

Hacia la Interoperabilidad Semántica para el Manejo Inteligente y Sostenible de Territorios de Alta Biodiversidad usando SmartLand-LD

Nelson Piedra¹, Juan Pablo Suárez¹

{nopiedra, jpsuarez}@utpl.edu.ec

¹ Universidad Técnica Particular de Loja, Ecuador

DOI: 10.17013/risti.26.104-121

Resumen: Un reto presente en los actuales esfuerzos de desarrollo sostenible es potenciar la interoperabilidad entre variables e indicadores almacenados en sistemas de información y proveedores de servicios de datos autónomos, distribuidos y heterogéneos. El acceso restringido a fuentes de datos y servicios limita las capacidades de un territorio inteligente y sostenible para mejorar los aspectos sociales, ambientales y económicos. Este trabajo presenta un enfoque de Web Semántica que apoya la extracción, la transformación de datos vinculados, la integración y la explotación de datos masivos y heterogéneos, con aplicación en la toma de decisiones relacionadas con el desarrollo inteligente y sostenible de los territorios. Motivado por el carácter no escalable de los métodos de gestión de datos sostenibles convencionales, así como por los desafíos de interoperabilidad presentes en los silos de datos contemporáneos, el marco descrito, acuñado como SmartLand-LD, aprovecha el potencial combinado de diversas fuentes de datos. Estos incluyen transmisiones de sensores, datos geográficos, bases de datos científicas, recursos relacionados disponibles en repositorios de conocimiento y otras fuentes de datos abiertos estructurados y no estructurados. A través de la apertura, la construcción de gráficos de conocimiento, la interoperabilidad semántica y las técnicas colaborativas de ciencia abierta, el marco propuesto permite el mapeo e integración de datos sobre indicadores de desarrollo sostenible en territorios de alta biodiversidad..

Palabras-clave: Web Semántica, Datos Enlazados, Datos Abiertos, Interoperabilidad, Grafos de Conocimiento.

Towards Semantic Interoperability for Smart and Sustainable Management of High Biodiversity Territories using SmartLand-LD

Abstract: A present challenge in today's sustainable development efforts is to empower interoperability between variables and indicators across autonomous, distributed and heterogeneous information systems, and data service providers. Restricted access to data sources and services limits the capabilities of a smart and sustainable territory to improve social, environmental and economic aspects. This work presents a Semantic Web approach that supports the extraction, linked data transformation, integration, and exploitation of large-scale and

heterogeneous territories data, with application to decision-making related with smart and sustainable development of territories. Motivated by the non-scalable character of conventional sustainable data management methods, as well as by the interoperability challenges present in contemporary data silos, the described framework– coined as SmartLand-LD – leverages the combined potential of diverse data sources. These include sensor streams, geo-data, scientific databases, related resources available from knowledge repositories, and other structured and non-structured open data sources. Through openness, knowledge graphs building, semantic interoperability, and open science collaborative techniques, the framework proposed enables the mapping and integration of data about sustainable development indicators in high biodiversity territories.

Keywords: Semantic Web, Linked Data, Open Data, Interoperability, Knowledge Graph.

1. Introducción

La convergencia de las tecnologías de la información y la comunicación están produciendo cambios en la gestión de los entornos urbanos y rurales. Estos territorios son diferentes de cualquier otra realidad que se haya experimentado hasta el día de hoy. Durante gran parte del siglo XX, la idea de que un territorio podría considerarse inteligente era simplemente una cuestión de ciencia ficción. Sin embargo, la expectativa de que una ciudad o territorio se gestione de forma inteligente y en armonía con el ambiente, convierte esta idea en una nueva realidad impulsada desde diferentes visiones y perspectivas.

Las diferentes visiones de desarrollo y sostenibilidad se pueden consensuar a través de acuerdos y colaboración que permitan avanzar hacia la creación y consolidación de ciudades y territorios, inteligentes y sostenibles. En cuanto al monitoreo y evaluación de los esfuerzos de desarrollo sostenible, se requiere una gran cantidad de indicadores de calidad que midan la sostenibilidad ambiental, así como el impacto en aspectos sociales y económicos (Cutter, Osborn, Romano, & Ullah, 2015; Hák, Janoušková, & Moldan, 2016) world leaders embark on the notion of universe global Sustainable Development Goals (SDGs. Por lo general, estos datos e indicadores están depositados en diferentes sistemas de información. En este sentido, un reto presente en las acciones de desarrollo sostenible está en mejorar la medición de los avances o retrocesos de la sostenibilidad y alcanzar la interoperabilidad con la ayuda de tecnologías que conecten los diferentes silos de información.

Crear un ecosistema para la integración e interoperabilidad de los datos es una tarea compleja. De esta forma,

En este trabajo, los autores presentan un enfoque para el manejo inteligente y sostenible de territorios de alta biodiversidad: la Iniciativa SmartLand. SmartLand es sobre la gestión inteligente de las interacciones entre personas y territorios. SmartLand dispone de un marco de trabajo, SmartLand-LD (descrito como trabajo en progreso en (Nelson Piedra & Suárez, 2017) we investigate the core challenges faced when consuming multiple data sources for smart management of territories of high biodiversity using the Linked Data approach. We design a framework to achieve better data interoperability and integration by republishing real-world data into linked data: the SmartLand-LD framework. SmartLand-LD will work as a flexible and distributed ecosystem of autonomous and

heterogeneous data sources (in content, technology, storage and structure), basado en el enfoque de Web Semántica (Berners-Lee, Hendler, & Lassila, 2001) y las tecnologías de Datos Vinculados (en inglés: *Linked Data*, LD) que habilitan la capacidad de construir un entorno global de información (Heath & Bizer, 2011) alineado con el desarrollo sostenible. Este enfoque reduce la duplicación de esfuerzos al describir los recursos de información y sus atributos, y fomenta la creación de un grafo de conocimiento (en inglés: *Knowledge Graph*, KG) global con capacidad de abarcar de forma masiva la información necesaria para realizar consultas y acciones complejas (W3C, 2011). La primera sección describe de forma general la gestión inteligente de ciudades y territorios; a continuación se presenta la organización de la iniciativa SmartLand. Luego se discute el poder de la Web Semántica (SW) y los Datos Enlazados (LD) como una forma de alcanzar la interoperabilidad semántica de datos y desarrollar la Web de Conocimiento. A continuación se describe el marco propuesto para la publicación de datos enlazados de algunos conjuntos de datos existentes y una aproximación sobre cómo definir y reutilizar recursos ontológicos para representar KG. Finalmente, la última sección cubre las conclusiones respectivas y áreas futuras de investigación.

2. Gestión inteligente de territorios - SmartLand

2.1. La gestión inteligente de las zonas urbanas

En las últimas dos décadas, el concepto de ciudad inteligente (Albino, Berardi, & Dangelico, 2015) se ha vuelto cada vez más popular en la literatura científica y las políticas internacionales. La razón principal parece estar relacionada con el papel principal de las ciudades y los territorios en los aspectos sociales y económicos de las personas en todo el mundo, y en el gran impacto sobre la sostenibilidad ambiental (Mori & Christodoulou, 2012).

Las ciudades son el núcleo del consumo de recursos globales. El número y la proporción de la población urbana crecerán en el futuro. Se estima que las ciudades globales ocupan aproximadamente el 2% de la superficie terrestre del planeta, consumen entre 60-80% de la energía y son responsables del 75% de las emisiones de carbono. Además, muchas de las decisiones llevadas a cabo por los habitantes de las ciudades afectan directamente la biodiversidad y los ecosistemas. A la luz de esto, la tendencia actual es crear ciudades inteligentes (o *Smart Cities*).

La idea de ciudades inteligentes se refiere a la promoción de ciudades más inteligentes, sostenibles e inclusivas a través de la promoción de innovaciones tecnológicas apropiadas (Hollands, 2008). Las ciudades y los territorios se consideran el elemento clave en los planes estratégicos de los países (Nam & Pardo, 2011) (Deakin & Waer, 2011) (Dirks, Gurdgiev, & Keeling, 2010). El ciudadano debe estar en el centro de atención. Sin embargo, la ciencia y la ingeniería desempeñan un papel crucial en la innovación y la creación de nuevos servicios públicos, así como en el desarrollo del concepto de gestión inteligente y sostenible (Ahvenniemi, Huovila, Pinto-Seppä, & Airaksinen, 2017).

2.2. La gestión inteligente de los territorios

Los territorios de alta biodiversidad no existen como una esfera separada de las acciones, ambiciones y necesidades humanas. Territorio y desarrollo son inseparables (Kates,

Parris, & Leiserowitz, 2005), y su gestión plantea un desafío contemporáneo de alcance global (Acosta, Piedra, & Suárez, 2017). El Convenio sobre la Diversidad Biológica establece tres vastos y ambiciosos objetivos: la diversidad biológica de conservación, el uso sostenible de sus componentes, y la creación de propuestas que busquen la distribución justa y equitativa de los beneficios derivado del uso de los recursos del territorio (Balmford et al., 2005). Por otro lado, desde el 2016, arrancó la Agenda 2030 y sus Objetivos de Desarrollo Sostenible, ODS, asumidos por 193 naciones comprometidas con resolver los problemas sociales, económicos y ambientales que aquejan al mundo (United Nations General Assembly, 2015), marcarán la agenda de desarrollo mundial durante los próximos.

Los ODS son un conjunto de 17 objetivos y 169 metas que reflejan un balance entre las dimensiones económicas, sociales y ambientales (Cutter et al., 2015) world leaders embark on the notion of universe global Sustainable Development Goals (SDGs), destinadas a fortalecer la paz, erradicar la pobreza en todas sus formas, combatir las desigualdades, construir sociedades pacíficas, justas e inclusivas, proteger los derechos humanos, promover la igualdad entre los géneros y el empoderamiento de las mujeres y las niñas, así como garantizar una protección duradera del planeta y sus recursos naturales ((UNSG), 2016).

3. SmartLand, una visión para el desarrollo inteligente y sostenible de territorios megadiversos

3.1. Contexto

Sobre un territorio de alta biodiversidad hay muchas partes interesadas, cada una con diferentes visiones, intereses, y opiniones variadas sobre objetivos, políticas, prioridades de inversión, y esfuerzos por el desarrollo y la sostenibilidad. Las partes interesadas son instituciones educativas, organizaciones gubernamentales, inversores, agencias reguladoras y grupos de la sociedad civil que llevan a cabo diferentes acciones que afectan el desarrollo sostenible. En particular, esta maleabilidad inherente del desarrollo sostenible, plantea el reto de mejorar las estrategias de articulación de mediciones sobre el desarrollo sostenible.

A la luz de los desafíos por lograr un mejor equilibrio entre las dimensiones económica, social y ambiental, y como parte de su responsabilidad de distribución social del Conocimiento (Acosta et al., 2017), la Universidad Técnica Particular de Loja (UTPL) promueve la iniciativa institucional SmartLand para la gestión sostenible e inteligente de los territorios de alta biodiversidad (ver <http://smartland.utpl.edu.ec>). SmartLand promueve la visión de un uso más sostenible e “inteligente” de la biodiversidad, y la distribución justa y equitativa de los beneficios derivados del uso de recursos del territorio. SmartLand impulsa la construcción de una visión consensuada sobre territorio y desarrollo; promueve alianzas para el desarrollo sostenible, la participación activa, el uso de conocimiento científico, e innovaciones tecnológicas para una gestión inteligente de la interacción entre los seres humanos y los territorios biodiversos. De acuerdo a la visión de SmartLand, los esfuerzos por el desarrollo inteligente de los territorios deben abarcar indicadores de impacto que midan la sostenibilidad ambiental, económica y social.

3.2. Apertura, Colaboración, Datos y Desarrollo sostenible

En la misión de SmartLand está medir el progreso relativo o la falta de progreso hacia el desarrollo sostenible. En este sentido, un tema crucial en el desarrollo de ciudades y territorios inteligentes y sustentables es el uso de estándares y enfoques para garantizar la interoperabilidad, de modo que los sistemas producidos por diferentes proveedores trabajen juntos sin interrupciones, y que brinden la capacidad reducir costos a través de economías de escala.

La apertura y colaboración son fundamentales para conectar recursos de información autónomos relacionados con la iniciativa SmartLand. Sin capacidad de interoperabilidad, los esfuerzos de monitoreo de indicadores a menudo toma la forma de largas y heterogéneas listas de indicadores, muchos de los cuales se enfocan en el presente o en el muy corto plazo. Además, las listas de indicadores resultantes generan silos de información (el concepto de silo está explicado en (Ensor, 1988; Piedra Pullaguari, 2016)), que reflejan las diversas aspiraciones del conjunto diverso de partes interesadas y demandan la necesidad de integrar los esfuerzos aislados de medición.

La perspectiva de medición de SmartLand es la de monitoreo permanente y de interoperar silos de información. SmartLand trabaja en la creación de indicadores que faciliten la planificación a través de múltiples escalas de tiempo, aumentando la posibilidad de hacer que los territorios sean más inteligentes a largo plazo, mediante el monitoreo de variables que describan el territorio. Con el fin de mejorar la comprensión de los indicadores, la iniciativa SmartLand promueve estrategias y herramientas tecnológicas que son necesarias para la recopilación, interoperabilidad, integración, análisis de datos, visualización y preservación de grandes colecciones de datos e información. Se busca alcanzar la interoperabilidad y el reuso de variables e indicadores, así como el compromiso de monitoreo a largo plazo, necesario para reflejar el horizonte temporal del ahora y del futuro en relación al desarrollo sostenible.

3.3. Organización de la Iniciativa SmartLand

SmartLand se ejecuta a través de programas/proyectos de investigación multidisciplinarios, organizados en 12 paquetes de trabajo, 4 objetivos comunes, y una plataforma tecnológica para describir e intercambiar datos. El pilotaje de la iniciativa se ejecuta en tres provincias del sur de Ecuador, que abarcan los ecosistemas de la Costa, los Andes y la Amazonía montañosa. Esta región es la más pequeña del territorio de Ecuador, pero también la más biodiversa. Se aplica un enfoque de innovación social para vincular a beneficiarios, docentes investigadores y estudiantes en mejorar el proceso de toma de decisiones públicas y privadas y mejorar la gestión de desarrollo sostenible.

Los paquetes de trabajo (WP, *Working Packages*) ayudan a contribuir al cumplimiento exitoso de los objetivos centrales de la iniciativa a través de proyectos de investigación propuestos por investigadores de diferentes dominios de conocimiento (Ver figura 1-a)). Cada paquete de trabajo está compuesto por proyectos de investigación que mejoran la comprensión científica del medio ambiente y agregan valor a los indicadores sociales, biológicos, ambientales y de infraestructura.

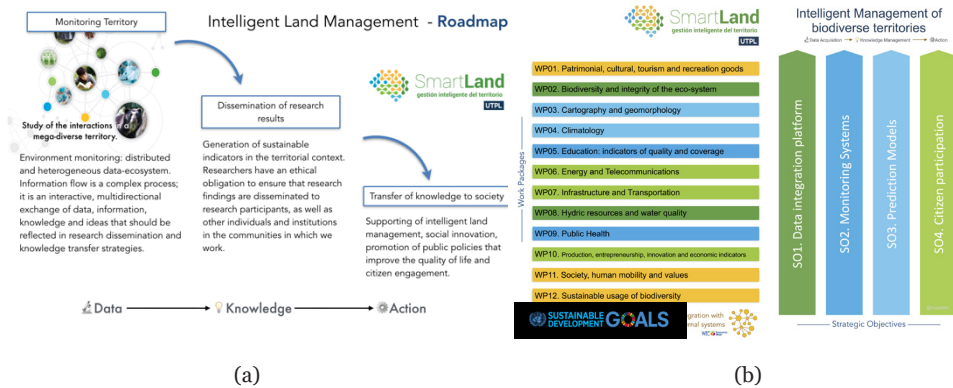


Figura 1 – Organización general de la iniciativa SmartLand. (a) Roadmap sobre la gestión del conocimiento. (b) Paquetes de trabajo y objetivos estratégicos en SmartLand

Los paquetes de trabajo comparten transversalmente cuatro objetivos estratégicos (SO, *Strategic Objectives*) para la gestión de la información, que se reflejan en la disponibilidad de una arquitectura de datos web abierta que abarca la recopilación, preservación de datos, reutilización, análisis, modelado predictivo y la visualización de datos (Ver figura 1-b). Cada grupo de investigación trabaja en sus propias preguntas de investigación, mientras comparte cuatro Objetivos Estratégicos y busca contribuir a los ODS.

SO-1. Consolidar un sistema para la preservación del conocimiento que garantice el acceso libre. SmartLand promueve la creación de una infraestructura tecnológica que, desde un enfoque distribuido, ayuda a recopilar, monitorear y reutilizar datos e información de alta heterogeneidad en diversos campos del conocimiento: biodiversidad, ecosistemas, economía, territorio, emprendimiento, sociedad, valores e identidad, entre otros.

SO-2. Mejorar nuestra comprensión de los indicadores y su monitoreo permanente. Se promueve una estrategia de recolección sostenible de datos, que se basa en redes de sensores y sistemas de monitoreo de indicadores a largo plazo.

SO-3. Mejorar la gestión de los territorios a través del uso eficiente de los recursos, la preservación de su patrimonio y el uso sostenible. SmartLand busca explotar, modelar, monitorear y visualizar diversas variables, así como desarrollar escenarios probabilísticos, ordenanzas territoriales virtuales y mejorar el análisis de la toma de decisiones.

SO-4. Fortalecer las capacidades y potencialidades de los ciudadanos. Se promueve la participación responsable de la ciudadanía mediante el fortalecimiento de la identidad y el valor de los seres humanos con el medio ambiente.

Aspectos destacados de la iniciativa. Se proponen acciones clave en cada WP. De esta forma, el iniciativa complementa las orientaciones más específicas que se espera que desarrollen los actores locales, cada una de las cuales refleja las necesidades y prioridades únicas de sus electores y al mismo tiempo comparte la visión y los objetivos comunes

descritos aquí. La iniciativa se desarrolla desde el 2014. SmartLand ejecutó 108 proyectos con aproximadamente 300 investigadores involucrados (84 internacionales) y 2000 estudiantes. Se han implementado herramientas de medición que pueden ser utilizadas como apoyo en la toma de decisiones relacionadas con el desarrollo sostenible, ya que proporcionan información sobre el progreso hacia los objetivos definidos. SmartLand genera de indicadores a través de sistemas de monitoreo consolidados y diferentes observatorios sobre ecosistemas, conflictos socio-ambientales, turismo, movilidad, comunicación estratégica, empresas familiares, y territorio y economía.

4. Interoperabilidad de Datos y Web Semántica

4.1. Desafío: un ecosistema para la interoperabilidad semántica

La interoperabilidad implica llegar a acuerdos, compartir estrategias y tecnologías que integren e interoperen todos los elementos de información que conforman el alcance de la iniciativa, y que de otra manera actuarían como silos de información. La arquitectura semántica de SmartLand debe garantizar la interoperabilidad de diferentes tipos de datos y sistemas (ver Figura 2).

En consecuencia, se deben abordar los siguientes desafíos: (i) heterogeneidad en la representación de datos y semántica, que da como resultado una variedad de modelos de datos y formatos (ii) La heterogeneidad de los sistemas que gestionan los datos y que admiten varios modos de acceso y consulta. (iii) El desafío de procesar los grandes volúmenes de datos y la solución de muchos problemas en la sociedad, los negocios y la academia; (iv) La pérdida del contexto interpretativo o la distorsión del significado de los datos cuando influyen en otros límites semánticos en los que los significados, la terminología y el vocabulario son diferentes: (v) Veracidad, diversidad de escalas, privacidad, mal uso y / o abuso de material publicado, entre otros.

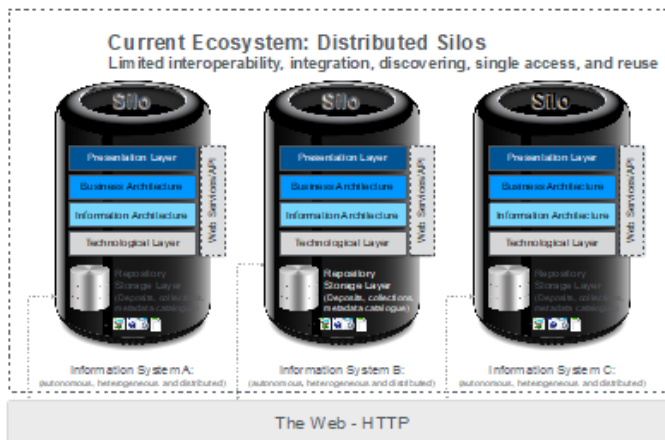


Figura 2 – El desafío de SmartLand: Conectar silos de datos, mejorar el reuso y la integración de información que apoye la construcción de indicadores sobre desarrollo Sostenible

4.2. Aportar datos a un Espacio Global de Información

La Web Semántica (*Semantic Web*, SW) se concibe como la evolución de la Web actual hacia un espacio de información global (Heath & Bizer, 2011), autónomo, y descentralizado. La SW, liderada desde el World Wide Web Consortium (W3C), potencia las posibilidades de compartir a través de la Web datos y significados, en formatos procesables directamente por las máquinas, y alcanzar una mayor interoperabilidad e integración, y con un mínimo de costos técnicos.

Según Tim Berners-Lee, el inventor de la web y el fundador del proyecto LOD, la SW no solo busca publicar datos en la web, sino que también busca crear vínculos entre los datos, lo que ayudaría a explorar el espacio global de información, tanto para usuarios humanos, como por usuarios máquinas (Berners-Lee et al., 2001). Con los datos de enlace, cuando consultamos un recurso, podemos acceder a otros recursos de información publicados en la Web (Heath & Bizer, 2011).

Esta visión presenta cinco desafíos principales: (1) la identificación global de recursos (W3C, 2008); (2) el uso de ontologías formales como una forma de especificar acuerdos específicos de contenido para compartir y reutilizar el conocimiento entre seres humanos y entidades de software (Gruber, 1995; W3C OWL Working Group, 2012); (3) el modelado distribuido del mundo con un modelo de datos compartidos (Cyganiak, Wood, & Lanthaler, 2014; Harris & Seaborne, 2013; W3C, 2011); (4) la infraestructura donde los datos, esquemas y ontologías formales se pueden publicar, compartir, encontrar y reutilizar para apoyar las actividades de intercambio de conocimientos; y, (5) calidad y *provenance* de la información (N. Piedra, Chicaiza, Lopez, & Caro, 2017). Los Datos Vinculados brindan la oportunidad para que los editores contextualicen sus datos, agregando riqueza y profundidad. La aplicación de Linked Data está documentada en diversos trabajos, ver (Guillén, Lloret, & Gutiérrez, 2016; Hallo, Luján-Mora, & Morga, 2017; N Piedra et al., 2015).

En (Berners-Lee, 2006) se plantean 4 principios de diseño para enlazar datos por medio de la Web: (1) Usar URIs (Identificadores de Recursos Uniformes) como nombres para las cosas; (2) Usar URIs HTTP para que las personas puedan buscar esos nombres; (3) Cuando alguien busca un URI, proporcionar información útil, usando los estándares RDF* (Cyganiak et al., 2014) y SPARQL (Harris & Seaborne, 2013)); (4) Incluir enlaces a otros URIs para descubrir más cosas; además, desde nuestra visión, se incluye un punto adicional: (5) el uso de ontologías (W3C OWL Working Group, 2012) para desarrollar la SW.

5. Framework para integración semántica en SmartLand

5.1. Niveles de Interoperabilidad en SmartLand-LD

En SmartLand-LD se abordan diferentes aspectos de coordinación, producción y uso de datos para el desarrollo sostenible. Esta sección presenta la arquitectura del marco semántico SmartLand que se ha diseñado en base a los requisitos definidos en las secciones previas. El concepto subyacente de SmartLand es de una constelación de componentes de varias escalas y tipos (tanto digitales como analógicos), que están

conectados por medio de múltiples redes, y que proporcionan datos continuos con respecto a los factores ambientales, biológicos, sociales y económicos.

El Framework SmartLand-LD implementa cuatro niveles de apertura e interoperabilidad, basado en (Piedra Pullaguari, 2016). La armonización de estos niveles permite alcanzar los objetivos de interoperabilidad semántica pretendida, la misión y apoyar el logro de la visión de futuro planteada en el contexto de este trabajo. A través de estos niveles de interoperabilidad, y los desafíos subyacentes, se puede convertir el potencial semántico de los datos abiertos en una realidad que amplía el descubrimiento y reuso de variables e indicadores relacionados con el desarrollo inteligente y sostenible. Estos niveles se integran de manera que es posible alcanzar objetivos más complejos.

En la Figura 3 se describen los cuatro niveles de interoperabilidad del marco de trabajo de SmartLand-LD. El **nivel de interoperabilidad legal** se refiere a habilitar globalmente el máximo grado de acceso, descubrimiento, uso, y reuso de datos, variables, indicadores y otros contenidos digitales; sino existe una legislación en este contexto, se debe motivar su creación. Se fundamenta en la legislación internacional, regional y nacional relacionada con el movimiento de acceso abierto a los datos y el acceso abierto al conocimiento.

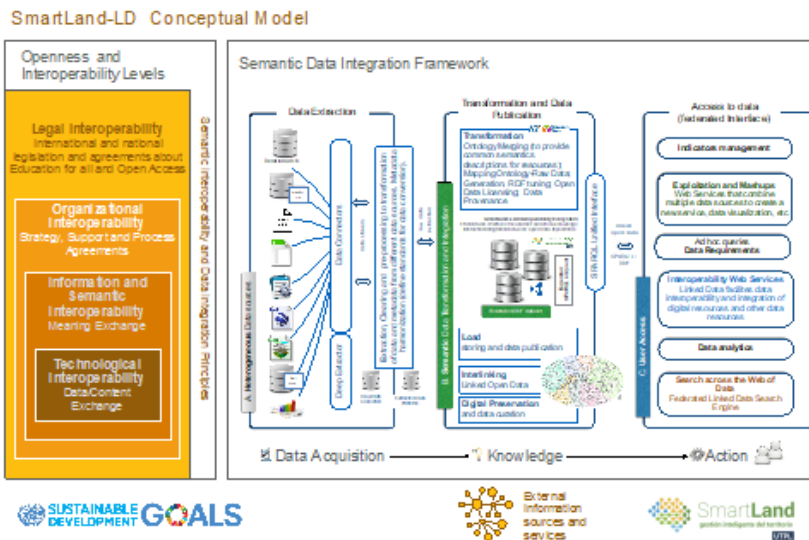


Figura 3 – Framework para la generación de grafos de conocimiento en SmartLand.

El siguiente **nivel de interoperabilidad es organizacional**. Este nivel trata sobre mejorar la eficiencia y la coordinación de esfuerzos, tanto al interior de los proyectos e iniciativa SmartLand, como con otros actores claves relacionados con el Desarrollo Sostenible. Este nivel es clave porque minimiza la sobrecarga (no razonable) de descubrir, usar y re-usar datos abiertos. Es importante establecer la visión de apertura

e interoperabilidad, y el soporte requeridos. Los acuerdos de interoperabilidad pueden, por ejemplo, tomar la forma de memorandos de entendimiento, o niveles de acuerdo de servicios que especifiquen las obligaciones de cada parte participante en los procesos de apertura e interoperabilidad; así como los niveles de servicio esperados, los procedimientos de soporte/escalamiento, datos de contacto, información de procedencia de datos (*provenance*), niveles de calidad de la información, haciendo referencia, cuando sea necesario, a los acuerdos subyacentes en los niveles semánticos y técnicos. Sobre este nivel se implementan **el nivel de interoperabilidad semántica** que habilita el intercambio de datos y significados, se cubren módulos diversos de extracción, desambiguación, modelado semántico, transformación de RDF, y publicación de información en grafos de conocimiento abierto. A **nivel de interoperabilidad tecnológica**, se cubre la funcionalidad que permite el intercambio de datos y contenidos a través de la Web. La interoperabilidad es implementada a través de diferentes componentes de extracción de datos, transformación de datos, publicación de *Linked Data*, y la capa de aplicaciones de usuarios basadas en datos; por ejemplo, un mecanismo de búsqueda federada de datos, un sistema de recomendación de indicadores para apoyar la toma de decisiones informada, aplicaciones de visualización de datos y geográficas, entre otras. A continuación se describen varios componentes tecnológicos del marco de trabajo:

5.2. Recolector de Datos

La información contenida en las colecciones de datos relacionadas con desarrollo sostenible es vital para la toma de decisiones, por lo que almacenar, indexar y recuperar información específica de estas colecciones es clave (Vázquez-Reyes, de León-Sigg, Velasco-Elizondo, Villa-Cisneros, & Briceño-Muro, 2017). El recopilador de datos es un componente central en las tareas de recolección, almacenamiento y preservación de diferentes conjuntos de datos de la iniciativa SmartLand. Los datos recopilados son preservados en el repositorio de datos abiertos Ambar (ver <http://ambar.utpl.edu.ec>). Ambar está basado en CKAN (CKAN, n.d.) que proporciona herramientas colaborativas y abiertas para agilizar la publicación, el intercambio, la búsqueda y el uso de datos, y establecer diferentes períodos de retención de datos. Es posible importar datos de una variedad de fuentes primarias, y no está limitado a datos estructurados. Ambar es extensible a través de su API. Cada vez que se detectan datos nuevos, el componente automáticamente realiza tareas de transferencia de datos hacia al proceso ETL. Los datos recolectados luego son integrados a SmartLand-LD.

5.3. Extracción de Datos

El Proceso de Extracción de datos es un aspecto muy importante de SmartLand-LD. La extracción de datos es un proceso que implica la recuperación de todos los formatos y tipos de patrones de datos. Para cumplir con este requisito, se han desarrollado varias herramientas para manejar este proceso. El proceso de extracción de datos en general se realiza a través del módulo del conector de datos, con funcionalidad personalizada para cada tipo y naturaleza de las fuentes de datos. La fase de extracción implica la recuperación (extracción) de datos brutos de una fuente de datos dada utilizando el complemento apropiado, según la tecnología de la fuente. Los patrones de datos se importan en un

sistema de extracción intermedio antes de procesarse para la transformación de datos, donde posiblemente se rellenen con metadatos antes de exportarse a otra etapa en el flujo de trabajo de transformación y publicación de LD. Los datos extraídos se agregan a una base de datos relacional, cuya flexibilidad se garantiza con esquemas en forma de triples. La extracción de datos se ejecuta constantemente o en un cronograma definido por el usuario. Como el componente necesita hacer otros procesos y no solo extraerlos, se ha implementado funcionalidad que verifica constantemente nuevas actualizaciones de datos. Estos datos se usarán posteriormente para la transformación semántica.

5.4. Identificación de Recursos Global

Con el intercambio creciente de metadatos, diferentes sistemas de identificación colisionarán en un ecosistema global de datos. En SmartLand-LD, cada identificador es un URI cool y persistente (N. Piedra et al., 2017; W3C, 2008). Un URI es una cadena de caracteres utilizada para identificar el nombre de un recurso. Dicha identificación permite la interacción con representaciones del recurso a través de una red, típicamente la Web, usando protocolos específicos. Los esquemas que definen cada URI que especifican una sintaxis concreta y los protocolos asociados (Gregorio & et. al, 2012). La resolución es el mecanismo clave que permite a un sistema localizar y acceder al objeto identificado o a la información relacionada con este en la Web. Resolver un URI significa convertir una referencia de URI relativa a forma absoluta, o desreferenciar una referencia de URI o URI intentando obtener una representación del recurso que identifica. El componente 'resolver' en el software de procesamiento de documentos generalmente proporciona ambos servicios. Cualquier tipo de identificador persistente adicional es útil para enlazar y obtener más información contextual

5.5. Ontologías base

Un ontología es una conceptualización formal, sistematizada, explícita y consensuada de un dominio de trabajo. Una conceptualización es una visión simplificada y abstracta del mundo que deseamos representar con algún propósito. Cada base de conocimiento, sistema basado en conocimiento o agente de nivel de conocimiento está comprometido con alguna conceptualización, explícita o implícitamente (Gruber, 1995). Las ontologías centrales (*Core-Ontologies*) en SmartLand, son modelos de datos, explícitos, simplificado, consensuados, reutilizables y extensibles que capturan las características fundamentales de conocimiento (propiedades, relaciones y restricciones), en lo posible de una manera neutral al contexto. Estos vocabularios principales se usan en acuerdos de interoperabilidad que permiten que los sistemas de información, desarrollados de forma independiente, intercambien información, lo que hace posible que la información atraviese fronteras y dominios. Están diseñados para convertirse en la base de nuevas ontologías de dominio que se utilizarán para intercambiar información. También se usan como vocabularios base en procesos de mapeo y alineamiento con otros vocabularios. Los vocabularios centrales utilizados en SmartLand-LD son:

DCAT, esta especificación recomendada por el W3C facilita la interoperabilidad entre los catálogos de datos. Su caso de uso básico es permitir la búsqueda de información en el portal de datos SmartLand. Además, se mejora la publicación descentralizada

de catálogos, la visibilidad y se permite que las aplicaciones consuman fácilmente los metadatos de múltiples catálogos.

SDMX, se usa en procesos de intercambio de metadatos y datos estadísticos (en inglés: *The Statistical Data and Metadata eXchange*) entre organizaciones internacionales y miembros individuales.

RDF Data Cube, se usa como base central que admite la extensión de vocabularios que describan diferentes aspectos de flujos de datos estadísticos u otros conjuntos de datos multidimensionales. El modelo que sustenta el vocabulario Data Cube es compatible con el modelo de cubo que subyace a SDMX.

DCMI Metadata Terms, es una especificación autorizada de todos los términos de metadatos mantenidos por Dublin Core Metadata Initiative (DCMI), que proporciona elementos, refinamientos de elementos, esquemas de codificación y términos de vocabulario, una especificación.

Geodata, define cómo describir la información geográfica, además desde una perspectiva global se usarán los estándares ISO 19115: 2003, ISO 19139 y 19115: 20141.

FOAF, permite describir redes de colaboración, a través de tres tipos de red: redes sociales de colaboración humana, amistad y asociación; redes representacionales en términos fácticos, y redes de información que usan enlaces basados en la Web para compartir descripciones publicadas de forma independiente.

VOID, permite expresar el conjunto de datos RDF. Además, VOID es vital en acciones de descubrimiento y catalogación de datos en la web.

PROV, proporciona una base para representar la información de procedencia sobre las entidades, actividades y actores involucrados en la producción de los elementos de datos.

5.6. Flujo de Publicación de Datos Abiertos y Enlazados

El objetivo de este componente la publicación de los metadatos extraídos como *Linked Data*. El primer paso es diseñar y asignar URIs persistentes a los datos y describirlos en RDF utilizando vocabularios consensuados. En el caso de los datos estructurados (por ejemplo, almacenados en bases de datos relacionales, archivos CSV, Excel o XML), la funcionalidad de transformación de RDF está basada en RDF wrappers y APIs. Mientras que, para los datos no estructurados, se han empleado técnicas de minería de textos que actúan como extractores de entidades, claves para descubrir las entidades de datos que serán descritas como Linked Data. El proceso de transformación SmartLand-LD consiste en los siguientes pasos:

Pre-procesamiento de datos - desambiguación y limpieza de datos. Este componente aporta a la calidad de los contenidos que se incorporan al flujo de datos del framework. Se realizan tareas de limpieza, remplazamientos, y desambiguación de datos de acuerdo con diferentes métodos de resolución de conflictos (Nelson Piedra, Chicaiza, Lopez-Vargas, & Caro, 2016).

Modelado de datos. El ciclo de vida para el desarrollo de nuevos recursos ontológicos se describe en la figura 4.



Figura 4 – Ciclo de vida para la construcción de recursos ontológicos en SmartLand.

Conversión de datos extraídos a RDF. Los mapeos de vocabulario se expresan usando un lenguaje de mapeo interno. El lenguaje proporciona transformaciones simples, así como transformaciones estructurales más complejas (1-a-n y n-a-1) y transformaciones de valores de propiedades, se normalizan diferentes unidades de medida, y es posible ejecutar manipulaciones complejas de datos.

Especificar licencias. De acuerdo a las características de cada conjunto de datos se define una licencia del grafo RDF. Ejemplo de licencia CC BY-SA.

Enlazar datos a otros conjuntos de datos. Los grafos de conocimiento descritos en RDF, se vinculan con otros datos, a fin de proporcionar contexto y enriquecer su significado. Con respecto al proceso de vinculación, se dispone de un módulo de enlazado y validación desarrollado a medida (se usó como base Silk Link Discovery Framework). Entre las fuentes externas están: datos del Banco Mundial (<http://data.worldbank.org>), GeoNames (<http://geoname.org>), DBpedia (<http://dbpedia.org>), y DBpedia Latinoamérica (<http://es-la.dbpedia.org>). Para especificar la condición que debe ser cierta para que dos entidades se consideren duplicadas, el usuario puede aplicar diferentes métricas de similitud, como métodos de comparación de cadena, fecha o URI, a valores de propiedades múltiples de una entidad o entidades relacionadas. El proceso de resolución de identidad descubre los alias de URI en los datos de entrada y los reemplaza con un único URI basado en la heurística de coincidencia proporcionada por el usuario. Además, se agregan enlaces owl:sameAs que apuntan a los URI originales, lo que hace posible que las aplicaciones vuelvan a referirse a las fuentes de datos originales en la Web.

Publicación. Una vez que los datos extraídos se han transformado en RDF, estos datos son almacenados y publicados en un triplestore. En nuestro escenario, decidimos que los datos convertidos, se describen con VoID, se anota la procedencia con PROV y se cargan en el Virtuoso, edición abierta. VoID.. Publicado el LD, cualquier sistema externo es capaz de ejecutar consultas semánticas (utilizando el lenguaje de consulta

SPARQL) para extraer y reutilizar la información que considere oportuno. Los avances hacia compartir indicadores en la forma de datos abiertos enlazados ofrecen una serie de ventajas: habilitar la interoperabilidad semántica, incrementa la integración y el reuso de datos con un pequeño impacto en los sistemas de información actuales; apoya la creatividad y la innovación a través de la participación activa de los ciudadanos. A través de una URI, las máquinas y los usuarios humanos deben poder recuperar descripciones semántica de los recursos de conocimiento. Los usuarios máquina obtienen datos RDF y los humanos obtienen una representación legible de la información en formatos HTML. La comunicación tanto entre los usuarios humanos, como los usuarios máquinas, y el RDF store se realiza en forma de consultas SPARQL.

5.7. Guías sobre Gestión de Datos para alcanzar la Interoperabilidad

La gestión de información de proyectos SmartLand se desarrolla en un entorno distribuido, autónomo y heterogéneo, y para alcanzar la interoperabilidad e integración de datos, se han acordado las siguientes buenas prácticas de datos:

1. Usar datos abiertos y plataformas abiertas para movilizar el conocimiento colectivo. Para que la iniciativa SmartLand funcione de manera efectiva, y para que la sociedad obtenga todos los beneficios de los esfuerzos científicos, es crucial que los datos científicos sean abiertos y legibles por máquina. Los datos abiertos tienen el potencial de ayudar a las comunidades a comprender, evaluar e implementar soluciones tecnológicas adaptadas a las necesidades y objetivos específicos de la comunidad.
2. Definir las licencias de uso de datos. Todos los datos o colecciones de datos que se comparten deben publicarse con una declaración clara y explícita sobre los permisos y restricciones que se han aplicado.
3. Transformar datos existentes en datos semánticamente enlazados. Establecer la semántica y las ontologías que deben ser reutilizadas en un dominio particular. Además, dado que se usarán varias ontologías, será necesario que los conceptos de diferentes ontologías se alineen y relacionen entre sí.
4. Satisfacer el requisito 'Cool URI' de SW, de modo que mejore la interoperabilidad entre sistemas independientes y haga que los URI sean lo más consistentes y persistentes posible.
5. Hacer que los potencialmente "datos enlazables" realmente se vinculen y permitan consultas complejas en un espacio de datos global y distribuido. La esencia de LD es permitir que los datos se conecten entre sí.
6. Habilitar una forma de consumir datos y semántica de manera que más usuarios puedan usarlos. Esto mejoraría la situación actual en la que la mayoría de los datos enlazados se obtienen a través de puntos de acceso SPARQL, que suponen que los usuarios estén familiarizados con las ontologías de dominio que se han usado para describir los datos.
7. Implementar un procedimiento para rastrear la procedencia de las piezas de datos. Dado que un proceso de publicación de datos enlazados transformará las fuentes de datos existentes en datos enlazados, descritos en RDF, es necesario establecer la trazabilidad de la transformación.

6. Conclusiones y Trabajos Futuros

La calidad y la información oportuna son vitales para que los gobiernos, el sector académico, las organizaciones internacionales, la sociedad civil, el sector privado y el público en general puedan tomar decisiones informadas. Por ello, la Gestión Inteligente y Sostenible del Territorio, SmartLand, requiere la recopilación, el procesamiento, el análisis y la difusión de una cantidad sin precedentes de datos e indicadores a nivel local, nacional, regional y mundial y de múltiples partes interesadas. Por lo tanto, la creación de capacidad en gestión de datos e indicadores es fundamental.

Con el fin de monitorear la implementación el desarrollo sostenible, es clave mejorar la disponibilidad y calidad de los datos e indicadores sectoriales a nivel local, alineados a los esfuerzos de organizaciones internacionales relevantes. El esfuerzo modernización y fortalecimiento planteado a través de SmartLand-LD requiere el compromiso pleno, activo y centrado, así como la confianza de las partes interesadas para el funcionamiento eficaz de una iniciativa de datos abiertos e indicadores sobre desarrollo sostenible.

El marco de trabajo describe los pasos necesarios para modernizar y fortalecer la interoperabilidad en los sistemas de indicadores. SmartLand-LD aborda acciones, con un enfoque particular en la construcción de la infraestructura semántica y la capacidad necesaria para respaldar los requisitos de datos locales, que agreguen valor a la agenda nacional y global de desarrollo sostenible.

En esta trabajo se reconoce el papel crucial de la cooperación y la apertura entre las partes interesadas en apoyar los planes y esfuerzos de los territorios en la creación de capacidades. La experiencia y las habilidades de estas partes interesadas son recursos esenciales clave para el progreso, la sostenibilidad y la modernización. La producción de indicadores alineados con el desarrollo inteligente y sostenible requiere una interacción integral entre los proveedores de datos, las partes interesadas y los usuarios. La coordinación y la racionalización de estas actividades son necesarias para evitar la duplicación de esfuerzos y canalizar los esfuerzos hacia promover una Gestión Inteligente y Sostenible del territorio.

Como trabajo futuro está abordar los desajustes de coordinación que existen entre las variables e indicadores de los proyectos SmartLand y los indicadores que están alineados a la Agenda 2030, y avanzar hacia su representación semántica. La Agenda 2030 llama explícitamente a mejorar la creación de capacidad en la gestión de datos e indicadores para apoyar los planes nacionales para implementar los objetivos de desarrollo sostenible. Es esencial que se aborden esas deficiencias a fin de permitir una mejor utilización de las indicadores generadas en el cálculo de los indicadores nacionales, regionales y mundiales de los ODS. SmartLand está diseñado como un documento vivo, abierto a posibles ajustes en una etapa posterior para dar cuenta del desarrollo futuro, por lo tanto, periódicamente, SmartLand revisará y, si es necesario, actualizará su estrategia para alcanzar su efectividad.

Agradecimientos. Los autores expresan su gratitud a UTPL por todo el apoyo institucional a SmartLand, en especial, expresan su reconocimiento a José Barbosa Corbacho, Rector, y a todas las personas involucradas en esta iniciativa.

Referencias

- Acosta, S., Piedra, N., & Suárez, J. P. (2017). The Social Distribution of Knowledge: University Ethical Commitment to the Intelligent Management of the Territory – SmartLand. In *Higher Education in the World 6. Towards a Socially Responsible University: Balancing the Global with the Local*. (pp. 330–337). GUNi SERIES ON THE SOCIAL COMMITMENT OF UNIVERSITIES. Retrieved from http://www.guninetwork.org/files/download_full_report.pdf
- Ahvenniemi, H., Huovila, A., Pinto-Seppä, I., & Airaksinen, M. (2017). What are the differences between sustainable and smart cities? *Cities*, 60, 234–245. <https://doi.org/10.1016/j.cities.2016.09.009>
- Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1), 3–21.
- Balmford, A., Bennun, L., Ten Brink, B., Cooper, D., Côté, I. M., & Crane, P. (2005). The convention on biological diversity's 2010 target. *Science*, 307(5707), 212–213.
- Berners-Lee, T. (2006). Linked data-design issues. Retrieved from <https://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 34–43.
- CKAN. (n.d.). The open source data portal software. Retrieved from <https://ckan.org/>
- Cutter, A., Osborn, D., Romano, J., & Ullah, F. (2015). *Sustainable Development Goals and integration: Achieving a better balance between the economic, social and environmental dimensions*. German Council for Sustainable Development. Retrieved from www.unmultimedia.org/photo/
- Cyganiak, R., Wood, D., & Lanthaler, M. (2014). RDF 1.1 Concepts and Abstract Syntax. *W3C Recommendation 25 February 2014*, (February), 263–270. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Deakin, M., & Waer, H. Al. (2011). From intelligent to smart cities. *Intelligent Buildings International*, 3(3), 133–139. <https://doi.org/10.1080/17508975.2011.586673>
- Dirks, S., Gurdgiev, C., & Keeling, M. (2010). *Smarter cities for smarter growth: How cities can optimize their systems for the talent-based economy*. Somers, NY, USA: IBM Global Services,
- Ensor, P. (1988). The Functional Silo Syndrome. *AME Target*. <https://doi.org/ame.org/sites/default/files/documents/88q1a3.pdf>
- Gregorio, J. (2012). *RFC 6570 - URI Template*. Retrieved from <http://tools.ietf.org/html/rfc6570>
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human - Computer Studies*, 43(5–6), 907–928. <https://doi.org/10.1006/ijhc.1995.1081>

- Guillén, A., Lloret, E., & Gutiérrez, Y. (2016). TLH Suite: Semantic information annotation tool. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, (18), 99-103. <https://doi.org/10.17013/risti.18.99-113>
- Hák, T., Janoušková, S., & Moldan, B. (2016). *Sustainable Development Goals: A need for relevant indicators*. *Ecological Indicators*. <https://doi.org/10.1016/j.ecolind.2015.08.003>
- Hallo, M., Luján-Mora, S., & Morga, A. M. (2017). Cuadros de mando para gestionar el uso de bibliotecas digitales sobre datos enlazados. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, (22), 57-72. <https://doi.org/10.17013/risti.22.57-72>
- Harris, S., & Seaborne, A. (2013). *SPARQL 1.1 Query Language. W3C Recommendation*. <https://doi.org/citeulike-article-id:2620569>
- Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. (J. Hendler & F. Van Harmelen, Eds.) (1st ed.). São Rafael, CA, USA: Morgan & Claypool Publishers. <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>
- Hollands, R. G. (2008). Will the real smart city please stand up? *City*, 12(3), 303-320. <https://doi.org/10.1080/13604810802479126>
- Kates, R. W., Parris, T. M., & Leiserowitz, A. A. (2005). What is sustainable development? Goals, indicators, values, and practice. *Environment*, 47(3), 8-21. <https://doi.org/10.1080/00139157.2005.10524444>
- Mori, K., & Christodoulou, A. (2012). Review of sustainability indices and indicators: Towards a new City Sustainability Index (CSI). *Environmental Impact Assessment Review*, 32(1), 94-106.
- Nam, T., & Pardo, T. A. (2011). Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th annual international digital government research conference: digital government innovation in challenging times* (pp. 282-291).
- Piedra, N., Chicaiza, J., Lopez-Vargas, J., & Caro, E. T. (2016). Guidelines to producing structured interoperable data from Open Access Repositories. In *Proceedings of Frontiers in Education Conference, FIE* (Vol. 2016-Novem). <https://doi.org/10.1109/FIE.2016.7757660>
- Piedra, N., Chicaiza, J., Lopez, J., & Caro, E. T. (2017). A rating system that open-data repositories must satisfy to be considered OER: Reusing open data resources in teaching. In *IEEE Global Engineering Education Conference, EDUCON*. <https://doi.org/10.1109/EDUCON.2017.7943089>
- Piedra, N., Chicaiza, J., Quichimbo, P., Saquicela, V., Cadme, E., López, J., & Tovar, E. (2015). Framework for the integration of digital resources based-on a semantic web approach. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, (E3), 55-70. <https://doi.org/10.17013/risti.e3.55-70>

- Piedra, N., & Suárez, J. P. (2017). SmartLand-LD: A Linked data approach for integration of heterogeneous datasets to intelligent management of high biodiversity territories. In *Advances in Intelligent Systems and Computing* (Vol. 688, pp. 207–218). https://doi.org/10.1007/978-3-319-69341-5_19
- Piedra Pullaguari, N. O. (2016). *Marco de trabajo de interoperabilidad semántica para la Educación Abierta: un enfoque para la Integración de Sistemas de Información de Recursos Educativos Abiertos autónomos, distribuidos y heterogéneos*. Madrid: Universidad Politécnica de Madrid.
- United Nations General Assembly. (2015). *Transforming our world: The 2030 agenda for sustainable development*. Retrieved from: <https://sustainabledevelopment.un.org/content/documents/7891Transforming%20Our%20World.pdf>. <https://doi.org/10.1007/s13398-014-0173-7.2>
- UNSG (2016). Progress towards the Sustainable Development Goals. *United Nations Economic and Social Council*, (July), 28. <https://doi.org/10.1017/S0020818300011322>
- Vázquez-Reyes, S., de León-Sigg, M., Velasco-Elizondo, P., Villa-Cisneros, J., & Briceño-Muro, S. (2017). Information retrieval through inverted index in Be Intelligent system. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, (21), 85-98. <https://doi.org/10.17013/risti.21.85-98>
- W3C. (2008). Cool URIs for the Semantic Web. *W3C Interest Group Note*, 49(December 2008), 1–15. Retrieved from <http://www.w3.org/TR/cooluris/>
- W3C. (2011). *Library Linked Data Incubator Group: Use Cases*. Retrieved from <https://www.w3.org/2005/Incubator/lld/XGR-ld-usecase-2011025/>
- W3C OWL Working Group. (2012). OWL 2 Web Ontology Language Document Overview. *OWL 2 Web Ontology Language*, (December), 1–7. Retrieved from <http://www.w3.org/TR/owl2-overview/>