




Enfoque de datos empresariales enlazados aplicado en una empresa cubana

Linked enterprise data approach applying at a cuban company

Mavis Lis Stuart-Cárdenas^I

 <https://orcid.org/200000-0002-0461-7118>

Tatiana Delgado-Fernández^I

 <https://orcid.org/0000-0002-4323-9674>

Mercedes Delgado-Fernández^{II}

 <https://orcid.org/200000-0003-2556-1712>

Diana Rosa Prieto-del Río^{III}

 <https://orcid.org/0000-0002-6122-509X>

Julio Quial-Sotolongo^{IV}

 <https://orcid.org/200000-0003-2444-2842>

^I Universidad Tecnológica de la Habana, CUJAE. La Habana, Cuba
correo electrónico: mavis@ind.cujae.edu.cu, tdelgado@ind.cujae.edu.cu

^{II} Escuela Superior de Cuadros del Estado y del Gobierno, Cuba. La Habana, Cuba
correo electrónico: mercedes@esceg.cu

^{III} Empresa de Informática y Automatización para la Construcción, AICROS. La Habana, Cuba
correo electrónico: drprieto@aicros.cu

^{IV} Empresa de Mensajería y Cambio Internacional, EMCI. La Habana. La Habana, Cuba
correo electrónico: jquial@mci.correos.cu

Recibido: 17 de septiembre del 2019.

Aprobado: 28 de octubre del 2019.

RESUMEN

El enfoque de datos enlazados resulta particularmente novedoso para la integración de datos en el contexto empresarial, al incorporar los beneficios derivados del empleo de estándares de la web semántica y los principios de los datos enlazados en el sistema de información de la empresa. En este artículo se aplica dicho enfoque para incidir en la mejora de la gestión de información, orientada a solucionar la carencia de interoperabilidad entre las diversas fuentes de datos que se involucran durante el proceso de planificación de los recursos de la planta de operaciones en la Empresa de Mensajería y Cambio Internacional. Se presenta un procedimiento de Datos Empresariales Enlazados, basado en aspectos característicos del contexto empresarial, siendo esta la principal distinción del método propuesto, respecto a otros consultados en la literatura. La viabilidad del procedimiento diseñado se muestra con su aplicación al proceso de prealertas en la empresa estudiada.

Palabras Clave: integración de información, datos enlazados, datos empresariales enlazados, Web Semántica.

ABSTRACT

The Linked Data approach is particularly novel to integrate data in the enterprise settings. When incorporating, the benefits stem from the use of semantic web standards and the data linked principles at the enterprise information systems. In the article, this approach was applied to influence the improvement of information management, aiming to solve the lack of interoperability among the various data sources involved in the resource planning process of the operation plant in La Empresa de Mensajería y Cambio Internacional. A tailored Linked Enterprise Data procedure to the enterprise features was presented, being this its primary distinctive attribute, compared to those consulted in the literature. The procedure viability demonstrated through its application to the pre-warnings process of the studied enterprise.

Keywords: information integration, Linked Data, Linked Enterprise Data, Semantic Web.

I. INTRODUCCIÓN

En el contexto de integración de datos, ha surgido en la última década, el enfoque de Datos Empresariales Enlazados (LED por sus siglas en inglés) o “*Linked Data in Business*” como lo nombran unos pocos [1, 2, 3, 4, 5]. LED, derivación para la empresa del término *Linked Data* acuñado por Berners Lee, es un enfoque novedoso para la integración de datos en el contexto empresarial, que integra los datos con independencia de su procedencia, para crear un espacio de información global y unificado [4, 6, 7]. Al igual que la web tradicional enlaza páginas web, LED enlaza los datos de la empresa en el nivel más detallado, considerando su significado y contenido [9, 10].

LED refiere un marco para incorporar los beneficios de la web semántica en el sistema de información de la empresa [11, 12, 13, 14]. Específicamente utiliza las herramientas y técnicas de la Web Semántica para conectar, publicar y usar datos de la empresa; y publica los datos, siguiendo los principios de los Datos Enlazados o *Linked Data* [1, 8, 13]. El “backbone” técnico de este enfoque, lo constituyen las tecnologías URI (*Uniform Resource Identifiers*), RDF (*Resource Description Framework*)¹, Ontologías, *Linked Data* y SPARQL (*SPARQL Protocol y RDF Query Language*)².

El URI proporciona una forma genérica de identificar algo en la web [14]. El modelo RDF es un modelo de datos para la web que permite expresar información sobre recursos que ha sido adoptado como modelo de intercambio de datos para la web semántica y facilita la tarea de integrar datos procedentes de múltiples fuentes [14, 15]. Las ontologías pueden captar eficazmente la semántica de los datos, así como integrar datos de forma eficiente [16]. *Linked Data* crea enlaces explícitos entre los conjuntos de datos y proporciona un conjunto de buenas prácticas para la publicación y vinculación de los datos estructurados en la web, conocido como los principios de linked data [6, 15, 17, 18]. SPARQL es un lenguaje estandarizado para la consulta de grafos RDF [17].

LED proporciona beneficios relacionados con el acceso y utilización de los datos en el contexto empresarial [1]. LED aporta una manera de utilizar los datos de forma eficiente y entre sus beneficios está la integración de aplicaciones existentes en el nivel de datos, que no se logra fácilmente en enfoques anteriores [19, 20]. Permite crear de forma ágil, nuevas aplicaciones que utilicen los datos existentes sin alterar los modelos de datos residentes en las aplicaciones que funcionan [21]. Con LED se mejora la interoperabilidad de los datos y el valor de cada información mejora por sus enlaces a datos complementarios, con lo cual se potencia la toma de decisiones bien fundamentadas [22, 23].

Otros beneficios de utilizar este enfoque en la integración de los datos en la empresa son:

- Define de forma precisa la semántica contenida en los datos empresariales [12]

¹ <https://www.w3.org/RDF/>

² <https://www.w3.org/TR/rdf-sparql-query/>

ENFOQUE DE DATOS EMPRESARIALES ENLAZADOS APLICADO EN UNA EMPRESA CUBANA

- Facilita la creación de taxonomías empresariales con estándares como SKOS³ para la estructuración de grandes cantidades de documentos que se producen diariamente como correos electrónicos, directivas empresariales, entre otros [24]
- Mejora la búsqueda de información al interior de la empresa, con resultados más precisos en relación al tema buscado [24]
- Facilita el acceso a los datos vía web [24].
- Facilita la integración de los datos estructurados con datos no estructurados y semiestructurados [22].
- Permite la creación de nuevas informaciones a partir inferencias utilizando los datos existentes [22].

Todos los beneficios anteriormente descritos, tributan a una mejor gestión de información en las empresas y contribuyen a una toma de decisiones más rápida y con más fundamentación, tributando entonces a una mejor adaptabilidad y sostenibilidad de las empresas en el mercado que compiten.

La problemática de integración de datos está presente en la Empresa de Mensajería y Cambio Internacional (EMCI), siendo esta, uno de los principales problemas detectados en un diagnóstico reciente sobre gestión de información [25].

La EMCI integrada al Grupo Empresarial Correos de Cuba, brinda servicios de intercambio y tratamiento internacional de correspondencia y encomiendas postales del servicio postal universal; así como servicios de mensajería y paquetería. La actividad postal constituye una de sus principales actividades y genera diariamente un gran volumen de información asociado al manejo de las operaciones relacionadas con los conceptos de Operaciones Postales, Telegráfica, y de Prensa. El Grupo Empresarial Correos de Cuba ha desarrollado algunas soluciones que responden a la evolución de los requerimientos informacionales de la organización; sin embargo, existen algunas carencias informacionales, como es el caso de la información para la planificación de los recursos [26].

La Dirección de Operaciones de la EMCI (DOEMCI) recibe diariamente un volumen considerable de información adelantada, procedente de las 36 agencias internacionales de viaje con las que se tienen contratos, referente a los bultos que serán transportados en el próximo período. Las prealertas o Manifiesto, como se nombra esta información, constituyen la base para planificar los recursos en la planta de operaciones.

Las prealertas se reciben en la empresa de forma dispersa, con variados formatos y con variación en el contenido, según la agencia de viaje que la envíe. La empresa no cuenta con ninguna herramienta que procese de forma integrada los datos contenidos en las prealertas con la consecuencia negativa que no se puede prever para un período de tiempo cuál sería el volumen de operación para la planta y por tanto no se planifican proactivamente los recursos para las operaciones. Esta situación trae como consecuencia que la planta de procesamiento se sobresature en algunas oportunidades, siendo esta una de las causas que no se cumplan los tiempos de entrega para los envíos, afectando finalmente los indicadores de calidad establecidos para esta actividad.

Considerando las ventajas del enfoque LED para lograr una efectiva integración de los datos y que a pesar de sus beneficios su aplicación en las empresas todavía es limitado existiendo autores que lo declaran como una idea abstracta o ideal [7, 16, 27]; así como las dificultades de integración de prealertas existentes en la DOEMCI, se declara el objetivo del artículo como aplicar el enfoque LED en la EMCI para integrar los datos de las prealertas. En el método se expone el Procedimiento empleado para la aplicación del enfoque LED y en los resultados y la discusión se describen los pasos que se han llevado a cabo en dicha empresa, lo que demuestra la aplicabilidad y beneficios que genera.

³ Simple Knowledge Organization System; Sistema Simple de Organización de Conocimiento

II. MÉTODOS

Para aplicar el enfoque de los Datos Empresariales Enlazados en la EMCI se utiliza un procedimiento diseñado para tales efectos, que cumple de forma general con el ciclo de vida de LD, incluyendo las actividades que de forma regular se ejecutan en este ciclo y se le han incorporado aspectos propios del contexto empresarial, tales como [28, 29, 30]:

1. Estar orientado a procesos desde el inicio: A diferencia de la otras metodologías de Datos Enlazados que suelen estar basados en el análisis de los datos en sí mismos o el conocimiento del dominio que tengan las comunidades, enlazar datos en una empresa debe seguir un riguroso enfoque orientado a procesos[1]. Dicho enfoque debe garantizar la alineación de la capa de información a los objetivos estratégicos del negocio; así como, entre los intereses de las personas con las políticas organizacionales.
2. Considerar tanto datos internos de la organización como externos (accesibles desde la web).
3. Aunque se interpreta con frecuencia, en etapas tempranas de este proceso, que enlazar datos empresariales implica a los recursos informacionales internos solamente, la propia capacidad del enfoque enlazado posibilita enriquecer los datos que se generan en la empresa con otros datos del cliente, el proveedor o el contexto donde se desenvuelve la misma. Con esto mejorar la capacidad del proceso de decisión basado en datos. En estos entornos se recomienda usar la anotación semántica para enlazar otros recursos de datos que complementan los datos estructurados en forma relacional en la empresa. Esos otros datos (externos) pueden estar públicos como datos enlazados (ej. el proyecto *Linked Open Data*) o en múltiples fuentes abiertas de la Web (ej. la información proveniente de redes sociales).
4. Aprovechar enfoques clásicos de integración de datos a nivel empresarial. Sin alejarse de los principios de los Datos Enlazados, este requerimiento busca utilizar los enfoques que proporcionan los sistemas de información ejecutivos de la empresa para simplificar la tarea de crear redes de ontologías o enlazar los RDF generados. Un ejemplo se puede ilustrar con la potencialidad que pudiera ofrecer un ETL⁴ Semántico al enlazado de datos empresariales, un concepto heredado de la Inteligencia de Negocio.

Estructuralmente, el procedimiento consta de cinco etapas como se muestran en la figura 1.

⁴ Extract-Transform-Load (ETL), como enfoque para la integración de datos, según 34. Chakraborty J, Padki A, Bansal SK. Semantic ETL – State-of-the-art and open research challenges. IEEE 11th International Conference on Semantic Computing. 2017. ISSN 00002017.

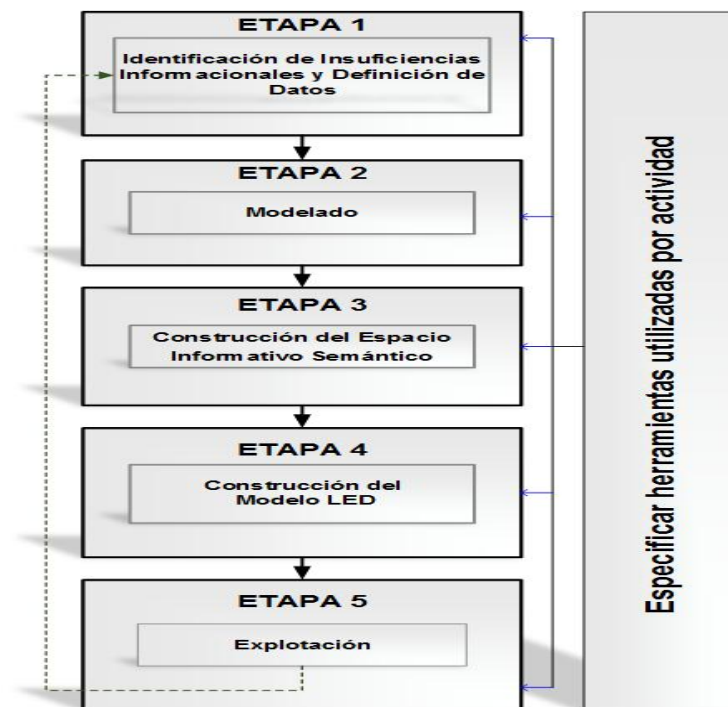


Fig. 1. Procedimiento LED para la aplicación en la EMCI

Las etapas están integradas por diferentes actividades, que representan los pasos que deben realizarse en el proceso de integración. A continuación, se describe de forma general el objetivo de cada etapa y se precisan algunas de las herramientas que pudieran utilizarse.

Etapa 1: Identificación de Insuficiencias Informacionales y Definición de Datos

La primera etapa del procedimiento tiene como objetivo precisar carencias informacionales que limitan el funcionamiento de uno o varios procesos de la empresa, así como especificar las fuentes de datos que deben integrarse para dar solución a la problemática de integración. Se debe precisar qué informaciones no se obtienen en uno o varios procesos de la empresa y que para su obtención se deba consultar datos dispersos en las diferentes fuentes existentes. Se define el para qué de la integración, siempre considerando mejorar la operatividad de los procesos. La etapa también identifica y selecciona las fuentes de datos que deben ser integradas para dar solución a las carencias detectadas. Se trata del con qué de la integración, a partir de considerar sistemas informáticos y fuentes de datos en general, que se utilicen en el funcionamiento de los procesos. Las herramientas que se pueden utilizar en la etapa son: entrevistas, análisis de documentos, análisis del mapa de procesos, inventario de los sistemas de información, Modelaciones BPM (*Business Process Management*), diagnósticos de gestión de información efectuados en la organización, Modelo de Negocios, entre otras.

Etapa 2: Modelado

Se modela la semántica contenida en las fuentes de datos, partiendo primero de modelar el dominio de actuación y su posterior representación ontológica. Da como resultado un modelo ontológico a utilizar. Se incluye el mapeo lógico entre las fuentes de datos identificadas en la primera etapa y el modelo ontológico resultante; así como el diseño de los URI para identificar los recursos en la web. En la etapa se deben utilizar herramientas para el desarrollo y edición de ontologías, así como herramientas de modelación de información.

Etapa 3: Construcción del Espacio Informativo Semántico

El espacio informativo semántico es la representación en el modelo RDF de cada fuente de datos especificada en la etapa 1, definir la semántica contenida en los datos de las diferentes fuentes. El objetivo de la etapa es transformar al modelo RDF cada fuente de datos involucrada. Tal como ocurre en un ETL semántico, que se incorpora las tecnologías semánticas en la fase de

transformación del proceso ETL [30]. En esta etapa se toma como base el modelo ontológico definido en la etapa anterior, para transformar cada fuente de datos a su correspondiente modelo RDF, haciendo corresponder las tablas y los campos de las fuentes de datos, con las clases y propiedades del modelo ontológico. Inicialmente, en la etapa se analizan los formatos y la información contenida en las fuentes de datos, para preparar tanto como sea posible los conjuntos de datos (eliminar inconsistencias; completar datos, extraer datos, eliminar duplicados, etc) previo a la transformación al modelo RDF. Luego con la herramienta de transformación que se utilice se ejecuta la transformación. En esta etapa se deben utilizar las herramientas que permitan transformar o preparar, de ser necesario previo a su transformación, los conjuntos de datos; así como las herramientas para la Transformación a RDF de las fuentes de datos.

Etapa 4: Construcción del Modelo LED

Esta etapa tiene como objetivo enlazar los modelos RDF obtenidos en la etapa anterior, teniendo en cuenta las necesidades informacionales que se definieron en la primera etapa de este procedimiento, obteniendo un conjunto de grafos RDF enlazados, nombrado en la literatura sobre el tema como "Dataset" [26], [27], [28]. En esta etapa, debe considerarse el enlace con un modelo LOD (Linked Open Data) exterior a la empresa. Por último, se almacena y se publica el *Dataset*, en un almacén de tripletas RDF o "triple store" para su posterior consumo y/ explotación [31], [26], [19]. Con esta etapa quedan integrados los datos en la empresa. Se utilizan herramientas para enlazar los modelos RDF y herramientas para almacenar y publicar el modelo RDF enlazado.

Etapa 5: Explotación

Esta es la etapa con el objetivo de utilización o consumo de la capa de datos enlazados para proporcionar fácil acceso a la información integrada y potenciar la toma de decisiones en la empresa, desarrollando por ejemplo aplicaciones Web. 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, tal como se refleja en [7]. Se utilizan herramientas como *frameworks* de desarrollo de software, CMS semánticos, API REST.

III. RESULTADOS

La aplicación de procedimiento se enmarca dentro del contexto de los envíos postales que son gestionados en la Dirección de Operaciones de la EMCI (DOEMCI) y tiene como objetivo integrar la información de las prealertas que se reciben de las diferentes agencias de viaje, para obtener datos relevantes que contribuyan al proceso de planificación de las operaciones de la planta de procesamiento.

Etapa 1: Identificación de Insuficiencias Informacionales y Definición de Datos

La DOEMCI recibe diariamente un volumen considerable de información adelantada referente a los bultos que serán transportados en el próximo período. Las prealertas o manifiesto, como se nombra esta información, se envían por las diferentes agencias de viaje y contiene la individualización de cada una de las Guías Courier que transporta un mensajero internacional, sea por vía aérea o terrestre, mediante el cual las mercancías se presentan y entregan a la Aduana. Las Guías Courier contienen información sobre la cantidad de bultos que contiene el envío y su peso en kilogramo; la agencia que lo envía, la línea aérea que lo transporta, el número de vuelo y el país origen; de cada envío en particular contiene un número identificativo, su peso individual, fecha, datos del remitente como nombre y número de identidad; y datos del destinatario como nombre, carnet de identidad, dirección, provincia y municipio.

Las prealertas son enviadas por: los operadores internacionales, la DOEMCI y la dirección de la planta. A partir del recibo de las prealertas pueden realizar la planificación de los medios y recursos materiales y humanos (transporte, operarios, combustible, servicios aduanales y transitarios, entre otros) que garanticen en tiempo y forma el desarrollo exitoso de las operaciones. Se tiene en cuenta que los operadores postales deben cumplimentar con indicadores relacionados con el servicio de entrega de los productos.

A partir de un estudio informacional en la DOEMCI no existe ninguna herramienta que brinde una visión unificada de todas las prealertas. En la empresa, esta información se procesa de forma separada por distintos especialistas, y no se procesa de forma integrada. Además, que según sea la

ENFOQUE DE DATOS EMPRESARIALES ENLAZADOS APLICADO EN UNA EMPRESA CUBANA

agencia que envíe las prealertas, varía el formato y el conjunto de información que contienen. En específico se carece de datos relacionados con:

- Resúmenes de tipo de carga
- Resúmenes de vía de transportación (área, marítima)
- Total de envíos para un periodo
- Total de Envíos para una provincia
- Volumen total de operación
- Volumen total de operación por provincia
- Peso total para un periodo

Lo anterior, deriva en una consecuencia negativa para la empresa dado que no se puede prever para un período de tiempo cuál sería su volumen de operación y por tanto no planifica adecuadamente los recursos para sus operaciones. Esta situación trae como consecuencia que la planta de procesamiento se sobreesature, siendo esta una de las causas que no se cumplan los tiempos de entrega para los envíos.

Los problemas con la integración de la información de las prealertas, ocasionan carencias informacionales que limitan el proceso de planificación de los recursos para la planta de operaciones. La tabla 1 muestra las fuentes de datos de interés para el proceso de integración de información, así como el formato en el cual las mismas están disponibles.

Tabla 1. Fuentes de datos de interés para la integración de información

Fuente de datos	Procedencia	Formato
Prealertas de COURIER	Agencias Courier	xls
Prealertas de DHL	Agencia DHL	xml

Etapa 2: Modelado

En esta etapa se debe modelar información relacionada con las prealertas. Las prealertas, como se menciona anteriormente en general contiene información sobre los envíos, sus destinatarios y remitentes. La figura 2 refleja el modelo de dominio correspondiente a la situación anterior:

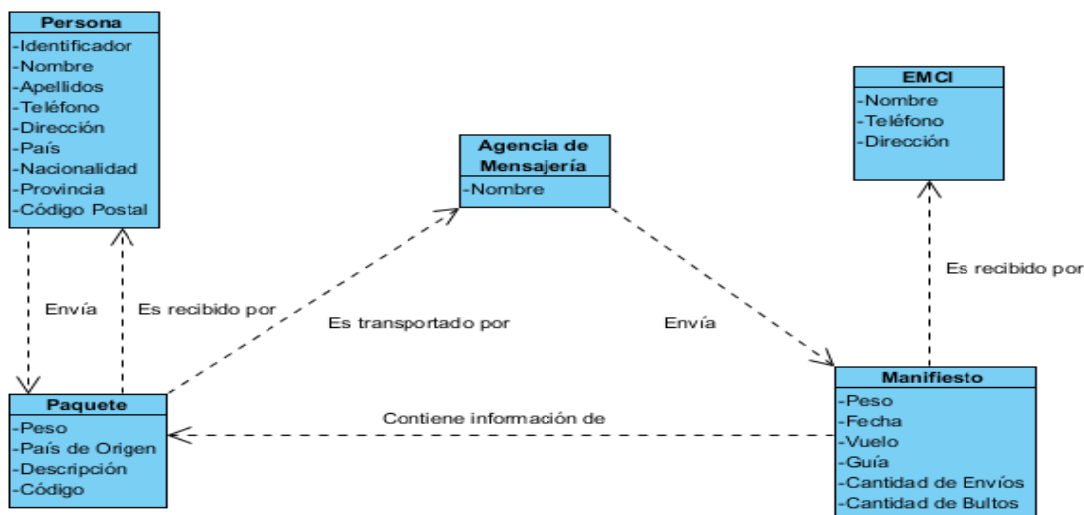


Fig. 2. Modelo de Dominio

Tomando en consideración el modelo de dominio, se crea una red de ontologías, a partir de la reutilización de clases y propiedades de los vocabularios *foaf* y *vcard* y de la creación de una ontología propia para estos efectos, la ontología *emciop*.

Del vocabulario FOAF5, se utilizan clases y propiedades en la modelación de personas que representan a los remitentes y destinatarios. Del vocabulario VCARD6, se utilizan las propiedades para completar la modelación de los datos de las direcciones de remitentes y destinatarios. La Tabla 2 muestra las clases y propiedades que se reutilizan de cada vocabulario.

Tabla 2. Clases y propiedades que se reutilizan de cada vocabulario

Vocabulario	Término	Tipo	Descripción
foaf	<i>Person</i>	Clase	Identifica a una persona
	<i>name</i>	Propiedad	Un nombre para alguna cosa
	<i>familyName</i>	Propiedad	Identifica los apellidos de una persona
vcard	<i>has_key</i>	Propiedad	Identificación de una persona
	<i>region</i>	Propiedad	Identifica la provincia de una persona
	<i>country_name</i>	Propiedad	Identifica el nombre de un país
	<i>has_postal_code</i>	Propiedad	Identifica el código postal
	<i>has_telephone</i>	Propiedad	Identifica el teléfono

Las clases propias de la ontología creada permiten modelar lo relacionado con paquete y manifiesto. El modelo ontológico resultante se muestra en la figura 3.

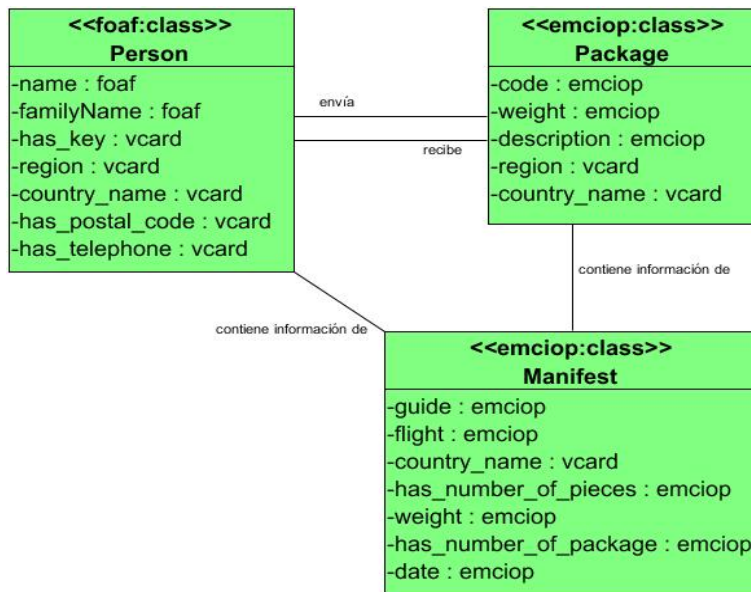


Fig. 3. Modelo Ontológico Utilizado

El resultado de la actividad relacionada con el mapeo lógico entre la fuente de datos XML y las clases del modelo ontológico utilizado según se muestra en la tabla 3. Una tabla similar se realizó para el mapeo de la fuente de datos XLS.

⁵ <https://es.wikipedia.org/wiki/FOAF>: es una ontología legible para las máquinas que describe a las personas, sus actividades y sus relaciones con otras personas y objetos.

⁶ <https://es.wikipedia.org/wiki/VCard>: es un formato estándar para el intercambio de información personal. Pueden contener nombre, dirección, números telefónicos, entre otras informaciones.

ENFOQUE DE DATOS EMPRESARIALES ENLAZADOS APLICADO EN UNA EMPRESA CUBANA

Tabla 3. Diseño lógico del mapeo Fuentes de datos Vs Ontologías para la fuente de datos en XML

Fuente de datos	Tipo de entidad en la Fuente de datos	Ontología/Vocabulario	Término	Tipo de entidad en la ontología	Descripción
Pack	Tabla	emciop	Package	Clase	Identifica a un paquete
AirwayBillNumber	Columna	emciop	code	Propiedad	Identifica el Número de guía.
ShipmentOriginCountry	Columna	vcard	country_name	Propiedad	Identifica el nombre de un país
ShipmentWeight	Columna	emciop	weight	Propiedad	Identifica el peso de un paquete
ShipperName	Columna	foaf	name	Propiedad	Un nombre para alguna cosa
ShipperPostCode	Columna	vcard	has_postal_code	Propiedad	Identifica el código postal
ShipperProvince	Columna	vcard	region	Propiedad	Identifica la provincia de una persona
ShipperPhoneNumber	Columna	vcard	has_telephone	Propiedad	Identifica el teléfono
ConsigneeName	Columna	foaf	name	Propiedad	Un nombre para alguna cosa
ConsigneePostCode	Columna	vcard	has_postal_code	Propiedad	Identifica el código postal
ConsigneeProvince	Columna	vcard	region	Propiedad	Identifica la provincia de una persona
ConsigneePhoneNumber	Columna	vcard	has_telephone	Propiedad	Identifica el teléfono

En relación al diseño de los URI para identificar las prealertas en la red, se sigue un patrón que combina un prefijo común <http://emci.correos.cu/> con las clases y propiedades del modelo ontológico, tal como se muestra en la Tabla 4.

Tabla 4. Patrón URI

Descripción	Patrón	Ejemplo
Vocabulario	http://{uri-base}v/{project-name}/schema#	http://emci.correos.cu/operaciones/schema#
Clases	http://{uri-base}/{project-name}/schema#{class}	http://emci.correos.cu/operaciones/schema#Manifest
Propiedades	http://{uri-base}/{project-name}/schema#{property}	http://emci.correos.cu/operaciones/schema#has_number_of_package

Etapa 3: Construcción del Espacio Informativo Semántico

En esta etapa se transforman al formato RDF las fuentes de datos de las prealertas en formato XLS y las prealertas en formato XML. Dada la variedad en el formato, se decide primero transformar ambas fuentes a un formato relacional, para lo cual se utiliza como administrador de bases de datos Navicat 7. Se extraen en un fichero sql, sólo los campos que resultan de interés para el caso que se ejecuta.

⁷ <https://www.navicat.com/es/>

Para la transformación se utiliza la plataforma D2RQ ⁸, herramienta que exporta el contenido de una base de datos relacional a un grafo RDF [18]. Para cada fuente de dato, se ejecuta el flujo de trabajo según se muestra en la figura 4, obteniéndose el correspondiente grafo RDF de cada fuente y dando lugar al Espacio Informativo Semántico.

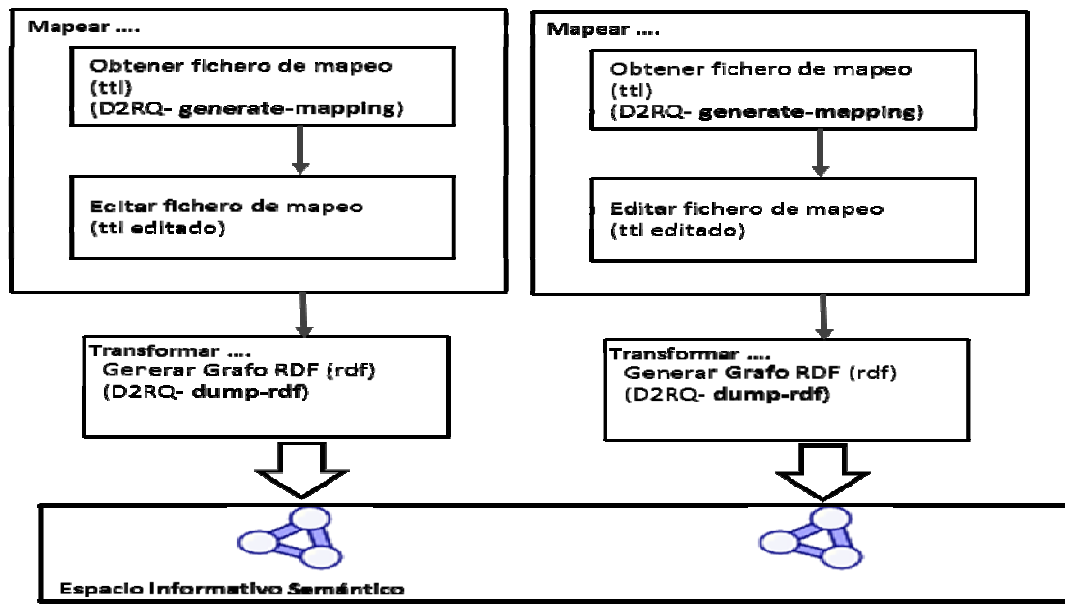


Fig. 4. Flujo de trabajo para obtener el grafo RDF

Etapa 4: Construcción del Modelo LED

La actividad de enlazado se realiza con la herramienta Silk (Link Discovery Framework)⁹, herramienta para el descubrimiento de enlaces entre grafos RDF [18]. Las relaciones seleccionadas para el enriquecimiento de los datos son: vcard: region, vcard: country_name, rdf:type. En la tabla 5, se presentan algunas de las triplas enlazadas y sus relaciones.

Tabla 5. Ejemplo de Tripletas enlazadas

<http://localhost:2020/resource/pack/9>	vcard:region	<http://localhost:2020/resource/paquete/CG910083211PA>
<http://localhost:2020/resource/pack/2>	vcard:country_name	<http://localhost:2020/resource/paquete/CG910085212PA>
<http://localhost:2020/resource/pack>	rdf:type	<http://localhost:2020/resource/paquete>

La publicación del conjunto de datos resultante se publica con el almacén de triplas Apache Jena Fuseki, que además proporciona un *endpoint SPARQL*, desde donde se pueden realizar consultas sobre las triplas almacenadas.

En esta aplicación del enfoque *LED* no se enriquece el modelo *LED* con datos externos, quedando esto como recomendación para próximas aplicaciones.

Etapa 5: Explotación

⁸ <http://d2rq.org/>

⁹ <https://app.assembla.com/wiki/show/silk>

Para el uso y exploración de la capa de datos enlazados se ha desarrollado una aplicación web, con el objetivo de mostrar de forma amigable, información adelantada sobre los envíos que deben recibirse por la empresa en un determinado período de tiempo. Las funcionalidades de una primera versión de la aplicación, de forma general son:

- Datos informativos generales de la EMCI
- Nombres y datos informativos de las Agencias u Operadores Internacionales
- Volumen de operación Total (cantidad de bultos; cantidad de envíos; peso en kg) de un período determinado
- Listar las provincias que deben recibir envíos en el período de un período determinado
- Navegar por los Envíos de una Provincia de un período determinado
- Volúmenes de operación previstos por Provincia.

IV. DISCUSIÓN

La aplicación del enfoque LED, a través del procedimiento, ha permitido integrar de forma factible la información de las prealertas procedente de orígenes distintos y con distintos formatos. Posibilitando obtener una herramienta informática que ofrece datos globales sobre los volúmenes de operación para un período determinado y facilita por tanto el proceso de planificación de los recursos para las operaciones en la planta. La mejor planificación de los recursos impacta positivamente en el cliente final de los servicios de mensajería y paquetería dado la mejoría que se obtiene en los tiempos de entrega de los envíos.

Para la EMCI, el enfoque LED abre el uso de las tecnologías semánticas en la empresa, lo cual incide en la problemática de integración de información, sin afectar el despliegue actual de aplicaciones y tecnologías informáticas en uso por la misma. Ello ha sido posible con la generación de una capa independiente de datos enlazados de prealertas provenientes de diferentes fuentes y con distintos formatos, quedando disponibles dichos datos para ser consumidos por otras aplicaciones, lo cual posibilita la obtención de nuevas informaciones para mejorar el proceso de toma de decisiones.

En general se han tenido que abordar algunas barreras que limitaban su aplicación, tales como:

Escasa cultura para acceder a los modelos de datos subyacentes a las aplicaciones informáticas.

Poca disponibilidad de conjuntos de datos externos relevantes al contexto empresarial para enlazar y enriquecer el modelo LED interno de la organización.

Es preciso facilitar una adecuada gestión de conocimientos hacia el cliente interno sobre las herramientas y tecnologías que son esenciales en el procedimiento para que el procedimiento sea gestionado habitualmente en la empresa. Con la finalidad de habilitar este objetivo en un futuro cercano, se cuenta con un grupo de investigación de la empresa que ha estado inmerso en las bases conceptuales y prácticas que sustentan el procedimiento. Dicho grupo tiene la responsabilidad de desplegar las herramientas informáticas asociadas, así como ofrecer las capacitaciones en los niveles requeridos.

V. CONCLUSIONES

1. LED es un enfoque novedoso y actual para la integración de datos e información en el contexto empresarial. Su aplicación en la EMCI, ha permitido la integración de los datos de las prealertas, proporcionando datos globales para el proceso de planificación de las operaciones en la planta. Para lo cual se diseñó un procedimiento, que transita desde la identificación de las insuficiencias informacionales, el modelado, la construcción del espacio informativo semántico y del modelo LED, hasta la explotación.
2. La aplicación desarrollada proporciona a la empresa una nueva capacidad en relación a la gestión de información: analizar información adelantada para realizar una mejor planificación de los recursos según el volumen de operación que se prevé. Esta nueva capacidad, tributa directamente a mejorar los tiempos de entrega de los envíos, cumplimentando con los indicadores de calidad establecidos para la actividad postal. Se pronostica la aplicación del enfoque LED en otras problemáticas de integración de la empresa, como por ejemplo la integración de datos relacionados con los operadores internacionales (contratos, facturación, etc).

3. El procedimiento puede ser aplicado en otras empresas que demanden también la integración de fuentes de datos heterogéneas. Como recomendación para trabajos futuros, se considera el tema de integrar los datos internos de la empresa con datos públicos en ese dominio de actuación. 🏠

VI. REFERENCIAS

1. Hu B, Svensson G. A Case Study of Linked Enterprise Data. In: 9th International Semantic Web Conference ISWC 2010; Shanghai, China: Springer-Verlag Berlin Heidelberg; 2010. p. ISBN 978-3-642-17746-0
2. Weichselbraun A, Streiff D, Scharl A. Linked enterprise data for fine grained named entity linking and web intelligence. In: 4th International Conference on Web Intelligence, Mining and Semantics (WIMS'14); Thessaloniki, Greece: ACM; 2014. p. ISBN 978-1-4503-2538-7
3. Afonso Pinto V, Silva Parreiras F. Enterprise Linked Data: A Systematic Mapping Study. In: 11th International Workshop on Web Information Systems Modeling (WISM14); Springer; 2014. p. 253-62. ISBN 978-3-319-12255-7
4. Gürdür D, Khoury JE, Nyberg M. Methodology for Linked Enterprise Data Quality Assessment Through Information Visualizations. Journal of Industrial Information Integration Elsevier BV. 2018 November. ISSN 2452414X.
5. Abramowicz W, Auer S, Heath T. Linked Data in Business. Business & Information Systems Engineering Springer 2016;58 (5):323-6. ISSN 1867-0202.
6. Berners-Lee T. Linked Data - Design Issues. W3C; 2006. [Citado Diciembre 2015]. Disponible en: <https://www.w3.org/DesignIssues/LinkedData.html>
7. Stuart ML, Prieto DR, Delgado T, Delgado M. Enfoque de Integración Basado en Datos Enlazados Empresariales. Revista Cubana de Administración Pública y Empresarial. 2018;Vol II(No 3 , septiembre-diciembre):pg. 268-79. ISSN 2664-0856.
8. Galkin M, Auer S, Scerri S. Enterprise Knowledge Graphs: A Backbone of Linked Enterprise Data. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI); Omaha, NE, USA: IEEE Computer Society; 2016. p. 497-502. ISBN 978-1-5090-4470-2
9. Galkin M, Auer S, Vidal ME, Scerri S. Enterprise Knowledge Graphs: A Semantic Approach for Knowledge Management in the Next Generation of Enterprise Information Systems. In: 19th International Conference on Enterprise Information Systems (ICEIS2017); Porto, Portugal: Springer International Publishing; 2017. p. 88-98. ISBN 978-3-319-93375-7
10. Silveira Gomes M, Visintin L, Freiberger Pereira L, Gauthier FAO, Macedo M. Uma Revisão de Literatura Acerca de Linked Enterprise Data. Anais do VII Congresso Internacional de Conhecimento e Inovação (ciKi) 2017;Vol 1 No 1 ISSN 2318-5376.
11. Galkin M. Strategies for Managing Linked Enterprise Data [Dissertation zur Erlangung des Doktorgrades.]. Mathematisch-Naturwissenschaftlichen Fakultät: RheinischenFriedrich-Wilhelms-Universität Bonn; 2018.
12. Nesrine L, Mimoun M, Ahmed L, Hamou RM. On Demand ETL of RDB to RDF Mapping for Linked Enterprise Data. International Journal of Strategic Information Technology and Applications (IJSITA). 2017;Volume 8 (Issue 3). ISSN 1947-3095.
13. Konstantinou N, Spanos D-E. Materializing the Web of Linked Data: Springer International Publishing Switzerland; 2015. ISBN 978-3-319-16073-3.
14. Ma L, Mei J, Pan Y, Kulkarni K, Fokoue A, Ranganathan A. Semantic Web Technologies and Data Management. W3C Workshop on RDF Access to Relational Databases 2007.
15. Bizer C, Heath T, Berners-Lee T. Linked Data - The Story so Far. En: Semantic services, interoperability and web applications: emerging concepts. 2011. p. 205-27. ISBN 9781609605933.
16. Heath T, Bizer C. Linked Data Evolving the Web into a Global Data Space. Claypool M, editor: Morgan & Claypool; 2011. ISBN 9781608454303.
17. Mihindikulasooriya N, García R, Esteban M. Linked Data Platform as a novel approach for Enterprise Application Integration. In: Fourth International Workshop on Consuming Linked Data (COLD2013) co-located with the 12th International Semantic Web Conference (ISWC 2013); 2013. p. ISBN 1613-0073

18. Mynarz J. Integration of public procurement data using Linked Data. Journal of Systems Integration. 2014;Vol 5 (4):pg. 19-31. ISSN 1804-2724.
19. Rao SS, Nayak A. LinkED: A Novel Methodology for Publishing Linked Enterprise Data. Journal of Computing and Information Technology. 2017 2017-09-00.25(No.3):pg. 191-209. ISSN 1846-3908.
20. Antidot. Linked Enterprise Data: leveraging the Semantic Web stack in a corporate IS environment. In: 11th International Semantic Web Conference, ISWC 2012 Boston, MA, USA: Springer-Verlag Berlin Heidelberg; 2012. p. ISBN 978-3-642-35173-0
21. FUJITSU. Linked Data, Connecting and Exploiting big data 2012. [Citado Diciembre 2015]. 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)
22. Hladky D, Maltseva SV. Linked Data Paradigm for Enterprises: Information Integration and Value Chain. БИЗНЕС-ИНФОРМАТИКА 2013 junio.No 2 (24). ISSN 2587-8166.
23. Pérez SJ, Novoa B. Diagnóstico de Gestión de Información del Servicio Postal Universal de la Empresa de Mensajería y Cambio Internacional con Enfoque de Arquitectura de Información Empresarial (AIE). [Trabajo de Diploma para optar por el título de Ingeniería Industrial]: Universidad Tecnológica de la Habana. Facultad de Ingeniería Industrial. CUJAE; 2018.
24. Fleites A. Sistema integrado postal para la gestión de los servicios del Grupo Empresarial Correos de Cuba. In: III Conferencia Internacional en Ciencias Computacionales e Informáticas (CICCI) Informática 2016; 2016. p. 9. ISBN 978-959-289-122-7
25. Moura Rd, Priscila E, Neiva MA, Macedo M, A. OF. Avaliação da Produção Científica sobre Enterprise Linked Data. In: LOD Brazil; Florianópolis, Brazil: 2014. p. ISBN 978-85-61115-09-8
26. Villazón B, Vilches LM, Corcho O, Gómez A. Methodological Guidelines for Publishing Government Linked Data. En: Linking Government Data. New York, NY: Springer New York; 2011. p. 27-49. ISBN 978-1-4614-1766-8 978-1-4614-1767-5.
27. Jovanovik M. Linked Data Application Development Methodology [Tesis de Doctorado]. Faculty of Computer Science and Engineering: Ss. Cyril and Methodius University; 2016.
28. Espinoza Arias P, García Delgado M, Corcho O, Vivas White P, Potti Manjavacas H. A sustainable process and toolbox for geographical linked data generation and publication: a case study with BTN100. Open Geospatial Data, Software and Standards. 2019; (2019) 4:2:8. ISSN 2363-7501.
29. Chakraborty J, Padki A, Bansal SK. Semantic ETL – State-of-the-art and open research challenges. IEEE 11th International Conference on Semantic Computing. 2017. ISSN 00002017.
30. Bansal SK. Towards a Semantic Extract-Transform-Load (ETL) framework for Big Data Integration 2014 IEEE International Congress on Big Data. 2014. ISSN 00002014.
31. Wood D, editor. Linking Enterprise Data: Springer; 2010. ISBN 978-1-4419-7664-2.

Los autores declaran que no hay conflictos de intereses de ningún tipo

Contribución de cada autor

Mavis Lis Stuart Cárdenas: diseño de la investigación, aplicación de la propuesta, análisis de los resultados y elaboración del documento.

Tatiana Delgado Fernández: diseño de la investigación, análisis de los resultados y revisión de artículo.

Mercedes Delgado Fernández: análisis de los resultados y revisión de artículo.

Diana Rosa Prieto del Río: diseño de la investigación, aplicación de la propuesta, análisis de los resultados y revisión de artículo.

Julio Quial Sotolongo: aplicación y validación de la propuesta y revisión de artículo.