

Methodological Guidelines for Publishing Library Data as Linked Data

Yusniel Hidalgo-Delgado, Reina Estrada-Nelson

*Programming Techniques Department
University of Informatics Sciences
Havana, Cuba
{yhdelgado, restradan}@uci.cu*

Boris Villazon-Terrazas

*Artificial Intelligence Research Division
Fujitsu Laboratories of Europe
Madrid, Spain
boris.villazon.terrazas@uk.fujitsu.com*

Bin Xu

*Department of Computer Science and Technology
Tsinghua University
Beijing, China
xubin@tsinghua.edu.cn*

Amed Leiva-Mederos

*Library and Information Sciences Department
Central University of Las Villas
Villa Clara, Cuba
amed@uclv.edu.cu*

Andrés Tello

*Computer Science Department
University of Cuenca
Cuenca, Ecuador
andres.tello@ucuenca.edu.ec*

Abstract—Publishing data as Linked Data increases the interoperability and discoverability of resources over the web space. This process involves several design decisions and technologies. However, there is no one-size-fits-all formula for publishing data as Linked Data. Also, the quality of linked data published is a key issue to take into account. In the library domain, the quality of linked data is a crucial point for improving the retrieval and use of the data. In this paper, we propose a set of methodological guidelines based on five activities for publishing library data as Linked Data. The proposed guidelines consider the quality of published data as a key issue. In this line, our approach includes a preprocessing task for data cleansing and normalization. The proposed approach has been applied in a use case for publishing bibliographic data from Open Access journals in Cuba. The results obtained show the applicability of the methodological guidelines proposed in a real environment.

Keywords—*bibliographic metadata; linked data; methodological guidelines; semantic web*

I. INTRODUCTION

The fast development of Information and Communication Technologies, the cheapening of storing technologies and increasing computer performance has improved the creation and access to scientific knowledge around the world. In this context, the digital libraries play an important role. According to the IFLA/UNESCO Manifesto for Digital Libraries [1], “a digital library is an online collection of digital objects, of assured quality, that are created or collected and managed according to internationally accepted principles for collection development and made accessible in a coherent and sustainable manner, supported by services necessary to allow users to retrieve and exploit the resources”. However, digital libraries themselves have of challenges and drawbacks.

Today, data integration and semantic interoperability in the context of the digital library are two open research issues. To address these issues, several protocols, standards and tools have been proposed by the research community in the last decades. Some of these protocols proposed are Z39.50 and OAI-PMH. Also, Digital Libraries Systems have adopted MARC21 and XML as standards for data interchange between them.

In the last five years, several international initiatives have promoted the publication and consumption of library data using semantic web technologies. The semantic web is an extension of current web where contents are described semantically. To publish data enriched semantically in the web space it is necessary to accomplish it with Linked Data principles.

Publishing data as Linked Data increases the interoperability and discoverability of resources over the web space. This process involves several design decisions and technologies. However, there is no one-size-fits-all formula for publishing data as Linked Data. Also, the quality of linked data published is a key issue to take into account. A recent survey [2] observes a widely varying data quality ranging from extensive curated datasets to crowdsourced and extracted data of relatively low quality. In the library domain, the quality of linked data is a crucial point for improving the retrieval and use of the data. In this paper, we propose methodological guidelines based on [3] for publishing library data as Linked Data improving the quality of the data published.

This paper is structured as follows: Section 2 presents related methodological guidelines for publishing Linked Data with particular emphasis on Library Linked Data domain. Section 3 presents methodological guidelines based on five activities for publishing library data. Section 4 presents the results obtained after applying the methodological guidelines proposed in a use

case in a real environment and finally, Section 5 presents main conclusions and future work.

II. REALTED WORK

The process to publish structured data as Linked Data involves several designs and technological decisions. The World Wide Web Consortium (W3C) is working on a series of best practices designed to facilitate development of open government data as linked open data [4]. In [5] is proposed a six-step “cookbook” to model, create, publish and announce government Linked Data. These steps could be applied to any project for publishing Linked Data. An important key to this approach is the role of data publishers to maintain up to date the data published.

A set of methodological guidelines is proposed in [3] for publishing Government Linked Data. It consists of five activities: (1) Specification, (2) Modelling, (3) Generation, (4) Publication and (5) Exploitation. These methodological guidelines could be specialized, with few changes, for publishing any type of data following the Linked Data principles. A similar publishing process was done in the generation of *datos.bne.es* dataset. This dataset makes the authority and bibliography catalogue available from the Biblioteca Nacional de España (BNE, National Library of Spain) as Linked Data [6].

A methodology of 15 steps is proposed in [7]. It can be applied to the library domain for publishing Linked Data. The dataset evaluation before publishing Linked Data is an interesting step in this methodology. Dataset evaluation is a crucial point to make sure that the Linked Data is published and accomplish standards.

Finally, the number of multilingual linked dataset has grown in the last years, so is needed a different perspective to publish these datasets following the linked data principles. Methodological guidelines mentioned above could be extended for this purpose. In [8] has proposed guidelines for publishing multilingual linked data that can help RDF publishers to deal with language barriers.

We found several approaches for publishing Library Linked Data in the literature, however, there is no consensus in the academic community and practitioners about what methodological guidelines should be used for publishing Library Linked Data. The Linked Data publishing process is highly dependent from nature of data sources, the expertise of the data publishers and the maturity of the existing tools for doing this complex process. A crucial point to take into account in all application domains is the quality of data published as Linked Data, however, this particular issue is poorly addressed by existing approaches published in the literature.

III. METHODOLOGICAL GUIDELINES

In this section we describe the methodological guidelines based in [3] for publishing library data as Linked Data. Our methodological guidelines do not intend to be exhaustive, neither applicable to all data domains. It consists of five activities: (1) data extraction, (2) data preprocessing, (3) data modelling, (4) data publishing, and (5) data exploitation and follows an iterative and incremental approach. The Fig. 1

provides an overview of the proposed methodological guidelines.

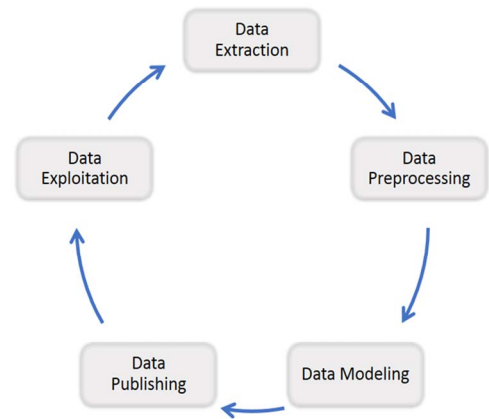


Fig. 1. Methodological guidelines for publishing library data as Linked Data

A. Data extraction

The goal of this activity is to extract and store library data from heterogeneous data sources. At this stage, it is necessary to analyze several issues from each data source, such as: data interoperability, quality of data, data schemas, update frequency, sustainability over time. The input of this activity is one or more data providers and the output is an intermediate database with the extracted library data. Two technological issues should be addressed (1) middleware design and (2) schema mapping.

In digital libraries domain there are data sources such as open access journals, conference proceedings, and even the web. These data sources are characterized by their heterogeneous and distributed manner. Middleware components are needed to achieve the integration and interoperability issues across these systems. Open access journals could disseminate their library data using OAI-PMH protocol, so we need an OAI-PMH Harvester component (middleware) for integrate and use these data.

A schema mapping is a specification that describes how data structured under one schema (the source schema) could be transformed into a data structured under a different schema (the target schema) [9]. In the digital libraries domain, several data sources may use diverse schemas for data representation (e.g. MARC and Dublin Core). A schema mapping approach could be useful to transform data structured under a source schema to a target schema.

At this stage the data storage layer is defined. There are two choices for storing data, independently of data domain. By one side, the relational database paradigm, spread out over the last decades and by the other side, the emerging paradigm named NoSQL. Both paradigms have advantages and drawbacks. The right choice depends of future uses of the data and technological requirements of the designed solution.

B. Data preprocessing

The quality of library data is a crucial point that significantly affects the visibility and discovery of resources described in a Linked Data context. The goal of this activity is to clean up and

normalize some fields of metadata improving considerably their quality. This activity includes data transformation for normalizing some metadata fields such as dates, volumes and numbers of journals. In these cases, a regular expression based approach could be enough. The input of this activity is the intermediate database previously generated and the output is the same database with the library metadata cleaned and normalized.

In this activity a common issue is Entity Resolution (ER). It refers to the issue of identifying and linking or grouping different manifestations of the same real world object [10], [11]. In the particular case of the library data, we identify two entities that could be identified and grouped, for example, the authors' names and their affiliations. The ambiguity in the representations of author's names is a common issue in data sources such as papers published in open access journals and conference proceedings distributed in compact disk. These issues arises due to lack of authority control systems. A similar issue arises with the affiliations of the authors. Author name disambiguation methods have been proposed in the literature based on unsupervised and supervised techniques. A brief survey of automatic methods for author name disambiguation has been proposed in [12].

C. Data Modeling

After the data preprocessing activity, is necessary to define an ontological model to share and annotate library data for both humans and computers. The goal of this activity is to determine the ontology or ontologies to be used for modelling library data. The most important recommendation in this context is to reuse as much as possible available vocabularies [13]. We recommend to follow the tasks and tools proposed in [3] for this activity. The input of this activity is the schema of the intermediate database previously generated and the output is an ontological model.

D. Data Publishing

The goal of this activity is to transform library data previously extracted, stored and modeled to RDF triples. The input of this activity is the ontological model previously defined and the intermediate database (schema and data). The output is one or more RDF graphs with the library data. This activity is divided into three tasks: transformation, linking and publication.

1) Transformation

According to the first Linked Data principle is necessary to use HTTP URIs as identifiers for things (abstract concepts or real world objects). In a Linked Data context "things" refers to the resources described by RDF triples. In this task, it is important the URI design. URIs plays an important role in the discoverability and interoperability of library data in the web space. In this sense, we propose to follow the guides proposed in [13] and [4].

2) Linking

The fourth Linked Data principle is to include/create RDF links to point out other data sources on the Web. External RDF links connect data islands in a global and interconnected data space, enabling other applications to discover additional data sources in a follow-your-nose [13]. The main goal of this task is to generate links between our dataset and other related datasets.

3) Publication

The goal of this task is to make RDF data, previously generated and interlinked, accessible over the web of data. This task has the following subtasks (1) dataset publication, and (2) metadata publication. There are three well-known ways for dataset publication: (1) SPARQL endpoints, (2) Linked Data Front-end, and (3) data dumps. An efficient approach has been formalized in [14] named Linked Data Fragments. The goal of Linked Data Fragments is to build servers that enable intelligent clients, reducing the low availability of public SPARQL endpoints. The metadata publication provides a way to describe RDF datasets using VoID, an RDF Schema vocabulary for expressing metadata about RDF datasets.

E. Data Exploitation

The goal of this activity is to develop real world applications to consume Linked Data previously published. In the particular case of library Linked Data domain, we could provide rich web applications for building digital libraries and value-added services. Several web applications have been developed using library data published as Linked Data. In [15] is proposed a Linked Data based digital library using full-text and faceted search paradigms. Another web application to consume Library Linked Data is on top of datasets published in Biblioteca Nacional de España.

IV. USE CASE: CUBAN JOURNALS

In Cuba, several academic journals disseminate their bibliographic data through OAI-PMH protocol. This protocol is an interoperability framework based on metadata harvesting and one of the most widely used today. In order to validate the applicability of the methodological guidelines proposed in this paper, we conducted an experiment with real data from open access journals. The data sources are a sample of nine Cuban journals with support for OAI-PMH protocol. These journals were selected according to three criteria: (1) available online by the journal over the time, (2) quality of the editorial process and (3) support for OAI-PMH protocol. Finally, we obtain a high-quality dataset with bibliographic metadata in RDF graph format.

A. Data Extraction

In this activity a middleware tool for extracting and storing bibliographic metadata from the selected journals was implemented. The first version supports OAI-PMH based data sources using Dublin Core as the metadata schema for the interchange. It supports both, selective and batch harvesting metadata processes. Also, it supports connections over proxy servers and stores harvested metadata into a Postgresql database server.

TABLE I. JOURNAL LIST OF THE USE CASE

Journal Name	URL OAI-PMH
Serie Científica	https://publicaciones.uci.cu/index.php/SC/oai
Revista Cubana de Ciencias Informáticas	http://rcci.uci.cu/index.php/rcci/oai
Revista Cub de Inform en Ciencias de la Salud	http://rcics.sld.cu/index.php/acimed/oai
Medisur	http://www.medisur.sld.cu/index.php/medisur/oai

Revista Finlay	http://www.revfinlay.sld.cu/index.php/finlay/oai
Rev. Cub. de Cardiolog. y Cirugía Cardiovasc.	http://www.revcardiologia.sld.cu/index.php/revcardiologia/oai
Revista Cubana de Oftalmología	http://www.revofthalmologia.sld.cu/index.php/oftalmologia/oai
Revista de Ciencias Médicas de P del Río	http://publicaciones.pri.sld.cu/index.php/publicaciones/oai
Revista Médico Científica	http://www.revistamedicocientifica.org/index.php/rmc/oai

The middleware implemented is able to extract metadata of four entities: journal, author, article and collection. Also, the middleware supports schema mapping between OAI_DC XML and relational schema, indeed, each XML tag in OAI_DC was mapped to entities and fields in the relational schema. The following table shows the schema mapping previously described. After configured and executed the middleware was obtained a database with 5203 authors, 2711 articles and 134 collections.

TABLE II. MAPPING BETWEEN RELATIONAL SCHEMA AND OAI DC XML SCHEMA

Entity	Attribute	OAI DC XML
Journal	title	<repositoryName>
	URL	<baseURL>
Article	title	<dc:title>
	abstract	<dc:description>
	date	<dc:date>
	source	<dc:source>
	identifier	<dc:identifier>
Author	name	<dc:creator>
	affiliation	<dc:creator>
Collection	title	<setName>

B. Data Preprocessing

In this activity were addressed two major issues: (1) remove of particles from author name and (2) author's name disambiguation. In the first case were removed particles, such as "MSc", "Dr", "Dra." and "Ing", following a regular expression approach. These particles are academic and scientific degrees obtained by authors and vary over the time. In the second case, a tool for author name disambiguation was implemented (see Fig. 2). The edit distance is used to compute the similarity values of author's names, authorship, affiliation and publication place. Next, these similarity values are used to build a vector for each author in the collection of documents from the bibliographic database. This vector is used to build clusters using k-means algorithm [16].

After executing author name disambiguation process was obtained a database with high-quality bibliographic metadata. The author name disambiguation is very important during the publication process of authors using linked data principles. Having a bibliographic database with ambiguous authors (resources) means that several URI for the same author could be published in the resultant RDF graph. It affects significantly the quality of RDF triples generated and the discoverability of the resources published on the web. Also, web information retrieval and building of information systems using RDF published could be affected.

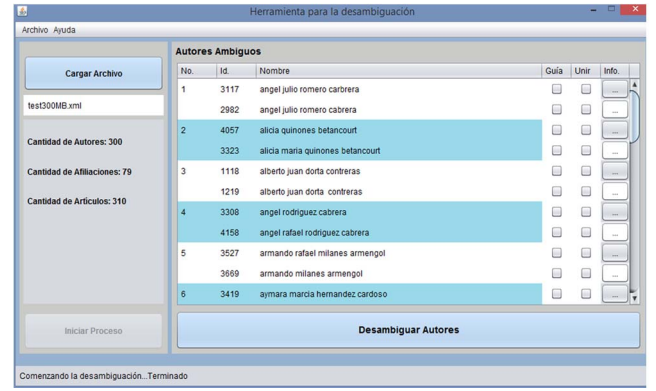


Fig. 2. Author's name disambiguation process

C. Data Modeling

After the aforementioned activities, we have a database with clean and normalized metadata. In this activity, we determine the domain ontologies used for modeling bibliographic metadata, in this case, research papers from scientific journals. The most important recommendation in this context is to reuse available vocabularies as much as possible. We use three well established ontologies: FABIO, SWRC and Dublin Core.

1) *FABIO* [17]: is the FRBR-aligned Bibliographic Ontology, an ontology for recording and publishing bibliographic records of scholarly endeavors on the Semantic Web. This ontology was developed with the minimum of restrictions in its classes and to the domains and ranges of its properties. This flexibility has the great advantage of allowing FaBiO to be used together with other ontologies. Three classes of this ontology were used to model Journal, Journal Article and Item Collection entities in our dataset (see Table 2).

2) *SWRC* [18]: is a well-established ontology for representing knowledge about researchers and research communities. This ontology generically models key entities relevant to typical research communities and the relations between them. In this case, we use the top level concept "Person" to model authors in our dataset.

Also, we use Dublin Core vocabulary to model some metadata fields in our dataset. Dublin Core is a well-established vocabulary to describe digital resources published on the web. It was developed and still maintained by the Dublin Core Metadata Initiative, an open organization supporting innovation in metadata design and best practices across the metadata standards.

TABLE III. MAPPING BETWEEN METADATA FIELDS AND ONTOLOGIES

Entity	Class	Metadata	Property
Journal	fabio:Journal	title	dc:title, rdfs:label
		URL	fabio:hasURL
Article	fabio:JournalArticle	title	dc:title, rdfs:label
		abstract	fabio:abstract
		date	dc:date
		source	dc:source
		identifier	fabio:hasURL
Author	swrc:Person	name	swrc:name
		affiliation	swrc:affiliation
Collection	fabio:ItemCollection	title	dc:title

D. Data Publishing

The data publishing activity is crucial intended to transform, linking and publication of the data sources following the linked data principles. Several standards and tools have been implemented in the last years by the research community and practitioners. In this use case, we reuse several of these standards and tools. In the next sections, we provide an in-deep explanation.

1) Transformation

In this task, we use D2RQ platform for transforming the relational database obtained in previous activities. First, we generate a schema mapping using D2RQ mapping language for aligning relational database fields with their corresponding classes and properties from domain ontologies selected. Second, we generate RDF triples using the dump-rdf bash script. Finally, these triples are stored in a Fuseki server instance. The Fuseki server instance, provides RDF storage capabilities and a SPARQL endpoint for querying RDF graph using SPARQL query language.

2) Linking

In this task, we generate links between the RDF graph generated in the previous task and others related RDF graphs, interconnecting data over the web space using the Silk framework. It provides the Silk - Link Specification Language (Silk-LSL), used to express heuristics for deciding whether there are semantic relationships between two related entities. After configured and executed the silk framework, we obtained semantic relationships using *owl:sameAs* to express that two entities are the same entity in our RDF graph and others related RDF graphs.

In our approach, we use the Levenshtein distance to compare the author names in our dataset with author names in the DBLP, ACM and IEEE datasets. A new set of RDF triples is generated from this process and mixed with the RDF graph originally created and uploaded into the Fuseki server instance.

3) Publication

The purpose of this task is to publish the RDF graph in the Web space, making it accessible to both humans and computers. This stage can be considered as a technical solution to the third linked data principle, which states that it is necessary to provide useful information about the published data using standards such as RDF and SPARQL. In our approach, we use the Pubby linked data interface in front of the SPARQL endpoint. Pubby rewrites URI dereferencing requests into SPARQL DESCRIBE queries, comparing them against the underlying RDF store, and handles 303 redirects and content negotiation.

E. Data Exploitation

It's not enough to publish bibliographic data following the linked data principles, it is necessary to develop information systems with value-added using the linked data published. In our approach, we developed a software component for detecting research communities from bibliographic metadata published as linked data. In the next section, we describe in details the software component and the method used for their implementation.

1) Community detection using linked data

Community detection refers to the problem to identify communities or partitions of nodes that share common properties in a network. The co-authorship networks are considered complex networks, where the nodes of the network are the authors and the edges between nodes provides the co-authorship relationships in one or more publications.

We implement a software component applying a method of three steps (1) modeling of the co-authorship network, (2) community detection and (3) visualization of the detected communities [19]. In the first step, we use SPARQL query language for building a GEXF file with the co-authorship network modeled. In the second step, we apply the Fast Unfolding [20] method for the community detection using the GEXF file previously generated. Fast Unfolding is a method based on the optimization of the modularity for extracting the structure of complex networks. Finally, we provide a visualization of the communities detected previously using the Gephi Toolkit in a web information system. Gephi Toolkit provides several algorithms for the visualization of communities previously detected. In our case, we use the Force Atlas 2, an algorithm based on force vector improving the analysis and visualization of complex networks. The web information system was implemented in the Grails framework, increasing the speed of the development and integration with third-part software and components.



Fig. 3. Interface of the software component implemented

V. CONCLUSIONS

Publishing library data as Linked Data increases the interoperability and discoverability of library resources over the web space. In this paper, we presented a methodological guidelines for publishing library data as Linked Data. Our approach has five activities: data extraction, data preprocessing, data modelling, data publishing and data exploitation. The approach follows an iterative, incremental life cycle model, and pays special attention to the data preprocessing activity, increasing the quality of library data published as Linked Data. Also, the use case developed shown the applicability of the linked data methodological guidelines in the publication of bibliographic metadata obtained from Open Access Cuban journals.

As future work, we will continue formalizing and refining the inputs and outputs of each activity defined in our methodological guidelines. We will keep working on the integration and alignment of our approach with best practices for software development, such as agile software development methodologies and tools and bringing the gaps between research and development of Linked Data based applications. Finally, we will also focus on the definition of data quality within the librarian context.

ACKNOWLEDGMENT

This work was partially supported by the China Scholarship Council (CSC), the University of Cuenca, Ecuador and the University of Informatics Sciences, Cuba, under the research project “Semantic Interoperability in Digital Libraries”.

REFERENCES

- [1] IFLA y UNESCO, «IFLA/UNESCO Manifesto for Digital Libraries», USA, 2011.
- [2] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, y S. Auer, «Quality assessment for linked data: A survey», *Semantic Web*, vol. 7, n.o 1, pp. 63–93, 2016.
- [3] B. Villazón-Terrazas, L. M. Vilches-Blázquez, O. Corcho, y A. Gómez-Pérez, «Methodological Guidelines for Publishing Government Linked Data», en *Linking Government Data*, D. Wood, Ed. New York, NY: Springer New York, 2011, pp. 27–49.
- [4] Bernadette Hyland, Ghislain Atemezang, y Boris Villazón-Terrazas, «Best Practices for Publishing Linked Data», *Best Practices for Publishing Linked Data*, 2014. [En línea]. Disponible en: <http://www.w3.org/TR/ld-bp/>.
- [5] B. Hyland y D. Wood, «The Joy of Data - A Cookbook for Publishing Linked Government Data on the Web», en *Linking government data*, Springer New York, 2011, pp. 3–26.
- [6] D. Vila-Suero, B. Villazón-Terrazas, y A. Gómez-Pérez, «datos.bne.es: a Library Linked Data Dataset», *Semantic Web J.*, vol. 4, n.o 3, pp. 307–313, 2013.
- [7] D. Zengenene, V. Casarosa, y C. Meghini, «Towards a Methodology for Publishing Library Linked Data», en *Bridging Between Cultural Heritage Institutions*, Rome, 2014, pp. 81–92.
- [8] A. Gómez-Pérez, D. Vila-Suero, E. Montiel-Ponsoda, J. Gracia, y G. Aguado-de-Cea, «Guidelines for Multilingual Linked Data», en *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics*, New York, NY, USA, 2013, p. 3:1–3:12.
- [9] R. Fagin, P. G. Kolaitis, L. Popa, y W.-C. Tan, «Composing Schema Mappings: Second-order Dependencies to the Rescue», *ACM Trans Database Syst*, vol. 30, n.o 4, pp. 994–1055, Diciembre 2005.
- [10] L. Getoor y A. Machanavajjhala, «Entity Resolution: Theory, Practice & Open Challenges», *Proc. VLDB Endow.*, vol. 5, n.o 12, pp. 2018–2019, 2012.
- [11] A. Gal, «Uncertain Entity Resolution: Re-evaluating Entity Resolution in the Big Data Era: Tutorial», *Proc VLDB Endow*, vol. 7, n.o 13, pp. 1711–1712, Agosto 2014.
- [12] A. A. Ferreira, M. A. Gonçalves, y A. H. Laender, «A Brief Survey of Automatic Methods for Author Name Disambiguation», *Acm Sigmod Rec.*, vol. 41, n.o 2, pp. 15–26, 2012.
- [13] Tom Heath y Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space*, 1.a ed., vol. 1. Morgan & Claypool, 2011.
- [14] R. Verborgh et al., «Querying Datasets on the Web with High Availability», en *Proceedings of 13th International Semantic Web Conference*, Italy, 2014, vol. 8796, pp. 180–196.
- [15] C. Cordoví-García, C. Hernández-Rizo, Y. Hidalgo-Delgado, y L. Reyes-Álvarez, «Using Search Paradigms and Architecture Information Components to Consume Linked Data», en *Proceedings of 1st Cuban Workshop on Semantic Web*, Cuba, 2014, vol. 1219, pp. 13–30.
- [16] L. Alonso-Sierra y Y. Hidalgo-Delgado, «Desambiguación del nombre de los autores en metadatos bibliográficos publicados como datos enlazados», en *Proceedings of 1st Cuban Workshop on Semantic Web*, Havana, Cuba, 2014, pp. 60–71.
- [17] S. Peroni y D. Shotton, «FaBiO and CiTO: ontologies for describing bibliographic resources and citations», *Web Semant. Sci. Serv. Agents World Wide Web*, vol. 17, pp. 33–43, 2012.
- [18] Y. Sure, S. Bloehdorn, P. Haase, J. Hartmann, y D. Oberle, «The SWRC Ontology – Semantic Web for Research Communities», en *Progress in Artificial Intelligence*, C. Bento, A. Cardoso, y G. Dias, Eds. Springer Berlin Heidelberg, 2005, pp. 218–231.
- [19] E. Ortiz Muñoz y Y. Hidalgo Delgado, «Detección de comunidades a partir de redes de coautoría en grafos RDF», *Rev. Cuba. Inf. En Cienc. Salud*, vol. 27, n.o 1, pp. 90–99, sep. 2015.
- [20] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, y E. Lefebvre, «Fast unfolding of communities in large networks», *J. Stat. Mech. Theory Exp.*, vol. 2008, n.o 10, p. P10008, 2008.