

UM ESQUEMA XSD GENÉRICO PARA EXPRESSÃO DE NORMAS MARC: princípios de conceção orientada para a Web Semântica

A GENERIC XSD SCHEMA FOR EXPRESSION OF MARC STANDARDS: design principles targeting the Semantic Web

António Tavares Lopes

Resumo: A criação de um esquema XSD genérico para suportar a expressão de normas MARC permite regular a serialização em XML da composição estrutural e semântica do UNIMARC bibliográfico. A partir da proposta inicial de Carvalho e Cordeiro, propõem-se princípios adicionais de conceção desse esquema genérico, em particular reorientando-o para a sua operação no processo de criação de um modelo para a expressão de conjuntos de dados bibliográficos na Web Semântica.

Palavras-chave: UNIMARC; Web Semântica; Informação bibliográfica; Normas MARC

Abstract: The creation of a generic XSD schema to support the expression of MARC standards regulates the XML serialization of the structural composition and semantics of bibliographic UNIMARC. Based on the initial proposal of Carvalho and Cordeiro, some principles to redesign this generic scheme are proposed, in particular by redirecting it towards more effective operation in the process of modeling bibliographic data sets for the Semantic Web.

Keywords: UNIMARC; Semantic Web; Bibliographic information; MARC standards

Apresentação

Este trabalho é componente de uma investigação em curso que tem por objetivo a construção de um modelo de interoperabilidade do UNIMARC, no quadro da Web Semântica¹. O propósito de tal modelo é o de permitir a integração das coleções de dados bibliográficos em redes de informação e conhecimento que transcendem o campo do controlo bibliográfico, entendido como o conjunto de processos que visa descrever e organizar os documentos que servem à memória cultural da humanidade, preservados ou referenciados pelas instituições de memória.

Para compreender o quadro operacional em que se pretende inscrever os instrumentos aqui referidos, e na apresentação dos princípios da sua conceção, é conveniente começar por uma exploração, ainda que sumária, do que é a Web Semântica e quais as balizas conceptuais e tecnológicas que imediatamente definem qualquer esforço de alinhamento.

¹A investigação decorre no âmbito do Programa Doutoramento em Informação e Comunicação em Plataformas Digitais (Universidades de Aveiro e Porto), sob a orientação da Professora Doutora Fernanda Ribeiro (Universidade do Porto).

Os blocos de construção da Web Semântica

A Web Semântica é uma nova camada da World Wide Web que acrescenta à Web de documentos “a web of data that can be processed directly or indirectly by machines” (BERNERS-LEE; FISCHETTI, 2000, p. 191). Aqui, as ligações são estabelecidas entre expressão de factos, elementos de informação que se declaram utilizando instrumentos formais de reduzida ou nenhuma ambiguidade, numa forma de expressão desenhada para ser consumida por sistemas automatizados, ou seja, por computadores. Assim, a Web Semântica tem a ver com o que sabemos mas igualmente com a forma como exprimimos o que sabemos para que máquinas o possam processar.

Ao conjunto de regras e instrumentos que formalizam os objetivos da Web Semântica dá-se genericamente o nome de *Linked Data*, que designa também o resultado da aplicação dessas regras e processos. A expressão faz-se acompanhar habitualmente do qualificativo *Open*, como em *Linked Open Data* (ou LOD, em sigla) para reforçar o carácter aberto que está na própria natureza desta camada e que a torna sustentável: uma vez que se trata de um conjunto de fundamentos formais e normativos que definem como os factos ou o conhecimento devem ser expressos para serem processáveis por sistemas computacionais a operar em rede, dificilmente se ganharia impulso no desenvolvimento e alargamento dessa estrutura se as ligações fossem sujeitas bloqueios ou restrições.

Tim Berners-Lee, o inventor da World Wide Web e também um dos principais impulsionadores da Web Semântica, propõe 4 regras – ou “expectativas de comportamento” – que devem ser seguidas na construção de dados como LOD (BERNERS-LEE, 2009):

- Usar *Uniform Resource Identifiers* (URI) como nome para as *coisas* de que falamos, sejam pessoas, acontecimentos, conceitos, objetos, relações, etc.
- Usar o HTTP – ou seja, servidores Web – para comunicar informação sobre essas *coisas*, atendendo de forma adequada a quem procura informação, distinguindo entre autómatos e agentes computacionais a operar em nome de leitores humanos.
- Usar normas estabelecidas na organização dessa informação (como o RDF, de *Resource Description Framework*, para a expressão, e o SPARQL, para a interrogação), de maneira a que os agentes computacionais disponham de uma base comum, uma linguagem, para o uso dos dados expressos.
- Ligar a outros URI para alargar o conhecimento, estabelecendo associações – de similitude, concordância, hierarquia, pertença, etc. – elas próprias definíveis por URI, como seria de prever.

A simplicidade destas regras antecipa igualmente a facilidade de implantação. Os URI, os elementos de base deste conjunto de procedimentos, são fáceis de construir e de comunicar, para além de se poderem ancorar em toda a estrutura de nomes e domínios já montada da Internet: se a Universidade de Coimbra pretender expor conjuntos de dados seguindo as especificações da LOD, pode identificar e organizar todos os seus elementos

(factos, conceitos, nomes) sob o seu domínio de rede uc.pt, sem perigo de conflito com outros e iniciando uma estrutura informacional de combinações potencialmente ilimitadas. É o que fazem já a Library of Congress, que a partir do seu domínio de rede loc.gov estabelece o espaço de nomes (*namespace*) <http://data.loc.gov>, ou a Bibliothèque Nationale de France, que de igual forma estabelece o seu <http://data.bnf.fr>.

Com os URI é possível referenciar de forma inequívoca elementos de informação que integram os repositórios LOD. Esta identificação dá aos agentes computacionais uma ferramenta para reduzir e eliminar a ambiguidade na expressão da informação, processo que os agentes humanos normalmente conduzem pela associação a um contexto enriquecido por outras informações.

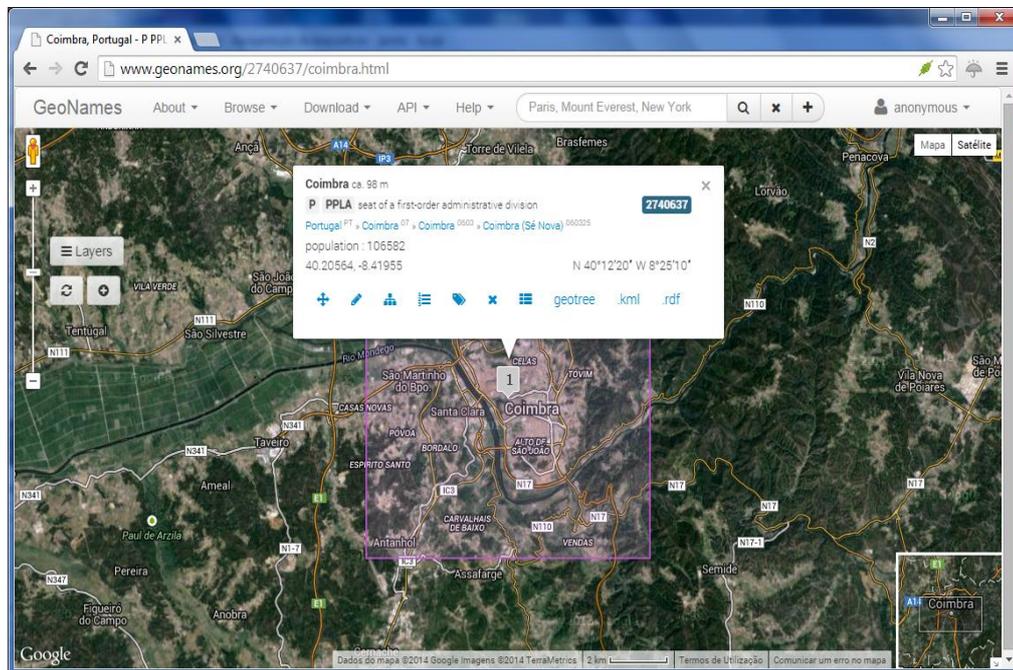
Por exemplo, todos os participantes deste *workshop* não tiveram qualquer dificuldade em identificar o seu local de realização, quando este lhes foi comunicado. Coimbra foi entendida rapidamente como sendo a cidade capital do distrito com o mesmo nome, localizada no centro de Portugal, e sede de uma das mais antigas universidades europeias. O termo “Coimbra”, no entanto, é apenas uma sequência de caracteres que pode ter muitos significados e referir-se a entidades distintas. Para além disso, não é a única sequência de caracteres que pode ser utilizada para identificar a cidade de Coimbra, tal como a definimos acima: alternativas poderão ser Коимбра, para um falante russo, ou コインブラ, para um falante japonês.

A utilização de um URI serve à identificação de um elemento de informação e à sua comunicação a outros agentes, computacionais e humanos. Recorrendo ao espaço de nomes do GeoNames (um serviço Web de referência toponímica que começa por ser um repositório de *Linked Open Data*), poder-se-á exprimir a localização do evento na cidade de Coimbra através do URI <http://www.geonames.org/2740637>, e fazê-lo de uma forma que é neutra, do ponto de vista do idioma. Este identificador permite distinguir os de outras referências como o da cidade de Coimbra, em Minas Gerais, Brasil (com o URI <http://www.geonames.org/6321278>), o do distrito de Coimbra, em Portugal (com o URI <http://www.geonames.org/2740636>), e até da sequência de caracteres “Coimbra” que pode integrar termos que não são de ordem geográfica (como o URI de António Coimbra Martins, <http://www.viaf.org/viaf/17346708/>).

Seguindo as boas regras na construção da LOD, todos estes URI que aqui se apresentam a título de exemplo *funcionam*, no sentido em que existe um servidor Web que honrará um pedido que lhe seja dirigido, através do protocolo HTTP. A informação que é devolvida ao iniciador do pedido depende, em primeiro lugar, da própria natureza desse sistema iniciador, determinada pelo serviço Web a partir dos parâmetros que compõem a mensagem HTTP que lhe foi dirigida:

- A um agente operado por um utilizador humano, o servidor retorna informação que um *browser* visualizará, integrando elementos interativos para exploração posterior (ver figura 1).

Figura 1: Resposta do serviço GeoNames ao URI relativo à cidade de Coimbra, em Portugal, quando iniciado e processado por um *browser*



- A um agente computacional com capacidade para consumir *linked data*, a informação devolvida é formatada segundo as normas usadas na Web Semântica, como, por exemplo, recorrendo a estruturas RDF expressas em XML (ver figura 2). Essas estruturas permitem aprofundar informação sobre Coimbra (a sua georreferenciação, por exemplo) ou identificar informação relacionada (na Wikipedia, por exemplo).

Figura 2: Resposta ao pedido do mesmo URI, quando iniciado por um agente semântico

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<rdf:RDF xmlns:cc="https://creativecommons.org/ns#" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:
  <gn:Feature rdf:about="http://sws.geonames.org/8010483/">
    <rdfo:isDefinedBy rdf:resource="http://sws.geonames.org/8010483/about.rdf"/>
    <gn:name>Coimbra</gn:name>
    <gn:alternateName xml:lang="ko">코임브라</gn:alternateName>
    <gn:alternateName xml:lang="ja">コインブラ</gn:alternateName>
    <gn:alternateName xml:lang="ch">科姆布拉</gn:alternateName>
    <gn:featureClass rdf:resource="http://www.geonames.org/ontology#A"/>
    <gn:featureCode rdf:resource="http://www.geonames.org/ontology#ADM2"/>
    <gn:countryCode>PT</gn:countryCode>
    <wgs84_pos:lat>40.21816</wgs84_pos:lat>
    <wgs84_pos:long>-8.45369</wgs84_pos:long>
    <gn:parentFeature rdf:resource="http://sws.geonames.org/2740636"/>
    <gn:parentCountry rdf:resource="http://sws.geonames.org/2264397"/>
    <gn:parentADM1 rdf:resource="http://sws.geonames.org/2740636"/>
    <gn:childrenFeatures rdf:resource="http://sws.geonames.org/8010483/contains.rdf"/>
    <gn:locationMap rdf:resource="http://www.geonames.org/8010483/coimbra.html"/>
    <gn:wikipediaArticle rdf:resource="http://en.wikipedia.org/wiki/Coimbra"/>
    <rdfo:seeAlso rdf:resource="http://dbpedia.org/resource/Coimbra"/>
  </gn:Feature>
  <foaf:Document rdf:about="http://sws.geonames.org/8010483/about.rdf">
    <foaf:primaryTopic rdf:resource="http://sws.geonames.org/8010483"/>
    <cc:license rdf:resource="http://creativecommons.org/licenses/by/3.0"/>
    <cc:attributionURL rdf:resource="http://sws.geonames.org/8010483"/>
    <cc:attributionName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GeoNames</cc:attributionName>
    <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2011-08-27</dcterms:created>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2011-08-27</dcterms:modified>
  </foaf:Document>
</rdf:RDF>
```

Alinhamento da informação bibliográfica à Web Semântica

Para chegar a equivalente ponto de usabilidade, a publicação de informação bibliográfica como *Linked Data* deverá orientar-se pelo mesmo conjunto de princípios. Operacionalmente, é um processo que requer a reunião de três tipos de componentes (por ordem de construção), tal como definidos no relatório final do Library Linked Data Incubator Group (BAKER *et al.*, 2011).

- Definição de conjuntos de elementos (*element sets*): tipos, classes e atributos usados na descrição dos recursos e do conhecimento (por exemplo, título, autor, tipo de material, etc.).
- Definição de vocabulários (*value vocabularies*): vocabulário controlado dos valores que podem ser usados nos elementos, e que têm por origem ficheiros de autoridade, listas de descritores, listas codificadas, sistemas de classificação, tesouros, etc.
- Produção de conjuntos de dados (*datasets*): informação bibliográfica que recorre aos conjuntos de elementos para a estruturação de metadados e aos vocabulários para (idealmente) grande parte do seu conteúdo.

Existe já em curso um esforço de produção de informação bibliográfica como *Linked Open Data*, e no mesmo relatório identificam-se manifestações de componentes contribuintes desse esforço, originárias ou não da comunidade biblioteconómica (em particular, as componentes estruturantes que se manifestam em formalizações RDF de conjuntos de elementos como o Dublin Core, a ISBD, os FRBR ou o CIDOC Conceptual Reference Model; de vocabulários como a CDU, o Virtual International Authority File, os GeoNames, o AGROVOC ou o EuroVOC).

Estará ainda por fazer uma análise crítica sistemática dos aspetos qualitativos das atuais iniciativas da construção de conjuntos de dados – o último patamar do processo de criação de repositórios bibliográficos como *Linked Open Data* – mas uma exemplificação singular permitirá evidenciar alguns dos atuais problemas mais imediatamente visíveis.

“Acto do Infante D. Pedro de Portugal, o qual andou as sete partidas do mundo”, é um recurso digital disponibilizado pela Biblioteca de Arte da Fundação Calouste Gulbenkian e integrada no portal Europeia, de onde pode ser acedido pelo endereço <http://data.europeana.eu/item/00101/E57D0044D5A95B061B67D9820F6D9C2FC2A789F2>. Uma comparação entre os metadados no catálogo do provedor original, expresso em UNIMARC (figura 3), e os que são comunicados em resultado da integração dessa informação no portal Europeia (figura 4), demonstra dois frustrantes problemas decorrentes da migração de um conjunto de elementos (*element set*) para outro e de uma limitação que desenha uma ainda longa distância até à satisfação de alguns requisitos basilares definidos pelas 4 regras para a *Linked Data*.

Figura 3: Registo bibliográfico em UNIMARC do "Acto do Infante D. Pedro de Portugal", na Biblioteca de Arte da Fundação Calouste Gulbenkian.

```
LDR: 00738cam 2200193 4500
001: 130466
003: http://www.biblartepac.gulbenkian.pt/ipac20/ipac.jsp?&profile=ba&uri=full=3100024~!130466~!0
100: $a19990720d1790 k y0pory0103 ba
101: 0 $apor
102: $aPT
200: 1 $aActo do Infante D. Pedro de Portugal, o qual andou as sete partidas do mundo$feito por Gomes de Santo Estevão,
    hum dos doze que foraõ em sua companhia, e
205: $aNovamente emendado nesta última impressão
210: 1 $aPorto$cOffic. de Antonio Alvarez Ribeiro,$d1790
215: $a29, [1] p.$d(21 cm)
300: $aNo final contém 1 p. com o catálogo das obras publicadas nesta Officina
321: $aCat. L. C.
620: $dPorto
675: $a821.134.3-2"17"
700: 1$aSanto Estevão,$bGomes de
702: 1$aRibeiro,$bAntónio Álvares$4610
956: $2BIBCTC
958: 1 $adoc=130466&img=10866&b&res=10&thb=1&pag=1$c&res=30&thb=1&pag=1$d (1578 KB)$90
966: 1 $sTC 155$930001001600450
999: $aVMJPF$cba-rf$d20000714$cbga-jpf$d20071115
```

Figura 4: Metadados LOD no portal Europeia, referentes ao mesmo recurso (removido o

envelope de agregação RDF)

```
<ore:Proxy a:about="http://data.europeana.eu/proxy/provider/00101/E57D0044D5A95B061B67D9820F6D9C2FC2A789F2">
  <dc:title>Acto do Infante D. Pedro de Portugal, o qual andou as sete partidas do mundo</dc:title>
  <dc:description>Monografia</dc:description>
  <dc:description>Monographic bibliographic item</dc:description>
  <dc:description>Novamente emendado nesta última impressão</dc:description>
  <dc:description>No final contém 1 p. com o catálogo das obras publicadas nesta Officina</dc:description>
  <dc:creator>Santo Estevão, Gomes de</dc:creator>
  <dc:contributor>Ribeiro, António Álvares</dc:contributor>
  <dc:subject>Porto</dc:subject>
  <dc:subject>821.134.3-2"17"</dc:subject>
  <ore:proxyFor a:resource="http://data.europeana.eu/item/00101/E57D0044D5A95B061B67D9820F6D9C2FC2A789F2"/>
  <edm:type>TEXT</edm:type>
  <dc:type>material textual, impresso</dc:type>
  <dc:type>language materials, printed</dc:type>
  <dc:format>29, [1] p.</dc:format>
  <dc:language>por</dc:language>
  <dc:publisher>Offic. de Antonio Alvarez Ribeiro</dc:publisher>
  <dc:publisher>Porto</dc:publisher>
  <ore:proxyIn a:resource="http://data.europeana.eu/aggregation/provider/00101/E57D0044D5A95B061B67D9820F6D9C2FC2A789F2"/>
  <dct:extent>(21 cm)</dct:extent>
  <dc:identifiier>130466</dc:identifiier>
  <dc:identifiier>CALLNUMBER-TC 155</dc:identifiier>
  <dct:issued>1790</dct:issued>
</ore:Proxy>
```

De uma expressão para outra é verificável um problema comum nas migrações de metadados, e que resulta em perda de expressividade (por exemplo, Porto é identificado, em Dublin Core, como conteúdo do elemento *publisher*, ao passo que no UNIMARC original se distingue entre local de publicação e editor; por seu lado, o elemento *subject* de Dublin Core recebe o conteúdo da notação CDU do registo UNIMARC, perdendo-se assim o referente do sistema de classificação).

Enquanto *Linked Open Data*, são igualmente visíveis limitações desta implementação (que ainda se encontra, de qualquer forma, em fase de piloto²): nenhuma conexão foi produzida e o uso de vocabulários de valores (apenas subentendido em elementos como *type*, do nome de espaços do Europeana Data Model) não se formaliza através de identificadores URI, dificultando a partilha de informação e dos domínios de conceito que são utilizados para a exprimir.

Estas limitações provocam dois efeitos que diminuem a perceção positiva dos esforços de expressão, na Web Semântica, da informação bibliográfica disponível em sistemas de informação de bibliotecas e outras instituições. Por um lado, não recompensam devidamente a produção histórica e corrente de informação bibliográfica; por outro, e ao mesmo tempo, o resultado produzido é pouco rico de conexões, inibindo ou restringindo o potencial do seu uso por agentes semânticos.

Propostas para expressões formais de normas bibliográficas

O enquadramento e exemplo até agora apresentados servirão para justificar o interesse da investigação que define as balizas deste trabalho: há uma necessidade de reforçar o

² Sítio Web do projeto em <http://pro.europeana.eu/linked-open-data>.

estudo das potencialidades e dificuldades da inscrição operativa de informação baseada em normas bibliográficas como o UNIMARC enquanto fonte de repositórios de dados utilizáveis pelas aplicações da Web Semântica, com vista ao desenvolvimento de ferramentas e procedimentos para uma migração eficiente da informação. Análise e desenvolvimento fazem-se a dois tempos:

- um primeiro incidindo sobre a norma bibliográfica, abstratamente, isto é, sobre o conjunto de regras, vocabulários, linguagens, códigos e formas que definem o UNIMARC e que se encontram plasmados nos seus manuais (nomeadamente, o bibliográfico e o de autoridades);
- e um segundo, aplicando a uma base de dados bibliográfica em concreto, originária da Bibliografia Nacional Portuguesa.

Instrumental para estes processos de análise e de desenvolvimento é a capacidade de operar fora das regras estritas das normas MARC, orientadas pela estruturação de dados definidas na norma ISO 2709, cuja quarta revisão foi publicada em 2008 (ISO, 2008). Numa primeira fase, essa capacidade foi assegurada pela fixação de uma serialização em XML dos registos bibliográficos MARC, definida pela norma MarcXchange, introduzida em 2006 a partir de desenvolvimentos prévios, e revista pela última vez em 2013 (ISO, 2006). O XML, de *eXtensible Markup Language*, é uma norma do consórcio W3C e desempenha o papel de *lingua franca* na estruturação, comunicação e interoperabilidade de dados nos atuais sistemas de informação em rede (e central, portanto, para o desenho da Web Semântica, que os exemplos acima de *linked open data* dos serviços GeoNames e Europeia ilustram).

O MarcXchange serve à informação bibliográfica da mesma forma que a ISO 2.709 o faz: define um suporte para a comunicação, transporte e registo da informação, sem impor qualquer regra de carácter semântico que uma norma bibliográfica – MARC21, UNIMARC ou qualquer outro MARC local ou nacional – em concreto estipula: ou seja, tal como a ISO 2.709 define que um registo bibliográfico se compõe de uma etiqueta de registo, campos e subcampos (para além das estruturas de controlo) sem especificar que informação se pode ou deve registar em cada um desses componentes, da mesma forma o MarcXchange define uma estrutura XML de elementos e atributos como *leader*, *controlfield*, *datafield* e *subfield* sem especificar regras para o conteúdo desses elementos.

A definição da estrutura de um registo MarcXchange é feita por um esquema XSD, de *XML Schema Definition*, um componente da família do XML que designa para cada tipo de documento XML os respetivos elementos, tipos de dados e regras de ocorrência³. Esta definição de esquema serve para validar um documento XML (documentos XML podem ser considerados *bem formados*, quando obedecem às regras sintáticas do XML, mas também *válidos*, quando obedecem às regras de estrutura e conteúdo definidas num esquema).

Quando se tornou evidente a necessidade de exprimir em XML os registos bibliográficos baseados em MARC, com o propósito de aumentar a integração da informação bibliográfica em outros ambientes, sistemas e processos aplicativos, pareceu um passo lógico recorrer a um XSD para formalizar, também em XML, as regras semânticas definidas pelos manuais das normas em uso. Genericamente, tal abordagem promove e facilita a automatização em cadeia de processos de transformação, integração e

³ O XSD da versão atual do MarcXchange está disponível em:
<http://www.loc.gov/standards/iso25577/marcxchange-2-0.xsd>.

reutilização, conceitos chave para a estruturação das mecânicas de interoperabilidade.

Os benefícios da representação dos formatos MARC, baseados na norma ISO 2709, numa estrutura formal em XML, nomeadamente através de um esquema XSD, foram apresentados numa proposta compreensiva por Joaquim de Carvalho, no Congresso Internacional da IFLA de 2005, em Oslo (CARVALHO, 2005), na continuidade de outros trabalhos anteriores em coautoria (CARVALHO *et al.*, 2004; CORDEIRO; CARVALHO, 2002). Estes trabalhos serviram de inspiração a Dimic e outros (DIMIC; MILOSAVLJEVIC; SURLA, 2010), para criar um esquema de representação do MARC21 e UNIMARC em XML para uso no sistema de catalogação BISIS, em particular para as funções de validação e de construção de listas de valores autorizados.

Na proposta apresentada em 2005, Carvalho definiu objetivos específicos para uma representação em XML do manual de uma norma MARC⁴:

- suportar a validação automatizada dos registos, aplicando as regras de ocorrência (obrigatoriedade e repetibilidade) e de conteúdo definido por listas de enumeração;
- representar os registos bibliográficos de diversas formas, não se limitando às diversas descrições bibliográficas mas antes permitindo descrições enriquecidas que sirvam à explicação do registo (veja-se um exemplo na figura 5);
- e servir de fonte para representação da norma bibliográfica em diversos formatos e suportes.

Figura 5: Segmento de um registo UNIMARC explicado através de uma transformação que o associa ao Manual UNIMARC em XML

⁴ Uma representação gráfica do esquema pode ser consultada em: <http://www.bookmarc.pt/unimarc/schema/FORMAT.html>.

Subfields	a : Title Proper	
	Content	Maria ou A filha de Bernardo
	Repeatable	Yes
	Mandatory	Yes
	Description	The chief title of the item, includes alternative title but excludes other title information (e.g. subtitles) and parallel titles. Repeatable for additional titles by the same author (EX 5). Mandatory for every record.
e : Other Title Information		
Content	comedia em tres actos ornada de musica	
Repeatable	Yes	
Mandatory	No	
Description	Subtitles and other title information that appear subordinate to the title proper on an item (EX 1, 2, 10). Repeatable for each segment of other title information (EX 10) and for parallel other title information (EX 13, 14). It does not include variant titles such as spine title, cover title, and other titles found outside the principal source of information. It may be subordinate to a title appearing in subfield \$a, \$c or \$d.	
b : General Material Designation		
Content	Manuscrito	
Repeatable	Yes	
Mandatory	No	
Description	Text of general material designation (EX 3, 11). This subfield contains a term indicating the broad class of material to which the item belongs given in the language and script of the agency preparing the record. It corresponds to the ISBD General Material Designation.	
f : First Statement of Responsibility		
Content	por Victor Ducange	
Repeatable	Yes	
Mandatory	No	
Description	The first statement of responsibility for a title appearing in subfield \$a, \$c or \$d, or for a numbered or named part of a work appearing in subfields \$h or \$i. Repeatable after each subfield \$a, \$c, \$d, \$h or \$i which has an accompanying statement of responsibility (EX 12) or for each parallel statement of responsibility (EX 6).	

Princípios de desenho para uma revisão

Os princípios de conceção que se propõem agora para uma revisão destes desenvolvimentos procuram responder às questões em aberto apresentadas por Carvalho mas, sobretudo, aos objetivos de especificação de um modelo de interoperabilidade que tem a Web Semântica como alvo e o UNIMARC como ponto de partida.

Identificação por URI. Em primeiro lugar, recorrer a URI para identificar todos os elementos funcionais e todos os vocabulários internos do UNIMARC, normalmente listas controladas de códigos com origem na própria norma (como os códigos de audiência) ou por ela referenciados (como os códigos de país).

Estabilização do MarcXchange. Em segundo lugar, uma adequação do esquema à estabilização do MarcXchange como forma da expressão XML de registos MARC/ISO 2709, que ainda não tinha ocorrido em 2005. Ou seja, o MarcXchange será usado como referência à expressão de registos segundo a semântica da norma bibliográfica.

Tipificação funcional. Em terceiro lugar, um enfoque na tipificação funcional da informação bibliográfica, cujos elementos serão identificados não pela sua formatação sintática (usando a nomenclatura do MarcXchange, *leader*, *controlfield*, *datafield*, *subfield* e *indicator*) mas pela sua funcionalidade semântica: funções de identificação, codificação, anotação, descrição, referência e controlo (para nomear as mais importantes).

Regulação da granularidade. Em quarto lugar, mas ainda associado ao anterior, a diminuição da irregularidade granular da estrutura de dados do UNIMARC. Nos pontos extremos desta irregularidade, fruto da marca do tempo no controlo editorial da norma, alguns segmentos de informação colocados no interior de sub-campos referem-se a

atributos do registo, ao passo que alguns sub-campos só têm valor informacional após serem agregados a outros.

Conexão a outros esquemas. Em quinto lugar, estabelecer mapas de interoperabilidade direta e indireta com outros esquemas e conjuntos de elementos sintáticos e semânticos (ISBD, Dublin Core, MARC21 / BIBFRAME, MODS, EAD), com o objetivo de guiar e acompanhar os processos de interoperabilidade, não apenas ou não apenas diretamente no sentido da Web Semântica, e os processos de remodelação da informação bibliográfica, sobretudo pela identificação dos elementos FRBR a partir dos MARC (AALBERG; HAUGEN; HUSBY, 2006; TAKHIROV; AALBERG; ŽUMER, 2011).

Conexão a vocabulários e conjuntos de dados. Em sexto lugar, estabelecer mapas de interoperabilidade direta e indireta com vocabulários e conjuntos de dados (CDU, VIAF, ISNI, GeoNames, e outros), com o objetivo de enriquecer de conexões os dados produzidos e assim satisfazer de forma mais completa os requisitos da criação de *Linked Open Data*.

Desvinculação do manual. Finalmente, abandono da intenção de uma versão eletrónica do manual, focando-se este novo esquema na expressão dos modelos de dados, regras sintáticas e vocabulários da norma e deixando de lado as suas descrições desenvolvidas, notas e exemplos, que não estarão supostamente licenciados para serem utilizados nesse contexto.

Considerações finais

Com a expectativa da criação de um modelo de interoperabilidade do UNIMARC para a Web Semântica, retoma-se uma linha de desenvolvimento de formalizações XML das normas bibliográficas utilizadas na composição dos catálogos de bibliotecas e outros serviços em Portugal.

Estas formalizações são instrumentais para o processo de produção de repositórios de dados bibliográficos ricos em conexões que vão para além da mera expressão em RDF dos registos. Os princípios da sua conceção são por isso e para isso realinhados ao ambiente operacional da *Linked Open Data*, facilitando a inscrição com valor acrescentado da informação bibliográfica histórica e corrente nas redes globais de informação e conhecimento.

Referências bibliográficas

AALBERG, Trond; HAUGEN, Frank Berg; HUSBY, Ole
2006 A Tool for converting from MARC to FRBR. In GONZALO, JULIO [et al.] ed. - *Research and advanced technology for digital libraries*. [S. l.]: Springer, 2006. ISBN 978-3-540-44636-1. p. 453-456.

BAKER, Thomas [et al.]
2011 *Library Linked Data Incubator Group Final Report* [Em linha], atual. 2011.

Disponível em: <http://www.w3.org/2005/Incubator/lld/XGR-lld/>

BERNERS-LEE, Tim

2009 *Linked Data* [Em linha], atual. 2009. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>

BERNERS-LEE, Tim; FISCHETTI, Mark

2000 *Weaving the Web: the past, present and future of the World Wide Web by its inventor*. [S. l.]: Texere, 2000.

CARVALHO, Joaquim

2005 An XML representation of the UNIMARC Manual: a working prototype. In *World Library and Information Congress* [Em linha]. Oslo: [s. n.], 2005. Disponível em: <http://archive.ifla.org/IV/ifla71/papers/199e-Carvalho.pdf>

CARVALHO, Joaquim Ramos de [et al.]

2004 Meta-information about MARC: an XML framework for validation, explanation and help systems. *Library Hi Tech*. ISSN 0737-8831. 22:2 (2004) 131-137. doi: 10.1108/07378830410524558.

CORDEIRO, Maria Inês; CARVALHO, Joaquim

2002 XML and bibliographic data: the TVS (Transport, Validation and Services) model. In *World Library and Information Congress* [Em linha]. Glasgow: [s. n.], 2002. Disponível em: <http://archive.ifla.org/IV/ifla68/papers/075-095e.pdf>

DIMIC, Bojana; MILOSAVLJEVIC, Branko; SURLA, Dusan

2010 XML schema for UNIMARC and MARC21. *The Electronic Library*. 28:144017 (2010) 245-262. doi: 10.1108/02640471011033611.

INTERNATIONAL STANDARD ORGANIZATION

2006 *Information and Documentation: MarcXchange*. 2006.

INTERNATIONAL STANDARD ORGANIZATION

2008 *Information and documentation: format for information exchange*. 2008.

TAKHIROV, Naimdjon; AALBERG, Trond; ŽUMER, Maja

2010 An XML-Based Representational Document Format for FRBR. In *Web Information Systems Engineering – WISE 2010 Workshops*.

António Tavares Lopes | atlopes@fl.uc.pt

Faculdade de Letras da Universidade de Coimbra