



Perspectivas de investigação

Um modelo de migração de dados UNIMARC para repositórios Linked Open Data

António Tavares Lopes

Universidade de Coimbra,
Faculdade de Letras
Portugal · atlopes@fl.uc.pt

Resumo: Um importante desafio se coloca ao controlo bibliográfico perante as oportunidades da Web Semântica, enquanto potencial fonte de informação para conjuntos de dados ligados: como produzir *linked open datasets* que sejam semanticamente expressivos e ricos de ligações? Este desafio associa-se, do ponto de vista operacional, a um outro: como realizar esse potencial a partir dos atuais e tradicionais produtos do controlo bibliográfico, sem alterar o fundamental das suas práticas e saber fazer mas, ao mesmo tempo, sendo capaz de integrar e interagir com desenvolvimentos que lhe são exteriores?

Em resposta a estas duas questões, este trabalho apresenta um modelo leve de expressão de dados bibliográficos MARC como *linked open data* e um exemplo da sua aplicação em dados reais em formato UNIMARC. A construção do modelo apoia-se nas tecnologias XML para a esquematização, expressão e transformação de dados, ainda que permitindo o recurso a outras tecnologias, quando a capacidade de transformação XSL não é suficiente.

No coração deste modelo está colocado um esquema que define o conjunto de elementos da norma bibliográfica. Baseando-se neste esquema, identificam-se, para cada um dos elementos que constitui o formato, os processadores que lhe estão associados e que têm funcionalidade variável: de identificação; de extração; de validação; de resolução; e de ligação. Tipicamente, estes processadores são constituídos por folhas de transformação mas que integram – em particular os processadores de ligação – o recurso a serviços Web para enriquecimento das conexões do conjunto de dados final.

Esta abordagem permite definir uma cadeia automatizada de processamento que parte da esquematização da norma bibliográfica para construir um controlador dos processadores que são aplicados a uma coleção de registos bibliográficos e que geram, no final, um repositório de *linked data*. Permite, igualmente, alargar as capacidades de expressão e ligação destes dados, através da adição de novos elementos e processadores à instância do esquema.

Palavras-chave: Web Semântica; Dados ligados; Controlo bibliográfico; UNIMARC; Modelo de migração.

Abstract: The Semantic Web movement poses an important challenge to bibliographic control, as a potential source for linked information: how to produce linked open datasets that are semantically expressive and rich in connectivity? Also, how to satisfy that goal by setting the current and traditional bibliographic products as a starting point, without changing the fundamentals of practices and know-how of bibliographic agencies, while integrating and interacting with outside developments?

This paper tries to address these two questions by presenting a lightweight model for the expression of MARC-based bibliographic data as linked open data and an example of its application to UNIMARC data. The model is built upon XML technologies for the formalization, expression and transformation of data, although allowing to resort to other technologies when the capabilities of the XSL transformation are not sufficient.

At the core of the model resides a schema that hosts the formalization of the bibliographic format element set. An instance of the schema associates the element set to processors that execute different types of functional actions: identification, extraction, validation, resolution, and linking. These processors are, typically, transformation stylesheets that may invoke Web Services, especially in the case of the linking processors that have the purpose of establishing semantic relations with other linked datasets.

This approach defines an automatized processing chain, built on the formalization of a bibliographic format and on the automatic generation of a processing controller, which ultimately produces a linked data repository. It also allows to extend the capabilities of the expression and connectivity of the data, by incorporating new elements and processors to the schema instance.

Keywords: Semantic Web; Linked Data; Bibliographic Control; UNIMARC; migration Model.

Introdução

A construção da Web Semântica constitui uma oportunidade para os agentes do controlo bibliográfico intensificarem as trocas de valor com outros domínios de representação da informação. Mesmo antes da Web, essas trocas de valor estavam longe de ser incomuns e faziam parte, como ainda fazem, da matriz de princípios do controlo bibliográfico. A definição de normas bibliográficas tem-se mostrado disponível para incorporar e hospedar, de forma convergente, códigos, linguagens e domínios do saber e do discurso de origem diferenciada. Essa incorporação resulta, efetivamente, no estabelecimento de conexões reais com outros esquemas e sistemas de descrição, identificação, representação e organização da informação:

- códigos, como os ISO-3166 para identificação de países ou os mantidos pela International Association of Music Libraries, Archives and Documentation Centres para a descrição de vozes e instrumentos usados na interpretação de uma obra musical;
- identificadores, como o ISBN para monografias ou o EAN para produtos vendido a retalho;
- sistemas de identificação de conteúdo, como a Classificação Decimal Universal ou os cabeçalhos de assunto da Library of Congress;
- normas descritivas e de catalogação, como a ISBD, as AACR2 ou as Regras Portuguesas de Catalogação.

A consciência desta capacidade para estabelecer conexões esteve presente na redação do relatório final do Library Linked Data Incubator Group, que reconhece nos princípios dos dados ligados uma “extensão natural dos modelos de partilha colaborativos, historicamente postos em prática pelas bibliotecas” (Baker et al., 2011). Estes *modelos de partilha* devem ser entendidos de duas formas, distintas mas complementares: por um lado, a troca de informação entre as instituições da memória e da organização do conhecimento cultural e científico, permitida e estimulada pela adoção de formatos comuns de representação; por outro, a incorporação de valor que resulta da aplicação de instrumentos exógenos, estabelecendo pontes de comunicação com outros agentes processadores que ultrapassam o restrito domínio das bibliotecas e agências bibliográficas. Neste trabalho, é sobretudo a segunda perspetiva que me interessa desenvolver.

A designação *modelo de migração* não é inócua. Trata-se da representação de um processo que parte dos atuais produtos do controlo bibliográfico para os colocar no espaço operacional da Web Semântica. Expresso desta forma, o sistema alvo do modelo é beneficiário de uma transferência de valor assumidamente desigual que, por isso e no contexto histórico do controlo bibliográfico, constitui apenas um início de negócio: migrar informação significa que a origem não recebe qualquer valor adicional direto. No futuro, o fluxo de informação deverá ocorrer nos dois sentidos mas, por enquanto, as vantagens são sobretudo prospetivas (Byrne & Goddard, 2010) e valores acrescentados correntes serão indiretos. Procurarei enunciar alguns dos mais evidentes e que motivam e justificam, ainda assim, o esforço necessário.

Disponibilizar repositórios de dados bibliográficos ligados alarga o potencial de inserção dos seus produtores em cadeias de valor da indústria da informação. Assentar a formatação dessa informação em protocolos e normas abertas e partilhadas permite responder à missão que a maior parte das bibliotecas abraça (Bérard, 2011), e simultaneamente evita o criticismo que se levanta relativamente à

privatização dessa informação, decorrente de acordos restritos e restritivos com empresas, como a Google (Darnton, 2009; Waller, 2009).

O desenvolvimento de capacidades para compreender, arquitetar e representar *linked open data*, torna mais próximas as capacidades para a processar e integrar. Esta capacidade de integração não tem por objetivo satisfazer o equilíbrio da balança entre produção e consumo, e atingir desta forma um saldo zero no trânsito de migrações, mas sim o de conseguir dar aos utilizadores dos sistemas de informação bibliográfica novos instrumentos para a sua descoberta, exploração e consumo (Coyle, 2012; Willer & Dunsire, 2013).

Finalmente, a modelação da Web Semântica obriga a repensar e a perspetivar criticamente os instrumentos para o processamento do controlo bibliográfico, a perceber-lhes os limites, formais e conceptuais, as contradições e inconsistências, e a prepará-los para novos contextos operativos (Alemu, Stevens, Ross, & Chandler, 2012). Tal necessidade da reavaliação não é nova e já se manifestou impactante em outros momentos, ainda antes da formulação da Web Semântica como conceito e prática. O desenvolvimento dos FRBR representa, precisamente, o mais significativo momento de reavaliação induzido por fatores externos, na convergência dos objetivos e valores do controlo bibliográfico com a evolução dos sistemas de informação, bebendo destes novos conceitos e terminologias, em particular na área da modelação dos sistemas de bases de dados (IFLA, 2009; Madison, 2005).

1. Princípios de desenho do modelo de migração

Os objetivos do modelo de migração que apresento são balizados por um alvo e por um constrangimento que implicitamente transporta um propósito de economia.

O alvo é a produção de repositórios de dados ligados de natureza bibliográfica, semanticamente expressivos e ricos de ligações. Acrescentar estas características – expressividade semântica e riqueza de ligações – pode parecer redundante, quando se fala de *linked open data* e de *linked open datasets*. Mas, ao salientá-las, pretendo, por um lado, destacar a necessidade de definir uma estrutura robusta de vocabulários e ontologias, criados originalmente ou apropriados de outros instrumentos de organização do conhecimento, e que sustentem a elaboração dos repositórios e a expressão dos factos descritos. Por outro, relevar a importância do estabelecimento imperativo de ligações, ao nível de conceitos e de dados, que alarga a usabilidade e o campo de aplicação dos dados expressos, permitindo a agentes semânticos inferir conhecimento a partir das expressões de outros domínios de discurso ou conjuntos de dados.

O constrangimento que delimita o desenho do modelo de migração é o de a sua operacionalização se poder realizar a partir dos produtos do controlo bibliográfico correntes, sem requerer alterações ao que são as práticas e o saber fazer dos agentes do controlo bibliográfico. Esta imposição do desenho responde à necessidade de não fazer pesar significativamente o processo de migração na economia das agências bibliográficas, visto na perspetiva de consumo dos seus recursos e de alinhamento às suas missões. Tal intenção é entendida numa perspetiva abrangente, procurando apresentar uma solução leve, de fácil desenvolvimento, distribuição e aplicação, e de reduzidas exigências de gestão.

Estes limites procuram enquadrar-se nos roteiros para a expressão de informação bibliográfica na Web Semântica que têm sido apresentados, e que ganharam força de manifesto programático nas recomendações do relatório do Library Linked Data Incubator Group (Baker et al., 2011). Outros roteiros, de caráter mais operativo, programáticos ou reportando casos de estudo, são também significativos credores da orientação dada a este trabalho, no que respeita ao planeamento (Willer & Dunsire, 2013), aos processos executivos (Dunsire, 2012; Hannemann & Kett, 2010; Vila-Suero & Gómez-Pérez, 2013) e a expressões de normas bibliográficas no espaço da Web Semântica (Dunsire & Willer, 2011; Dunsire, 2009; Styles, Ayers, & Shabir, 2008).

Genericamente, o desenho deste modelo procura satisfazer, ou contribuir para a satisfação, de três propósitos que estes roteiros identificam:

- definição de conjuntos de elementos;
- identificação e construção de vocabulários de valores relevantes;
- produção de conjuntos de dados ricos de ligações.

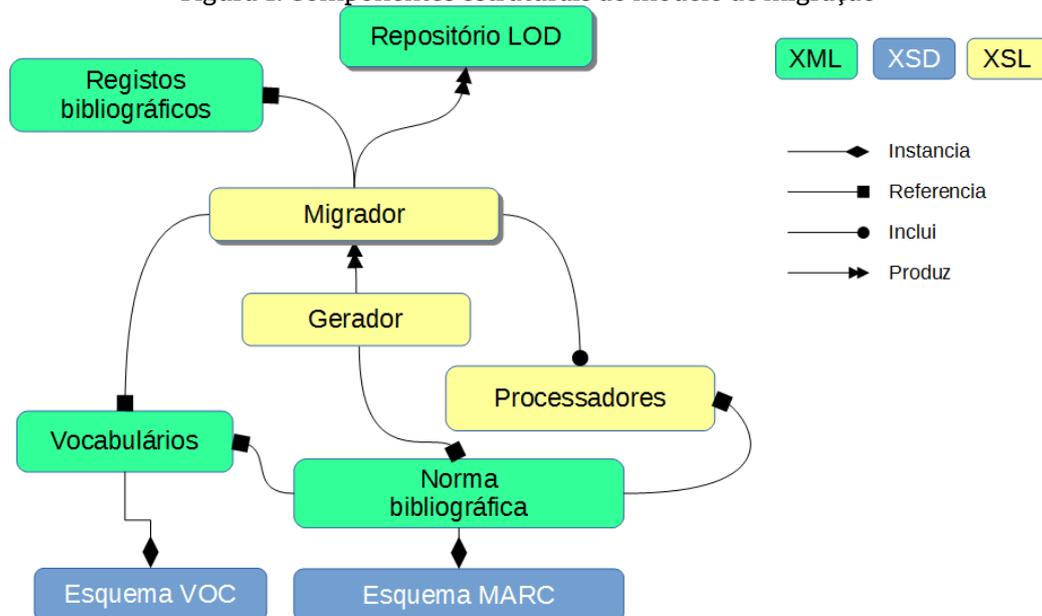
Sob estes princípios, este trabalho apresenta um modelo de expressão de dados bibliográficos MARC como *linked open data* (LOD) e um exemplo da sua aplicação em dados reais em formato UNIMARC.

2. Modelo para o processamento de dados MARC orientado para a Web Semântica

A construção do modelo de migração apoia-se nas tecnologias baseadas em XML (Bray et al., 2006) para a expressão e transformação de dados MARC. As potencialidades do recurso ao XML para processamento de informação bibliográfica têm sido geralmente salientada em diversos estudos e projetos, e o corpo de tecnologias XML, como é comum nos atuais sistemas de informação, tem sido integrado em produtos e serviços bibliográficos de forma central ou periférica (Aalberg, 2006; Carvalho & Cordeiro, 2002; Carvalho, 2005; Dimic, Milosavljevic, & Surla, 2010; Keith, 2004). Em particular, a capacidade de integração que o XML apresenta para a representação e o processamento de estruturas de dados complexas, correspondente ao seu propósito inicial, a um baixo custo e a operar num ambiente aberto e distribuído (Carvalho & Cordeiro, 2002).

A arquitetura do modelo de migração está esquematizada na Figura 1, que identifica os componentes (ou tipos de componentes) e as suas relações funcionais.

Figura 1: Componentes estruturais do modelo de migração



Na base do modelo encontra-se um esquema, definido com recurso à linguagem XML Schema (Gao, Sperberg-McQueen, & Thompson, 2012), comumente identificada por XSD, e que define a cadeia de processamento a que os dados estão sujeitos. O propósito deste esquema não é o de estruturar formalmente a informação, ou de representar em XML o formato MARC que está na base da informação bibliográfica que se pretende migrar, mas antes o de estruturar o processo da migração.

Uma instância do esquema, ou seja, a expressão sobre o esquema de uma norma bibliográfica como o MARC21 ou o UNIMARC, identifica o conjunto de elementos designadores de conteúdo estipulados pela norma e que serão objeto de migração, associando-lhes processadores que executam as funções necessárias a essa tarefa.

Para cada elemento, a instância do esquema identifica a fonte da informação através da especificação do tipo de designador de conteúdo e a respetiva localização:

etiqueta de registo, campos de controlo e de dados, subcampos, indicadores, ou uma composição (isto é, todo o registo).

Os elementos definidos na instância do esquema não têm de respeitar ou de seguir a granularidade da formatação dos designadores de conteúdo da norma bibliográfica de origem. Designadores de conteúdo que têm um caráter funcional e não semântico – como, por exemplo, um indicador que estipula a forma de uma nota – podem ser ignorados na produção do repositório LOD, em termos de declarações de informação. Adicionalmente, mais do que uma perspectiva de processamento se pode aplicar a um mesmo conjunto de segmentos de informação que se encontram reunidos sob uma mesma unidade de conteúdo, como acontece a um campo que reúne subcampos. A ênfase da aplicação do esquema é dada ao nível do modelo conceptual da norma bibliográfica, servindo o modelo lógico (o formato MARC e a atribuição de valor semântico aos diversos componentes da estrutura de dados) apenas como suporte e fonte da informação.

Os processadores associados aos elementos são, na sua base, modelos de transformação na linguagem XSL Transformation 2.0 (Kay, 2007), identificada por XSL ou XSLT, mas que podem recorrer a serviços Web referenciados por hiperligação, alargando a capacidade de processamento muito para além dos limites da tecnologia e linguagem de base.

Tais serviços não precisam de ser escritos numa linguagem de programação particular. Aliás, os processadores que os invocam não conhecem nem precisam de conhecer os detalhes de produção e funcionamento interno. Os únicos requisitos funcionais a que os serviços Web devem obedecer é que recebam a instrução do pedido pelo método GET do protocolo HTTP, e que devolvam o resultado em XML (em qualquer esquema). Mesmo se estes requisitos não forem possíveis de satisfazer, ou se outras restrições de acesso se aplicarem, é relativamente trivial construir uma aplicação *proxy* que envolva o acesso ao serviço de forma a fazer cumprir estes requisitos e torná-lo consumível pelos processadores. Esta camada intermédia pode ainda ser usada para otimizar a consulta do sistema de informação de destino, diminuindo o tráfego e o número de pesquisas através de um mecanismo de *cacheing* (se uma relação semântica foi estabelecida entre o identificador A e o identificador B, ela é replicável para todas as novas ocorrências do identificador A).

Apesar de os processadores terem funcionalidade variável, identificada por um tipo de ação, como se verá a seguir, obedecem a um interface aplicacional regular: independentemente da natureza da informação que é processada, os aspetos formais de contexto, parametrização e estrutura de resultado da execução dos processadores são constantes para cada uma das ações. Esta regularidade permite que a cadeia de processamento seja articulada, significando que os processadores podem passar os seus resultados para os seguintes ou obter informação dos anteriores, dependendo da ação que executam.

A lista de tipos de ações que os processadores suportam compreende:

- a identificação, que fornece um identificador Uniform Resource Identifier (URI) para o conteúdo do elemento ou, num caso especial, para todo o registo bibliográfico;
- a extração, que formata os dados observados relativos a um determinado elemento, ou a uma determinada perspectiva de um elemento, com vista à sua utilização na cadeia de processamento;
- a validação, para identificação de problemas de codificação dos dados observados;
- a tipificação, isto é, a definição do tipo das componentes de informação do elemento e dos seus predicados, permitindo o alinhamento explícito a outras ontologias e vocabulários de valores;
- a resolução, que estabelece as propriedades da informação registada no elemento;
- a ligação, com o estabelecimento de conexões semanticamente ricas da informação expressa no registo a outros vocabulários de valores, ontologias e conjuntos de elementos.

Os dois primeiros tipos de ação – identificação e extração – produzem valores que são usados na cadeia de processamento. Isto quer dizer que o resultado de um processo de identificação ou extração é transmitido a todos os processos seguintes que os requeiram, até que um novo processo produza novos valores e os passe para a continuação da cadeia. A validação produz informação de apoio ao processo e as restantes ações contribuem para gerar o documento final que constituirá o repositório LOD, expresso em RDF/XML (Gandon & Schreiber, 2014).

Excetuando o caso da identificação do registo bibliográfico como recurso de referência global, e que ocorre apenas uma vez na preparação da migração de um registo em particular, cada elemento definido no esquema pode suportar um número não limitado de processadores de qualquer tipo.

Tal capacidade introduz duas vantagens importantes no desenho da cadeia de processamento. Em primeiro lugar, permite que cada elemento seja processado segundo diversas perspetivas e também segundo diferentes níveis de granularidade (o que é particularmente relevante nos elementos de informação compostos, ou seja, os que agregam informação registada em diferentes designadores de conteúdo). Pode até dar-se o caso de o mesmo processador ser chamado repetidas vezes na mesma cadeia de processamento, para diferentes valores extraídos ou identificados por funções a montante.

Em segundo lugar, oferece um ambiente para o crescimento e enriquecimento das conexões, que podem ser consideradas autonomamente e integradas de forma modular na cadeia. A possibilidade de estabelecer ligações a novas ontologias e vocabulários de valores carece apenas da criação ou reutilização de um processador adequado e da sua inscrição na cadeia de processamento.

Para aliviar a carga de programação necessária para processar todos os designadores de conteúdo definidos na norma bibliográfica de origem, e em particular os dados codificados, a aplicação do esquema permite que um elemento dispense a associação explícita a um processador. Nestes casos, a fonte da informação e o vocabulário que lhe está associado são usados para produzir automaticamente o resultado das ações de tipificação, resolução e de ligação, sem necessidade de esforço de programação adicional. Os vocabulários são definidos também eles sobre um esquema XSD e são usados, acessoriamente, para a construção de serviços Web que comunicam o vocabulário a agentes humanos e computacionais.

A instância do esquema MARC, com todos os seus processadores e vocabulários associados, é consumida por um gerador, uma transformação XSL que produz automaticamente uma nova transformação. O gerador recebe por documento de entrada a aplicação do esquema MARC, e produz como documento de saída um migrador de registos bibliográficos para um repositório LOD (em RDF/XML). Os processadores referenciados implícita ou explicitamente pela aplicação do esquema são incluídos pelo migrador e invocados no processamento da coleção de registos objeto da migração.

O modelo pode ainda ser visto como uma sucessão de camadas, dispostas de baixo para cima. A primeira camada é a esquemática, e nela se definem os esquemas formais sintáticos. A segunda camada é a da semântica, onde todo o conhecimento sobre a norma bibliográfica, vocabulários e regras é formalizado segundo os esquemas definidos na camada inferior, e são desenvolvidos os processadores que interpretam e manipulam o modelo lógico da norma e identificam conexões a outras ontologias e vocabulários de valores. O principal do esforço de formalização e de desenvolvimento contínuo está concentrado nesta segunda camada. Trata-se de verter para o esquema apropriado todo o corpo de regras e informação vocabular de uma norma e de construir processadores que executam as ações necessárias ao tratamento dos elementos formalizados, com recurso a transformadores XSL ou a serviços Web que forneçam informação adequada e relevante.

A terceira camada, a do processamento, é, pelo contrário, completamente automatizada. Um pré-processador, desenvolvido como uma transformação XSL geradora de nova transformação, constrói o programa de migração capaz de consumir dados bibliográficos e de criar um repositório de dados ligados.

Para além das vantagens de extensibilidade e de incrementabilidade, já referidas acima, assentar este modelo nas tecnologias XML garante a sua fácil aplicação e distribuição a um baixo custo. O XML e tecnologias associadas são ubíquos nas atuais plataformas computacionais, e constituem um instrumento fundamental no processamento e na comunicação de informação num ambiente distribuído, como a Web. Incentiva, ainda, à partilha e à contribuição, em que todos os agentes podem participar, não apenas no desenvolvimento de processos, que exige competências específicas de programação, mas também, e começando, na formalização dos componentes semânticos.

Finalmente, o modelo garante um funcionamento transparente e transportável. Uma vez definidas as instâncias da norma e dos vocabulários e o desenvolvimento de processadores, e gerado um migrador, a sua execução é aplicável a qualquer conjunto de dados bibliográficos, criado segundo a mesma norma.

3. Uma aplicação: a migração de dados bibliográficos UNIMARC

O UNIMARC é um formato bibliográfico mantido e regulado pela IFLA, desenhado originalmente como um instrumento para a troca de informação entre diferentes formatos baseados em MARC, mas que foi adotado por mais de duas dezenas de agências nacionais como formato interno para codificação da informação bibliográfica (Galvão & Cordeiro, 2013, p. 152; IFLA, 2008, p. 13). Trata-se, na realidade, de uma família de formatos que integra ainda os formatos de autoridades, de classificação e de existências (IFLA, sem data).

Como os demais formatos baseados em MARC, a sua estrutura sintática corresponde a uma aplicação da norma ISO 2709, desenhada para a transferência de dados em fita magnética, mas que, por força de um uso generalizado, persistiu e permanece no atual contexto tecnológico de comunicação em linha (ISO, 2008). Num esforço de adaptação a esse contexto, o essencial da informação que os registos ISO 2709 suportam pode ser expresso, sem perda, em XML, segundo a norma ISO 25577, que tem o nome comum de MarcXchange (ISO, 2006).

Os exemplos que se seguem demonstram a aplicação do modelo de migração ao UNIMARC bibliográfico, em três aspetos relevantes:

- a formalização de designadores de conteúdos da norma bibliográfica como elementos do esquema MARC e o seu efeito na geração automática do migrador;
- a utilização do esquema VOC para definir o processo de migração e estabelecer mapeamentos conceptuais explícitos;
- e o funcionamento genérico dos processadores que executam funções de ligação e contribuem para o enriquecimento semântico do repositório.

Deve notar-se que todas as listagens inseridas nas Figuras se reportam a um determinado estado de desenvolvimento do modelo e da sua aplicação, que se mantém contínuo, e foram editadas para efeito desta apresentação.

3.1. Formalização de elementos

Para concretizar a aplicação do modelo de migração a dados bibliográficos UNIMARC, o primeiro passo consiste em formalizar a expressão dos diversos designadores de conteúdo que a norma compreende, segundo os esquemas MARC e VOC que alicerçam o modelo.

No esquema MARC, a expressão de um elemento é definida por um tipo complexo, como descrito na Figura 2 (não é pormenorizada a estrutura dos outros tipos referidos pelo tipo *marc:elementType*¹, mas o principal das suas funções será esclarecido nos exemplos de aplicação).

¹É usada a seguinte convenção tipográfica: *prefixo:nomeDoTipoDeElemento* referencia um tipo de elemento num determinado esquema, identificado pelo prefixo; *<prefixo:nomeDeElemento>* referencia um elemento numa instância de um esquema, identificado pelo prefixo; *@nomeDeAtributo* referencia um

Figura 2: Definição do tipo de elemento no esquema MARC

```
<xs:complexType name="elementType">
  <xs:sequence>
    <xs:element name="description" type="xs:string" minOccurs="1" maxOccurs="1"/>
    <xs:element name="source" type="dataSegmentType" minOccurs="0" maxOccurs="1"/>
    <xs:element name="vocabulary" type="vocabularyType" minOccurs="0" maxOccurs="1"/>
    <xs:element name="processing" type="processingType" minOccurs="0" maxOccurs="1"/>
  </xs:sequence>
  <xs:attribute name="name" type="xs:QName" use="required"/>
  <xs:attribute name="form" type="formOfElementType" use="required"/>
  <xs:attribute name="content" type="contentType" use="required"/>
  <xs:attribute name="occurrence" type="occurrenceType" use="required"/>
  <xs:attribute name="type" type="xs:NCName" use="optional"/>
</xs:complexType>
```

8

Na definição do esquema, os elementos *marc:elementType* estão agrupados numa sequência de nome *<marc:elements>*, iniciada por um *<marc:identifier>* que serve à identificação genérica do recurso, e completada por um número ilimitado de *<marc:element>*, correspondentes a cada um dos designadores de conteúdo identificados na norma, ou a uma sua composição, se o nível de granularidade assim o aconselhar.

No UNIMARC, o identificador de registo – que fornecerá a identificação do recurso – é inserido no campo de controlo 001. A sua formalização na instância do esquema MARC está descrita na Figura 3, e corresponde ao primeiro elemento da sequência *<marc:elements>*.

Figura 3: Formalização do identificador de registo, na instância UNIMARC do esquema MARC

```
<identifier name="resourceIdentifier" form="textual" content="controlfield" occurrence="single">
  <description>Record identifier</description>
  <source tag="001"/>
  <processing>
    <processor id="UNIMARC-IDS" action="identification" function="id"/>
  </processing>
</identifier>
```

Os atributos de *<marc:identifier>* (*@name*, *@form*, *@content* e *@occurrence*) especificam o nome dado ao elemento identificador de registo, com que o gerador compõe o nome do controlador de processadores, a forma (textual ou codificada), o tipo de conteúdo (no caso, um campo de controlo) e a regra de ocorrência (dado que o designador ocorre apenas uma vez no registo, é definida como de ocorrência singular). *<marc:source>* identifica a fonte de informação, a etiqueta do campo UNIMARC 001, seguindo-se a sequência de processadores que manipulam a informação, segundo uma determinada ação: neste caso, apenas a de identificação.

A informação reunida na composição do elemento, da fonte, do vocabulário e dos processadores é usada pelo gerador para preparar o contexto e demais parametrização de chamada das funções dos processadores. Tais funções fazem parte de folhas de transformação XSL, identificadas na instância do esquema. Como explicado acima, dependendo do tipo de ação, o resultado das funções é armazenado para uso em outros processadores na cadeia ou inserido diretamente no documento de saída (como declarações RDF/XML).

No exemplo do identificador de registo, o gerador referencia a função *f:id* da folha UNIMARC-IDS como uma função que devolve um URI, a partir da fonte especificada na definição do elemento. O código gerado, visível na Figura 4, é integrado no processamento para servir de identificador geral ao recurso que é migrado, ou seja, o identificador de cada um dos registos que fazem parte da coleção MarcXchange.

Figura 4: Controlador de processadores do identificador de registo, gerado a partir da definição do elemento na instância UNIMARC do esquema MARC

```

<!--Record identifier-->
<xsl:template name="process_resourceIdentifier">
  <xsl:param name="record"/>
  <xsl:param name="about"/>
  <xsl:for-each select="$record/mx:controlfield[@tag='001'][1]">
    <xsl:call-template name="subprocess_resourceIdentifier">
      <xsl:with-param name="record" select="$record"/>
      <xsl:with-param name="about" select="$about"/>
      <xsl:with-param name="context" select="."/>
      <xsl:with-param name="v0" select="."/>
      <xsl:with-param name="i0" select="$about"/>
    </xsl:call-template>
  </xsl:for-each>
</xsl:template>
<xsl:template name="subprocess_resourceIdentifier">
  <xsl:param name="record"/>
  <xsl:param name="about"/>
  <xsl:param name="context"/>
  <xsl:param name="v0"/>
  <xsl:param name="i0"/>
  <xsl:value-of select="p1:id($record,$context)"/>
</xsl:template>

```

Para cada elemento, o gerador produz dois níveis de processamento. No primeiro, o contexto da transformação que executa a migração do elemento é estabelecido a partir da especificação da *<marc:source>* de cada *<marc:element>*, e são inicializadas as variáveis que definem os valores de identificação e extração que são transmitidos na cadeia de processamento. Esta arquitetura permite que designadores de conteúdo que têm ocorrência múltipla, nas várias formas admitidas pela norma, possam ser processados pela mesma definição de cadeia, que é controlada pelo segundo nível definido pelo gerador.

O processador associado ao identificador de registo, inscrito no campo UNIMARC 001, é uma função simples que constrói um URI para o recurso, como se pode ver na Figura 5. A variável *\$resourcesURI* recolhe o seu valor de uma das *<marc:identity>* definidas na instância do esquema e que o gerador torna visíveis para todos os processadores, servindo à composição de URI sempre que necessário nas várias fases do processo. O parâmetro *\$context* representa o campo de controlo 001, identificado no primeiro nível do processador, passado para o segundo e, daí, para a função *f:id*.

Figura 5: Função que fornece um URI para identificar o recurso

```

<xsl:function name="f:id">
  <xsl:param name="record"/>
  <xsl:param name="context"/>
  <xsl:value-of select="concat($resourcesURI,$context)"/>
</xsl:function>

```

3.2 Vocabulários

Um segundo exemplo da formalização de designadores de conteúdo UNIMARC esclarece a utilização dos vocabulários para a transcodificação da informação bibliográfica.

No modelo de migração, o papel das instâncias do esquema VOC é o de facilitar a tipificação dos valores atribuídos aos designadores de conteúdo da norma bibliográfica e o de ligação a outros conceitos e tipos, definidos em ontologias disponíveis. Acessoriamente, as instâncias do esquema servem à construção

automática de representações de vocabulários que podem ser consumidas com recurso a tecnologias Web por agentes humanos (em HTML) e computacionais (em RDF).

No UNIMARC, o tipo de recurso que é descrito no registo bibliográfico é identificado por um valor colocado na etiqueta de registo, na posição 6 (de base 0). O valor é limitado por uma enumeração codificada, que designa, do ponto de vista tipológico, a característica dominante do recurso que se descreve (texto, registo sonoro, gráficos, etc.). A instância do esquema MARC define esse designador de conteúdo como se vê na Figura 6.

Figura 6: Formalização do tipo de registo, na instância UNIMARC do esquema MARC

```
<element name="typeOfRecord" form="coded" content="inleader" occurrence="single">
  <description>Type of resource described by the record</description>
  <source offset="6" length="1"/>
  <vocabulary href="unimarc/typeofrecord.xml"/>
</element>
```

O conteúdo está registado num segmento da etiqueta de registo ("inleader") e os atributos de *marc:source* especificam a sua posição e comprimento. Ao contrário do que acontece com o identificador de registo, na formalização do tipo de registo não se recorre a processadores, antes se faz referência a um vocabulário, que é, por seu lado, uma instância do esquema VOC. Este esquema estrutura-se como uma sequência de conceitos, identificados por URI, e eventualmente associados a outros conceitos de outros vocabulários, recorrendo às capacidades de expressão e de relacionamento semântico do padrão SKOS de representação de sistemas de organização do conhecimento (Miles & Bechhofer, 2009; Miles & Pérez-Agüera, 2007).

A enumeração dos códigos de tipo de registo é formalizada na instância do esquema VOC de forma a determinar o URI que identifica cada um dos conceitos. Quando em outro vocabulário exista um conceito que se possa de alguma forma associar ao que se está a identificar, o mapeamento é explicitado de acordo com a correspondência SKOS que mais bem se aplique (exata, próxima, relacionada, ou hierarquizada de subordinação ou superordenação).

Figura 7: Formalização do conceito "material textual, impresso", do vocabulário UNIMARC "Tipo de registo"

```
<id notation="a">
  <label lang="en">language materials, except manuscript</label>
  <label lang="pt">material textual, impresso</label>
  <map match="narrow">http://id.loc.gov/vocabulary/contentTypes/txt</map>
  <map match="exact">http://marc21rdf.info/terms/formofmaterial#a</map>
</id>
```

A Figura 7 representa a formalização do conceito "material textual impresso", na instância UNIMARC do esquema VOC. O atributo *@notation* de *voc:id* é concatenado ao URI de base do corpo de conceitos, definido num elemento *voc:scheme* que antecede a lista de conceitos, para compor o URI de cada conceito individual.

Na formalização do conceito são estabelecidas duas relações semânticas com outros dois conceitos externos. Com o primeiro, publicado pela Library of Congress, a relação é de subordinação, ou seja, o conceito identificado no vocabulário pode ser usado como um conceito específico do identificado pelo URI <http://id.loc.gov/vocabulary/contentTypes/txt> (na composição do seu vocabulário de tipos de conteúdo, a Library of Congress optou por reunir num único conceito todos os materiais textuais, independentemente de serem manuscritos ou impressos). Com o segundo, publicado pelo grupo Metadata Management Associates, a relação é exata, o que permite a utilização intermutável de ambos.

Para compor a informação sobre o tipo de registo, o migrador é gerado integrando uma chamada a *vocabularyProcessor* com os parâmetros que decorrem da

formalização do designador de conteúdo, em particular o vocabulário (como corpo de conceitos) e a fonte de informação no registo (cujo conteúdo servirá para identificar o conceito, de entre todos que integram o vocabulário). Uma vez que a composição do `<marc:element>` que formaliza o tipo de registo não declarou processadores, como se pode ver na Figura 6, nenhum será invocado na respetiva cadeia de processamento.

O resultado da migração de um registo UNIMARC de tipo de registo marcado como material textual impresso referencia todos os conceitos identificados no vocabulário, seja o vocabulário próprio ao UNIMARC, sejam os que estão semanticamente relacionados, como se pode ver na Figura 8. O sujeito da primeira declaração RDF é o registo (no exemplo, identificado pelo URI `http://lcci.uc.pt/data/bnportugal.pt/resource#1849861`), e o das seguintes é o conceito "material textual impresso", do vocabulário UNIMARC de tipos de registo (identificado por `http://lcci.uc.pt/vocabulary/typeOfRecord#a`).

Figura 8: Resultado da migração de informação sobre tipo do item, relativa a um texto impresso

```
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/resource#1849861">
  <typeOfRecord xmlns="http://lcci.uc.pt/vocabulary/typeOfRecord"
    rdf:resource="http://lcci.uc.pt/vocabulary/typeOfRecord#a"/>
</rdf:Description>
<rdf:Description rdf:about="http://lcci.uc.pt/vocabulary/typeOfRecord#a">
  <skos:narrowMatch rdf:resource="http://id.loc.gov/vocabulary/contentTypes/txt"/>
</rdf:Description>
<rdf:Description rdf:about="http://lcci.uc.pt/vocabulary/typeOfRecord#a">
  <skos:exactMatch rdf:resource="http://marc21rdf.info/terms/formofmaterial#a"/>
</rdf:Description>
```

Uma instância do esquema VOC permite ainda tipificar os recursos que são descritos, identificando-os como uma instância de uma classe, segundo o modelo vocabular RDF Schema, uma extensão do RDF básico com objetivos de servir à modelação da informação através de mecanismos de classificação e relacionamento de recursos (Brickley & Guha, 2014).

No UNIMARC, os agentes responsáveis pela produção do item que é descrito num registo são identificados nos campos do bloco 7 (bloco de responsabilidade), sendo distinguidos, na norma, o nível de responsabilidade (primária ou secundária) e a personalidade do agente (pessoa individual, coletiva, ou família). Na instância do esquema MARC, a formalização da informação relativa à responsabilidade primária de uma pessoa individual refere simultaneamente um vocabulário, que servirá às funções de tipificação, e os processadores associados (Figura 9).

Figura 9: Formalização do campo de responsabilidade primária individual na instância UNIMARC do esquema MARC

```
<element name="personalNamePrimary" form="textual" content="datafield" occurrence="alternative"
  type="primaryResponsibility">
  <description>Name of the person considered to have primary responsibility for a work</description>
  <source tag="700"/>
  <vocabulary href="unimarc/personalresponsibility.xml"/>
  <processing>
    <processor id="UNIMARC-IDS" action="identification" function="personId"/>
    <processor id="UNIMARC-PERSON" action="extraction" function="personalName"/>
    <processor id="UNIMARC-PERSON" action="resolution" function="person"/>
    <processor id="UNIMARC-PERSON" action="typification" function="types"/>
    <processor id="UNIMARC-LABELS" action="linking" function="label"/>
    <processor id="UNIMARC-PERSON" action="linking" function="isni"/>
    <processor id="UNIMARC-PERSON" action="linking" function="dublinCore"/>
  </processing>
</element>
```

Neste caso, o vocabulário referido em `<marc:vocabulary>` não integra uma lista de conceitos, como acontece no exemplo anterior com o vocabulário que descreve tipos de itens. Em vez disso, `<voc:class>` identifica as classes de recursos de que o conceito é

membro, como ilustra a Figura 10. Da mesma forma que os elementos `<voc:map>` servem para estabelecer relações semânticas entre conceitos pertencentes a diferentes vocabulários, os elementos `<voc:type>` explicitam quaisquer outras classes de que determinado recurso é membro.

Figura 10: Definição dos recursos de tipo "personal" da classe "responsibility", com associação a outras classes de que são membros

```
<class type="personal" baseURI="http://lcci.uc.pt/class/responsibility#">
  <label lang="en">Personal responsibility</label>
  <label lang="pt">Responsabilidade pessoal</label>
  <type>http://xmlns.com/foaf/0.1Person</type>
  <type>http://xmlns.com/foaf/0.1Agent</type>
</class>
```

No processamento da migração, as ocorrências do campo UNIMARC 700 geram, por ação de *f:types*, incluída na respetiva lista de processadores, declarações que explicitam os tipos associados a uma pessoa individual com participação na produção do item descrito pelo registo (como no exemplo da Figura 11).

Figura 11: Explicitação das classes de que é membro um recurso que identifica uma pessoa individual, gerada a partir das instâncias dos esquemas MARC e VOC, de acordo com o vocabulário RDF Schema

```
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/name#1402853">
  <rdf:type rdf:resource="http://lcci.uc.pt/class/responsibility#personal"/>
</rdf:Description>
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/name#1402853">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1Person"/>
</rdf:Description>
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/name#1402853">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1Agent"/>
</rdf:Description>
```

Os alinhamentos semânticos que são estabelecidos por um e outro processo – explicitação de relações de conceitos e instanciação dos recursos como membros de uma classe – permitem que um agente semântico que não conheça o UNIMARC, enquanto sistema de representação de informação, possa ainda assim consumir informação declarada no repositório de dados. Neste último exemplo, se tal agente domina a ontologia FOAF, reconhecerá o recurso identificado por `http://lcci.uc.pt/data/bnportugal.pr/name#1402853` como se referindo a uma pessoa, tal como entendida na especificação da ontologia (Brickley & Miller, 2014).

3.3. Ligações de valores

A instanciação da informação relativa à responsabilidade primária individual compreende, como já apresentou a Figura 9, uma lista de processadores da qual fazem parte processadores de ligação, cuja ação é executada pelas funções *f:label*, *f:dublinCore* e *f:isni*. Seguindo as especificações do modelo, estas funções recebem por parâmetros o identificador geral do recurso e o identificador do objeto que está a ser descrito, determinados por ações de identificação, executadas anteriormente na cadeia de processamento. Recebem, ainda, os valores determinados por ações de extração anteriores e o contexto de transformação.

A primeira função, *f:label*, é uma função genérica executada no processamento de vários elementos, e produz uma declaração RDF que define o valor literal correspondente a uma designação do recurso legível por humanos (em vez de um URI). Aqui, a chamada à função é usada para a expressão do nome, composto pela ação de extração antecedente na cadeia de processamento, e referenciado como uma propriedade `<rdfs:label>` do modelo vocabular RDF Schema.

A segunda função, *f:dublinCore*, produz uma declaração RDF que identifica os produtores do recurso, enquanto `< dct:creator >` ou `< dct:contributor >`. Consoante o nível

de responsabilidade, que é possível derivar pela informação do contexto, a função escolhe um ou outro dos termos Dublin Core. Na prática, o desempenho destas duas funções é semelhante ao das funções de resolução, mas para maior clarificação prefere-se identificar estas ações como de ligação porque correspondem a um processo de ligação a um conjunto de elementos e vocabulários que são externos ao UNIMARC.

A terceira função, *f:isni*, exemplifica o estabelecimento automático de ligações a vocabulários de valores. O objetivo da ação que executa é o de identificar no sistema ISNI (International Standard Name Identifier) uma correspondência SKOS para o identificador local, isto é, o identificador que é dado à pessoa com responsabilidade na produção do item que se descreve.

A função começa por compor uma chamada ao serviço Web do ISNI, criando uma *query string* a partir do nome da pessoa registada no campo UNIMARC, valor que lhe é fornecido pela antecedente ação de extração. Em resposta ao pedido, o serviço Web devolve um documento XML que é processado pela função: se cumpre os requisitos de confiabilidade, uma relação de equivalência semântica entre o identificador local e o identificador ISNI é declarada.

Um exemplo real do resultado da execução destas três funções, juntamente com a de resolução e tipificação, produzido no processo de transformação, pode ser visto na Figura 12.

Figura 12: Resultado da execução das funções de resolução, tipificação e ligação, aplicadas a uma responsabilidade primária individual

```
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/resource#1849861">
  <primaryResponsibility xmlns="http://lcci.uc.pt/class/responsibility#"
    rdf:resource="http://lcci.uc.pt/data/bnportugal.pt/name#142132"/>
</rdf:Description>
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/name#142132">
  <rdf:type rdf:resource="http://lcci.uc.pt/class/responsibility#personal"/>
</rdf:Description>
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/name#142132">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1Person"/>
</rdf:Description>
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/name#142132">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1Agent"/>
</rdf:Description>
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/name#142132">
  <rdfs:label>Belo, Duarte</rdfs:label>
</rdf:Description>
<rdf:Description rdf:about="http://lcci.uc.pt/data/bnportugal.pt/name#142132">
  <skos:exactMatch rdf:resource="http://isni.org/isni/0000000108675908"/>
</rdf:Description>
```

Esta capacidade de ligação a distintos vocabulários de valores, que cresce à medida que novos processadores se associam a qualquer elemento, é fator de enriquecimento das ligações de um repositório LOD e, conseqüentemente, do seu valor de uso no quadro da Web Semântica.

No entanto, os riscos de homonímia que este processo e outros semelhantes podem apresentar são evidentes, e algum cuidado deve ser colocado no estabelecimento das relações automatizadas de equivalência entre valores de vocabulários, ainda que se trate normalmente de vocabulários controlados em ambos os lados da operação. Uma prevenção que se pode introduzir no processo é a elevação dos parâmetros de confiabilidade, isto é, de aceitação de uma equivalência positiva. No exemplo do consumo do serviço Web do ISNI, tal prevenção faz passar o resultado por filtros que diminuem o risco de falsos positivos (considerando documentos de resposta com apenas um registo, por exemplo), à custa, no entanto, do aumento de falsos negativos.

Uma outra forma de tornar estes dados mais fiáveis consiste em facilitar o controlo, após o processamento, dos resultados produzidos automaticamente. Com o RDF, tal controlo pode ser feito através de metadeclarações, declarações sobre as declarações,

uma técnica a que se dá o nome de reificação e que permite comunicar, entre outra informação, a proveniência de uma declaração num determinado documento RDF.

O vocabulário RDF de reificação é usado no modelo de migração para fornecer informação sobre os processos automatizados de ligação, documentando positivos (isto é, ligações que se puderam estabelecer com sucesso) e negativos (ligações não estabelecidas). A informação integra o repositório LOD produzido e pode ser, por isso, consumida pelos agentes semânticos que o processam, para efeitos de rastreabilidade, ou apenas para apoio à validação, como exemplificado no relatório que a Figura 13 visualiza.

Figura 13: Relatório relativo aos resultados de uma ação de ligação automatizada a vocabulários de valores (dois casos de sucesso e um de insucesso, com explicação do motivo)

Linking Report			
Subject	Predicate	Object	Fail
http://unimarc.aftlopes.com/unimarc-persons#isni			
http://lci.uc.pt/data/bnportugal.pt/name#142132 (Belo, Duarte)	http://www.w3.org/2004/02/skos/core#exactMatch	http://isni.org/isni/0000000108675908	
http://lci.uc.pt/data/bnportugal.pt/name#23883 (Cordeiro, Maria Inês)	http://www.w3.org/2004/02/skos/core#exactMatch	http://isni.org/isni/0000000108769795	
http://lci.uc.pt/data/bnportugal.pt/name#1402853 (Roldão, Helena)	http://www.w3.org/2004/02/skos/core#exactMatch		Weak Confidence

Uma vez que esta informação consta do repositório, uma aplicação que a consuma pode escolher não utilizá-la ou fazê-lo sob um qualquer tipo de reserva. Dispõe, de qualquer forma, dos dados que lhe permitem distinguir as declarações que foram produzidas em ambiente controlado das que resultam de uma execução não supervisionada.

Conclusão

O modelo de expressão de dados bibliográficos UNIMARC como *linked open data* que se apresenta neste trabalho permite criar conjuntos de dados semanticamente expressivos e enriquecidos, reaproveitando eficaz e eficientemente o esforço da produção da informação bibliográfica. Não se trata apenas de exprimir o UNIMARC em RDF, mas o de reforçar nesse processo o alinhamento semântico e a ligação a vocabulários de valores, e de fazê-lo como um subproduto da atividade corrente do controlo bibliográfico.

O esforço de formalização e de desenvolvimento necessários à sua implementação é absorvível pela potencialidade da sua fácil distribuição e aplicação por distintos agentes a distintas fontes de informação bibliográfica que necessitam apenas de coincidir na utilização daquela norma.

Referências

- Aalberg, T.** (2006). A Process and Tool for the Conversion of MARC Records to a Normalized FRBR Implementation. Em *LNCS: Digital Libraries: Achievements, Challenges and Opportunities* (pp. 283–292). doi:10.1007/11931584_31
- Alemu, G., Stevens, B., Ross, P., & Chandler, J.** (2012). Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. *New Library World*, 113(11/12), 549–570. doi:10.1108/03074801211282920
- Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., ... Zeng, M.** (2011). Library Linked Data Incubator Group Final Report. Obtido em 1 de Setembro de 2015, de <http://www.w3.org/2005/Incubator/lld/XGR-lld/>
- Bérard, R.** (2011). Free library Data? *LIBER Quarterly*, 20(3-4), 321–331. Obtido em 1 de Setembro de 2015, de <http://liber.library.uu.nl/index.php/lq/article/view/7997>
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., Yergeau, F., & Cowan, J.** (2006). Extensible Markup Language (XML) 1.1. Obtido em 1 de Setembro de 2015, de <http://www.w3.org/TR/xml11/>
- Brickley, D., & Guha, R. V.** (2014). RDF Schema 1.1. Obtido em 1 de Setembro de 2015, de <http://www.w3.org/TR/rdf-schema/>

- Brickley, D., & Miller, L.** (2014). FOAF Vocabulary Specification 0.99. Obtido em 1 de Setembro de 2015, de <http://xmlns.com/foaf/spec/>
- Byrne, G., & Goddard, L.** (2010). The Strongest Link: Libraries and Linked Data. *D-Lib Magazine*, 16(11/12). doi:10.1045/november2010-byrne
- Carvalho, J.** (2005). An XML representation of the UNIMARC Manual: a working prototype. Em *World Library and Information Congress: 71th IFLA General Conference and Council*. Oslo: IFLA. Obtido em 1 de Setembro de 2015, de <http://archive.ifla.org/IV/ifla71/papers/199e-Carvalho.pdf>
- Carvalho, J., & Cordeiro, M. I.** (2002). XML and bibliographic data: the TVS (Transport, Validation and Services) model. Em *68th IFLA Council and General Conference*. Glasgow. doi:10.1177/034003528100700212
- Coyle, K.** (2012). Semantic Web and Linked Data. Em *Linked data tools: connecting on the Web* (pp. 10–15).
- Darnton, R.** (2009, Dezembro 17). Google and the New Digital Future. *The New York Review of Books*. Obtido em 1 de Setembro de 2015, de <http://www.nybooks.com/articles/archives/2009/dec/17/google-and-the-new-digital-future/>
- Dimic, B., Milosavljevic, B., & Surla, D.** (2010). XML schema for UNIMARC and MARC21. *The Electronic Library*, 28(144017), 245–262. doi:10.1108/02640471011033611
- Dunsire, G.** (2009). UNIMARC, RDA and the Semantic Web. Em *World Library and Information Congress*. Milan: IFLA.
- Dunsire, G.** (2012). Linked data for manuscripts in the Semantic Web. Obtido em 1 de Setembro de 2015, de <http://www.gordondunsire.com/pubs/docs/LinkedDataForManuscripts.pdf>
- Dunsire, G., & Willer, M.** (2011). UNIMARC and linked data. Em *World Library and Information Congress. 77th IFLA General Conference and Assembly*. San Juan, Puerto Rico: IFLA.
- Galvão, R. M., & Cordeiro, M. I.** (2013). UNIMARC - Understanding the past to envision the future. *IFLA Journal*, 39(2), 151–161. doi:10.1177/0340035213486410
- Gandon, F., & Schreiber, G.** (2014). RDF 1.1 XML Syntax. Obtido em 1 de Setembro de 2015, de <http://www.w3.org/TR/rdf-syntax-grammar/>
- Gao, S., Sperberg-McQueen, C. M., & Thompson, H. S.** (2012). W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures. Obtido em 1 de Setembro de 2015, de <http://www.w3.org/TR/xmlschema11-1/>
- Hannemann, J., & Kett, J.** (2010). Linked Data for Libraries. Em *World Library and Information Congress* (pp. 1–12). Gothenburg: IFLA.
- IFLA. (sem data). UNIMARC strategic programme. Obtido em 1 de Setembro de 2015, de <http://www.ifla.org/unimarc>
- IFLA. (2008). *Manual UNIMARC formato bibliográfico*. (R. M. Galvão & M. P. Lopes, Eds.). Lisboa: Biblioteca Nacional de Portugal.
- IFLA. (2009). *Functional requirements for bibliographic records: final report*. Obtido em 1 de Setembro de 2015, de http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf
- ISO. (2006). Information and Documentation - MarcXchange. Obtido em 1 de Setembro de 2015, de http://www.loc.gov/standards/iso25577/ISO_DIS_25577__E_.pdf
- ISO. (2008). Information and documentation: Format for information exchange. ISO.
- Kay, M.** (2007). XSL Transformations (XSLT) Version 2.0. Obtido em 1 de Setembro de 2015, de <http://www.w3.org/TR/xslt20/>
- Keith, C.** (2004). Using XSLT to manipulate MARC metadata. *Library Hi Tech*, 22(2), 122–130. doi:10.1108/07378830410524549
- Madison, O. M. A.** (2005). The Origins of the IFLA Study on Functional Requirements for Bibliographic Records. *Cataloging & Classification Quarterly*. doi:10.1300/J104v39n03_02
- Miles, A., & Bechhofer, S.** (2009). SKOS Simple Knowledge Organization System Reference. Obtido em 1 de Setembro de 2015, de <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>
- Miles, A., & Pérez-Agüera, J. R.** (2007). SKOS : Simple Knowledge Organisation for the Web. *Cataloging & Classification Quarterly*, 43(3-4), 69–83. doi:10.1300/J104v43n03

Styles, R., Ayers, D., & Shabir, N. (2008). Semantic marc, MARC21 and the semantic web. *CEUR Workshop Proceedings*, 369.

Vila-Suero, D., & Gómez-Pérez, A. (2013). datos.bne.es and MARiMbA: an insight into library linked data. *Library Hi Tech*, 31(4), 575–601. doi:10.1108/LHT-03-2013-0031

Waller, V. (2009). The relationship between public libraries and Google: Too much information. *First Monday*, 14(9). doi:10.5210/fm.v14i9.2477

Willer, M., & Dunsire, G. (2013). *Bibliographic information organization in the Semantic Web*. Oxford: Chandos Publishing.