



Inovações tecnológicas no compartilhamento de metadados para elaboração de referências

Erika Alves dos Santos^a, Marcos Luiz Mucheroni^b

^a*Fundação Jorge Duprat Figueiredo de Segurança e Medicina do Trabalho – Fundacentro /
Universidade de São Paulo – USP, Brasil, erikasantos@usp.br*

^b*Universidade de São Paulo – USP, Brasil, mucheroni.marcosl@gmail.com*

Resumo

Trata de relato de investigação inicial, derivado de tese de doutorado onde são problematizadas as normas de referências. Discute-se a normalização de referências no contexto Brasileiro, a partir da análise da estrutura de apresentação da norma ABNT NBR 6023 – Informação e documentação: Referências - elaboração. A partir da análise desta norma, que identificou diversas deficiências no contexto da normalização de referências no Brasil, a discussão avança para o contexto mundial e, evidencia as referências como faceta da representação descritiva e da recuperação da informação. Com fundamentação nas relações de metadados indicadas pelo FRBR, é proposto o relacionamento e o intercâmbio de metadados entre VIAF e OpenCitations, como forma de simplificar o processo de elaboração de referências, do ponto de vista dos pesquisadores e, estabelecer um formato único de representação da informação na forma de referências, de modo a consolidar o princípio da normalização e favorecer a recuperação da informação, inclusive com perspectiva de utilização de recursos semânticos para interligação de metadados relacionados.

Palavras-chave: Referências, Metadados, Interoperabilidade, VIAF, OpenCitations.

Introdução

A democratização tecnológica evidenciou o desequilíbrio na relação entre a oferta e a procura pela informação. Sem entrar no mérito das questões conceituais sobre organização da informação, nem tampouco sobre aspectos cognitivos essenciais ao processo de conversão de dados em informação, desprezar os fundamentos da análise documental, enquanto prática e, as linguagens documentárias, enquanto instrumentos para a organização da informação, configura negligência, pois, ambas constituem os pilares da recuperação da informação.

Também o tratamento destinado à informação também é determinante para a contextualização e atribuição de valor a esta e, deve ser considerado sob duas vertentes: a primeira, sob a visão de quem procede ao tratamento da informação e a segunda, sob a visão de quem a recebe e determinará o seu uso ou descarte, de acordo com a pertinência e relevância ao trabalho do pesquisador.

Nesse contexto, entende-se por tratamento adequado à informação aquele em que ao leitor (pesquisador) são oferecidos todos os elementos necessários à identificação da informação representada, de modo a evidenciar as abordagens contidas em um determinado documento.

No entanto, os recursos tecnológicos, especificamente da Web Semântica, atribuíram um caráter relativista a tal definição, já que há contextos em que o tratamento adequado da informação além de identificar, também promove a interação e informações sobre a localização da informação estabelecendo, inclusive, relações com conteúdos correlatos.

Todavia, o tratamento da informação é uma tarefa de caráter não exclusivo aos profissionais de Ciência da Informação, considerando a publicação científica como o principal meio de comunicação de achados científicos. E em sendo assim, a representação descritiva, na forma de referências, desempenha uma função relevante nos aspectos de encontrabilidade da informação. Uma vez que a redação de referências normalmente é conduzida pelos próprios pesquisadores nas diversas áreas do conhecimento, é importante que os padrões de apresentação de metadados para a elaboração de referências sejam linguagens de domínio comum.

A incorporação das tecnologias semânticas ou de compartilhamento e intercâmbio (interoperabilidade) de metadados para a representação descritiva, no âmbito da elaboração de referências não é uma realidade final, há esforços que apontam caminhos.

Mais do que uma questão de processo, o atual panorama conduz ao entendimento de que a comunidade científica está alheia aos benefícios, e talvez até à possibilidade, da interoperabilidade aplicada à elaboração de referências.

Método

Trata de estudo descritivo e exploratório, que integra pesquisa de doutorado em Ciência da Informação, em desenvolvimento na Universidade de São Paulo (USP), Brasil. A discussão apresenta considerações críticas sobre a norma ABNT NBR 6023 – Informação e documentação: Referências: elaboração, publicada pela Associação Brasileira de Normas Técnicas (ABNT). A norma, amplamente adotada no Brasil e cuja edição mais recente data de 2002 é baseada nas normas ISO 690:1987 e ISO 690-2:1997, da International Standardization Organization (ISO), ambas canceladas e substituídas pela norma ISO 690:2010 em 14 de junho de 2010. Além do levantamento bibliográfico e a partir da análise crítica de conteúdo das normas supra, procedeu-se avaliação da contribuição das propostas apresentadas pelo VIAF e pelo Open Citations no estabelecimento do intercâmbio de metadados para a elaboração de referências em caráter mundial, baseado nas tecnologias da Web Semântica.

Resultados

As análises sugeriram que a norma ABNT NBR 6023 – Informação e documentação: Referências: elaboração é deficitária no cumprimento de seu propósito de fornecer subsídios para a redação de referências de forma clara, concisa e, principalmente, padronizada. A norma não instrui sobre os procedimentos descritivos adequados para alguns tipos de documentos, a exemplo das entrevistas. Isto favorece a ocorrência de redações de referências distintas para um mesmo documento. Também inexistem indicações de procedimentos descritivos para documentos nos quais não consta data de publicação. Também são deficitárias as orientações normativas sobre a ordenação da lista de referências: não há menção procedimental para a ordenação correta de duas ou mais referências de documentos de mesma autoria e sem indicação de data de publicação, nem tampouco procedimentos para referenciar partes distintas, porém não seguidas e de mesma autoria, na mesma obra.

Estas e outras questões concorrem para a não completude do alcance dos objetivos estabelecidos pela ABNT NBR 6023:2002, no sentido de “orientar a preparação e compilação de referências do material utilizado para a produção de documentos e para a inclusão de bibliografias, resumos, resenhas, resenhas e outros.” (ABNT, 2002, p. 1).

Além disso, a norma ABNT NBR 6023:2002 determina, em caráter obrigatório em alguns casos, a descrição de metadados dispensáveis para a recuperação da informação. Um exemplo é a indicação da cidade de publicação de publicações periódicas. Ora, o local de publicação de periódicos é uma informação que nem sempre é tratada com destaque pelos editores e pode, inclusive, demandar pesquisas em fontes externas ao documento em si. Tal obrigatoriedade, além de gerar dificuldades

para o pesquisador, que nem sempre sabe onde localizar tais informações, demonstra que a norma ABNT NBR 6023:2002 não está alinhada com os recursos tecnológicos para a recuperação da informação. Outro sinal de sua obsolescência é o fato de que dos 4 documentos prescritivos à referida norma, pelo menos 1 está obsoleto: a norma ABNT NBR 10522:1988 Abreviação na descrição bibliográfica, foi cancelada, sem substituição, em 30 de setembro de 2003.

As normas ISO 690:1987 e ISO 690-2:1997, nas quais está baseada a ABNT NBR 6023:2002, foram ambas substituídas pela norma ISO 690:2010. Outro instrumento prescritivo é o Código de Catalogação Anglo Americano (CCAA2), (em inglês, Anglo American Cataloging Rules - AACR2), que tende a ser gradativamente substituído pelo Readable Description and Access (RDA). Conste, entretanto, que a ABNT NBR 6023:2002 permaneceu em processo de revisão de março de 2009 até março de 2017. O primeiro projeto de revisão da norma em pauta circulou em consulta nacional entre agosto e outubro de 2015, enquanto o segundo projeto de revisão, que data de junho de 2018, estará disponível para consulta pública até 16 de julho de 2018.

Embora tal projeto de revisão não configure o foco desta reflexão, algumas observações merecem destaque: as omissões de informações nas orientações descritivas, tendem a persistir, face a adoção frequente de termos e expressões genéricas como “entre outros”. É compreensível que em alguns casos, o detalhamento da informação se torna inviável pela sua extensão, porém, deixar de fazê-lo pode representar óbices aos objetivos normativos de alcançar padrões mínimos de uniformidade em suas áreas de competência, na medida em que tais omissões dão abertura para interpretações distintas em diferentes contextos, que podem culminar em aplicações múltiplas de uma mesma instrução normativa. A revisão da norma também não esclarece a diferença entre bibliografia e referência. Outra aparente novidade é a possível previsão de extração de metadados a partir de outras fontes de informações além do próprio documento, o que já é uma tendência na representação descritiva, analogamente às prescrições do RDA. Em contrapartida, o projeto de revisão retirou a indicação dos códigos de catalogação como documentos prescritivos, o que pode caracterizar um retrocesso, ao dissipar a aproximação entre a catalogação e a elaboração de referências, enquanto facetas da representação descritiva. Também pode ser alvo de questionamento a efetividade de uma norma desenvolvida (ou revisada) em um cenário no auge de um processo de redesenho na organização da informação, tal como ocorre atualmente com a representação descritiva e, sobretudo, considerando a notável irreflexão de tais conceitos no conteúdo da norma.

Cumulativamente, tais pontos fomentam múltiplas interpretações das diretrizes para a elaboração de referências, que culminam no aumento da probabilidade de ocorrência de referências distintas para um mesmo documento. As deficiências identificadas na representação da informação na forma de referências no cenário brasileiro são passíveis de ocorrer também em outros países, tendo como agravante a existência de cerca de 8 centenas de estilos bibliográficos distintos. Uma vez que pesquisadores não se limitam a publicar trabalhos apenas em seus países de origem, o conhecimento sobre as diretrizes estabelecidas pelos estilos bibliográficos vigentes em cada país (ou área do conhecimento) para os quais venham a submeter suas contribuições se torna um requisito.

Em resposta a tal cenário, alguns editores científicos disponibilizam instrumentos de elaboração automática de referências, baseados nos inúmeros estilos bibliográficos vigentes pelo mundo. Tais instrumentos demandam o preenchimento, de forma prévia e manual, dos campos de metadados correspondentes à publicação a ser referenciada, os quais são armazenados apenas localmente. Não há uma base de dados para armazenamento compartilhado e reutilização destes metadados, de modo que a cada vez que um mesmo documento é referenciado, os campos de metadados devem ser novamente preenchidos, em contradição aos princípios do compartilhamento de dados e da interoperabilidade. Desconsiderem-se no âmbito dessa discussão o tempo necessário para adequação desses softwares às

revisões das normas de referências em todo o mundo e, também, a fundamentação conceitual e os documentos prescritivos de cada uma das normas de referências vigentes no mundo, sob o risco de adentrar questionamentos que extrapolam a delimitação desta discussão. Porém, se as referências são uma faceta da representação descritiva e, se as normas são instrumentos delimitadores de suas formas de apresentação, visando sobretudo a padronização, a variância no conjunto documental que fundamenta os estilos bibliográficos é questionável e contradiz os princípios da normalização.

Neste cenário, se propõe o relacionamento e o compartilhamento de metadados entre VIAF e OpenCitations, como forma de simplificar o processo de elaboração de referências. Em sendo o VIAF um consórcio de cooperação internacional que visa estabelecer entradas padronizadas para autoridades (autores pessoais, corporativos e eventos), e o OpenCitations, um repositório de acesso aberto para o compartilhamento de metadados referentes a citações e referências, a interoperabilidade entre as duas iniciativas por meio de um modelo RDF é potencialmente favorável à universalização das formas de registro de referências e ao desencargo da comunidade científica no que se refere ao manuseio de inúmeros estilos bibliográficos.

Entende-se que a aplicação das tecnologias semânticas, pautadas nos princípios de relações de metadados de que trata o FRBR, às referências, além de lhes agregar valor enquanto fontes de informações, também tornaria injustificável a multiplicidade de estilos bibliográficos. Assim, a referência assumiria efetivamente o papel de fonte de informação, como uma extensão do catálogo bibliográfico.

Discussão

Como instrumentos de controle e acesso à informação, catálogos bibliográficos e referências são abrangidos pelo conceito de representação descritiva e se constroem a partir de diretrizes estabelecidas por padrões e normas de amplo espectro de reconhecimento e uso. Contudo, o conjunto documental que norteia a representação descritiva, especialmente as diretrizes para a elaboração de referências, se apresenta segundo a linguagem específica da Ciência da Informação com todo o seu arsenal de abreviaturas, siglas e conceitos, mesmo em alguns dos padrões bibliográficos de maior utilização. Inclusive, há estilos bibliográficos que apontam as diretrizes de descrição documental determinadas pelo Anglo American Cataloging Rules (AACR2) como instrumentos prescritivos e que recomendam a sua consulta para o fomento de decisões mediante situações não abrangidas pelo próprio estilo bibliográfico. É o caso da norma ABNT NBR 6023:2002. Tal cenário evidencia a fragilidade na interpretação e uso da norma em pauta, que se revela como uma das possíveis causas para a elaboração de padrões próprios, ou versões interpretadas dos estilos bibliográficos pelos editores científicos e instituições de ensino. Sobre a existência de diversos estilos bibliográficos e suas múltiplas interpretações, Bertholino e Silva (2008, p. 40), pontuam que:

“[...] há uma miscelânea de tipos de apresentação. Os artigos encontrados nas revistas apresentam padrões que em determinadas situações seguem interpretações dos editores.

Um exemplo eminente é a adoção, pela área de Ciências da Saúde ou Biomédicas, da norma Vancouver que difere em forma de apresentação das referências preconizadas pela ABNT.

Surgem, então, questionamentos e dúvidas e uma grande confusão de interpretação pelos alunos e/ou autores que querem padronizar sua produção científica.

[...]

Seria extremamente interessante que algum tipo de consenso pudesse ser alcançado a fim de que as comunidades científicas, acadêmicas e editores trabalhassem como uma única norma evitando as disparidades, hoje observadas, na normalização.”

Apesar da factibilidade na correspondência entre catalogação e elaboração de referências, a restrição de sua materialização apenas por meio dos instrumentos documentais de apresentação das diretrizes, em ambas as atividades, pode representar óbices para o entendimento da forma adequada de

apresentação de metadados em citações e referências, principalmente por pesquisadores de outras áreas que não a Ciência da Informação. Em contrapartida, a normalização requer noções de documentação nem sempre familiares aos pesquisadores.

“Em uma realidade em que, cada vez mais se tende a manipular grandes massas de dados, dentro do princípio cibernético de que todo esforço a mais na entrada do processo (aqui visto amplamente, como a geração do conhecimento) redunde em uma esperada economia na saída (recuperação e uso da informação), a normalização surge como fator não só de qualidade, mas como facilitador da transferência da informação científica.” (Rodrigues, Lima e Garcia, 1998, p.155)

A codificação da informação em representações passíveis de processamento automático é um dos meios que permitem que as máquinas façam o processamento que atualmente, na web sintática, exige a intervenção humana. O processamento automático da informação pode ser um grande aliado no sentido de prover informações relevantes no tempo e espaço correto e, pode trazer benefícios em diversos aspectos, desde o comércio eletrônico, até o gerenciamento do conhecimento e sua aplicação pelo usuário final, por meio da indexação de informações em mecanismos de busca e da utilização de agentes de softwares que agem de maneira independente e têm autonomia para atuar em nome dos interesses de seus usuários. (BREITMAN, 2014). Isto reforça a necessidade de repensar as formas de representação da informação que a partir da aplicação de funcionalidades semânticas, considerando que pode estabelecer uma aproximação exitosa entre a catalogação e a apresentação de referências, no sentido de agregar informações de forma independente e proativa.

“O que é necessário é uma forma de representar o conhecimento que permita aos computadores tanto interpretá-lo no sentido tradicional de exibi-lo na tela em um formato legível por humanos, quanto compreendê-lo em um nível de computador, permitindo assim que o computador reaja de forma autônoma a esse conhecimento.” (KÜCK, 2004, tradução nossa).

Neste sentido, o compartilhamento e intercâmbio de metadados entre VIAF e Open Citations representa a materialização do modelo entidade-relacionamento proposto pelo FRBR, e pode trazer benefícios ao contexto da normalização de referências, seja pelo controle das formas de registro das autoridades, seja pela atribuição do caráter inteligível e semântico às referências pelas máquinas ou, pelo desencargo da comunidade científica em interpretar as diretrizes normativas estipuladas por cada estilo bibliográfico. O trabalho cooperativo e o intercâmbio de metadados são iniciativas consagradas na Ciência da Informação, porém, é admirável o fato de que embora inúmeros estudos demonstrem a efetividade e os benefícios oriundos da aplicação dessas filosofias de trabalho para a organização e encontrabilidade da informação, parece não haver movimentação efetiva para a aplicação de tais técnicas em favorecimento da representação da informação na forma de referências. Tal fato, somado à complexidade envolvida no processo da estruturação de metadados, denuncia uma postura (involuntária) da Ciência da Informação que além de depreciar e contradizer o propósito da organização e recuperação da informação, ainda sugere que o compromisso com o registro da informação é uma atribuição exclusiva dos bibliotecários, até mesmo porque, os instrumentos reguladores dessa atividade, tal como as normas e guias de utilização de estilos bibliográficos se apresentam e em uma linguagem técnica, vulgarmente chamada de “biblioteconomês”, que dificulta a interpretação de pesquisadores de áreas além da competência da Ciência da Informação. Contudo, não se pode ignorar que a redação de referências integra ativamente o cenário da organização e recuperação da informação, em alguns casos de forma mais efetiva do que os próprios catálogos bibliográficos e que, a sua redação e uso extrapola a atuação profissional do bibliotecário, sendo exercida por grande parte dos pesquisadores.

O compartilhamento de metadados para a elaboração de referências pode ser uma alternativa para tornar mais efetiva a normalização na redação de referências. Inclusive, havendo um repositório compartilhado de acesso aberto no qual as referências possam ser recuperadas, extingue-se a necessidade da existência de estilos bibliográficos múltiplos em prol da garantia da uniformidade na

apresentação de referências. De fato, a forma de registro de metadados é determinante para a recuperação, ou não, de uma informação codificada.

“[...] a relativa falta de experiência de muitos acadêmicos neste novo contexto internacional propicia alguns problemas que afetam a visibilidade de seus trabalhos. A isto há que se somar as dificuldades decorrentes das diferenças culturais entre os criadores das principais bases de dados [...] e os diversos contextos nacionais aos que se projetam como expressão de sua globalização. Produz-se assim uma grande quantidade de desencontros e erros que afetam a qualidade e eficácia dos processos de codificação e recuperação da informação.

[...] Os problemas com a correta identificação dos autores preocupam não somente aos envolvidos, mas também às grandes empresas que gerenciam as principais bases de dados bibliográficos que também pretendem reduzir o número de erros e desenvolver ferramentas que permitam buscas mais precisas” (ALIAGA; CORREA, 2011, tradução nossa).

Dois iniciativas promissoras nesse contexto são o VIAF e o OpenCitations. O primeiro, trata de um consórcio internacional do qual participam mais de 50 instituições, na sua maioria bibliotecas nacionais, que visa o intercâmbio de dados de autoridades para pessoas, entidades e eventos, bem como o relacionamento com suas respectivas obras. O OpenCitations, por sua vez trata de um repositório de metadados de referências. O compartilhamento de metadados entre VIAF e OpenCitations, pode ser uma alternativa na unificação das formas de apresentação de referências e, uma vez consolidado, tornaria injustificável a existência de centenas de estilos bibliográficos distintos. Ter-se-ia então efetivado a efetiva normalização na apresentação de referências, que poderia inclusive, se apresentar nos documentos na forma de um link onde as informações seriam descritas na sua forma completa. Além disso, a proposta além de conferir maior visibilidade às publicações, por meio da uniformidade na apresentação de referências, também possibilita o tratamento semântico da informação contida em tais instrumentos de compartilhamento de metadados. Isto pode, inclusive, ser considerado como elemento de ampliação de suas funções enquanto plataformas de relacionamento de metadados, com efetiva contribuição positiva para a recuperação da informação.

Nesse contexto, a atribuição de significados, o compartilhamento e a ligação de dados compõem a tríade (sujeito, predicado e objeto), que rege a organização da informação na era pós-semântica e consolida as relações interdisciplinares entre a Ciência da Informação e a Informática.

Conclusões

As contribuições da tecnologia para a Ciência da Informação no que se refere à organização e gestão da informação são cada vez mais presentes e evidentes. Entretanto, não há evidências de uma discussão organizada sobre a utilização de dispositivos semânticos para a gestão e recuperação da informação. Embora as tecnologias tenham conquistado popularidade e reconhecimento social, ainda há muito que se avançar no âmbito da Ciência da Informação. A indústria do comércio eletrônico e das mídias de redes sociais está explorando cada vez mais as funcionalidades semânticas e de dados ligados (*linked data*) (BIZER; HEATH; BERNERS-LEE, 2009), ironicamente, de forma mais evidente e eficaz do que a Ciência da Informação, em alguns aspectos.

No que se refere à recuperação da informação por meio de referências, a simples supressão ou inclusão de instruções nas normas pode tornar o cenário da gestão da informação confuso para a maioria dos pesquisadores, a quem a complexidade do tema pode parecer ininteligível, porém, o uso das tecnologias em favor do gerenciamento de dados e metadados visando a organização e recuperação da informação pode representar uma alternativa para o ambiente confuso do universo normativo, do ponto de vista dos pesquisadores.

É fundamental, portanto, não apenas a discussão da integração dos instrumentos de gestão de metadados e da representação da informação, mas também o uso das tecnologias na inclusão crítica e didática de instrumentos que favoreçam o rigor às normas e a facilidade de uso pelos pesquisadores, sobretudo em um cenário em que mudanças ocorrem continuamente, decorrentes tanto do

desenvolvimento de novas mídias como da complexidade que envolve o tratamento e recuperação da informação.

O uso das tecnologias não pode ser ignorado nesse contexto, sobretudo considerando as funcionalidades da Web Semântica que, em face dos dados ligados, suscita questionamentos sobre a real necessidade do dispêndio de esforços para a normalização de referências, considerando, além da infinidade de estilos bibliográficos distintos, a possibilidade de ampliar os seus índices de encontrabilidade, por meio do tratamento semântico e da interligação de metadados por meio dos dados ligados (linked data).

Referências

ALIAGA, F. M.; CORREA, A. D. (2011) - Tendências em la normalización de nombres de autores em publicaciones científicas. *RELIEVE* [em linha], Vol 17, Nº. 1. [Consult. 20 Fev. 2018]. Disponível na Internet: <http://www.uv.es/RELIEVE/v17n1/RELIEVEv17n1_0.pdf>. ISSN 1134-4032

BERTHOLINO, M. L. F.; SILVA, V. L. B. da. (2008) - Normas técnicas de informação e documentação: ABNT versus Vancouver. *Publicatio UEPG: Ciências Biológicas e da Saúde* [Em linha]. Vol. 14, Nº. 2. [Consult. 03 Jan. 2018]. Disponível na Internet: <<http://www.revistas2.uepg.br/index.php/biologica/article/view/1275/920>>. ISSN 1676-8485

BIZER, C.; HEATH, T.; BERNERS-LEE, T. (2009) - Linked data: the story so far. *International Journal on Semantic Web and Information System*. ISSN 1552-6283. Vol. 5, Nº 3, p. 1-22.

BREITMAN, K. K. (2014) - *Web semântica: a internet do futuro*. Rio de Janeiro: LTC. 190 p. ISBN 978-85-216-1466-1

KÜCK, G. (2004) - Tim Berners-Lee's semantic web. *South African Journal of Information Management* [em linha]. Vol. 6, Nº. 1. [Consult. 15 Mar. 2018]. Disponível na Internet: <<https://sajim.co.za/index.php/sajim/article/view/297/288>>. ISSN 1560-683X

RODRIGUES, M. E. F.; LIMA, M. H. T. de; GARCIA, M. J. de O. (1998) - A normalização no contexto da comunicação científica. *Perspectivas em CI* [em linha]. Vol. 3, Nº 2 [Consult. 15 Mar. 2018]. Disponível na Internet: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/603/372>>. ISSN 1981-5344