

## ENFOQUE DE INTEGRACIÓN BASADO EN DATOS ENLAZADOS EMPRESARIALES

### APPROACH OF INTEGRATION BASED ON LINKED ENTERPRISE DATA

Mavis Lis Stuart Cárdenas

Universidad Tecnológica de la Habana “José Antonio Echeverría”, La Habana, Cuba,

[mavis@ind.cujae.edu.cu](mailto:mavis@ind.cujae.edu.cu)

Diana Rosa Prieto del Rio

Empresa de Informática y Automatización para la Construcción, La Habana, Cuba, [drprieto@aicros.cu](mailto:drprieto@aicros.cu)

Tatiana Delgado Fernández

Unión de Informáticos de Cuba, La Habana, Cuba, [tatiana.delgado@uic.cu](mailto:tatiana.delgado@uic.cu)

Mercedes Delgado Fernández

Escuela Superior de Cuadros del Estado y del Gobierno, La Habana, Cuba, [mercedes@esceg.cu](mailto:mercedes@esceg.cu)

*Recibido: 12/11/2018*

*Aceptado: 21/12/2018*

#### Resumen

La información es considerada hoy como uno de los principales recursos que posee la empresa. Muchas veces esta información se encuentra dispersa en diferentes fuentes de datos que conforman las aplicaciones de negocios de la empresa. Ante el reto aún presente en la mayoría de las organizaciones de alcanzar una efectiva integración de los datos, resulta imperioso que las empresas opten por mecanismos cada vez más avanzados que le permitan utilizar de manera favorable los recursos y fuentes de información para el logro de sus objetivos. En este artículo las autoras discuten la utilidad de los Datos Empresariales Enlazados para lograr una efectiva integración de los datos en la empresa, así como un marco general para su aplicación. Se presenta también el estudio de algunas herramientas tecnológicas que permitirán la implementación del marco general de integración de datos basada en LED.

**Palabras Clave:** Información, Integración de datos, Datos Empresariales Enlazados, Marco de integración.

#### Abstract

Information is considered today as one of the main resources that the company has. General times this information is scattered in different data sources that make up the business applications of the company. Given the challenge still present in most organizations to achieve effective data integration, it is imperative that companies opt for increasingly advanced mechanisms which allow them to use resources and sources of information in a favorable manner for the achievement of your objectives. In this article the authors will discuss the usefulness of the Linked Enterprise Data to achieve an effective integration of the data in the company, as well as a general framework for its application. A study of some technological tools for the implementation of the general framework of data integration based on Linked Enterprise Data will also be presented.

**Keywords:** Information, Data integration, Linked Enterprise Data, Integration framework.

### Introducción

En la actualidad continúa siendo la valoración de la información un recurso importante de la empresa,<sup>1y2</sup> y todo lo que se haga en función de su acceso en tiempo y forma constituyen acciones claves para garantizar el éxito empresarial.<sup>3</sup> Muchas veces la información se encuentra dispersa en diferentes fuentes de datos contenidas en las aplicaciones de la empresa, como los sistemas de Planificación de Recursos Empresariales (ERP, por sus siglas en inglés), los sistemas de Gestión de Relaciones con el Cliente (CRM por sus siglas en inglés), los almacenes de datos, las Intranets corporativas, y otros sistemas de información empresariales.<sup>4-6</sup> Esta diversidad de aplicaciones ha ocasionado que los datos se encuentren dispersos en toda la organización, sin una relación explícita entre ellos, limitando la usabilidad y el aprovechamiento de su potencial.<sup>1</sup> De ahí que los intentos para lograr la integración de todas estas aplicaciones sea un tema recurrente en los últimos años y un reto a vencer por las empresas.<sup>3y5</sup>

En la última década, los enfoques de integración se basaron principalmente en XML, Servicios Web y Arquitecturas Orientadas a Servicios (SOA).<sup>1,3y5</sup> En este sentido, se aprecia un marcado énfasis en la integración de las aplicaciones empresariales, mientras es menos empleada la integración de los datos de la empresa bajo una vista unificada a partir de las relaciones entre los propios datos. Por otra parte, la integración a nivel de aplicaciones y sistemas informáticos existentes no es suficiente para resolver, en última instancia, los retos de la integración e interoperabilidad de datos en las empresas.<sup>3</sup> Hoy en día existen tecnologías más eficientes para la integración de estos datos, en particular, cabe destacar el uso de los Datos Enlazados,<sup>3</sup> popularmente conocidos por su terminología en inglés "Linked Data" (LD).

LD se refieren a un conjunto de buenas prácticas para la publicación y vinculación de datos estructurados en la Web.<sup>7y8</sup> Los datos provienen de diferentes fuentes de datos, como bases de datos de diferentes sistemas en la misma organización o de dos organizaciones diferentes. Siguiendo los principios de LD se crean vínculos entre los datos,<sup>8</sup> dando lugar a un espacio global de datos<sup>7y9</sup>, que facilita la tarea de compartirlos<sup>7</sup> y ayuda por tanto a disminuir la dispersión de los mismos.<sup>9</sup>

Más formalmente, LD se refiere a datos que se publican en la web de tal manera que son legibles por la computadora, su significado se define explícitamente, están vinculados a otras bases de datos externas, y a su vez pueden ser vinculados desde conjuntos de datos externos (8). La tecnología de datos vinculados tiene una fuerte potencialidad por su capacidad de integrar datos provenientes de múltiples fuentes de datos con variada estructura lo que posibilita revelar partes ocultas, extraer nuevos datos, crear perspectivas de análisis que antes no estaban disponibles, y facilitar nuevos escenarios para el procesamiento de estos datos. Todo lo cual mejora la usabilidad de los datos.<sup>10y11</sup>

La motivación detrás de la vinculación de datos de la empresa es evidente. Se han realizado intentos cautelosos para experimentar los principios de Linked Data y evaluar los beneficios, dando lugar al denominado paradigma de "Datos Empresariales Enlazados" (LED, por sus siglas en inglés), la contrapartida de Linked Data en el dominio empresarial.<sup>12y13</sup>

En los Datos Empresariales Enlazados se usan herramientas y técnicas de la Web Semántica para conectar, exponer y usar datos de la empresa.<sup>1y14</sup> Al igual que la web tradicional enlaza páginas web, los Datos Empresariales Enlazados enlazan los datos de la empresa en el nivel más detallado. A su vez estos enlaces son como uniones en una base de datos relacionales, pero a través de todos los sistemas de la empresa, convirtiéndose así en una inmensa base de datos distribuida de información y conocimiento del contexto en cuestión.

Este artículo tiene como objetivo evidenciar las potencialidades de los Datos Empresariales Enlazados para lograr una efectiva integración de datos en el contexto empresarial. Primeramente se presentan las definiciones básicas y el enfoque contemporáneo de Datos Empresariales Enlazados. Posteriormente se muestra un marco general para integrar información empresarial mediante el modelo de Datos Empresariales Enlazados. Por último se muestra la perspectiva tecnológica para la implementación de este marco general.

### Definiciones básicas

Las decisiones eficaces se basan en el análisis de los datos y de la información, por lo que para las empresas constituye una meta contar con datos precisos y coherentes para administrarlos de un modo estratégico, con el fin de conseguir un mayor rendimiento y competitividad.

Los datos, según el Decreto 281 del 2011 de Cuba,<sup>15</sup> representan hechos reales u objetivos que por sí mismos no tienen la facultad de comunicar un significado. Su importancia radica en la capacidad de asociarse dentro de un contexto para convertirse en información. En este mismo Decreto se define información como *un conjunto organizado de datos procesados que constituyen un mensaje sobre un determinado ente o fenómeno. Proporciona significado o sentido a las cosas y su uso racional es la base del conocimiento, facilitando la solución de problemas y la toma de decisiones.*

En los últimos años se ha vuelto estratégico para las empresas contar con una solución informática que integre los datos de las diferentes fuentes de información con las que trabaja. Para ello se han desarrollado múltiples soluciones, aplicaciones y toda una infraestructura con una visión más integral y dinámica de cada área del negocio que facilita que la información contribuya a la toma de decisiones clave. La idea y la búsqueda de la integración de las aplicaciones y sus datos no es algo nuevo. Lo diferente es el enfoque y las ideas que la integración abarca y las técnicas que utiliza.<sup>16</sup>

Gartner define la integración de datos (Data Integration) y al mismo tiempo la gestión e integración de datos (DMI -Data Management and Integration-) como *prácticas, técnicas arquitectónicas y herramientas para lograr un acceso y entrega consistentes de datos a través del espectro de áreas de datos y tipos de estructura de datos en la empresa, para satisfacer los requisitos de consumo de datos de todas las aplicaciones y procesos empresariales.*<sup>17y18</sup>

La integración de datos, según IBM, es *la combinación de procesos técnicos y empresariales utilizados para combinar datos de fuentes dispares en información significativa y valiosa.* Una solución completa de integración de datos proporciona datos confiables de una variedad de fuentes.<sup>19</sup> La integración favorece la búsqueda, el acceso, la replicación, la transformación y el análisis coherente mediante una vista unificada de activos de información para así poder satisfacer todas las necesidades de los

negocios. Por tanto la integración de datos es sinónimo de rapidez y eficacia en el procesamiento y la obtención de información de la empresa.

Históricamente han existido modelos empresariales para integrar los datos de una empresa, como los ERP, la Inteligencia de negocio (ETL – Almacenes de datos) y la Gestión de Datos Maestros (MDM). A diferencia de sus predecesores, los Datos Empresariales Enlazados constituye un modelo de integración independiente de las plataformas informáticas, aunque co-existen con ellas, y permiten integrar y vincular cualquier tipo de datos de cualquier fuente (por ejemplo, sensores, redes sociales y fuentes de datos transaccionales), lo cual lo hace particularmente flexible y adaptable a las condiciones actuales en las que se desenvuelve la empresa moderna, en la era de Internet de las Cosas (IoT), la hiperconectividad y las redes sociales.

### **Datos empresariales enlazados como un enfoque contemporáneo para la integración**

Los datos empresariales no pueden actuar como islas de información dentro de la empresa. Para dar solución a tal problemática, existe la posibilidad de migrar al uso de la tecnología web semántica a través de los principios de Linked Data; con la misión de obtener información automatizada y vinculada a conjuntos de datos de dominios relacionados a las diversas áreas del contexto empresarial.

El término de Linked Data se refiere a un conjunto de buenas prácticas para publicar y enlazar datos estructurados en la web. Estas prácticas fueron introducidas por Tim Berners-Lee y son conocidas como los principios de Linked Data.<sup>7</sup> Los llamados “principios de linked data” proporcionan una receta básica para la publicación y conexión de los datos, respetando la arquitectura y los estándares de la infraestructura de la Web. Específicamente, se usa mediante el empleo de las URIs HTTP para identificar los recursos, el protocolo HTTP como mecanismo de recuperación de información, y el modelo de datos RDF para representar descripciones de recursos.<sup>8</sup> LD, representa los datos en formato RDF (Resource Description Framework), que constituye un modelo de datos basado en gráficos, es expresivo, flexible y robusto.<sup>14</sup>

La tecnología de Linked Data ofrece un enorme potencial para las empresas,<sup>5</sup> dando lugar al denominado paradigma de "Datos Empresariales Enlazados" (LED, por sus siglas en inglés), aplicado el dominio empresarial.<sup>12y13</sup> LED es un enfoque de integración de los datos de la empresa, que integra los datos con independencia de su procedencia, para crear un espacio de información global y unificado con el que se facilita la creación de nuevas informaciones del negocio para resolver necesidades operacionales.<sup>1</sup>

Los datos se publican siguiendo los “principios de linked data” y se convierten en parte de una capa o un único espacio de datos o web de datos,<sup>8</sup> donde los datos están relacionados.<sup>7</sup> LED tiene en cuenta todos los tipos de datos empresariales e interconecta estos datos, creando un almacén unificado y coherente que expone y comparte nuevos objetos de conocimiento adaptados a la toma de decisiones y las acciones requeridas por las áreas del negocio.<sup>1</sup>

LED aplica las tecnologías de la web semántica al sistema de información empresarial como respuesta al desafío de crear un sistema de información ágil y de alto rendimiento: fuentes de datos internas se

enlazan entre sí y finalmente se consolidan con datos externos, se recombina la información existente en nuevas e interesantes formas que potencia la generación de información valiosa y oportuna.<sup>1</sup>

El modelo LED aporta una manera de utilizar los datos de forma eficiente y aún cuando todavía su aplicación sea un reto<sup>1</sup> y se necesiten impulsar los factores que viabilizan su difusión,<sup>20</sup> LED aporta al sistema de información beneficios como los siguientes:<sup>1</sup>

- Una alternativa para crear nuevas aplicaciones de forma rápida que utilicen los datos existentes sin alterar los modelos de datos residentes en las aplicaciones que funcionan.
- Enlazar explícitamente los datos de diferentes fuentes que de forma implícita estén relacionados.
- Vincular los datos existentes, ya sean internos o externos.
- Incorporar datos externos a las aplicaciones existentes, sin desestabilizar el sistema de información de la empresa.
- Reducir la complejidad, ofreciendo una metodología unificada para el intercambio de datos entre aplicaciones. Ya sea que los datos sean internos o extraídos de la Web, el LED abre una serie de posibilidades para consolidar toda la gama de información disponible.

A pesar de su potencial, LED todavía está poco utilizado en las empresas, o es lenta su aplicación.<sup>5,13y21</sup> El estudio de los factores que viabilizan o impiden la difusión de LED en el entorno empresarial, es un tema actual de análisis.<sup>20</sup> En este artículo se presenta un marco general de integración de datos basado en LED, con vistas a facilitar su adopción en las empresas cubanas. Entre los factores que viabilizan se encuentran la interoperabilidad y el apoyo a la toma de decisiones de la gestión. Los que impiden se refieren a la cultura organizacional, el alineamiento estratégico, la confusión en cuanto a datos abiertos y conectados y la calidad de los datos.

### **Marco general de integración de datos basado en LED**

La meta de un modelo o sistema de integración de datos es combinar los datos que residen en diferentes fuentes y proveer una vista unificada de estos para los usuarios. Generalmente esta vista unificada es representada por un esquema de mediación o meta-modelo de referencia. Los principales componentes de un sistema de integración de datos, además del esquema de mediación, son las fuentes de datos y los mapeos de las fuentes de datos con el esquema de mediación, también llamados descripciones de la fuente.<sup>22</sup>

El marco general para la integración de datos que se propone en este artículo, para posibilitar la aplicación de LED en el entorno empresarial, considera elementos de la integración semántica de los datos,<sup>22</sup> así como, el flujo de trabajo necesario para el desarrollo de LD en cualquier dominio<sup>23</sup> y se compone de los elementos siguientes: las fuentes de datos, un esquema de mediación para la modelación semántica de los datos y los mapeos de las fuentes de datos con el esquema de mediación. En la figura 1 se muestra el marco general para la integración de datos basado en LED.

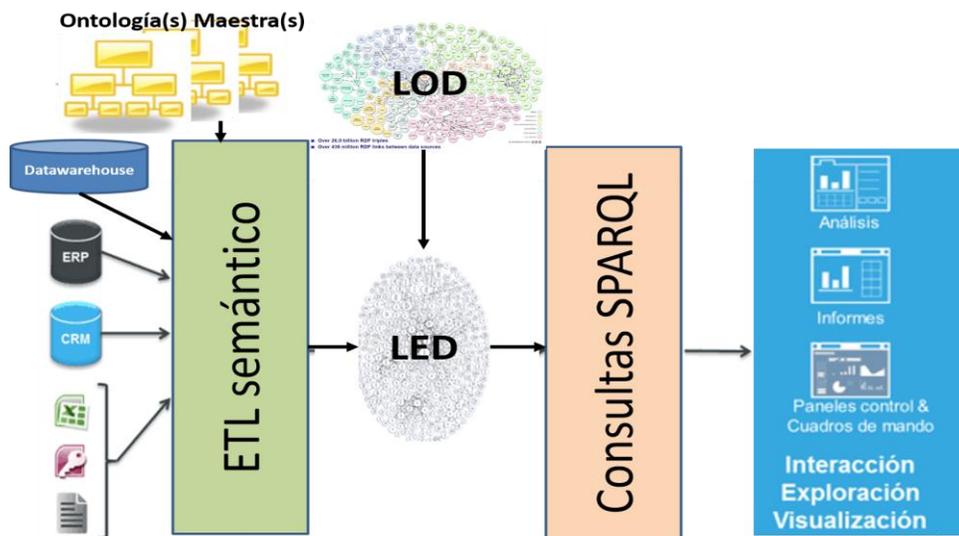


Figura 1. Marco general de integración de datos basado en LED. (fuente: elaboración propia)

El objetivo del marco de integración propuesto es combinar los datos que residen en diferentes fuentes y proveer una vista unificada de los mismos, según las necesidades informativas de los usuarios. Para ello, contempla los elementos siguientes:

1. El uso de ontologías maestras (meta-modelos y ontologías empresariales existentes).
2. Mapeo basado en reglas en el componente ETL (extraer, transformar y cargar) para obtener las nuevas ontologías (RDF).
3. Obtención de una vista unificada LED.
4. Enriquecimiento del modelo LED a partir de los conjuntos de datos disponibles en LOD (Linked Open Data).
5. Generación de consultas SPARQL sobre el LED obtenido.
6. Visualización de los resultados de las consultas.

A los efectos del alcance de este artículo, se describen algunas características de los componentes hasta la obtención de la vista unificada LED.

### Uso de ontologías maestras

Las ontologías empresariales se han estado abordando en la última década cada vez más frecuentemente, existiendo en la actualidad varias de ellas que constituyen en sí mismas meta-modelos de los cuales se pueden heredar ontologías de dominio de datos vinculados, particulares para un sector o una empresa en cuestión. En<sup>24</sup> se ofrece una sistematización de ontologías de dominio empresarial de los sistemas de ERP, gestión de cadenas de suministro (SCM) y CRM, entre otras.

Construido sobre las bases conceptuales de los marcos de arquitectura empresarial TOGAF, DODAF y Zachmann, el meta-modelo Essential es una representación ontológica que agrupa los conceptos de la

arquitectura empresarial en cuatro capas –Negocio, Información, Aplicaciones e Infraestructura de TI–, cada una de las cuales es especificada en tres niveles de profundidad: conceptual, lógico y físico,<sup>25</sup> el cual ha sido utilizado para el diagnóstico de capacidades de gestión organizacional, entre otros escenarios de uso.

El uso de una ontología maestra, como entidad central de la definición de los términos y relaciones presentes dentro del proceso ETL y sus diversas fuentes de información, es también presentado en la metodología Onto-ETL.<sup>26</sup>

### Componente ETL. Mapeo basado en reglas.

Las ontologías soportan mapeos de Bases de Datos Relacionales, Excel, CSV, y otros formatos de extensión definidos por el usuario hacia RDF.<sup>22</sup> Se emplea el método estático de extraer, transformar y cargar (ETL) para la implementación del mapeo entre los variados métodos de generación de reglas de mapeo y transformación de datos de acuerdo a las reglas establecidas. La desventaja principal de este método es la dificultad de actualizar las reglas para la inclusión de nuevas estructuras de datos al modelo de datos semántico; sin embargo, estos mismos autores han demostrado su idoneidad para ejecutar el análisis y procesamiento basado en datos o inferencias.

Otros trabajos parten de los propios almacenes de datos existentes en la organización para crear las redes de ontologías o modelo de datos enlazados. En<sup>27</sup> se propone una metodología dirigida a transformar los modelos de datos, desde cubos multidimensionales hacia esquemas de ontologías RDF.

### Obtención de la vista unificada LED

Se genera el modelo LED a partir del enlace de los conjuntos de datos (RDF) generados por el método ETL, siguiendo la lógica del meta-modelo empleado y de la dinámica organizativa de la empresa en cuestión. La vista unificada semántica de los datos de la empresa basada en LED es enriquecida, en caso necesario, con el proyecto de Datos Abiertos Enlazados de la Web (LOD), para diseñar una web empresarial extendida de datos, que mediante consultas SPARQL y usando los estándares RDF/XML, permita incrementar la interoperabilidad y accesibilidad de los datos de la empresa; y proporcione nuevas informaciones del negocio para las necesidades operacionales y la toma de decisiones.

### **Perspectiva metodológica y tecnológica para la implementación del marco de integración LED**

Para la implementación del modelo, se han estudiado un grupo de plataformas y herramientas tecnológicas que serán descritas en este acápite. El marco de enlazado Silk proporciona un conjunto de herramientas para descubrir y mantener enlaces entre recursos de la Web de Datos. Silk está compuesto por tres componentes:

- El motor de búsqueda de enlaces, basado en la similitud entre recursos o identificadores compartidos.
- Herramienta para evaluar los enlaces generados.
- Protocolo para hacer frente a los posibles cambios producidos en los enlaces entre datasets.

Silk utiliza el lenguaje declarativo Silk - LSL (Link Specification Language) con el que el usuario puede establecer el tipo de enlaces RDF que deben ser descubiertos entre los diferentes conjuntos de datos y las condiciones que deben cumplir. Para la problemática de la integridad de los enlaces entre

los diferentes datasets, Silk utiliza el denominado *Web of Data - Link Maintenance Protocol (WOD-LMP)*. Este protocolo se basa en el patrón publicación-subscripción para propagar los cambios que un conjunto de datos pueda tener entre sus suscriptores. Dispone de una consola para realizar enlaces entre dos conjuntos de datos y un servidor http capaz de recibir datos e introducirlos en el flujo RDF a través de ítems de datos. La flexibilidad y calidad de los datos que gestiona el lenguaje Silk-LSL permite adaptarlos a la filosofía de trabajo de Datos Enlazados.

Por su parte LIMES (LInk discovery framework for MEtric Spaces) es un marco de descubrimiento de enlaces para la Web de Datos. Implementa enfoques de tiempo eficiente para el descubrimiento de enlaces a gran escala basados en las características de los espacios métricos. Es fácilmente configurable a través de un archivo de configuración; así como, a través de una interfaz gráfica de usuario. LIMES se puede descargar como herramienta independiente para llevar a cabo el descubrimiento de enlaces o como una biblioteca Java.<sup>28</sup>

Fuseki es un servidor de SPARQL que proporciona consultas mediante HTTP. Se encuentra integrado dentro del conjunto de servicios de Apache Jena, un entorno de código abierto para el desarrollo de la web semántica mediante Java. La arquitectura de esta plataforma incluye:

- Una API para trabajar con ontologías RDF y OWL.
- Motor de inferencia para razonar sobre ontologías RDF y OWL.
- Estrategias de almacenamiento flexible para almacenar triples RDF en memoria o en fichero.
- Motor de consultas compatible con SPARQL.

Virtuoso<sup>29</sup> es un Data Store híbrido que combina las funcionalidades de los gestores de bases de datos enlazados (RDBMS), de los sistemas gestores de bases datos objeto-relacional (ORDBMS), bases de datos virtuales, RDF, XML y aplicaciones web. Entre las características más relevantes se pueden citar las siguientes:

- Posee licencia pública general (GPL) para el producto OpenLinkVirtuoso y también licencia pagada para otras versiones del producto.
- Posee un diccionario de datos en donde se almacena toda la información de los objetos de los usuarios.
- Trabaja en los sistemas operativos Windows, Mac, Linux y Unix.
- Soporta los formatos RDF y XML.
- Soporta SPARQL.
- Soporta ISQL, lenguaje que permite una fácil administración y mantenimiento de la base de datos.
- Mecanismos de seguridad basados en privilegios y roles.
- Provee conexiones a fuentes de datos en formatos, XAML, ODBC, JDBC, ADO.NET y OLE.DB.

Pubby es otro servidor RDF que agrega una interfaz de datos vinculados a los almacenes tripletas existentes con SPARQL. Tiene las características:

- Proporciona una interfaz de datos vinculados a servidores de protocolo locales o remotos SPARQL.
- Proporciona URI desreferenciables al volver a escribir los URI encontrados en el conjunto de datos expuesto a SPARQL en el espacio de nombres del servidor Pubby.

## ENFOQUE DE INTEGRACIÓN BASADO EN DATOS ENLAZADOS EMPRESARIALES

- Proporciona una interfaz HTML sencilla que muestra los datos disponibles sobre cada recurso.
- Se ocupa de manejar 303 redirecciones y negociación de contenido.
- Compatible con contenedores de servlets Tomcat y Jetty.
- Incluye una extensión de metadatos para agregar metadatos a los datos proporcionados.

Sesame es un marco de código abierto que permite almacenar, navegar y analizar los datos RDF a través de SPARQL. En mayo de 2016, Sesame oficialmente bifurcó en un proyecto de Eclipse llamado RDF4J. Este último proporciona funcionalidad para almacenamiento, consulta y razonamiento eficaces y escalables con datos RDF y un conjunto de librerías (APIs) de acceso neutral para proveedores de bases de datos RDF. Además de su uso principal como un conjunto de bibliotecas Java, RDF4J también proporciona una aplicación web de servidor a la que se puede invocar como un servicio web para acceso a bases de datos RDF y una aplicación web Workbench que proporciona una interfaz de usuario cliente un servidor RDF4J.

A su vez, las aplicaciones que utilizan o consumen los datos enlazados publicados pueden ser desarrolladas con el uso de diferentes librerías, según sea el lenguaje de programación:

Lenguaje de programación	Librerías	Descripción
<b>PHP</b>	EasyRDF	Es una biblioteca que facilita el uso de consultas SPARQL en aplicaciones PHP.(«EasyRdf - RDF Library for PHP», 2011)
	Graphite	Librería para el trabajo con RDF, SPARQL y Linked Data <sup>30</sup>
	sparqllib.php	Librería sencilla y gratuita para el trabajo con consultas SPARQL, no hace caso de lenguajes ni de la información de tipos de datos <sup>31</sup>
<b>Java</b>	Jena	Es un proyecto de Apache que ofrece un marco de Java de fuente abierta para construir aplicaciones de web semántica y datos vinculados.
<b>Python</b>	RDFLib	Es una librería Python para el trabajo con RDF que incluye analizadores y serializadores, interfaz gráfica, soporta consultas SPARQL 1.1 <sup>32y33</sup>
<b>C#</b>	dotNetRDF	Es una librería .NET para RDF y de código abierto para el análisis, la gestión, la consulta y la escritura RDF. Permite el trabajo con almacenes de tripletas RDF como AllegroGraph, Jena, Stardog y Virtuoso <sup>34</sup>

Tabla 1. Librerías para desarrollar aplicaciones bajo el enfoque de LED (fuente: elaboración propia)

Estudios revelan que en las empresas los datos no estructurados o semi-estructurados superan el 70%<sup>35</sup>; sin embargo la mayoría de ellos existen como “silos de datos” con un potencial a la interoperabilidad dentro y fuera de la empresa. Una de las metodologías descritas para la aplicación del LED<sup>36</sup> propone las etapas y fases siguientes, teniendo como premisa a garantizar que solo usa datos estructurados, además recomienda el uso de algunas de las tecnologías descritas en la Tabla 1:

1. Limpieza (OpenRefine).
2. Triplificación. Se usan las fases a) Extracción, b) Denotación, c) Enriquecimiento y Mapeo y d) Conversión. La tecnología recomendada es D2R Server.
3. Entrelazado (SILK Framework).
4. Almacenamiento (Triple Stores).
5. Visualización (Sparql Endpoint/Sgvizler).

El uso de estas metodologías, herramientas y plataformas en el marco de integración de datos basado en LED es una tarea que está siendo desplegada en una empresa cubana donde se está experimentando la implementación del marco general propuesto.

### Conclusiones

El enfoque de los Datos Empresariales Enlazados ofrece múltiples ventajas para la organización, permitiendo mejoras en el acceso y visibilidad de toda la información de la empresa desde una vista unificada.

La principal contribución de este artículo es el análisis, la discusión y la difusión del modelo de Datos Empresariales Enlazados como enfoque efectivo de integración de datos complejos en una organización. Un marco general de integración de los datos basados en LED requiere desplegar las etapas mínimas de ETL, Entrelazado, Consultas y Visualización.

En el método ETL expuesto, los datos relacionales de la empresa son transformados en RDF usando reglas de mapeo y de almacenamiento de los datos en tripletas RDF que heredan características de ontologías superiores, meta-modelos y otras conceptualizaciones de dominio.

El trabajo inmediato está orientado al diseño de un procedimiento derivado del marco general de integración basado en LED y su implementación en una empresa cubana.

### Referencias bibliográficas

1. Antidot. Linked Enterprise Data. Principles, uses and benefits. [Internet] 2012 [consultado 2015 Dic] Disponible en: <https://www.antidot.net/wp-content/uploads/2012/11/LinkedEnterpriseData-WP-en-v2.2.pdf>.
2. Ahmad M. Semantic Derivation of Enterprise Information Architecture from Riva-based Business Process Architecture [Tesis de Doctorado]: University of the West of England. Faculty of Environment and Technology, 2016.
3. Frischmuth P, Klímek J, Auer S, Tramp S, Unbehauen J, Holzweißig K, et al. Linked Data in Enterprise Information Integration. IOS Press Journal. Semantic Web 0. [Internet] 2012; 1-0. Disponible en: [http://svn.aksw.org/papers/2012/SWJ\\_LinkedDataInEnterpriseInformationIntegration/public.pdf](http://svn.aksw.org/papers/2012/SWJ_LinkedDataInEnterpriseInformationIntegration/public.pdf)

4. Allemang D. Semantic Web and the Linked Data Enterprise. In: Wood D, editor. Linking Enterprise Data: Springer, Boston, MA; 2010.
5. Hladky D, Maltseva SV. Linked Data Paradigm for Enterprises: Information Integration and Value Chain. БИЗНЕС-ИНФОРМАТИКА 2013; 2(24):3-12.
6. Afonso V, Sousa G, Ziviani F, Silva F, editors. Supporting Competitive Intelligence with Linked Enterprise Data. 17th International Conference on Enterprise Information Systems (ICEIS-2015); 2015.
7. Berners-Lee T. Linked Data - Design Issues. W3C; [Internet] 2006 [actualizado septiembre 2016]; Disponible en: <https://www.w3.org/DesignIssues/LinkedData.html>.
8. Bizer C, Heath T, Berners-Lee T. Linked Data - The Story so Far. Journal on Semantic Web and Information Systems. 2009; 5(3):1–22.
9. Mihindikulasooriya N, García-Castro R, Esteban M, editors. Linked Data Platform as a novel approach for Enterprise Application Integration. Fourth International Workshop on Consuming Linked Data (COLD2013) co-located with the 12th International Semantic Web Conference (ISWC 2013); 2013.
10. FUJITSU. Linked Data, Connecting and Exploiting big data. [Internet]. 2012 [consultado 2015 Abr] Disponible en: [https://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-\(v1.0\).pdf](https://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-(v1.0).pdf).
11. Mynarz J. Integration of public procurement data using Linked Data. Journal of Systems Integration. 2014; 5(4):19-31.
12. Servant F-P. Linking Enterprise Data. 2008.
13. Moura Rd, Priscila E, Neiva MA, Macedo M, A. OGF. Avaliação da Produção Científica sobre Enterprise Linked Data. 2014.
14. Feigenbaum L. Semantic Web Technologies in the Enterprise 2006.
15. Decreto Ley No. 281 Sobre la continuidad y el fortalecimiento del sistema de dirección y gestión empresarial cubano. (2011).
16. Florence Lin. Enterprise Application Integration (EAI) Techniques. [Internet] [consultado 2005 Marzo] Disponible en: <http://www0.cs.ucl.ac.uk/staff/ucacwx/lectures/3C05-04-05/EAI-Essay.pdf>.
17. Gartner. Data Integration - Gartner IT Glossary. [Internet] [consultado 2017 Jun 29]; Disponible en: <http://www.gartner.com/it-glossary/data-integration-tools/>.
18. Glossary G. DMI (data management and integration) - Gartner IT Glossary. [Internet] [consultado 2017 Jun 29]; Disponible en: <http://www.gartner.com/it-glossary/dmi-data-management-and-integration>.
19. IBM. IBM Data Integration - IBM Analytics. [Internet] [consultado 2017 Jun 29]; Disponible en: <https://www.ibm.com/analytics/us/en/technology/data-integration/>
20. Silveira M, Visintin L, Ostuni FÁ. Os fatores que viabilizam e impedem a difusão do Linked Data no âmbito empresarial: uma revisão da literatura. Informação & Tecnologia (ITEC). 2017; 4(1,5):129-14.
21. Abramowicz W, Auer S, Heath T. Linked Data in Business. Bus Inf Syst Eng Springer Fachmedien Wiesbaden. 2016; 58(5):323–6. Epub August 2016.

22. Olivé A. A Universal Ontology-based Approach to Data Integration. Special Issue on Conceptual Modelling in Honour of Heinrich C. Mayr. International Journal of Conceptual Modeling. 2018.
23. Auer S, Lehmann J, Ngonga Ngomo A-C, Zaveri A. Introduction to Linked Data and its Lifecycle on the Web. [Internet] 2011:1-95. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.654.7040&rep=rep1&type=pdf>.
24. Westerski A, Iglesias CA. Exploiting Structured Linked Data in Enterprise Knowledge Management Systems: An Idea Management Case Study. 15th IEEE International Enterprise Distributed Object Computing Conference Workshops; 2011.
25. Ortega YC., Pavón Y, Delgado M, Blanco J, Infante MB, Pérez N, Fuentes E. Ontologías para la integración organizacional con tecnologías de la información. 18 Convención Científica de Ingeniería y Arquitectura CUJAE, La Habana; 2016. p.1-10.
26. Villanueva J. Marco de trabajo basado en ontologías para el proceso ETL. [Tesis de Maestría]. México, DF: Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional; 2011. [Internet] Disponible en: <https://www.cs.cinvestav.mx/TesisGraduados/2011/TesisJoelVillanueva.pdf>
27. Serra F, Delgado T. DW2RDF4SDG–Ontology modeling from multi-dimensional cubes for Sustainable Development Goals. Sistemas & Telemática. 2018; 16(44):9-24
28. AKSW. LINES. Research Group Agile Knowledge Engineering and Semantic Web (AKSW). [Internet] [consultado 2016]; Disponible en: <http://aksw.org/Projects/LINES.html>.
29. OpenLink S. Openlink Virtuoso. Disponible en: <https://virtuoso.openlinksw.com/>
30. Gutteridge C. Graphite PHP Linked Data Library. University of Southampton. [Internet] 2010 [2017/01/04/16:52:55]; Disponible en: <http://graphite.ecs.soton.ac.uk/>.
31. Gutteridge C. SPARQL RDF Library for PHP. University of Southampton. [Internet] 2010 [consultado 2017 Ene 4]; Disponible en: <http://graphite.ecs.soton.ac.uk/sparqllib/>.
32. Python Software F. Python Package Index. [Internet] 2015 [consultado 2018 Ene 4]; Disponible en: <https://pypi.python.org/pypi/rdflib>.
33. Gomez B, Vasconcelos A, Sousa P, editors. Using an Enterprise Ontology Approach in the Definition of the Enterprise Information Architecture 5th European Conference on Information Management and Evaluation, ECIME; 2011.
34. dotNet RDFP. dotNetRDF. [Internet] 2010 [consultado 2018 Ene 4]; Disponible en: <https://dotnetrdf.github.io/>
35. Holzinger A, Stocker C, Ofner B, Probaska G, Brabenetz A, Hofman R. Combining HCI, Natural Language Processing, and Knowledge Discovery - Potential of IBM Content Analytics as an Assistive Technology in the Biomedical Field. In: Proceedings of the Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data, Maribor, 2013; 13–24. [Internet] Disponible en: [http://dx.doi.org/10.1007/978-3-642-39146-0\\_2](http://dx.doi.org/10.1007/978-3-642-39146-0_2).
36. Suresh S, Nayak A. LinkED: A Novel Methodology for Publishing Linked Enterprise Data. In: CIT. Journal of Computing and Information Technology. September, 2017; (25)3:191–209, doi: 10.20532/cit.2017.1003477.