

## Publicación de los datos enlazados iraníes de la Biblioteca Nacional y Archivo de Irán

*Traducción al español del documento original: "Publishing Persian linked data of National library and Archive of Iran"*

### Saeedeh Eslami

Departamento de Tecnologías de la Información y Comunicación, Biblioteca Nacional y Archivo de Irán, Teherán, Irán.

Dirección de correo electrónico: s-eslami@nlai.ir

### Mohammad Hossein Vaghefzadeh

Subdirector de Investigación, Planificación y Tecnología de la Información, Biblioteca Nacional y Archivo de Irán, Teherán, Irán.

Dirección de correo electrónico: m-vaghefzadeh@nlai.ir

TRADUCTORA: Ana M<sup>a</sup> Herranz . Biblioteca Nacional de España



Esto es una traducción al español de " *Publishing Persian linked data of National library and Archive of Iran* " Copyright © 2013 por: **Ana M<sup>a</sup> Herranz**. Este trabajo está disponible en los términos de la licencia Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

### Resumen:

*Actualmente las instituciones intentan publicar, compartir e interrelacionar sus datos mediante el uso de tecnologías de la web semántica, especialmente de Datos Enlazados. Esta tecnología presenta un gran potencial para las bibliotecas e instituciones culturales y les ayuda a complementar sus datos al vincularlos con otras fuentes de información externas. Los datos enlazados bibliotecarios derivan de registros bibliográficos que, al basarse en normativa internacional, serán de gran importancia. Por tanto, la Biblioteca Nacional y Archivo de Irán (NLAI) tiene previsto pasar sus datos a RDF siguiendo los principios de Datos Enlazados propuestos por Tim Berners Lee. Este trabajo es el primer experimento de NLAI en el campo de la publicación de datos enlazados. Utilizando los modelos FR de IFLA, se ha estado definiendo un modelo que genere la versión en datos enlazados de los ficheros de autoridad de NLAI, los cuales se basan en IRANMARC. Se analizan las dificultades que nos encontramos durante la experiencia, especialmente los problemas del idioma persa. Describimos cómo se puede facilitar este proceso.*

**Palabras clave:** Datos enlazados bibliotecarios, Bibliotecas, Ficheros de autoridad, Ontologías, RDF, URI.

## 1 ANTECEDENTES

Las bibliotecas están produciendo datos más grandes y complejos que nunca. Resulta imprescindible que estos productos se compartan y se gestionen con eficacia. Datos mejores: mejor descritos, más conectados, más integrados y organizados, más accesibles y accesibles más fácilmente. Desgraciadamente, los datos de bibliotecas aún no constituyen una parte integral de la web. Jan Hannemann (2010) se dio cuenta de que esto se debe principalmente al bajo nivel de vinculación entre los conjuntos de datos de bibliotecas y los datos procedentes de otros ámbitos, pero, que también se debe a los actuales procesos y formatos de recopilación de información que -naturalmente- se centran en los escenarios clásicos de uso para las bibliotecas. Para conseguir esto, serían útiles los datos enlazados (Linked Data = LD).

Linked Data es una de las tecnologías de la Web Semántica. Berners-Lee, et. al (2009) señaló que el término “Linked Data” hace referencia a un conjunto de buenas prácticas para la publicación y conexión de datos estructurados en la web. Esto anima a las instituciones a publicar, compartir e interrelacionar sus datos. En Linked Data, los “datos” se expresan como enunciados sencillos que utilizan el Resource Description Framework<sup>1</sup> (un método genérico para describir los vínculos entre datos estructurados dentro de un modelo de datos de base gráfica) y se conectan mediante los Uniform Resource Identifier (URI)<sup>2</sup>, identificadores web-direccionables legibles por máquina que identifican entidades. Una declaración RDF es una estructura sujeto-predicado-objeto de tres partes comúnmente conocida como “triple”. La base de un triple es su predicado, que se representa como una propiedad RDF, y el tema y objeto específicos de un triple están representados como miembros de clases RDF. Las clases describen cosas, y las propiedades describen la relación entre cosas; las clases y las propiedades son los tipos básicos de elementos en RDF. Dunsire y Willer (2009) indicaron que el uso de los estándares bibliotecarios actuales como base de nuevas tripletas, así como la extracción de tripletas a partir de registros heredados, requiere la representación de tales normas en RDF ya sea mediante la creación de los apropiados elementos RDF o mediante el mapeo de los elementos existentes. Esto no sólo permitirá que la Web Semántica se beneficie de los metadatos bibliotecarios, sino que también debería mejorar la interoperabilidad entre entidades bibliográficas, atributos y relaciones descritos por estándares diferentes, pero, relacionados. Las propiedades RDF pueden elegirse a partir de diferentes estándares y mezclarse en una única aplicación para satisfacer sus requisitos funcionales (G. Dunsire, 2011).

Visto que los LD ayudan a las bibliotecas a aumentar su presencia y visibilidad en la web, la Biblioteca Nacional y el Archivo de IRÁN (NLAI) está enriqueciendo sus datos vinculándolos con otras fuentes de información externas. NLAI, como biblioteca cabecera de Irán, es responsable de la distribución e intercambio de los datos de las colecciones bibliotecarias. Por tanto, decidimos publicar los datos de los ficheros de autoridad como LD. Las principales contribuciones de este artículo son: 1) presentar la primera experiencia sobre publicación de ficheros de autoridades de NLAI como LD 2) tipificar las principales dificultades al publicar los datos enlazados bibliotecarios persas (LLD)<sup>3</sup> 3) proponer posibles soluciones para estos desafíos y 4) ser pioneros al mapear el formato IRANMARC con las ontologías y vocabularios existentes y al escoger colecciones para su vinculación externa en lo relativo a lenguas iraníes.

---

<sup>1</sup> Resource Description Framework (RDF): Marco de Descripción de Recursos

<sup>2</sup> Uniform Resource Identifier (URI): Identificador Uniforme de Recursos

<sup>3</sup> Library Linked Data (LLD): Datos Enlazados Bibliotecarios

## 2 OBRAS RELACIONADAS

Tim Berners-Lee (2006), director del World Wide Web Consortium, acuñó este término en una nota de diseño en la que se exponían cuestiones sobre todo el proyecto de la Web Semántica. LD promueve la integración de datos mediante la conexión de recursos acerca de temas relacionados. Nosotros clasificamos las obras relacionadas en dos categorías: investigaciones sobre LD en general e investigadores sobre LLD. En general, LD se refiere a aquellos estudios que aprovechan el uso de LD en el contexto de las aplicaciones no-bibliotecarias como “spamming”<sup>4</sup>, sistemas de búsqueda de respuestas, sistemas de recomendación de contenido, etc.

En el área de los datos enlazados bibliotecarios, el World Wide Web Consortium con el Library Linked Data Incubator Group<sup>5</sup> trata los datos bibliotecarios como el pilar central de la Web de Datos y enfatiza el papel primordial de los LD bibliotecarios para la Web Semántica en su conjunto (Baker, et al. 2011). Por otro lado, el Library of Congress (LOC) Bibliographic Framework Initiative General Plan demostró que la web semántica y el relacionado modelo LD albergan interesantes posibilidades tanto para las bibliotecas como las instituciones de patrimonio cultural. Por lo tanto, Keßler, et al. (2009) manifiesta que las bibliotecas reconocen que están a la vanguardia del movimiento de LD. Durante estos años, gran cantidad de bibliotecas han publicado sus catálogos como Linked Open Data y lo han hecho con un alto grado de normalización comparable al de los vocabularios aceptados.

DBpedia constituye el núcleo del acceso a LD, otros proyectos de LD se han desarrollado y conectado a DBpedia gradualmente. En el ámbito de LD en las bibliotecas nacionales, La Biblioteca Nacional de Francia (BNF), la Biblioteca Nacional de Hungría (NSZL), la Biblioteca Nacional de España (BNE), la Biblioteca Nacional de Alemania (DNB) (Hauser, 2012), la Biblioteca Nacional de Suecia (LIBRIS), la Biblioteca del Congreso (LCSH [sic]), y la Biblioteca Nacional Británica (BNB) han publicado sus datos bibliográficos y de autoridad como LD. Además, otra institución colabora en la mejora de los LD bibliotecarios: OCLC viene experimentando con LD desde hace bastante tiempo. Por ejemplo, en agosto de 2012, se publicaron los datos enlazados bibliográficos de cerca de 1,2 millones de registros de WorldCat. El Australian National Data Service (ANDS) creó una colección nacional de recursos para la investigación en Australia.

Crear LD resulta difícil porque requiere el uso de vocabularios y reglas exactos para elaborar datos RDF y vincular conjuntos de datos RDF. Algunos trabajos tratan sobre las aplicaciones que facilitan el proceso. Por ejemplo, Jung y Park (2011) presentaron un sistema que ayuda fácilmente a crear documentos RDF y LD a los usuarios poco expertos. Mediante este sistema, el usuario puede generar sin dificultad documentos RDF y puede agregar nuevos vínculos entre entidades RDF sin un conocimiento completo de la gramática y los vocabularios RDF. En cualquier caso, publicar LD bibliotecarios depende tanto de la normalización bibliotecaria como de los datos en sí. Los datos bibliográficos de bibliotecas son más complejos que elaborar un plan integral para publicarlos como LD, por lo tanto, modelar la ontología y seleccionar el vocabulario resulta el punto más importante. Distintas bibliotecas han venido utilizando vocabulario y ontologías diversos para describir sus recursos como LD. Por ejemplo, a DNB usa Bibo<sup>6</sup> ISBD<sup>7</sup>, Dublin Core<sup>8</sup>, FOAF<sup>9</sup>, la BNE utiliza

---

<sup>4</sup> Spamming: envío de correo electrónico publicitario no solicitado

<sup>5</sup> Véase: <http://www.w3.org/2005/Incubator/ld/>

<sup>6</sup> Bibliographic Ontology (BIBO): Bibliografía Ontológica, <http://purl.org/ontology/bibo/>

<sup>7</sup> International Standard Bibliographic Description (ISBD): Descripción Bibliográfica Internacional Normalizada

<sup>8</sup> <http://purl.org/dc/terms/>

FRBR<sup>10</sup>, ISBD, Dublin Core.

Tanto el Grupo de Revisión de FRBR como el Grupo de Trabajo ISBD/XML de IFLA han trabajado en el desarrollo de representaciones de las normas de la IFLA como recursos RDF. Por consiguiente, se han publicado en RDF los espacios de nombres para los metadatos bibliográficos de la familia de Requisitos Funcionales (FR<sup>11</sup>) de IFLA. Los modelos comprenden FRBR, Requisitos Funcionales para Datos de Autoridad (FRAD), y Requisitos Funcionales para Datos de Autoridades de Materia (FRSAD). Los vocabularios del conjunto de elementos FR incluyen las clases y propiedades RDF correspondientes a entidades, atributos y relaciones FR. Cada clase y cada propiedad tienen un URI para su uso en las tripletas de datos de la Web Semántica. Philippe Le Pape (2011), miembro del PUC<sup>12</sup>, ha afirmado que UNIMARC es plenamente compatible con la estructura FRBR. Con el modelo FRBR, la comunidad bibliotecaria se ha introducido en un nuevo marco conceptual para datos bibliográficos. Aalberg, et al. (2011) sustentan que la coordinación del estándar UNIMARC con FRBR es importante para facilitar la puesta en marcha de la publicación de LD y para facilitar el intercambio de datos bibliográficos basados en el modelo FRBR para una amplia gama de aplicaciones semánticas.

Al desarrollarse la versión RDF de la normativa de IFLA, las bibliotecas comenzaron a utilizarla para crear modelos de ontología fundamental. Los proyectos que se centran en la extracción de entidades y relaciones FRBR entre registros basados en MARC han demostrado en este nuevo contexto muchas de las potencialidades y muchos de los problemas de uso de los formatos MARC (Dunsire, 2011). Dunsire analizó algunos de los problemas que existen al codificar FRBR en UNIMARC. El problema de asignación de identidad y atributo tiene su origen principalmente en las reglas de catalogación y está causado por falta de información (por ejemplo, la falta de títulos que identifiquen correctamente las obras) o por campos de datos que son ambiguos en el contexto de FRBR. Aparte de las ontologías y los vocabularios utilizados en LD, para promover la identificación en diferentes repositorios, los descriptores de repositorios de meta-nivel deberían representarse utilizando el vocabulario VoID. K. Alexander, et al (2009) exponen el diseño y los casos de uso de VoID (Vocabulary Of Interlinked Datasets<sup>13</sup>) para describir conjuntos de datos enlazados. En enero de 2008, publicaron el vocabulario VoID para comunidades de datos enlazados.

### 3 NUESTRO ENFOQUE

De acuerdo con las ya mencionadas obras relacionadas, decidimos explotar los modelos RDF de IFLA, en concreto, FRBR, FRAD e ISBD para transformar registros IRANMARC en datos enlazados RDF. El conjunto de datos basado en IRANMARC de NLAI así como su infraestructura son ISBD y están representados como XML. En lo que respecta a esta investigación, NLAI pretende crear una colección nacional coherente de registros bibliográficos iraníes, lo cual supone un plan a largo plazo. El presente artículo describe lo que proyectamos hacer para lograr este objetivo. En la actualidad, la base de datos de NLAI incluye 982.892 registros de autoridad que se utilizan en 3.175.125 registros bibliográficos, ambos codificados de acuerdo con IRANMARC

---

<sup>9</sup> <http://xmlns.com/foaf/spec/>

<sup>10</sup> Functional Requirements for Bibliographic Records (FRBR): Requisitos Funcionales de los Registros Bibliográficos

<sup>11</sup> Véase: <http://iflstandards.info/ns/fr/>

<sup>12</sup> Permanent UNIMARC Committee (PUC): Comité Permanente sobre UNIMARC

<sup>13</sup> Vocabulario de Conjuntos de Datos Intervinculados

En marzo de 2013, el plan de investigación, "Publicación de las Colecciones NLAI como LD para unirse a la Web de Datos" se presenta al Subdirector de Investigación, Planificación y Tecnología en NLAI. A pesar de que se ha de verificar en el consejo de investigación, hemos comenzado nuestro modelo. En su primer nivel de implementación se ha optado por un acercamiento progresivo, por tanto, nos centramos intencionadamente en un ámbito pequeño al escoger los datos de autoridad para comenzar. Nuestra elección se debe a la ausencia de datos persas en LOD<sup>14</sup> y a que los ficheros de autoridades persas son los datos más relacionables al publicar LD iraníes. Publicar los ficheros de autoridad garantizaría una mejor persistencia para publicar datos bibliográficos persas en el futuro. Por tanto, se seleccionó una hoja de trabajo de nombres de autoridad como punto de partida. Actualmente, los ficheros de autoridades de NLAI se almacenan en bases de datos relacionales de Oracle (RASA) en formato IRANMARC. La figura 1 muestra una panorámica de la arquitectura del proceso de publicación de LD persa.

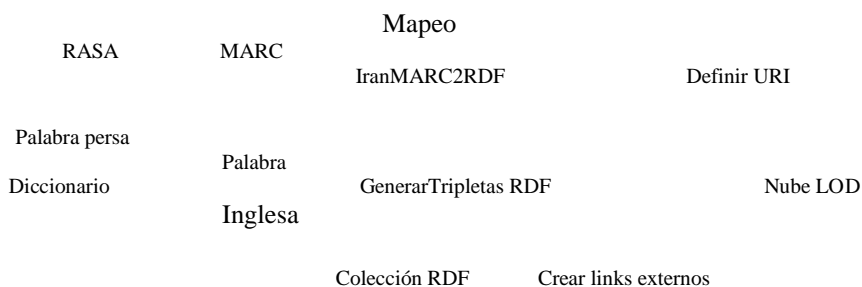
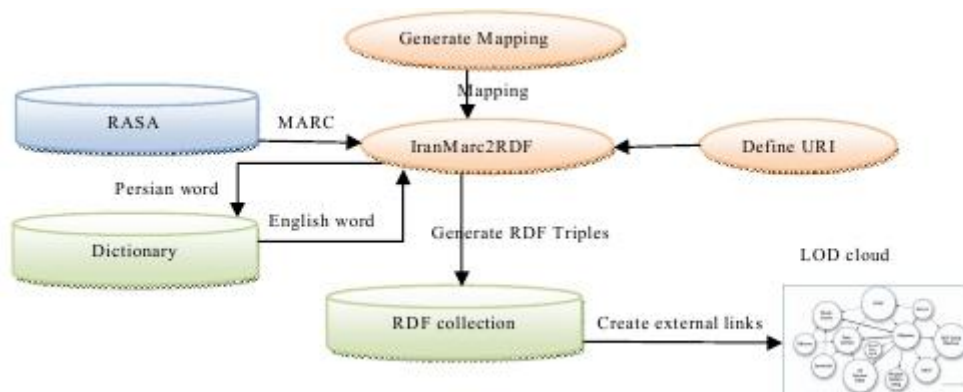


Figura 1: Visión de la arquitectura del proceso de publicación de LD persa. Utiliza los ficheros de autoridad en IRANMARC de NLAI.

Para cada tipo de fichero de autoridad se genera un mapeo, los registros de autoridad se extraen de la base de datos relacional RASA y, posteriormente, en la fase IranMarc2RDF, se crea un archivo RDF y se clasifica en el dispositivo de almacenaje de tripletas (colección RDF) de acuerdo con el mapeo elaborado por un bibliotecario especialista. Durante el proceso de mapeo se utilizan diferentes vocabularios y ontologías. Un paso importante al publicar un conjunto de datos como LD es el diseño de un esquema de URLs para localizar las entidades que se van a publicar. Estos URI se utilizan para crear enlaces externos e internos. Los enlaces internos entre nuestros datos se establecen durante la generación del archivo RDF. La generación de enlaces se expondrá más adelante.

Para crear los modelos de datos utilizamos diversas ontologías existentes que responden

<sup>14</sup> Linking Open Data (LOD): Enlace de datos Abiertos

a las hojas de trabajo de nuestros ficheros de autoridades. Generamos un mapeo manual que define la ontología correspondiente para cada parte de los ficheros de autoridades IRANMARC. Al procesar cada registro IRANMARC, el mapeo indica las entidades de registros IRANMARC, los atributos de dichas entidades y cómo tales entidades se conectan unas con otras, así como un equivalente para cada una de ellas en el modelo FRBR. Por lo tanto, los registros bibliográficos y de autoridad de NLAI se componen de entidades, atributos de estas entidades (propiedades) y relaciones de las entidades.

De acuerdo con los espacios de nombre de IFLA, pensamos comenzar con las importantes propiedades que se muestran en la Figura 2 y, posteriormente, pensamos ir ampliando el campo de cobertura de modo gradual.

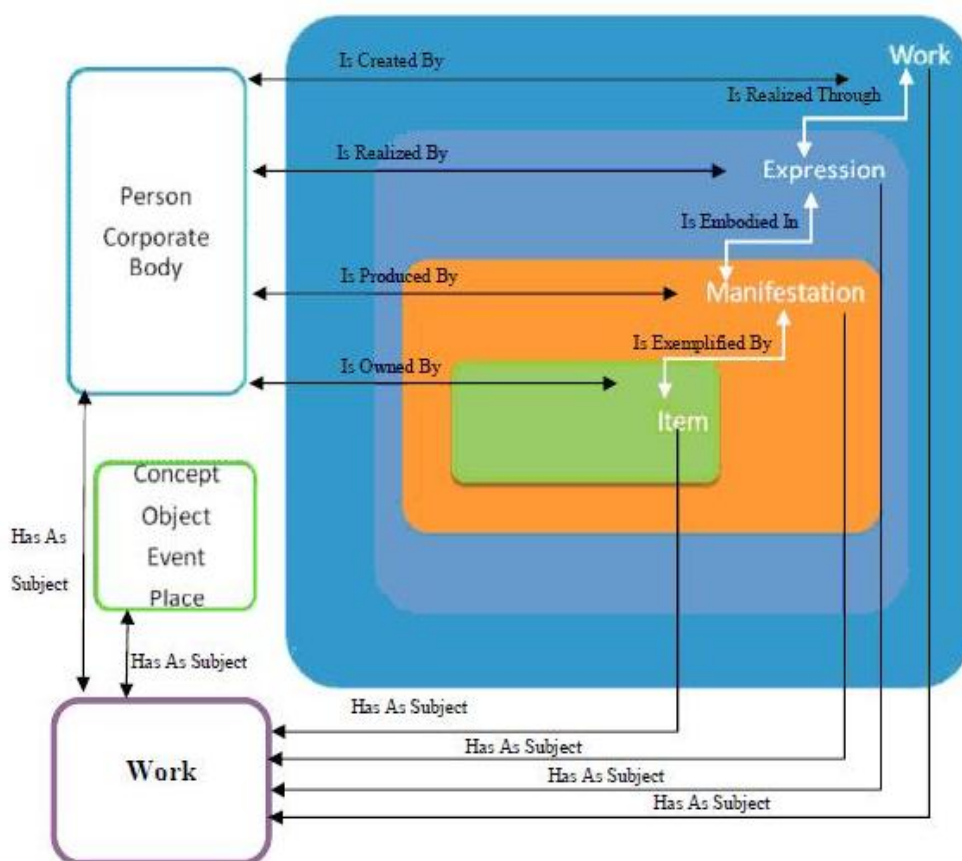


Figura 2– Modelo FRBR y propiedades de objeto utilizadas en nuestro mapeo

En los datos de autoridad, cada registro se equipara a una instancia RDF de tipo de Persona, Entidad Corporativa, Obra o Expresión. En registros bibliográficos, cada registro se equipara a una instancia RDF de tipo Manifestación. Entonces, las propiedades RDF de IFLA se utilizan: para establecer relaciones entre las clases y sus instancias (una obra “Is Created”

por una Persona/Entidad Corporativa e “Is Realized” a través de una Expresión que “Is Embodied” en una Manifestación que, a su vez, “is Exemplified” por un ejemplar); para establecer la relación entre las propiedades de entidades FRBR utilizadas (por ejemplo, اندلوم es creador de “يونتئم”); para describir las instancias de cada registro de autoridad y las propiedades RDF de ISDB utilizadas (por ejemplo, el recurso “hastitleproper” حرشد يونتئم). Resumiendo, las propiedades RDF de FBBR se utilizan en el proceso de mapeo para describir obras, expresiones, entidades del Grupo 1 y personas y entidades corporativas de Grupo 2 y las propiedades RDF de ISDB se utilizan para describir manifestaciones. En el proceso de mapeo manual, se debería tener en cuenta que estos conceptos no nos llevan al cotejo de todos los campos IRANMARC. Teniendo en cuenta la figura 2, el reto principal es de qué manera podríamos extraer entidades de los recursos IRANMARC. Actualmente se hace de forma manual y no para todos los campos IRANMARC

De acuerdo con los principios de LD, cada instancia RDF debe identificarse mediante un único URI. Para la identificación de los recursos RDF de NLAI fijamos el esquema de URI <http://linkeddata.nlai.ir/rasa/identifier>. El campo de tratamiento de datos generales (001) se utiliza como parte local en el URI especificado, por ejemplo, <http://linkeddata.nlai.ir/rasa/148350> se asigna como URI para “تيداده قداص” que es una instancia de persona en FRBR. La tabla 1 muestra las principales clases (i.e., entidades) URI de espacios de nombres de IFLA que tenemos previsto utilizar.

Tabla 1 – Clases conforme a los estándares de IFLA

Entidad (Clase)	Espacio de nombre (URI)
Obra	<a href="http://iflstandards.info/ns/fr/frbr/frbrer/C1001">http://iflstandards.info/ns/fr/frbr/frbrer/C1001</a>
Expresión	<a href="http://iflstandards.info/ns/fr/frbr/frbrer/C1002">http://iflstandards.info/ns/fr/frbr/frbrer/C1002</a>
Manifestación	<a href="http://iflstandards.info/ns/fr/frbr/frbrer/C1003">http://iflstandards.info/ns/fr/frbr/frbrer/C1003</a>
Persona	<a href="http://iflstandards.info/ns/fr/frbr/frbrer/C1005">http://iflstandards.info/ns/fr/frbr/frbrer/C1005</a>
Entidad corporativa	<a href="http://iflstandards.info/ns/fr/frbr/frbrer/C1006">http://iflstandards.info/ns/fr/frbr/frbrer/C1006</a>

Para ejemplificar el proceso de mapeo, la tabla 2 muestra algunos campos de autoridades de nombre en IRANMARC, la entidad equivalente que se asigna y también el URI correspondiente para representar la relación.

Tabla 2 – Etiquetas IRANMARC representadas por sus correspondientes propiedad y clases

Campo IRANMARC de Nombre de Autoridad	Representado por		
001	Tratamiento de datos generales	Utilizado como parte local del URI de una persona	
101\$a	Idioma de la entidad	Propiedad	<a href="http://purl.org/dc/terms/language">http://purl.org/dc/terms/language</a>
200\$a 200\$b	Elemento de entrada Parte del nombre distinta del elemento de entrada	Tipo de entidad	Persona
		Igual a	Owl:sameAs
		Propiedad	hasNameOfPerson
		URL	<a href="http://iflastandards.info/ns/fr/frbr/frbrer/P3039">http://iflastandards.info/ns/fr/frbr/frbrer/P3039</a>
200\$f	Fecha	Propiedad	hasDatesOfPerson
		URL	<a href="http://iflastandards.info/ns/fr/frbr/frbrer/P3040">http://iflastandards.info/ns/fr/frbr/frbrer/P3040</a>
400\$a	Punto de acceso alternativo – nombre de persona-	Propiedad	<a href="http://iflastandards.info/ns/fr/frbr/frbrer/P4031">http://iflastandards.info/ns/fr/frbr/frbrer/P4031</a>
Campo bibliográfico IRANMARC	Representado por		
001	Identificador de registro	Utilizado como parte local del URI de una manifestación	
200\$a	Título propio	Propiedad	<a href="http://iflastandards.info/ns/isbd/elements/P1099">http://iflastandards.info/ns/isbd/elements/P1099</a>

En cuanto a la conectividad exterior y el número de enlaces a bases de datos externas, hemos considerado la propiedad de objeto “owl:sameAs” con el fin de remitir a los enlaces de equivalencia. En lo que respecta a la vinculación con conjuntos de datos externos (y dado que son raras las fuentes de datos externas persas) decidimos usar fuentes de datos conocidas ya existentes tales como VIAF, LCSH, DBPedia y, por consiguiente, utilizar un diccionario para vincular los datos persas con su correspondiente término en inglés.

Como ya se ha mencionado en otros proyectos y estudios, implementar FRBR en un sistema bibliotecario significa mucho más que simplemente cambiar el formato o el modelo de datos subyacente. Por tanto, tenemos un largo camino por recorrer. Para su valoración, otros trabajos similares principalmente se han centrado al señalar la calidad del trabajo en indicadores tales como el número de tripletas RDF o el número de enlaces a bases de datos externas. Estimamos poder hacerlo tras subir la versión beta del conjunto de datos enlazados de NLAI.

#### 4 DIFICULTADES Y SOLUCIONES

En este proceso, los modelos RDF de IFLA nos ayudan a aprovechar las ventajas de las semánticas que se encuentran tras los modelos FR. Sin embargo, todavía hay problemas para generar una solución global que permita la conversión de los registros IRANMARC de NLAI a LD. En cualquier caso, trabajos anteriores demostraron que los modelos RDF de IFLA facilitan tanto el proceso de transformación de datos como el desarrollo de interfaces para la



navegación de los usuarios mediante el “**Error! Reference source not found**”, pese a lo cual, durante este proyecto se identifican diferentes problemas y dificultades. Aquí, exponemos estos problemas y recomendamos algunas soluciones. En concreto, un problema habitual es la falta de experiencia para asentar los LD persas.

- Problemas con los enlaces

Una cuestión importante al publicar LD es decidir qué ontologías deberían utilizarse para describir los recursos. La mayoría de las veces la selección de una ontología depende de su popularidad. Hay algunas ontologías que se convirtieron en estándares para ámbitos específicos (ej. Dublín Core, FOAF). Sin embargo, la popularidad no es un criterio adecuado para la selección de una ontología. Por ejemplo, hay ontologías para disciplinas específicas poco conocidas. Con el objetivo de obtener una interoperabilidad eficiente, para la descripción de nuestros datos tenemos que elegir vocabularios aceptados. Necesitamos métodos que ayuden a los editores de LD a determinar qué ontologías son adecuadas. En el campo bibliotecario, escoger las ontologías debe considerarse el paso más importante puesto que podría disminuir la precisión y la calidad de los resultados, especialmente en conjuntos de datos grandes, dinámicos y complejos tales como los datos bibliográficos bibliotecarios. Por tanto, uno de los retos de LD es la falta de un método normalizado para la elección de ontologías. Una solución consiste en aplicar las experiencias de otras instituciones.

- Interrelaciones de datos

La tarea de vincular datos con recursos externos puede llevarse a cabo con aplicaciones. Algunas aplicaciones utilizan técnicas de correspondencia para detectar la relación semántica entre dos entidades. La BNE desarrolla MARiMbA con este fin (Daniel Vila-Suero, et al. 2012). Lamentablemente, NLAI aún está en pañales en lo relativo a la publicación de LD, no obstante, tenemos la intención de mejorar estos procesos mediante el desarrollo de este tipo de herramientas.

- Dificultades con los datos persas

La falta de datos y los datos incompletos o incorrectos en la base de datos original suponen otro problema al publicar LD. Algunos registros IRANMARC están incompletos, están en otra escritura, o en otro tipo de formato. Estos problemas llevan a dificultar la vinculación con recursos externos. Para hacer frente a este reto, se deberían analizar con precisión los datos originales para descubrir los problemas existentes. Para remediarlo, podríamos emplear técnicas de limpieza. Por ejemplo, es posible implementar un algoritmo que convierta los diferentes formatos de algunos campos a un único formato. Como la mayoría de los datos de la nube LOD se publican en inglés, resulta difícil para nosotros vincular un conjunto de datos iraní a conjuntos de datos externos. En los sistemas multilingües en los que los datos son generados por los usuarios finales, algunos usuarios eligen su lengua materna, mientras que otros utilizan el inglés para introducir sus datos. Por ejemplo, en lugar de 'John' alguien podría meter ‘نجد’. Por poner otro ejemplo, en la base de datos existen términos persas idénticos en diferentes formas de inglés, por ejemplo, un solo nombre persa "ي م ل س ا" se introduce tanto "Eslami" como "Islami". Tales problemas, causados por los datos multilingües, introducen desafíos en la búsqueda de datos externos para los recursos relacionados que deben enlazarse y disminuyen la calidad del conjunto de datos publicado. Por ello, debería haber un mecanismo para encontrar los equivalentes en inglés de los términos persas. Como solución para encontrar el equivalente apropiado del término requerido pensamos utilizar un diccionario local. Así pues, todos los equivalentes de ‘ي م ل س ا’ (ie. Islami, Eslami) se cotejarían al buscar datos externos.

- Mantenimiento de links

Resulta importante mantener los enlaces y también la calidad de los datos en LD. La actualización de bases de datos conectadas entre sí puede originar enlaces no válidos. Puede

requerir que se actualicen los links existentes. Si los conjuntos de datos externos fijan su última modificación, nosotros podríamos decidir cuándo actualizar los nuestros. Por otra parte, ya que el conjunto de datos original ha cambiado, es necesario actualizar nuestros LD. Por tanto, informaciones tales como “momento de creación”, “momento de modificación” deberían publicarse junto con el conjunto de datos. Para hacer frente a este requisito usamos la especificación de nuestra base de datos dcterms: created y dcterms:modified en VoID (Vocabulary Of Interlinked Datasets) .

## 5 CONCLUSIÓN Y FUTUROS TRABAJOS

Este trabajo se hizo de acuerdo con los principios de LD. En el futuro, nos gustaría crear bases de datos de conocimiento disponibles en RDF a partir de los registros bibliográficos y de autoridad de la biblioteca. Se necesita tiempo para investigar sobre la coordinación de IRANMARC con estándares como FRBR, ISBD, FRAD. NLAI ha planeado usar el sistema de gestión de contenidos Drupal para la representación de LD. La presentación de los conjuntos LD de NLAI ofrecerá servicios eficientes para los usuarios finales y los investigadores. La implementación de una interfaz para la búsqueda del conjunto de datos enlazados de NLAI también se considera uno de nuestros próximos trabajos. Además, estos datos se interconectarían con otras bases de datos de conocimiento y con otros conjuntos de datos enlazados existentes en el mundo tales como DBPedia/Wikipedia. Por cierto, esto se hace bastante fácilmente una vez que se tengan los datos en RDF. Por otro lado, tratamos de definir progresivamente un modelo de datos completo que incluya todas las entidades importantes para publicar en la web de datos los registros bibliográficos IRANMARC. Algunas de las razones que obligan a NLAI a seguir adelante son: hay terceros que se podrían enlazar a nuestros datos más fácilmente, nosotros podríamos desarrollar nuevos servicios basados en nuestro contenido, y por último, podría convertirse en un "punto de referencia de normalización" tanto a nivel local como internacionalmente.

## 6 BIBLIOGRAFÍA

Hyosook Jung, Seongbin Park, (2011). A System for Linked Data Creation, Studies in Informatics and Control, vol. 20 . Issue 4. Disponible en: [http://sic.ici.ro/sic2011\\_4/art04.php](http://sic.ici.ro/sic2011_4/art04.php)

Jan Hannemann & Jürgen Kett (2010). Linked Data for Libraries, in world library and information congress: 76th IFLA general conference and assembly, Gothenburg, p. 12.

Bizer, Christian, Tom Heath, Tim Berners-Lee (2009). Linked data: the story so far ,International Journal on Semantic Web and Information Systems (IJSWIS), vol. 5, issue 3. Disponible en: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

Dunsire, Gordon, Mirna Willer (2011). UNIMARC and Linked Data, IFLA Journal 37, 4, 314-326. Disponible en: <http://conference.ifla.org/past/ifla77/187-dunsire-en.pdf>

Thomas Baker, Emmanuelle Bermès, Karen Coyle, Gordon Dunsire, Antoine Isaac, Peter Murray, Michael Panzer, Jodi Schneider, Ross Singer, Ed Summers, William Waites, Jeff Young, and Marcia Zeng. Library Linked Data Incubator Group Final Report. W3C Incubator Group Report. Disponible en: <http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025>.

Carsten Keßler, Mathieu d'Aquin, and Stefan Dietze (2012). Linked Data for Science and Education. *Semantic Web Journal*, vol. 4, no 1.

Julia Hauser (2012). Linked Data Service of the German National Library. Disponible en: [http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata\\_node.html](http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html)

Library of Congress. Subject Headings. Disponible en: <http://id.loc.gov/authorities/subjects.html> . Último acceso, 2012-08-26.

Trond Aalberg, Jan Pisanski and Maja Žumer (2011). UNIMARC and FRBR - can we have both? *Advancing UNIMARC: alignment and innovation*, IFLA. UNIMARC and linked data.

Philippe Le Pape (2011). Expressing FRBR in UNIMARC Yes we can! *Advancing UNIMARC: alignment and innovation — IFLA*.

Daniel Vila-Suero, Boris Villazón-Terrazas, Asunción Gómez-P.,(2012). *datos.bne.es: a Library Linked Data Dataset*. In *Semantic Web Interoperability, Usability, Applicability an IOS Press Journal*

Tim Berners-Lee (2006-07-27). "Linked Data—Design Issues". W3C. Obtenido el 2013-05-18.

Keith Alexander, Michael Hausenblas (2009). Describing linked datasets - on the design and usage of void, the 'vocabulary of interlinked datasets, In *Linked Data on the Web Workshop (LDOW 09)*, in conjunction with 18th International World Wide Web Conference.

**\*Saeedeh Eslami** : Profesora de Ciencia  
Departamento de Tecnologías de la Información y Comunicación, Biblioteca Nacional y  
Archivo de Irán (NLAI)  
Correo: [s-eslami@nlai.ir](mailto:s-eslami@nlai.ir)  
Tf.: +98 21 81622440

*Saeedeh Eslami nació en 1983 en Teherán, Irán. Acabó el título BA en Ingeniería de Software Informático en 2006 y, posteriormente, se graduó en 2010 en Maestría de Ingeniería de Software Informático. Forma parte de la plantilla de NLAI y ha trabajado como especialista en software desde 2005 participando destacadamente en proyectos de software en NALI. Es analista y programadora del Grupo de Diseño y Arquitectura de Software. Es profesora de universidad y enseña en la Universidad Islamic Azad y en NLAI. Su campo de investigación se centra en la Web Semántica con interés específico en el desarrollo de linked data, ontologías y topic maps, también en el desarrollo de software Free Open Source, interoperabilidad, etc. Ha publicado diversos artículos sobre tales temas.*

**Mohammad Hossein Vaghefzadeh** : Profesor de Ciencia  
Subdirector de Investigación, Planificación y Tecnología de la Información, Biblioteca Nacional y Archivo de Irán (NLAI)  
Correo: [m-vaghefzadeh@nlai.ir](mailto:m-vaghefzadeh@nlai.ir)  
Tf.: +98 21 81623270

*Mohammad Hossein Vaghefzadeh nació en 1967 en Teherán, Irán. Se graduó en Maestría de Ingeniería de Software Informático. Fue Director General del Departamento de TIC hasta mayo de 2013, y ahora es adjunto de Investigación y Desarrollo, Planificación y Tecnología de la Información, así como director general de Rahyab, Company in Ports Designing. Es profesor universitario, enseña en diversas universidades. Su campo de investigación se centra en Simulación, Ingeniería Económica y Planificación Estratégica. Ha publicado varios artículos y el libro titulado "Programación con Turbo Pascal" en 199.*

*Idioma de la ponencia: inglés  
Puede contactarse con los autores en:  
[eslami.saeedeh@gmail.com](mailto:eslami.saeedeh@gmail.com) , [s-eslami@nlai.ir](mailto:s-eslami@nlai.ir)*