



Contributos para uma biblioteca do conhecimento

Maria Gorete Barroso Afonso^a, Jorge Gustavo Rocha^b

^a*Universidade Portucalense, Portugal, gorette.afonso@gmail.com*

^b*Universidade do Minho, Portugal, jgr@di.uminho.pt*

Resumo

O objectivo deste trabalho é abrir e disponibilizar o catálogo de uma biblioteca de modo a que o mesmo possa ser apropriado por outros. Para isso, não basta disponibilizar o catálogo na Internet, algo que já está a ser feito. É preciso transformar a informação encerrada no mesmo em conhecimento que possa ser explorado por terceiros, à luz dos conceitos de dados abertos e dados ligados [Berners-Lee 2001], que são as pedras angulares da web semântica. Propõe-se nesta comunicação um fluxo de transformação de um catálogo tradicional de uma biblioteca em recursos da web semântica, representados por triplos RDF (Resource Description Framework) que estabelecem relações entre os oriundos do catálogo e recursos externos provenientes de outras fontes, em particular com o VIAF. Estes triplos estabelecem relações bem conhecidas, isto é, relações definidas em ontologias de grande utilização na web semântica. Estes dados abertos e ligados, com identificadores únicos e universais (URIs), ficam disponíveis para exploração a partir de uma plataforma CKAN instalada para o efeito.

Palavras-chave: Dados Abertos, Web Semântica, Requisitos Funcionais dos Registos Bibliográficos

Introdução

A Web Semântica é uma evolução da web atual, com o propósito de atribuir um significado (semântica) aos conteúdos publicados na Internet de modo que sejam perceptíveis tanto pelo humano como pelo computador. O computador é uma ferramenta excelente para processar grandes volumes de dados, mas é muito limitado na interpretação do significado da informação. Se melhorarmos essa capacidade semântica, o computador será ainda mais eficaz em nos fornecer o que realmente queremos, da forma que queremos. Uma web semântica deveria poder responder a uma questão como: “Queremos fazer umas férias na Toscana este verão, com um orçamento de 4 000 euros. Temos um filho de 8 anos”. Uma questão daria origem a várias pesquisas, que depois teriam que ser combinadas para gerar várias soluções, que correspondessem ao solicitado.

A criação da web semântica pressupõe a integração de mais tecnologias [Synak 2009], como XML, RDF, metadados, ontologias, agentes autónomos e outras que garantam a interoperacionalidade das diversas ferramentas.

Pressupõe também a disponibilização de dados abertos, como um dos pilares da web semântica. Nesse sentido, é importante lembrar alguns dos princípios dos dados abertos. Os dados podem ser considerados abertos quando é permitida a sua livre utilização, reutilização e redistribuição

por qualquer pessoa, eventualmente obrigado a referir a autoria e origem dos mesmos, mas sem qualquer outra limitação.

Devem estar acessíveis a um público o mais amplo possível, documentados e estruturados para possibilitar o seu processamento de forma mais ou menos automática e obedecendo aos seguintes princípios:

Disponibilização e acesso - os dados devem encontrar-se acessíveis como um todo e a um custo não superior ao da sua produção, de preferência acessíveis através da internet. Devem estar num formato adequado.

Reutilização e Redistribuição - os dados fornecidos devem estar disponíveis de forma a permitir a sua reutilização e a junção com outros conjuntos de dados.

Participação - a utilização deve ser permitida a qualquer entidade, seja individual ou colectiva.

Interoperacionalidade - a capacidade de diversos sistemas e organizações poderem trabalhar em conjunto e utilizar diversos múltiplos conjuntos de dados, de acordo com standards.

Inspirados pelos princípios de dados abertos e com o objectivo de potenciar a web semântica, neste trabalho descreve-se o processo de disponibilizar o catálogo bibliográfico como dados abertos, num formato RDF, com ligações a outros conjuntos de dados, e recorrendo a ontologias estabelecidas para dar significado às relações estabelecidas entre os termos dos triplos RDF.

Método

Nesta secção descrevemos o processo de transformação do catálogo existente para RDF. Não se trata de uma mera transformação de formatos. Para além da necessária transformação de formatos, ao cruzar e associar a informação existente com outras fontes, estamos a acrescentar mais informação.

Conversão de MARC para RDF

A generalidade dos catálogos bibliográficos está codificada numa das múltiplas variantes do formato MARC, sendo o MARC 21 e o UNIMARC as mais utilizadas, respectivamente na América do Norte (EUA e Canadá) e na Europa. Estes dois formatos são implementações específicas do standard ISO 2709, que veio substituir o standard ANSI Z39.2 para a troca de informação bibliográfica entre catálogos MARC. A generalidade das ferramentas ILS (Integrated Library System) usadas nas bibliotecas suporta uma destas variantes do formato MARC, e a importação e exportação no formato ANSI Z39.2 e/ou ISO 2709. Como estes standards já sofreram várias revisões, assim como os formatos específicos MARC 21 e UNIMARC, nem sempre é fácil as soluções de ILS suportarem todas as variantes. Isso traduz-se, na prática, em algumas dificuldades quando se quer trocar informação bibliográfica entre diferentes soluções ILS.

No caso da Biblioteca de Montalegre (BMM) não foi possível usar a exportação do catálogo para o formato ISO 2709, devido a pequenas incompatibilidades entre o software utilizado pela biblioteca e as ferramentas de processamento utilizadas a jusante. Após um processo iterativo de tentativa e erro, chegámos à conclusão que a codificação mais flexível era a representação em MARC XML.

UNIMARC para MARC XML

O primeiro passo de transformação consiste em pegar num registo no formato UNIMARC e transformá-lo em MARC XML. No extracto seguinte, transcrevemos literalmente um registo em UNIMARC, tal como é apresentado pelo ILS da BMM.

```
MFN 5950
Leader 00926nam 2200289 450
001
005 20
021 ^aPT^b7013/85
100 ^a20090702d1986 k y0pory0103 ba
101 0 ^apor
102 ^aPT
200 1 ^aMensagem de Fernando Pessoa^fApresentação crítica e linhas de
leitura de Silvina Rodrigues Lopes
205 ^a1ª ed
210 ^aLisboa^cComunicação^d1986
215 ^a98, [5] p.^d19 cm
225 2 ^aTextos literários^v47
600 1^aPessoa^bFernando^f1888-1935
606 ^aLiteratura portuguesa
608 ^aCrítica literária
608 ^aPoesia
675 ^a82.09
675 ^a82-1
700 1^aLopes^bSilvina Rodrigues^f1956-^4450
701 1^aPessoa^bFernando^f1888-1935^4100
801 0^aPT^bBiblioteca Municipal de Montalegre^gRPC
856 ^ulocateMedia.aspx?cod=20090702162643&ext=jpg^zCAPA
966 ^a0011229^1BMM^sSCG 82.09 LOP men 47
909 ^a1
```

O registo anterior segue a tradicional estrutura dos registos MARC, com os campos bibliográficos de 001 a 909. A mesma informação pode ser codificada em XML. Esta codificação introduz marcas explícitas para estruturar hierarquicamente a informação. Em paralelo com a estrutura hierárquica, usam-se pares de atributos e valores para complementar a informação de uma marca. Vamos recorrer a um pequeno exemplo para introduzir a codificação XML. Considere-se o campo 600 (NOME DE PESSOA USADO COMO ASSUNTO) do registo bibliográfico anterior, codificado em UNIMARC, que voltamos a transcrever:

```
600 1^aPessoa^bFernando^f1888-1935
```

Em XML, um campo é indicado com a marca <datafield>. Para indicar que se trata do campo 600, usa-se o atributo tag com o valor "600". Assim sendo, este campo dará origem a algo com a seguinte estrutura:

```
<datafield tag="600"> (...) </datafield>
```

A restante informação segue o mesmo esquema: uma parte é encaixada hierarquicamente dentro da marca <datafield> e outra parte é colocada em atributos. A transformação deste campo dá origem ao seguinte XML:

```
<datafield tag="600" ind1=" " ind2="1">
<subfield code="a">Pessoa</subfield>
<subfield code="b">Fernando</subfield>
<subfield code="f">1888-1935</subfield>
</datafield>
```

Nesta transformação, para além do atributo tag="600", associado à marca <datafield> surgem também os atributos ind1=" " e ind2="1". Cada campo tem dois indicadores, embora estes nem sempre sejam muito evidentes na representação textual do UNIMARC, até porque o primeiro destes indicadores é um espaço (um caráter não visível). Para o campo 600, o indicador 1 deve ser branco (não definido) e o indicador 2 pode ser 0 ou 1.

Os subcampos são anotados com a marca <subfield>, com o atributo code a ser utilizado para indicar o tipo de subcampo. No exemplo, são utilizados os subcampos "a", "b" e "f".

Aplicando estes princípios de transformação a todo o registo, obtém-se o XML que se segue.

```
<?xml version="1.0" encoding="UTF-8"?>
<record>
<leader>00926nam 2200289 450</leader>
<controlfield tag="001">5950</controlfield>
<controlfield tag="003">
>http://montalegre.opacnet.com/OPAC/default.aspx?ContentAreaControl=Sh
owRegHoldings.aspx&RegNo=16662</controlfield>
<controlfield tag="005">20</controlfield>
<datafield tag="021" ind1=" " ind2=" ">
<subfield code="a">PT</subfield>
<subfield code="b">7013/85</subfield>
</datafield>
<datafield tag="100" ind1=" " ind2=" ">
<subfield code="a">20090702d1986 k y0pory0103 ba</subfield>
</datafield>
<datafield tag="101" ind1="0" ind2=" ">
<subfield code="a">por</subfield>
</datafield>
<datafield tag="102" ind1=" " ind2=" ">
<subfield code="a">PT</subfield>
</datafield>
<datafield tag="200" ind1="1" ind2=" ">
<subfield code="a">Mensagem de Fernando Pessoa</subfield>
<subfield code="f">Apresentação crítica e linhas de leitura de Silvina
Rodrigues
Lopes</subfield>
</datafield>
<datafield tag="205" ind1=" " ind2=" ">
<subfield code="a">1ª ed</subfield>
</datafield>
<datafield tag="210" ind1=" " ind2=" ">
<subfield code="a">Lisboa</subfield>
<subfield code="c">Comunicação</subfield>
<subfield code="d">1986</subfield>
</datafield>
<datafield tag="215" ind1=" " ind2=" ">
<subfield code="a">98, [5] p.</subfield>
<subfield code="d">19 cm</subfield>
</datafield>
<datafield tag="225" ind1="2" ind2=" ">
<subfield code="a">Textos literários</subfield>
<subfield code="v">47</subfield>
</datafield>
<datafield tag="600" ind1=" " ind2="1">
<subfield code="a">Pessoa</subfield>
<subfield code="b">Fernando</subfield>
<subfield code="f">1888-1935</subfield>
</datafield>
<datafield tag="606" ind1=" " ind2=" ">
```

Contributos para uma biblioteca do conhecimento

```
<subfield code="a">Literatura portuguesa</subfield>
</datafield>
<datafield tag="608" ind1=" " ind2=" ">
<subfield code="a">Critica literária</subfield>
</datafield>
<datafield tag="608" ind1=" " ind2=" ">
<subfield code="a">Poesia</subfield>
</datafield>
<datafield tag="675" ind1=" " ind2=" ">
<subfield code="a">82.09</subfield>
</datafield>
<datafield tag="675" ind1=" " ind2=" ">
<subfield code="a">82-1</subfield>
</datafield>
<datafield tag="700" ind1=" " ind2="1">
<subfield code="a">Lopes</subfield>
<subfield code="b">Silvina Rodrigues</subfield>
<subfield code="f">1956-</subfield>
<subfield code="4">450</subfield>
</datafield>
<datafield tag="701" ind1=" " ind2="1">
<subfield code="a">Pessoa</subfield>
<subfield code="b">Fernando</subfield>
<subfield code="f">1888-1935</subfield>
<subfield code="4">100</subfield>
</datafield>
<datafield tag="801" ind1=" " ind2="0">
<subfield code="a">PT</subfield>
<subfield code="b">Biblioteca Municipal de Montalegre</subfield>
<subfield code="g">RPC</subfield>
</datafield>
<datafield tag="856" ind1=" " ind2=" ">
<subfield
code="u">locateMedia.aspx?cod=20090702162643&amp;ext=jpg</subfield>
<subfield code="z">CAPA</subfield>
</datafield>
<datafield tag="909" ind1="&#9;" ind2=" ">
<subfield code="a">1</subfield>
</datafield>
<datafield tag="966" ind1=" " ind2=" ">
<subfield code="a">0011229</subfield>
<subfield code="l">BMM</subfield>
<subfield code="s">SCG 82.09 LOP men 47</subfield>
</datafield>
</record>
```

Esta mudança de codificação não introduz nem causa qualquer perda de informação. É apenas uma mudança de codificação.

No entanto, a codificação XML é lida e processada por uma enorme diversidade de ferramentas. Em particular, esta codificação permite-nos utilizar estilos XSLT (EXtensible Stylesheet Language Transformations) para transformações sucessivas, mantendo a codificação XML.

Para esta transformação, usámos a interface de consulta disponível ao cidadão, que tem uma vista Unimarc. O programa de transformação usa essa vista para apresentar os registos, e extraindo a informação da vista apresentada (HTML) é retirada a informação que interessa e gerado o XML. Usando esta técnica de parsing da página HTML apresentada, podemos aplicar todo o processamento que fizemos com os dados de qualquer outra biblioteca.

Enriquecimento da representação

A transformação descrita anteriormente, a primeira, é meramente sintática. A partir desta, irão ser aplicadas transformações sucessivas, mas que já não são meramente sintáticas: usam novas regras para organizar a informação e acrescentam valor.

O modelo relacional FRBR

A segunda transformação trata de converter uma sequência de registos MARC em entidades conceptualmente mais próximas do leitor, como seja a noção de obra, autor, etc. As entidades escolhidas são as contempladas no modelo FRBR proposto pela IFLA.

O modelo FRBR (Functional Requirements for Bibliographic Records) é um modelo conceptual desenvolvido pela IFLA para representar registos bibliográficos. Este modelo rompe com a tradição de recorrer a um formato da família MARC, a favor do modelo de entidades e relacionamentos, vulgar em aplicações informáticas.

Esta segunda transformação consiste, por isso, em transformar uma sucessão de registos MARC em 4 tipos de entidades distintas, que se relacionam entre si. As entidades que se consideraram relevantes para este exercício são: Pessoa, Obra, Expressão e Manifestação.

Após esta segunda transformação, obtém-se um conjunto de registos, já divididos em 4 tipos diferentes. Por exemplo, um registo do tipo autor, passa a ter a seguinte configuração:

```
<record f:id="8d725ed7-6128-359f-adc8-85fe4995b8bc"
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/C1005">
<datafield tag="700" ind1=" " ind2="1">
<subfield code="a"
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P3039">Cervantes</
subfield>
<subfield code="b"
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P3039">Miguel
de</subfield>
<subfield code="f"
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P3040">1547-
1616</subfield>
</datafield>
<f:relationship
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
f:href="b606bdc6-8a10-318d-9eed-de998e4d9d37"/>
<f:relationship
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
f:href="fc082a3e-935a-334e-8e4a-e945d6d02589"/>
<f:relationship
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
f:href="14f3462f-a9ce-3501-ad1f-4d4c9192e623"/>
<f:relationship
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
f:href="875fcae2-988e-30bd-919d-2b18cceb44ae"/>
<f:relationship
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
f:href="80ca8011d-39ce-3c83-a8f4-67c77babedea"/>
<f:relationship
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
f:href="80ca5807-93da-31e0-83f7-afa4ef0f92e0"/>
<f:relationship
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
f:href="fa9b8782-6eb7-358b-8f3e-7048b93b136e"/>
<f:relationship
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
```

```
f:href="e753b7b8-191c-3694-a093-cf61c30208f4"/>
<f:relationship
f:type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
f:href="b0f5ec55-0829-3f09-b7e9-2831ed38f2dc"/>
<f:label>Miguel de Cervantes</f:label>
</record>
```

Como se pode observar, o registo é unicamente sobre a informação que existe sobre o autor. Para já, a informação sobre o mesmo é mínima, mas será enriquecida a jusante. Para além da pouca informação sobre o autor, o registo tem as relações que esta entidade autor tem com as suas obras.

Isto exemplifica, de uma forma simplificada, o resultado do segundo passo de transformação. Passamos a ter a informação mais organizada, separada em quatro conceitos relacionados entre si, como aparece representado na Figure 1.

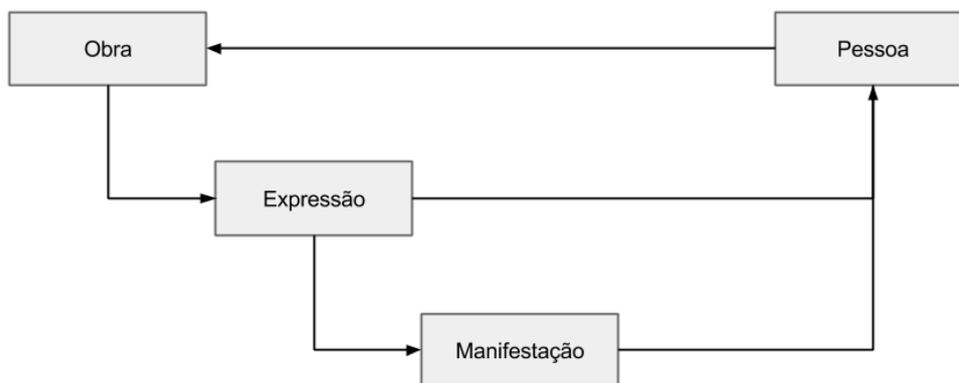


Figure 1 -Informação bibliográfica organizada segundo o modelo relacional FRBR.

A transformação do catálogo em formato XML para este modelo relacional FRBR foi conseguida através de um estilo XSLT que foi gerado a partir de um conjunto de regras.

A título de exemplo, as regras utilizadas para recuperar a informação sobre um autor são as seguintes:

```
<entity type="http://iflstandards.info/ns/fr/frbr/frbrer/C1005"
label="Person" templatename="MARC21-700-Person">
<note>Criador gerado a partir da informação do campo 700</note>
<anchor tag="700"/>
<attributes>
<datafield tag="700">
<subfield code="b"
type="http://iflstandards.info/ns/fr/frbr/frbrer/P3039" label="has
name of person"/>
<subfield code="a"
type="http://iflstandards.info/ns/fr/frbr/frbrer/P3039" label="has
name of person"/>
<subfield code="f"
type="http://iflstandards.info/ns/fr/frbr/frbrer/P3040" label="has
dates of person"/>
<!-- <subfield code="e"/> -->
</datafield>
</attributes>
<!-- relacionamento entre person e work -->
<relationships>
<relationship type="http://iflstandards.info/ns/fr/frbr/frbrer/P2010"
itype="http://iflstandards.info/ns/fr/frbr/frbrer/P2009">
```

```

<target entity="MARC21-200-Work"/>
</relationship>
</relationships>
<label select="string-join( (*:datafield[@tag='700']/*:subfield[@code
= 'b'], *:datafield[@tag='700']/*:subfield[@code = 'a']], ' ' )"/>
<!-- gerar uma chave para única para Person: nome próprio x apelido x
data -->
<key order="1">
<element>*:datafield[@tag = '700'][1]/*:subfield[@code =
'b'][1]</element>
<element>*:datafield[@tag = '700'][1]/*:subfield[@code =
'a'][1]</element>
<element>*:datafield[@tag = '700'][1]/*:subfield[@code =
'f'][1]</element>
</key>
</entity>

```

Embora a sintaxe do XML possa intimidar, estas regras dizem claramente o que se pretende retirar do catálogo para criar uma entidade autor. Neste caso, indicamos que queremos ir buscar os subcampos a, b e f do campo 700. O mesmo tipo de regras (um pouco mais extensas) foi utilizado para recuperar a informação relativa a obras, manifestações e expressões. No exemplo seguinte, ilustram-se as regras utilizadas para recuperar uma manifestação:

```

<entity type="http://iflstandards.info/ns/fr/frbr/frbrer/C1003"
label="Manifestation" templatename="MARC21-001-Manifestation">
<note>Manifestation</note>
<anchor tag="001"/>
<attributes>
<controlfield tag="003"/> <!-- N° interno de registo substring-
after("http://montalegre.opacnet.com/OPAC/default.aspx?ContentAreaCont
rol=ShowRegHoldings.ascx&RegNo=16662", "RegNo=") -->
<datafield tag="010"> <!-- ISBN -->
<subfield code="a" />
</datafield>
<datafield tag="021"> <!-- Depósito legal -->
<subfield code="a"/>
<subfield code="b"/>
</datafield>
<datafield tag="200"> <!-- Título -->
<subfield code="a"/>
</datafield>
<datafield tag="205"> <!-- Sobre a edição, p. ex "2a ed" -->
<subfield code="a"/>
</datafield>
<datafield tag="210"> <!-- Editora -->
<subfield code="a"/>
<subfield code="c"/>
<subfield code="d"/>
</datafield>
<datafield tag="215"> <!-- características físicas -->
<subfield code="a"/>
<subfield code="d"/>
</datafield>
<datafield tag="225"> <!-- séries -->
<subfield code="a"/>
<subfield code="v"/>
</datafield>
</attributes>
<label select="string-join(*:datafield[@tag = '245']/*:subfield[@code
= ('a', 'b', 'c')], ' / ' )"/>
<key order="1">

```

```
<element>if (*:datafield[@tag="010"][1]/*:subfield[@code='a'][1]) then
*:datafield[@tag="010"][1]/*:subfield[@code='a'][1] else
*:controlfield[@tag='001']</element>
</key>
<relationships/>
</entity>
```

No caso das manifestações, temos que ir recuperar informações a uma maior diversidade de campos. Este trabalho de criação de regras só é possível conhecendo bem o sistema de catalogação utilizado em Portugal, para se poder fazer o mapeamento para o modelo FRBR.

A partir destas regras, é gerado um estilo XSLT que percorre todo o catálogo em XML e gera os diferentes registos para autores, obras, expressões e manifestações. O estilo XSLT foi gerado recorrendo ao projecto de código aberto **MARC to FRBR**, disponibilizado na plataforma Source Forge. Este projecto foi desenvolvido por Trond Aalberg, da Universidade de Ciência e Tecnologia da Noruega (NTNU). Desta forma, só nos concentrámos na escrita das regras, que são o essencial desta transformação, e o código da transformação em si foi gerado pela ferramenta.

O resultado desta transformação deu origem a um novo documento XML, records.xml. Para facilitar o processamento posterior, para além deste documento com toda a informação, para cada uma das entidades também foi gerado um documento XML distinto (autor.xml, obra.xml, expressao.xml e manifestacao.xml).

Cruzamento com o VIAF

Em dois passos, conseguimos passar de um catálogo baseado no formato MARC para uma organização da informação dividida por entidades, entre as quais estão definidas relações.

Agora que temos a informação organizada de uma forma mais conveniente, vamos enriquecer cada uma das entidades. Este passo cria uma ponte entre os conteúdos enclausurados no catálogo (um sistema fechado) e a multiplicidade de conteúdos que existem fora do catálogo. É uma ponte fundamental na senda da biblioteca do conhecimento pois o conhecimento faz-se a partir de relações. Queremos demonstrar que o acervo da biblioteca fica mais rico quando ligado ao resto da informação que já existe fora desta biblioteca e fora do mundo das bibliotecas.

Neste exercício de enriquecimento, tirámos partido da informação que já existe em viaf.org. Este VIAF começou por ser um registo de autoridades conjunto de três grandes bibliotecas - United States Library of Congress (LC), Deutsche Nationalbibliothek (DNB) e Bibliothèque Nationale de France (BnF) – que depois se abriu à participação de muitas outras bibliotecas. No final de 2014, o VIAF agregava registo de 35 bibliotecas nacionais. Também participa neste registo conjunto a Biblioteca Nacional de Portugal, que contribui com o registo nacional de autoridades.

O cruzamento foi feito à custa de um robot que pesquisa cada um dos recursos no VIAF. Se a pesquisa for bem sucedida, estabelece-se uma ligação entre o nosso recurso e os recursos encontrados no VIAF. Dado o carácter mais técnico deste passo, optamos por uma ilustração deste enriquecimento na secção seguinte, em detrimento de uma descrição do processo informático.

Resultados

Do processo de transformação e enriquecimento descrito anteriormente resultaram conjuntos de dados referentes a obras, expressões, manifestações e autores. Cada um destes conjuntos de dados é disponibilizado num documento RDF.

Nesta seção aproveitamos para ilustrar, a partir do exemplo dos autores, como funciona o enriquecimento e qual é o resultado do mesmo.

É importante perceber o propósito deste VIAF, já que o mesmo está em linha com o mesmo propósito de servir de rampa de lançamento para a web semântica. Nesse sentido, este registo de autoridade tenta criar um único registo, com um identificador único, que agrega os vários registos existentes nas bibliotecas nacionais, referentes à mesma entidade, permitindo a coexistência de variações na língua, formas de escrita e de abreviatura.

Para exemplificar o propósito deste VIAF, ilustra-se o resultado da pesquisa pelo autor Miguel de Cervantes, na Figure 2.

The screenshot shows the VIAF (Virtual International Authority File) search interface. At the top, it says 'VIAF Virtual International Authority File'. Below that is a search box with three fields: 'Select Field:' (set to 'All Headings'), 'Select Index:' (set to 'All VIAF'), and 'Search Terms:' (containing 'Miguel de Cervantes Saavedra 1547 1616'). A 'Search' button is to the right. Below the search box, it states '742 headings found for Miguel de Cervantes Saavedra 1547 1616'. A table of results is shown with columns 'Heading' and 'Type'. The first result is '1 Cervantes Saavedra, Miguel de 1547-1616' with a 'Personal' type. The heading is followed by various language and script versions of the name, each with a small icon representing the language or script.

Heading	Type
1 Cervantes Saavedra, Miguel de 1547-1616 	Personal

Figure 2 - Registo VIAF para o autor Miguel de Cervantes

Como se pode verificar no registo VIAF, o autor a quem chamamos Miguel de Cervantes, na verdade, no seu país de origem é identificado com um nome ligeiramente diferente. Todas as variantes apresentadas referem-se ao mesmo autor. Por esta razão, na prossecução de uma web semântica, interessa-nos remeter todas as múltiplas ocorrências das variantes de Miguel de Cervantes para o mesmo registo. Dessa forma, os mecanismos informáticos que manipulam a componente semântica saberão que se trata do mesmo conceito.

A existência de um registo de autoridade como o VIAF facilitou imenso este trabalho, assim como alavanca muitos outros alinhados com a criação de uma web semântica. A nossa rotina de enriquecimento da informação associada aos autores baseia-se na consulta e cruzamento da informação que temos no catálogo com a informação existente no VIAF. Como a Biblioteca Nacional de Portugal também participa neste registo de autoridades agregador, a probabilidade de encontrarmos as entradas do nosso catálogo nesse registo de autoridades é grande. Essa consulta e cruzamento foi automatizado, com uma rotina que percorre todos os nossos registos de autores e faz uma procura do mesmo em viaf.org. Para cada autor, é recuperado o seu viafid. Tendo esse viafid, a pesquisa seguinte já se faz por esse identificador.

Após correr a rotina que recolhe a informação complementar para descrever os autores, o registo do autor passa a ter uma informação complementar essencial para se ligar o nosso registo com o mundo exterior.

```
<record f:id="8d725ed7-6128-359f-adc8-85fe4995b8bc"
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/C1005">
<datafield tag="700" ind1=" " ind2="1">
<subfield code="a"
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P3039">Cervantes</
subfield>
<subfield code="b"
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P3039">Miguel
de</subfield>
<subfield code="f"
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P3040">1547-
1616</subfield>
</datafield>
<f:relationship
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P2010"
f:href="b606bdc6-8a10-318d-9eed-de998e4d9d37"/>
<f:relationship
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P2010"
f:href="fc082a3e-935a-334e-8e4a-e945d6d02589"/>
<f:relationship
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P2010"
f:href="14f3462f-a9ce-3501-ad1f-4d4c9192e623"/>
<f:relationship
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P2010"
f:href="875fcae2-988e-30bd-919d-2b18cceb44ae"/>
<f:relationship
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P2010"
f:href="80a8011d-39ce-3c83-a8f4-67c77babedea"/>
<f:relationship
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P2010"
f:href="80ca5807-93da-31e0-83f7-afa4ef0f92e0"/>
<f:relationship
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P2010"
f:href="fa9b8782-6eb7-358b-8f3e-7048b93b136e"/>
<f:relationship
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P2010"
f:href="e753b7b8-191c-3694-a093-cf61c30208f4"/>
<f:relationship
f:type="http://iflastandards.info/ns/fr/frbr/frbrer/P2010"
f:href="b0f5ec55-0829-3f09-b7e9-2831ed38f2dc"/>
<f:label>Miguel de Cervantes</f:label>
<extra>
<ptbnp>44322</ptbnp>
<viaf>17220427</viaf>
<wikipedia>http://en.wikipedia.org/wiki/Miguel_de_Cervantes</wikipedia
>
<lccn>http://www.worldcat.org/wcidentities/lccn-n79100233</lccn>
<bne>http://datos.bne.es/autor/XX1718747</bne>
<bnf>http://catalogue.bnf.fr/ark:/12148/cb118957747</bnf>
</extra>
</record>
```

Veja-se, por exemplo, que conseguimos associar este autor à sua página na Wikipédia, assim como liga-lo à BNE e BNF.

Discussão

Enquanto os conteúdos das bibliotecas estiverem confinados a soluções proprietárias e em formatos como o MARC, pouco adequados ao tratamento informático, dificilmente alguém tira partido dos mesmos e lhes acrescenta valor. Neste trabalho apenas nos preocupamos em transformar e enriquecer os dados bibliográficos para potenciar a sua utilização, disponibilizando-os numa plataforma CKAN.

Para que realmente alguém se possa apropriar, tirar partido e acrescentar valor aos dados disponibilizados, é preciso decidir qual a forma mais adequada de disponibilização destes dados.

Havendo vários caminhos possíveis para a prossecução deste objectivo, fizemos um estudo das práticas que se vêem nas grandes bibliotecas, para encontrar as melhores práticas e partilhar as mesmas.

Conseguimos de uma forma o mais automatizada possível¹ transformar um catálogo MARC em RDF onde os recursos aparecem relacionados com outros recursos online. Julgamos que desta forma conseguimos concretizar algumas das realizações que eram esperadas pelo consórcio W3C [W3C 2011] no que se refere a registos bibliográficos.

Conclusões

As bibliotecas vão-se transformando e evoluindo para responderem às necessidades da sociedade envolvente. Vivemos agora num mundo digital onde o papel da biblioteca tem que ser constantemente reinventado.

Para se aproveitar todo o valor da informação das bibliotecas, e em particular do seu catálogo e registos de autoridade, é preciso abrir e disponibilizar essa informação. Este pequeno passo é fundamental para se potenciar a web semântica e construir uma sociedade do conhecimento.

Neste trabalho descrevemos todo um processo de transformação e enriquecimento de dados bibliográficos em dados abertos, de acordo com os princípios de dados abertos. A metodologia de transformação baseou-se numa sequência de pequenas transformações que se podem replicar em qualquer outra biblioteca.

Os dados transformados e enriquecidos, estão disponibilizados no formato RDF, a partir de uma plataforma CKAN. Escolheu-se esta plataforma por ser a mais conhecida plataforma de disponibilização de dados abertos.

Os dados disponibilizados são imediatamente manipuláveis através de ferramentas da web semântica. É possível, por exemplo, explorar os dados através da linguagem SPARQL.

¹ Código disponível em <https://github.com/jgrocha/BiblioCognosis>.

Referências

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*. doi:10.1038/scientificamerican0501-34

Synak, M., Dabrowski, M., & Kruk, S. R. (2009). Semantic web and ontologies. In *Semantic Digital Libraries* (pp. 41–54). doi:10.1007/978-3-540-85434-0_3

W C Library Linked Data Incubator Group. (2011). *Library Linked Data Incubator Group: Use Cases*. Library. Retrieved from <http://www.w3.org/2005/Incubator/lld/XGR-lld-usecase-20111025/>