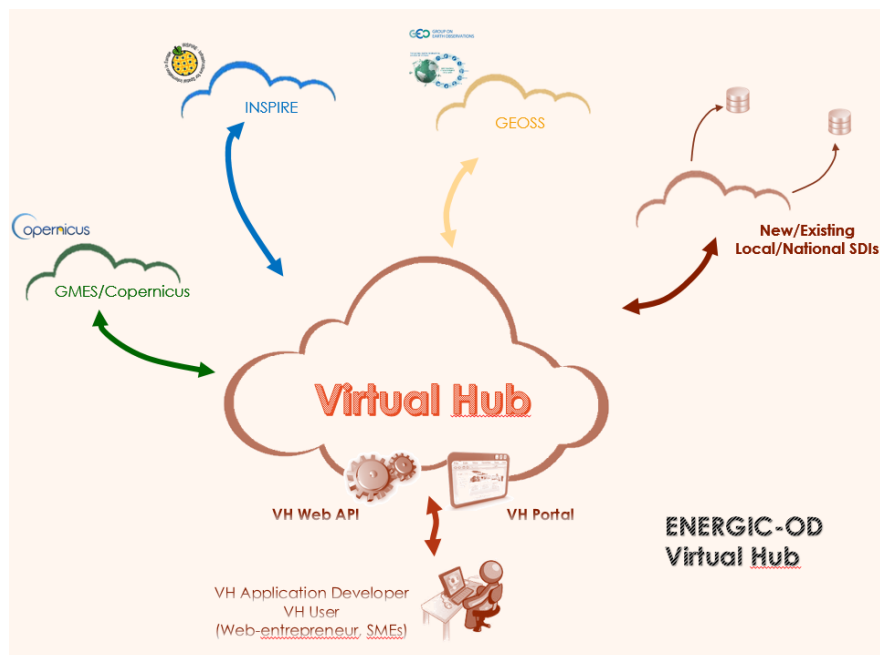


Figura 2. Interacción de los usuarios con los *hubs* virtuales de ENERGIC OD

DESCRIPCIÓN TECNOLÓGICA

Como proyecto de innovación, ENERGIC OD puede dedicar solamente un pequeño esfuerzo a actividades de investigación. Por tanto, debe adoptar un enfoque de desarrollo basado en la integración de herramientas existentes. ENERGIC OD ha seleccionado las tecnologías potenciales que utilizará a través de las propuestas de los socios del consorcio y por medio de un estudio de proyectos de investigación relacionados con ENERGIC OD (financiados principalmente a través del Séptimo Programa Marco de Investigación e Innovación de la Unión Europea) [13].

Como se explicado en la sección anterior, ENERGIC OD ha decidido adoptar una arquitectura de mediación para construir su sistema de sistemas. Por lo tanto, los intermediadores tienen un papel fundamental en la implementación de los *hubs* virtuales de ENERGIC OD. De hecho, la propuesta del proyecto, fue construida en torno a un conjunto de *brokers* desarrollados por el CNR y mejoradas en el transcurso de varios proyectos del Séptimo Programa Marco como EUROGEOS, UncertWeb y GEOWOW: el *GI-suite Brokering Framework* [14].

El *GI-suite Brokering Framework* implementa mecanismos de descubrimiento y acceso intermediados a más de cuarenta tipos de fuentes de datos diferentes y la publicación de los mismos más de diez interfaces distintos. También implementa servicios de valor añadido tales como:

- servicios semánticos implementados a través de la expansión de consultas utilizando bases de conocimiento externo (accesible a través de una interfaz de SPARQL/SKOS) [15];
- armonización de datos a través de transformaciones de subconjuntos, reproyecciones y codificaciones de formato, utilizando algoritmos internos o servicios externos (accesibles a través de interfaces WPS de OGC) [16].

El *GI-suite Brokering Framework* se complementa con la *GEO API*, una implementación en JavaScript de una API web que facilita la creación de aplicaciones web y móviles [17].

El *GI-suite Brokering Framework* está siendo utilizado actualmente en uno de los componentes principales de la Infraestructura Común de GEOSS, donde ha demostrado ser capaz de interconectar numerosos sistemas heterogéneos y gestionar gran cantidad de datos geoespaciales [18]. Está conectado a un servicio semántico gestionado por el Centro

Común de Investigación de la Comisión Europea (EC-JRC), que publica una base de conocimiento compuesta por un conjunto de tesauros y ontologías alineadas que incluye GEMET para el apoyo al multilingüismo.

Durante el transcurso del proyecto, ENERGIC OD integrará en el GI-suite Brokering Framework otras herramientas y marcos para el cumplimiento de otros requisitos (por ejemplo, la publicación de datos para aplicaciones de *crowdsourcing*, control de acceso, etc.). ENERGIC OD también extenderá el API web para proporcionar un acceso cómodo a los desarrolladores de *software*.

DESARROLLO

El desarrollo de los *hubs* virtuales de ENERGIC OD tiene una duración de tres años, con lanzamientos anuales de nueva funcionalidad. La primera versión, basada en el GI-suite Brokering, ha sido entregada en julio de 2015.

La primera versión de los *hubs* virtuales ENERGIC OD se están desplegando a nivel nacional y regional:

- El hub virtual italiano (VH-IT) se ha desplegado en la nube comercial de Amazon. Se conecta a varias fuentes de datos abiertas geospaciales principalmente de administraciones públicas italianas (municipios, provincias y autoridades regionales)
- El hub virtual alemán (VH-DE), también se ha desplegado en Amazon. Un *hub* regional, considerando solo el área de Berlín también ha sido desplegado por uno de los socios del proyecto.
- El hub virtual español (VH-ES) ha sido desplegado en la Universidad de Zaragoza en un entorno *virtualizado*. Está previsto que durante su primer año, sirva datos de la Plataforma de Datos del Ayuntamiento de Zaragoza Abrir y la infraestructura de Datos Espaciales de Zaragoza (IDEZar).
- El hub virtual francés (VH-FR) ha sido instalado localmente en sus instalaciones.
- El hub virtual polaco (VH-PL) está siendo desplegado en un entorno *virtualizado*.

Para valorar y evaluar la eficacia del concepto e implementación de los *hubs* virtuales, el proyecto ENERGIC OD está desarrollando diez aplicaciones piloto para web o para móviles. Se caracterizan por abarcar una amplia gama de dominios de problema distintos (evolución del paisaje, agricultura, biodiversidad y temas más urbanos como la contaminación acústica y la eficiencia energética), lo que permite poner a prueba las capacidades y funcionalidades de los *hubs* virtuales. Cada aplicación se desplegará en uno o más *hubs* virtuales en función de la disponibilidad de los datos que cada una de ellas necesite, haciendo uso de los mismos. Esto proporcionará retroalimentación durante el desarrollo de los *hubs*. El desarrollo de las aplicaciones piloto se ha iniciado en septiembre de 2015, con el primer lanzamiento previsto para septiembre de 2016.

RESULTADOS

En estos momentos, tan solo es posible acceder a los *hubs* virtuales a nivel interno, por parte de los miembros del consorcio del proyecto. Durante el segundo año del proyecto (2015-16), está previsto facilitar acceso libre a los portales contruidos sobre los *hubs* virtuales y la ejecución de demostraciones sobre los mismos, que comenzarán a enlazar datos abiertos procedentes de distintas fuentes, en particular, aquellas necesarias por las aplicaciones piloto, cuyo desarrollo principal también se llevará a cabo durante el segundo año del proyecto. Para el tercer año (2016-17) están previstas mejoras adicionales en el *software* de *hubs* virtuales, el aumento de las conexiones a fuentes de datos en los mismos, la creación de comunidades de usuarios finales y de desarrolladores y el establecimiento de planes de negocio y explotación tanto de los *hubs* como de las aplicaciones

BENEFICIOS ESPERADOS

Esperamos que los *hubs* virtuales del proyecto ENERGIC OD puedan facilitar la creación de nuevas oportunidades de negocio en la explotación de datos abiertos, proporcionando ventajas económicas significativas. La reducción de las barreras de entrada tanto para proveedores de datos geográficos abiertos como para desarrolladores usuarios ofrecerá beneficios directos e indirectos. De forma indirecta, al facilitarse la conexión de nuevas fuentes de datos sin que sea necesario realizar cambios relevantes en las infraestructuras que los ofrecen, los *hubs* virtuales pueden aumentar la cantidad de datos abiertos disponibles, así como reducir el grado de dificultad que plantea su reutilización. Al mismo tiempo, al facilitar el uso de datos abiertos, es previsible que aumente el número de desarrolladores de aplicaciones que utilicen información geográfica abierta. Esto podría iniciar un círculo virtuoso en el que el aumento de las fuentes de datos que atraerán a los desarrolladores de *software*, y por otro lado, un mayor número de desarrolladores

estimularán a los proveedores de información a publicar más datos. Facilitar el uso de datos abiertos también tiene el beneficio directo de reducir el tiempo de desarrollo de aplicaciones que utilicen los *hubs* virtuales: los programadores no tendrán necesidad de ser expertos en interoperabilidad geográfica para crear sus aplicaciones. La curva de aprendizaje se reducirá, y el desarrollador puede concentrarse en su idea de negocio. Obviamente, la simplificación de la utilización de los datos no procede de la nada, sino del desarrollo de los *hubs*.

Para hacer realidad estos beneficios esperados se requiere crear un mercado en torno al concepto de *hub* virtual, definiendo posibles modelos de negocio que sustenten la gestión de los *hubs* virtuales y el ciclo de vida de desarrollo de aplicaciones sobre los mismos. Para ello, se tendrán en cuenta experiencias de proveedores de servicios en la nube, tiendas de aplicaciones y del dominio de los datos abiertos. También deberán definirse los mecanismos de gobernanza de los *hubs* virtuales, con la identificación de los principales actores involucrados los procesos de negocio relativos a la administración de los *hubs*, su evolución, actividades de formación, etc. ENERGIC OD explorará algunas de estas cuestiones en un plan de explotación que se publicará al final del proyecto (septiembre de 2017).

CONCLUSIONES

El concepto de *hub* virtual, tal como fue concebido y está siendo desarrollado en el proyecto europeo ENERGIC OD, propone un enfoque innovador para el intercambio de datos geográficos abiertos. Basándose en resultados de investigaciones previas, en lugar de crear un nuevo sistema para el intercambio de datos, aprovechará los sistemas existentes sin requerir modificaciones, centrándose en la reducción de las barreras de acceso a los datos y su reutilización a través de intermediadores (*brokers*) entre los distintos formatos de datos y protocolos de acceso que utilizan los sistemas existentes.

La primera versión de los *hubs* virtuales, que proporcionan capacidades de búsqueda y el acceso a conjuntos de datos heterogéneos procedentes de servicios de datos dispares, ya muestra que el enfoque intermediado facilita enormemente el descubrimiento, el acceso y uso de los datos geoespaciales abiertos. El trabajo futuro durante los dos años que quedan para que finalice el proyecto se centrará en proporcionar nuevas capacidades, con el objetivo final de desplegar *hubs* virtuales totalmente funcionales que demuestran que muchas tareas complejas relacionadas con la interoperabilidad de datos y su facilidad de uso se pueden automatizar. Estos *hubs* virtuales facilitarán el uso de datos geográficos abiertos, permitiendo que los desarrolladores de aplicaciones web y para móviles se centren en sus propias aplicaciones de negocio, mejorando las oportunidades para la creación de nuevos mercados y oportunidades de negocio.

AGRADECIMIENTOS

ENERGIC OD has received funding from the European Union ICT Policy Support Programme (ICT PSP) under the Competitiveness and Innovation Framework Programme (CIP), grant agreement n° 620400. Este trabajo también ha sido parcialmente financiado por el Gobierno de España a través del proyecto TIN2012-37826-C02-01.

AUTORES

Miguel Ángel LATRE

latre@unizar.es

Universidad de Zaragoza

Departamento de Informática e Ingeniería de Sistemas

Francisco J. LOPEZ-PELLICER

fjlopez@unizar.es

Universidad de Zaragoza

Departamento de Informática e Ingeniería de Sistemas

Nargess KAMALI

nargess.kamali@geokomm.de

Verband der
Geoinformationswirtschaft
Berlin/Brandenburg

Michael BAUER

michael.bauer@geokomm.de

Verband der
Geoinformationswirtschaft
Berlin/Brandenburg

Mattia PREVITALI

mattia.previtali@polimi.it

Politecnico di Milano (POLIMI) -
dABC

Raffaella BRUMANA

raffaella.brumana@polimi.it

Politecnico di Milano (POLIMI) -
dABC

Stefan BRAUMANN
stefan.braumann@lup-umwelt.de

Luftbild Umwelt Planung GmbH

Helga KUECHLY
helga.kuechly@lup-umwelt.de

Luftbild Umwelt Planung GmbH

Paolo MAZZETTI
paolo.mazzetti@cnr.it

Consiglio Nazionale delle
Ricerche - Istituto
sull'Inquinamento Atmosferico
(CNR-IIA)

NATIVI, Stefano
stefano.nativi@cnr.it

Consiglio Nazionale delle
Ricerche - Istituto
sull'Inquinamento Atmosferico
(CNR-IIA)