

Algunas reflexiones sobre los datos abiertos enlazados en Cuba

Some Reflections on Linked Open Data in Cuba

Yusniel Hidalgo-Delgado,¹ Alejandro Jesús Mariño-Molerio,¹ Yarina Amoroso-Fernández,¹ Amed Abel Leiva-Mederos²

¹ Universidad de las Ciencias Informáticas. La Habana, Cuba.

² Universidad Central "Marta Abreu", de Las Villas. Cuba.

RESUMEN

Cuba está realizando una introducción intensiva de las tecnologías de la información y la comunicación en la sociedad, con el desarrollo y el despliegue de sistemas de información para aumentar la eficiencia de varios sectores económicos y sociales. En este contexto, varias instituciones gubernamentales están publicando datos; sin embargo, estos datos se publican utilizando diversos formatos y estándares, sin relaciones explícitas entre ellos, con alta dispersión y baja capacidad de integración de datos. En este artículo presentamos algunas reflexiones sobre la necesidad del uso de los datos abiertos enlazados en Cuba y algunos desafíos que afectan la adopción masiva de ellos en el país. Señalamos que la apertura de los datos abiertos enlazados en Cuba requiere el uso de mecanismos adecuados relacionados con una infraestructura tecnológica actualizada, un marco legal adecuado y una sinergia adecuada entre los productores y los consumidores de los datos abiertos enlazados en el país.

Palabras clave: Datos enlazados; web semántica; gestión de la información; sistemas de información.

ABSTRACT

Cuba is carrying out an intensive introduction of the ICT in the society, developing and deploying information systems to increase the efficiency of several economic and social sectors. In this context, several government institutions are publishing open data; however, these data are published using diverse formats and standards, without explicit relations between them, with high dispersion and low data integration

capabilities. In this paper, we state some reflections about the use of the linked open data in Cuba and some challenges affecting the massive adoption of them in the country. We point out that the openness of the linked open data in Cuba requires the use of adequate mechanisms related to an updated technological infrastructure, an adequate legal framework and an adequate synergy between the producers and consumers of the linked open data in the country.

Key words: Linked Data; Semantic Web; Information Management; Information Systems.

INTRODUCCIÓN

Cuba está efectuando una introducción intensiva de las tecnologías de la información y la comunicación (TIC) en la sociedad, mediante el desarrollo y el despliegue de sistemas de información para aumentar la eficiencia de varios sectores económicos y sociales. Alineado con el proceso de informatización, se realizan importantes esfuerzos para la creación y el mantenimiento de la infraestructura tecnológica que soporte el flujo masivo de datos en la red nacional y su interconexión con la red internacional.

Este escenario propicia mayores posibilidades de interconexión, así como el incremento del número de usuarios conectados y la generación e intercambio de datos en todas las esferas de la sociedad. Los datos constituyen la materia prima fundamental para la toma de decisiones a todos los niveles, tanto local, regional como nacional; de ahí que su adecuada gestión y su procesamiento constituyen una tarea prioritaria e impostergable para el país.

En la última década ha emergido un movimiento internacional llamado "Open Data". Los objetivos fundamentales de este movimiento son similares a otros movimientos como son *Open Source*, *Open Science* y *Open Access*. Según *Marijn Janssen* y otros autores,¹ los datos abiertos se definen como datos sin restricciones de privacidad, datos no confidenciales, los cuales fueron producidos con fondos públicos y están disponibles sin restricciones en su uso o distribución. Los datos abiertos contribuyen a la construcción de productos y servicios de valor añadido, atendiendo a la fuente del dato y su calidad. Algunos tipos de datos abiertos más comunes son los datos meteorológicos, geográficos, estadísticos, e incluso, datos provenientes del sector de la salud.

En Cuba existen actualmente diversas fuentes de datos que poseen un valor extraordinario tanto para los ciudadanos como para los órganos y organismos gubernamentales, tales como la Oficina Nacional de Estadística e Información (<http://www.one.cu>), del Ministerio de Salud Pública, entre los que se encuentra el Formulario Nacional de Medicamentos (<http://www.infomed.cu>). Existen, además, datos climatológicos proporcionados por el Instituto de Meteorología y otros, de entidades como el sitio de la Infraestructura de Datos Espaciales (<http://www.iderc.cu>) o la Empresa de Telecomunicaciones de Cuba a través de las Páginas Amarillas (<http://www.paginamarilla.cu>). Estos datos incumplen con frecuencia requisitos que deben caracterizar la apertura de datos para que puedan ser utilizados, ya que son publicados utilizando formatos y estándares diferentes, sin que

existan relaciones explícitas entre ellos, con alta dispersión y baja capacidad de integración a nivel del dato. Todos estos problemas afectan la capacidad de utilización de estos, tanto por los humanos como por los sistemas informáticos. En este artículo se presentan algunas reflexiones en torno al uso de los datos abiertos enlazados en Cuba y algunos desafíos que limitan la adopción masiva de estos en el país.

LOS DATOS ABIERTOS ENLAZADOS

En el año 2006 surgieron los Datos Abiertos Enlazados (*Linked Open Data*, en *Inglés*), una iniciativa internacional que logra de manera efectiva resolver algunos de los problemas de los datos abiertos mencionados anteriormente. El primer proyecto de Datos Abiertos Enlazados a escala planetaria surgió en el año 2007 y se denominó DBPedia.² Este proyecto se concibe con el objetivo de transformar los datos existentes en la Wikipedia a un formato único procesable automáticamente por las computadoras. Los Datos Abiertos Enlazados constituyen la base sobre la que se sustenta la próxima generación de la Web, conocida como la web semántica. El objetivo de la web semántica es extender la web actual y convertirla en una enorme base de datos interconectados semánticamente.

A diferencia de la web de los documentos, donde los enlaces hipertextos se crean entre páginas web, en la web de los datos o datos enlazados, los enlaces se establecen entre piezas de datos. La diferencia fundamental entre estos dos enfoques radica en la posibilidad de asociar una relación semántica entre las piezas de datos interconectados. La semántica o significado de la relación está definida mediante la utilización de clases y propiedades existentes en ontologías de dominio. Las ontologías de dominio, como forma de representación del conocimiento, constituyen un elemento importante para dotar la web de significado, y mejorar el procesamiento automático de los datos tanto por humanos como por algoritmos computacionales.

El *World Wide Web Consortium* ha definido un grupo de estándares y tecnologías con el objetivo de potenciar el crecimiento de los datos enlazados en la web. Los estándares fundamentales son: *Resource Description Framework* (RDF), *Ontology Web Language* (OWL) y *SPARQL Query Language* (SPARQL). RDF es un marco de descripción de recursos en la web basado en grafos dirigidos. Se basa en tripletas que describen semánticamente los recursos publicados en la web. Las tripletas están formadas por tres componentes: sujeto, predicado y objeto. El sujeto define las URIs que identifican los recursos en la web. El predicado define las clases y propiedades existentes en ontologías de dominios y el objeto define el valor de las clases o propiedades definidas en el predicado (Fig.). Por otra parte, OWL es el lenguaje ontológico generalmente utilizado para la construcción de ontologías de dominio. Finalmente, SPARQL es un lenguaje de consultas similar a SQL, diseñado para consultar tanto grafos RDF como ontologías en OWL.

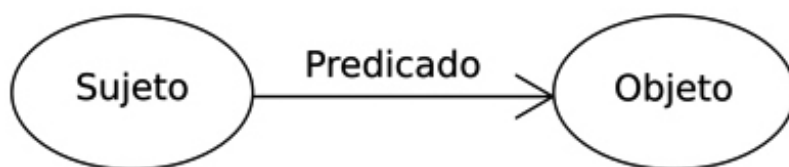


Fig. Estructura de las tripletas *Resource Description Framework* (RDF).

En Cuba se están desarrollando acciones importantes para la publicación y utilización de datos abiertos enlazados. Desde el punto de vista académico se han realizado tres talleres internacionales en Cuba con el objetivo de propiciar el intercambio científico entre académicos que trabajan en esta área de investigación. El III Taller Internacional de la Web Semántica, celebrado en el contexto del XV Congreso Internacional de Información INFO 2018, contribuyó a la socialización de diversos trabajos de investigación provenientes de Bélgica, Cuba, Ecuador, México y Estados Unidos.

LOS DATOS ABIERTOS ENLAZADOS EN CUBA

ALGUNOS PROYECTOS I+D QUE ACTUALMENTE UTILIZAN LOS PRINCIPIOS DE LOS DATOS ABIERTOS ENLAZADOS EN CUBA

Varios proyectos cubanos han comenzado a adoptar los principios de los datos enlazados atendiendo a los beneficios que estos reportan. La Biblioteca Nacional de Cuba, junto a la Biblioteca Central de la Universidad de La Habana y el Instituto de Historia de Cuba, están enfrascados en un proyecto peculiar que trata de convertir los viejos catálogos de fichas impresas en datos enlazados.³ Este proyecto fue presentado en el Congreso Internacional de la Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA) realizado en los Estados Unidos en agosto del año 2016, con muy buena acogida por parte de los participantes.

Por otra parte, se desarrolla en Cuba el proyecto VLIR RED Cuba.⁴ Esta red persigue gestar la universidad virtual cubana. Para esto fortalece el rol de las TIC en las universidades cubanas para el desarrollo de la sociedad. Es un proyecto VLIR que tiene el objetivo de desarrollar un programa educativo virtual nacional en red que compartirá el material de los cursos entre todas las universidades cubanas. Esta red está constituida, al mismo tiempo, por una red educativa virtual, un sistema de gestión de la información de biblioteca, una red de repositorios digitales y un Current Research Information Service (CRIS). Los mecanismos de interoperabilidad semántica deben contribuir al desarrollo de los datos enlazados y grafos de conocimiento. El estado de la cuestión de esta red aún está a nivel de modelos independientes para cada plataforma, aspecto que debe convertirse en un gran sistema basado en el paradigma de los datos abiertos enlazados. Los problemas más complejos están en la construcción de un modelo de interoperabilidad, la asunción de una normativa cubana para estos asuntos y la necesidad de un marco regulador de estas actividades en el país.

Otro de los proyectos cubanos relacionados con los datos abiertos enlazados es el referido a la concepción y desarrollo de una Biblioteca Digital Semántica Cubana.⁵ Este proyecto es el primero de su tipo en el país. Pretende desarrollar un modelo conceptual y su correspondiente implementación en un sistema informático que soporte la extracción, publicación y consumo de datos bibliográficos siguiendo los principios de los datos abiertos enlazados. El sistema propuesto se basa en la integración de diversas fuentes de datos heterogéneas y distribuidas para luego construir servicios de valor añadido para los usuarios. Entre los servicios implementados se cuenta con el servicio de detección de comunidades científicas⁶ y la desambiguación del nombre de los autores⁷ Ambos servicios han sido implementados utilizando técnicas y algoritmos existentes en el estado del arte.

Aunque se han obtenido algunos resultados importantes en materia de definición de proyectos vinculados con la publicación y utilización de datos abiertos enlazados en Cuba, aún siguen siendo insuficientes las acciones desarrolladas para la publicación masiva de datos abiertos enlazados en diversos sectores de la sociedad. La introducción y expansión de estas tecnologías en Cuba resultan necesarias,

atendiendo a las oportunidades y beneficios que estas reportan. A continuación, se describen algunos de los desafíos existentes en la comunidad de datos abiertos enlazados en Cuba que, a juicio de los autores, se deben resolver en el corto plazo para lograr en el país un despegue significativo en esta temática.

ALGUNOS DESAFÍOS DE LA COMUNIDAD DE DATOS ABIERTOS ENLAZADOS EN CUBA

Carencia de un marco regulatorio debidamente establecido para la apertura de Datos Abiertos Enlazados y su adecuada utilización por terceros en Cuba

El Consejo de Estado de la República de Cuba aprobó el Decreto Ley No. 281 en el año 2011, en virtud del cual fueron establecidos los principios de organización y funcionamiento del sistema de información de gobierno para la gestión y el uso de los datos.

Con la aprobación del Decreto Ley 335, que tiene por objeto la creación, organización y funcionamiento del Sistema de Registros Públicos de la República de Cuba, se encaminan los esfuerzos para facilitar la integración de servicios de cara al ciudadano e imprimirle mayor celeridad y eficiencia al flujo de datos para los propósitos de gestión de la propia Administración Pública. Si bien ambas disposiciones jurídicas contribuyen a visualizar un entorno superior de gobernanza en los temas de Gobierno Electrónico, en sí mismas no aluden a datos abiertos y enlazados.

El 28 de febrero del año 2017, el Consejo de Ministros aprobó la política integral para la informatización de la sociedad cubana, que es fruto del cumplimiento de los Lineamientos de la Política Económica y Social que sustenta la conceptualización del Modelo Económico en plena implementación y desarrollo en el país. De la Política y en correspondencia con los objetivos de estudio relacionamos los siguientes puntos a destacar:

- Fomentar la industria de aplicaciones informáticas, en correspondencia con las prioridades de informatización del país y orientada a fortalecer el crecimiento de nuestra economía.
- Establecer una plataforma nacional que incentive la generación de contenidos y garantice la posibilidad de socializarlos, dirigirlos a fortalecer la identidad y a preservar los valores de la sociedad cubana, así como a desarrollar y modernizar la infraestructura tecnológica, prestando especial atención al despliegue de la banda ancha y a la producción de dispositivos informáticos en Cuba.
- Preservar la superación del capital humano asociado y que se potencie el acceso de los ciudadanos al empleo de las nuevas tecnologías.
- Favorecer la comunicación entre las diferentes instituciones gubernamentales y los trámites a la población.

En correspondencia, se asevera que existen tres objetivos principales:

- La informatización para garantizar el desarrollo socioeconómico sostenible, ya sea aplicándolas tanto a la propia informática como industria, como a las restantes industrias y sectores productivos de la sociedad.

- La informatización para la administración pública que incluya una tramitación ágil y efectiva de cara al ciudadano, con una creciente virtualización e interoperabilidad.
- La utilización masiva de las TIC en el desarrollo de los procesos sociales (ciencia, educación, salud, ocio, entre otros) a escala nacional e internacional.

Se afirma entonces que es un documento que constituye una base programática para proyectar una informatización social consistente y coherente a partir del aprovechamiento, con efectividad, de las tecnologías de la información y las comunicaciones, además de ser el pivote para impulsar una política pública de datos abiertos y enlazados, dado que se puede asegurar que existe una relación inextricable entre el Gobierno Electrónico y el perfeccionamiento de los procesos de registro, recolección, tratamiento y entrega de datos públicos con apoyo de las TIC, ya que imprime celeridad, transparencia a la gestión administrativa y permite optimizar los servicios públicos. A la vez, está muy vinculado a las estrategias de anticorrupción y el empoderamiento para el ejercicio de la auditoría y el control social. También hace parte del ecosistema de información que nutre a la economía digital.⁸

Esta realidad exige alentar un enfoque de apertura de los datos cuyos principios y requisitos se encuentren establecidos legalmente, de modo que sea posible reclamar a la administración, en caso de incumplir el deber de informar y, además, generar un ambiente de transparencia en su gestión, así como favorecer la creación de fuentes de información que permitan la innovación y la cocreación entre las entidades públicas y los ciudadanos de nuevos servicios de valor agregado que generen bienestar para todos.

De manera que, unido a la adecuada infraestructura tecnológica y al perfeccionamiento del orden legal entre los problemas cruciales a resolver, está la necesidad de establecer la obligatoriedad de la publicación de datos abiertos y los principios que la rigen, con el objetivo de contribuir a generar un ecosistema de datos abiertos, regulado y sistemático que permita, a su vez, enlazarlos con otros datos abiertos y generar la cadena valor de datos enriquecidos para la innovación y la mejora de la toma de decisiones.

Para lograrlo, también se impone un cambio de mentalidad en cuanto al dato público y su disponibilidad de reutilización para proporcionar servicios públicos convergentes, el acceso a los bienes públicos de información y el incremento de la eficiencia operativa del gobierno. Todo esto supone, además, una revisión integral de la legislación de gestión, el acceso y la protección de la información; gestionar adecuadamente la resistencia cultural con un adecuado plan de acompañamiento y la acertada conducción de la gestión de proyectos y de cambio organizacional involucrado, así como establecer los instrumentos para la evaluación de impacto.⁹

INSUFICIENTE SINERGIA ENTRE LAS COMUNIDADES CIENTÍFICAS DE CIENCIAS INFORMÁTICAS Y CIENCIAS DE LA INFORMACIÓN EN CUBA

La sinergia entre las comunidades científicas de Ciencias de la Información y Ciencias Informáticas no está declarada en la praxis. Los nuevos desarrollos del proyecto VLIR en Cuba, mencionado anteriormente, han intentado sumar ambas comunidades de práctica al desarrollo de los datos abiertos. Para las ciencias de la información los problemas son más fuertes que para las ciencias informáticas. Los temas de la web semántica han entrado con la tendencia *iSchool* con celeridad en los currículos académicos, y pocas universidades cubanas enseñan que la web semántica es más que metadatos insertadas en un grafo RDF, donde las lógicas descriptivas y la matemática juegan un papel esencial.

Por otro lado, la comunidad de ciencias informática está centrada en el código sin mirar las aportaciones que pueden hacer los expertos en ciencias de la información con los conocimientos que poseen de lingüística documental. El estado de las cosas evidencia grandes problemas de estandarización de metadatos y de la web semántica. Nos hemos quedado en la descripción plana de recursos y anquilosada en el siglo XIX; nuestros datos no tienen hoy la consistencia ni la calidad para emplear los estándares emergentes. Esto acrecienta los problemas de interoperabilidad; de ahí la necesidad de las consabidas sinergias para lograr un desarrollo cualitativo y cuantitativo en el país.

CARENCIA DE UNA INFRAESTRUCTURA TECNOLÓGICA QUE FACILITE LA PUBLICACIÓN, EL INTERCAMBIO Y LA UTILIZACIÓN DE LOS DATOS ABIERTOS ENLAZADOS EN CUBA

Para lograr una efectiva publicación, intercambio y utilización de los datos abiertos en Cuba, es necesario contar con una adecuada infraestructura tecnológica que los soporte. Las infraestructuras de datos abiertos enlazados no son nuevas. En diversos países alrededor del mundo ya se cuenta con potentes infraestructuras de este tipo. Dos de los proyectos más importantes en este ámbito lo son sin dudas la infraestructura de datos abiertos del Reino Unido¹⁰ y la infraestructura de datos abiertos del Estados Unidos.¹¹ Ambos proyectos tienen como objetivo proveer una infraestructura para la publicación y el acceso de datos abiertos en diversos dominios. Los más comunes son los datos provenientes del sector de la salud, la educación, la energía y los datos sobre investigación y desarrollo.

En la región de América Latina y el Caribe existen diversas iniciativas que publican datos abiertos de gobierno. Las iniciativas más importantes son: Open Data Brasil^a, Open Data Argentina^b, Open Data Chile^c, Open Data Perú^d, Datos Abiertos Mexicanos^e y Datos Abiertos de Ecuador^f. A excepción de la iniciativa ecuatoriana, el resto de las plataformas de datos abiertos no utilizan los principios de los datos abiertos enlazados.

En el contexto cubano, aún no existe una iniciativa similar, donde las instituciones gubernamentales y empresariales pueden publicar y compartir de manera integrada los datos abiertos enlazados que contribuyan a la generación de productos y servicios de valor añadido para la población cubana. No obstante, Cuba cuenta con una infraestructura de datos espaciales¹² que tiene como función la de facilitar la creación y la publicación de los datos, metadatos y servicios geográficos en la Web, para incentivar la industria de la cartografía en el país. Nótese que esta infraestructura de datos geográficos no soporta la publicación de datos abiertos de otros dominios. Tampoco proporciona los datos geográficos publicados siguiendo los principios de los datos abiertos enlazados.

CONCLUSIONES

A pesar de que se han obtenido discretos avances, aún siguen siendo insuficientes las acciones desarrolladas en Cuba para la publicación y la utilización de datos abiertos enlazados. La apertura de los datos abiertos enlazados requiere del uso de adecuados mecanismos que incluyen la existencia de una infraestructura tecnológica actualizada, un marco regulatorio adecuado y la sinergia entre los actores que producen y utilizan los datos en el país.

En Cuba es necesario establecer jurídicamente el proceso de apertura de los datos y los principios de publicidad para generar un ecosistema de datos abiertos, regulado y sistemático que permita, a su vez, enlazarlos con otros datos abiertos y generar la cadena valor de datos enriquecidos para la innovación y la mejor toma de decisiones, favorecer la innovación y la cocreación entre las entidades públicas y los ciudadanos de nuevos servicios de valor agregado que generen bienestar para todos.

Se debe gestionar adecuadamente la resistencia cultural con un adecuado plan de acompañamiento y la acertada conducción de la gestión de proyectos y de cambio organizacional involucrado en correspondencia con la instrumentación de una Política Pública de Datos abiertos y Enlazados. Además, establecer los instrumentos para la evaluación de impacto.

Conflicto de intereses

Los autores declaran no presentar conflictos de intereses en la realización del estudio.

REFERENCIAS BIBLIOGRÁFICAS

1. Janssen M, Charalabidis Y, Zuiderwijk A. Benefits, adoption barriers and myths of open data and open government. *Inform Syst Manag.* 2012 [citado 16 de agosto de 2018];29(4):258-68. Disponible en: <https://doi.org/10.1080/10580530.2012.716740>
2. Sören A. DBpedia: A Nucleus for a Web of Open Data. En: *The Semantic Web. Lecture Notes in Computer Science.* Berlin, Heidelberg: Springer Berlin Heidelberg. 2007 [citado 16 de agosto de 2018]:722-35. Disponible en: https://doi.org/10.1007/978-3-540-76298-0_52
3. Urra González P. Web semántica, datos enlazados y bibliotecas. *Rev Cubana Inform Cienc Salud.* 2016 [citado 16 de agosto de 2018];27(4). Disponible en: <http://rcics.sld.cu/index.php/acimed/article/view/1058>
4. Goovaerts M. Metadata and authority systems for VLIRED Cuba. *Proceedings of the 3rd International Workshop on Semantic Web;* 2018 [citado 16 de agosto de 2018];35-49. Disponible en: <http://ceur-ws.org/Vol-2096/paper4.pdf>
5. Hidalgo-Delgado Y. Methodological guidelines for publishing library data as linked data. Ecuador: *International Conference Information Systems and Computer Science (INCISCOS);* 2017 [citado 16 de agosto de 2018];241-6. Disponible en: <https://doi.org/10.1109/INCISCOS.2017.17>
6. Ortiz-Muñoz E, Hidalgo-Delgado Y. Detección de comunidades a partir de redes de coautoría en grafos RDF. *Rev Cubana Inform Cienc Salud.* 2016;27(1):90-9.
7. Alonso-Sierra L, Hidalgo-Delgado Y. Desambiguación del nombre de los autores en metadatos bibliográficos publicados como datos enlazados. *La Habana: Proceedings of 1st Cuban Workshop on Semantic Web;* 2014 [citado 16 de agosto de 2018];60-71. Disponible en: <http://ceur-ws.org/Vol-1219/paper6.pdf>
8. Amoroso-Fernández Y, Costales-Ferrer D. Big Data: una herramienta para la administración pública. *Cienc Inform.* 2016;47(3):3-8.

9. Amoroso Y. Open Data. Breve referencia a las realidades y perspectivas en Latinoamérica. *Inform Dirit*. 2011;20(1-2):323-39.
10. Nigel Shadbolt N. Linked Open Government Data: Lessons from Data.Gov.Uk. *IEEE Intell Syst*. 2012;27(3):16-24.
11. Hendler J. US Government Linked Open Data: Semantic.data.gov. *IEEE Intell Syst*. 2012 [citado 16 de agosto de 2018];27(3):25-31. Disponible en: <https://doi.org/10.1109/MIS.2012.27>
12. Delgado-Fernández T. La infraestructura de datos espaciales de la República de Cuba, avances y perspectivas. Nueva York: Novena Conferencia Cartográfica Regional de las Naciones Unidas para América; 2009 [citado 16 de agosto de 2018]. Disponible en: http://mdgs.un.org/unsd/geoinfo/RCC/docs/rcca9/ip/9th_UNRCCA_econf.99_IP15.pdf

Recibido: 27 de junio de 2018.

Aprobado: 08 de noviembre de 2018.

Yusniel Hidalgo-Delgado. Universidad de las Ciencias Informáticas. La Habana, Cuba.
Correo electrónico: yhdelgado@uci.cu

a <http://dados.gov.br/>

b <http://datos.gob.ar/>

c <http://datos.gob.cl/>

d <https://odpe.org>

e <https://datos.gob.mx/>

f <http://data.utpl.edu.ec/>