

## METADADOS DE QUALIDADE E VISIBILIDADE NA COMUNICAÇÃO CIENTÍFICA

Silva Silvia Maria Puentes Bentancourt<sup>i</sup>  
Rafael Port da Rocha<sup>ii</sup>

**Resumo:** Produzir metadados com qualidade é um quesito fundamental para dar visibilidade a revistas eletrônicas, pois o modelo tecnológico atual para revistas de acesso livre na web está focado na interoperabilidade de metadados. Este trabalho investiga a qualidade dos metadados da Revista Eletrônica Em Questão para fins de colheita desses metadados com base no protocolo OAI-PMH. Realiza a colheita dos metadados da revista no formato Dublin Core, e os analisa a partir dos critérios de qualidade do Modelo de Qualidade de Dados ISO/IEC 25012 e pelas especificações, recomendações e melhores práticas para uso de Dublin Core. Conclui que há qualidade nos elementos Identificador, Idioma, Editor, Fonte, Título, Formato e Descrição. Nos demais, há comprometimentos com a qualidade, principalmente quanto à inconsistência (em Date, Colaborador, Cobertura e Criador, pela falta de critérios para os valores), incompletude (Colaborador, Cobertura e Tipo, por não ser usado em todos os recursos requeridos) e exatidão (tipo, por conter valores imprecisos ou errados).

**Palavras-chave:** Metadados. Dublin Core Metadata Initiative (DCMI). Acesso Livre (Open Access). Visibilidade de Periódico Científico. Qualidade de dados.

### *METADATA OF QUALITY AND VISIBILITY IN SCIENTIFIC COMMUNICATION*

**Abstract:** Metadata quality is fundamental to the visibility of electronic journals, since metadata interoperability is on the base of the current technical framework for open access on the web. This paper investigates the quality of metadata of the journal *Em Questão* for the purpose of metadata harvesting, which followed the OAI-PMH protocol. The quality of metadata harvested from the journal was analyzed using Dublin Core best practices, recommendations and specifications, and ISO/IEC 25012 Data Quality Model. It concludes that there is quality in the elements Identifier, Language, Publisher, Source, Title, Format and Description. The other elements have lack of quality, especially with regard to inconsistency (in Date, Contributor, Coverage, Creator; due to the lack of criteria for the values), incompleteness (Contributor, Coverage and Type; due to not be used in all the required resources), and accuracy (Type; due to contain inaccurate or erroneous values).

**Keywords:** Metadata. Dublin Core Metadata Initiative (DCMI). Open Access. Visibility of Scientific Journals. Data Quality.



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/)

<sup>i</sup> Universidade Federal do Rio Grande do Sul – [silviampb@gmail.com](mailto:silviampb@gmail.com)

<sup>ii</sup> Universidade Federal do Rio Grande do Sul – [rafael.rocha@ufrgs.br](mailto:rafael.rocha@ufrgs.br)

## 1 INTRODUÇÃO

As mudanças tecnológicas levaram a mudanças de hábitos e costumes que resultaram em transformações sociais, entre elas, a relação com a informação. A tecnologia favoreceu a disseminação da informação, mas provocou a proliferação desta que, em muitos casos, ocorre de forma desordenada, complexa e nem sempre confiável. A falta de infraestrutura qualificada para o tratamento da informação pode prejudicar a recuperação da informação, mesmo quando está armazenada em base de dados.

Para facilitar a disseminação e o acesso, a comunicação científica beneficiou-se de recursos tecnológicos, de softwares livres, como o *Open Journal System* (OJS)<sup>3</sup>, que viabilizaram o surgimento de periódicos digitais editados em centros de pesquisa vinculados a universidades e outras instituições similares. O Acesso Livre ampliou as opções e segundo ele, o importante é compartilhar para alcançar o crescimento conjunto.

Num cenário em que os usuários tendem a usar motores de busca para a recuperação da informação, como os portais de teses e dissertações BDTD<sup>4</sup> e DART<sup>5</sup>, e o portal de periódicos eletrônicos DOAJ<sup>6</sup>, a recuperação da informação armazenada numa base de dados depende dos metadados disponibilizados para esses motores. No entanto, só fornecer os metadados não é garantia de sucesso, ou seja, os dados transmitidos precisarão estar corretos, acessíveis no momento necessário, de forma completa e no formato certo. Portanto, a qualidade dos metadados disponibilizados para serviços de recuperação da informação é fundamental para que o processo se desenvolva a contento. Essa qualidade precisa ser garantida, seja por meio da definição e uso de Política de Metadados, como Normas de Procedimentos, treinamentos, avaliações, ou outros.

A reunião de periódicos em portais que viabilizam a recuperação e o acesso às suas edições os tornaram, a princípio, mais visíveis. Porém, o desafio da visibilidade permanece e requer um esforço editorial que faça girar o círculo virtuoso no qual revistas de qualidade atraem mais submissões e mais citações, o que resulta em maior prestígio e qualidade, reiniciando o ciclo. Assim, os editores não podem prescindir de buscar soluções técnicas para alcançar seus propósitos de visibilidade. A interoperabilidade envolvendo diferentes bases de dados é uma dessas soluções, na qual as bases disponibilizam metadados sobre os seus recursos informacionais que serão compartilhados. Padrões para prover interoperabilidade são

---

<sup>3</sup> Disponível em: <<http://pkp.sfu.ca/?q=ojs>>

<sup>4</sup> Biblioteca Digital Brasileira de Teses e Dissertações. Disponível em: <<http://bdttd.ibict.br/>>

<sup>5</sup> *Europe E-theses Portal*. Disponível em: <<http://www.dart-europe.eu/basic-search.php>>

<sup>6</sup> *Directory of Open Access Journals*. Disponível em: <<http://www.doaj.org/>>

promovidos pela Iniciativa dos Arquivos Abertos (OAI), que inclui o protocolo para colheita de metadados OAI-PMH<sup>7</sup>, e institui o *Dublin Core* como requisito mínimo de representação de metadados.

A qualidade dos metadados de um periódico, que proporcionarão a visibilidade, envolve um universo de decisões que inclui a escolha do *software* para disponibilizar o periódico. No Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) oferece um ambiente próprio para isso, o Sistema Eletrônico de Editoração de Revistas (SEER)<sup>8</sup>. Entre as decisões táticas envolvidas, estão escolhas que incluem padrões de interoperabilidade (como o protocolo OAI-PMH) e de representação dos metadados (como o padrão *Dublin Core*). No caso do *Dublin Core*, ainda há que qualificá-lo, ou seja, definir o perfil de aplicação mais adequado, com decisões embasadas em conhecimento dos procedimentos do *Dublin Core Metadata Initiative* (DCMI) e de meios de assegurar a qualidade dos metadados, entre outras.

Este trabalho investiga o uso de padrões do DCMI e a qualidade dos metadados de uma revista científica. Levanta a situação atual dos metadados disponibilizados na Revista Em Questão, da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul (FABICO/UFRGS), e os analisa com o propósito de identificar a sua aderência às recomendações de melhores práticas do *Dublin Core Metadata Initiative* (DCMI) mediante o levantamento das não conformidades com as recomendações do DCMI. Os metadados foram analisados por critérios de qualidade de dados para cada elemento que descrevem. Para unidades de análise, foram selecionados todos metadados disponibilizados na *Web* referentes às publicações de 2003 a 2010, totalizando 17 edições.

## 2 METADADOS NA VISIBILIDADE DA COMUNICAÇÃO CIENTÍFICA NA WEB

Weitzel (2006) aponta que o ambiente eletrônico provocou mudanças estruturais no fluxo de **comunicação científica** e que “[...] existe um mito no qual o discurso científico se perde na *internet*, tanto no sentido figurativo diante dos discursos possíveis na *internet*, quanto literalmente (WEITZEL, 2006, p. 86). Apenas disponibilizar na *Web* não garante a disseminação e o uso, são necessários mais elementos do que apenas a indexação dos assuntos; também se necessita da localização do documento, permissão de acesso, direitos de uso, entre outros. Ao discutir modelos emergentes da comunicação

---

<sup>7</sup> Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Disponível em: <<http://www.openarchives.org/pmh/>>

<sup>8</sup> Sistema Eletrônico de Editoração de Revistas (SEER). Disponível em: <<http://seer.ibict.br/>>

científica, Pavan (2008, p. 46) atenta que há uma forte presença da tecnologia e que esses “refletem os novos anseios das comunidades científicas, como o livre acesso ao conhecimento científico.” Nesse contexto, instituições de pesquisa buscam disseminar sua produção científica agilmente e aderem aos Repositórios Institucionais.

**O Movimento do Acesso Livre** visa a disponibilização da literatura científica de forma integral na *Web*, para ler, baixar os arquivos, partilhar, divulgar, em suma, para que esta seja utilizada ao juízo do usuário de forma ampla. O Acesso Livre envolve, além de questões políticas ou de normalização, questões relacionadas à cultura e aos costumes da comunidade científica, entre as quais, novamente, a questão tecnológica se faz presente. Apesar das mudanças, a prática da revisão por pares permaneceu. Mas o Acesso Livre pode ir além da comunicação científica; é uma ferramenta que pode atingir a disseminação da informação científica.

A partir daí surgiu a necessidade da **interoperabilidade** entre os repositórios e/ou periódicos. No movimento do Acesso Livre, a interoperabilidade ocorre mediante a troca de metadados, os quais representam as informações contidas nos documentos disponibilizados ou que dizem respeito a estes. Em 1999 foi criada a *Open Archive Initiative* (OAI) para promover e desenvolver a interatividade, a qual também foi responsável pela criação do *Open Archives Initiative Protocol for Metadata Harvesting* (OAI – PMH). Esse protocolo foi desenvolvido com o propósito de distribuir e coletar metadados, com foco em descritores de documentos. O protocolo OAI - PMH atua em uma arquitetura composta por provedores de serviços e provedores de dados e utiliza a técnica da colheita de metadados. Os provedores de dados são os gestores de arquivos abertos, que devem possuir, pelo menos, mecanismos de submissão de trabalhos, sistema de armazenamento a longo prazo, e mecanismo de exposição de metadados para colheita (KURAMOTO, 2006). Provedores de serviço são “as instituições ou serviços de terceiros que implementam os serviços com valor agregado a partir dos dados coletados junto aos arquivos *e-prints*, ou repositórios digitais” (KURAMOTO, 2006, p. 94). Assim, o protocolo de comunicação OAI-PMH possibilita a coleta a partir de determinado provedor de dados. Como o protocolo OAI-PMH adota o padrão *Dublin Core* como requisito mínimo para representação dos metadados, os provedores de dados devem obrigatoriamente dispor seus metadados representados em *Dublin Core*, embora possam utilizar outros padrões para representá-los, como o *Metadata Object Description Schema* (MODS) e o *Learning Object Metadata* (LOM).

O SEER, plataforma a partir da qual é publicada a Revista Eletrônica Em Questão, e o *Directory of Open Acces Journals* (DOAJ) são exemplos de provedor de dados e provedor de serviços,

respectivamente. O SEER visa agilizar o fluxo das informações. O DOAJ, segundo consta na sua página oficial na *Web*, se propõe a aumentar a visibilidade e facilitar o uso das revistas de Acesso Livre, sejam científicas ou acadêmicas, porém preocupadas com a qualidade do seu conteúdo. Afirma que qualidade do conteúdo pode ser resultado do trabalho do editor, equipe editorial e/ou pela revisão por pares.

Uma característica da comunicação científica é a **visibilidade**, estudada por Packer e Menegheni (2006). A visibilidade, conceituada por esses autores, “representa a capacidade de exposição que uma fonte ou fluxo de informação possui de, por um lado, influenciar seu público alvo e, por outro, ser acessada em resposta a uma demanda de informação.” (PACKER; MENEGHENI, 2006, p. 237). A visibilidade de um periódico está associada aos metadados coletáveis e a sua capacidade de representação dos artigos publicados em periódicos por ele registrados. Meleros (2011) aponta como instrumentos para a melhoria da visibilidade e uso das revistas científicas: a adoção de normas formais de qualidade, bem como a possibilidade de colher seus metadados, entre outros. Acrescenta que toda publicação presente em portais de acesso livre já implica na adoção de requisitos que estes determinam, no sentido da qualidade e de propiciar interoperabilidade. Finalmente, a autora afirma que cabe aos editores observar a qualidade dos conteúdos, imprescindíveis para a visibilidade. Aliado a isso, “os editores terão que conhecer melhor as normas de metadados e dos gerenciadores de editoração eletrônica para aproveitar ao máximo as vantagens das publicações eletrônicas [...]” (CETTO; ALONSO GAMBOA, 2011, p. 17).

### 3 A PRODUÇÃO DE METADADOS EM *DUBLIN CORE*

Os padrões de metadados servem para a representação dos documentos. O *Dublin Core* é um padrão de metadados que possibilita a catalogação de documentos eletrônicos na *Web*. Segundo Rocha (2004, p. 119), o “*Dublin Core* apresenta-se como língua mãe para metadados, que pode ser estendida, através da criação de Perfis de Aplicação, para contemplar necessidades de comunidades específicas.” O *Dublin Core* tem sido amplamente adotado, ele coexiste com outras normas de metadados com semânticas diferentes e o seu lema é “tornando mais fácil encontrar a informação” (DCMI, *on-line*).

A norma ISO 15836 aborda os 15 elementos básicos de *Dublin Core*, nomeadamente: Título (*title*), Criador (*creator*), Assunto (*subject*), Descrição (*description*), Publicador (*publischer*), Colaborador (*contributor*), Data (*date*), Tipo (*type*), Formato (*format*), Identificador (*identifier*), Fonte (*source*), Idioma (*language*), Relação (*relation*), Cobertura

(*coverage*), Direitos Autorais (*rights*). Estes elementos são apresentados no documento *Dublin Core Metadata Element Set*<sup>9</sup>.

Os 15 elementos ainda podem ser ampliados pelo uso de um refinamento, que diz respeito aos qualificadores *Dublin Core*. A adoção dos qualificadores possibilita representar a informação com maior precisão. O refinamento dos elementos consta nas orientações sobre o uso do padrão no documento *Using Dublin Core – Dublin Core Qualifiers*<sup>10</sup>, que é o resultado do processo de uso e busca de melhores práticas e não pretende ser exaustivo. A qualificação pode ser pelo Refinamento do Elemento ou pelo Esquema de Codificação. O refinamento procura especificar o elemento a fim de reduzir seu escopo. O Esquema de Codificação auxilia a interpretar o valor do elemento e segue instruções de vocabulários controlados, notações formais e regras de análise conhecidas.

O DCMI apresenta uma relação de instruções, denominadas Melhores Práticas, para orientar a implantação e uso do padrão. As Melhores Práticas foram organizadas em aspectos gerais e específicos. Os aspectos gerais, que se aplicam a todos os elementos, são:

a) cada um dos 15 elementos utilizados precisa ter seus valores conforme a definição e instruções (demais recomendações) desse elemento no documento que define os Termos de Metadados do *Dublin Core Metadata Initiative*<sup>11</sup>;

b) se algum dos elementos requer um refinamento, este deve ser utilizado;

c) quando necessário, optar pelo atributo *xsi:tipo* para especificar o esquema de codificação a fim de expressar o valor do registro com maior precisão;

d) quando necessário, utilizar o atributo *xml: lang* conforme valores do *Open Language Archives Community (OLAC) language*<sup>12</sup> para identificar o idioma em que está escrito o conteúdo do elemento; e

e) no caso de um elemento possuir mais de um valor, recomenda-se o uso de refinamento em detrimento de usar mais de um valor para um mesmo elemento.

Com a intenção de ter-se uma noção da importância desse esquema padronizado de metadados, Sené (2004) fez um levantamento da repercussão dos trabalhos do DCMI pela perspectiva do número de aprovações de normas internacionais logo após o congresso do DCMI em 2001. A autora listou 7 normas ISO em três anos.

#### 4 METADADOS COM QUALIDADE

<sup>9</sup> *Dublin Core Metadata Element Set*. Disponível em: <<http://dublincore.org/documents/2010/10/11/dces/>>

<sup>10</sup> *Using Dublin Core – Dublin Core Qualifiers*. Disponível em: <<http://dublincore.org/documents/2005/11/07/usageguide/qualifiers.shtml>>

<sup>11</sup> *DCMI Metadata Terms*. Disponível em: <<http://dublincore.org/documents/2010/10/11/dcmi-terms/>>

<sup>12</sup> OLAC. Disponível em: <<http://www.language-archives.org/>>

Apesar dos esquemas padronizados de dados, como o DCMI, ainda há a preocupação de assegurar que o valor atribuído ao metadado (campo) esteja correto. Como garantia, é apropriado o uso de conceitos de qualidade de dados e normas. A qualidade pode ser definida em função da conformidade a especificações, denominada qualidade de projeto, que no caso dos metadados refere-se à conformidade com as especificações dos padrões adotados. A norma ISO/IEC 25012 aborda especificamente a qualidade dos dados de forma multidimensional, segundo critérios classificados em duas categorias. A primeira categoria reúne critérios relativos à qualidade inerente dos dados, onde possa atender-se uma demanda informacional a partir dos dados em si. A segunda, por outro lado, dependente do sistema, presume que a qualidade do dado é melhorada ou dependente do sistema computacional em que opera. No entanto, também há casos em que ambas as categorias estão presentes.

Os 15 critérios inerentes ao sistema, denominados dimensões, são:

- **exatidão** (*accuracy*) - grau no qual o dado representa o valor correto do atributo intencionado de um conceito ou evento em um contexto específico de uso;
- **completude** (*completeness*) - grau ao qual o dado do sujeito associado com uma entidade tem valores para todos os atributos esperados e instâncias de entidades relacionadas em um contexto específico de uso;
- **consistência** (*consistency*) - grau no qual o dado tem os atributos que são livres de contradição e são coerentes com outros dados em um contexto específico de uso.
- **credibilidade** (*credibility*) - grau no qual o dado tem atributos que são considerados como verdadeiros e fidedignos pelos usuários em um contexto específico de uso; e
- **atualidade** (*currentness*) - grau no qual o dado tem os atributos que são do período correto em um contexto específico de uso.

Os critérios inerentes e dependentes do sistema são:

- **acessibilidade** (*accessibility*) - grau no qual o dado pode ser acessado em um contexto específico de uso, em particular por quem necessita do suporte de tecnologia ou de uma configuração especial decorrente de alguma necessidade especial ou incapacidade;

- **conformidade** (*compliance*) - grau no qual o dado tem atributos que se aderem a normas, convenções ou regulamentações vigentes e regras similares relacionadas com a qualidade de dados em um contexto específico de uso;
- **confidencialidade** (*confidentiality*) - grau no qual o dado tem os atributos que garantam que este é somente acessível e interpretável por usuários autorizados em um contexto específico de uso;
- **eficiência** (*efficiency*) - grau no qual o dado tem os atributos que podem ser processados e proporciona os níveis esperados de funcionamento (desempenho) utilizando as quantidades e os tipos de recursos apropriados em um contexto específico de uso;
- **precisão** (*precision*) - grau no qual o dado tem os atributos que são exatos ou que proporcionam a discriminação em um contexto específico de uso;
- **rastreabilidade** (*traceability*) - grau no qual o dado tem atributos que proporcionam um rasto de auditoria de acesso aos dados e de qualquer alteração ocorrida com os mesmos em um contexto específico de uso; e
- **entendibilidade**, (*undestandability*) - grau no qual o dado tem atributos que lhe permitem ser lidos e interpretados por usuários, e é expresso em linguagens apropriadas, símbolos e unidades em um contexto específico de uso.

Finalmente, os critérios dependentes do sistema são:

- **disponibilidade** (*availability*) - grau no qual o dado tem atributos que lhe permitem ser recuperados por usuários autorizados e/ou aplicações em um contexto específico de uso;
- **portabilidade** (*portability*) - grau no qual o dado tem os atributos que lhe permitem ser instalados, substituídos ou removidos de um sistema a outro conservando a qualidade existente em um contexto específico de uso; e
- **recuperabilidade** (*recoverability*) - grau no qual o dado tem atributos que lhe permitem manter e conservar um nível especificado de operações e qualidade, mesmo em caso de falhas (do sistema), em um contexto específico de uso.

Para finalizar, Senso e Piñero (2003, p. 100) destacam um questionamento que, segundo os autores, é comum quando se trata de metadados: “quem será o responsável de introduzir os documentos [na base]?”. Eles ressaltam que o questionamento não é feito sob o ponto de vista técnico, e sim pela conscientização da necessidade de adoção dos padrões.



Cabe destacar que a qualidade dos metadados não é abordada em profundidade nem em iniciativas que visam a interoperabilidade.

## 5 A QUALIDADE DOS METADADOS DA REVISTA “EM QUESTÃO”

A Revista Em Questão surgiu em substituição a Revista de Biblioteconomia & Comunicação. O periódico tem como objetivo publicar a produção intelectual de professores, alunos e técnicos-administrativos da Faculdade e de outros especialistas.

A primeira edição de 2004 foi lançada concomitantemente *on-line* (e-ISSN 1808-5245) e em formato impresso (ISSN 1807-8893), o que ocorre ainda hoje. O formato impresso é utilizado apenas para permuta entre bibliotecas. A disponibilização *on-line* é feita por meio do SEER, e em 2007 foram incluídos os metadados das edições a partir de 2003, o que possibilitou maior visibilidade ao disponibilizar seus metadados. A Revista é indexada em bases de dados nacionais e internacionais e participa da iniciativa do Acesso Livre, oferecendo seu conteúdo integral, gratuito e de forma imediata.

A qualidade dos metadados da Revista Em Questão foi investigada tendo como fonte os metadados da Revista, colhidos a partir do protocolo OAI-PMH. Para essa análise foram utilizados os critérios de qualidade da norma ISO/IEC 25012: conformidade, exatidão, completude, consistência e credibilidade. Os critérios acessibilidade, confidencialidade, eficiência, precisão, rastreabilidade, entendibilidade, disponibilidade, portabilidade e recuperabilidade não foram analisados, pois este estudo tem como objeto os metadados colhidos, e a análise desses critérios implicaria também na investigação do sistema da Revista e seu uso. Dentre os critérios dependentes do sistema, apenas o critério de conformidade é investigado, pois a verificação da conformidade com normas e padrões pode ser feita a partir dos metadados colhidos. O critério atualidade não foi investigado, pois o protocolo OAI-PMH não permite a colheita do histórico de atualizações de um metadado. A análise ocorreu quanto a sua adequação às definições, instruções e recomendações de melhores práticas de *Dublin Core* expressas nos documentos: Diretrizes de uso do Metadados OLAC<sup>13</sup>; Definições e propriedades dos elementos do DCMI<sup>14</sup>; e Usando o Dublin Core<sup>15</sup>.

O formato escolhido para colheita foi o OAI\_DC. OAI\_DC é um esquema de metadados composto pelos 15 elementos básicos do DCMI, sem inclusão de elementos do *Dublin Core* Qualificado. OAI\_DC especifica os elementos mínimos necessários para interoperabilidade e, por isso, é o único esquema exigido como obrigatório pelo protocolo

---

<sup>13</sup> Disponível em: <<http://www.language-archives.org/NOTE/usage-20080711.html#All%20elements>>

<sup>14</sup> Disponível em: <<http://dublincore.org/documents/dcmi-terms/#H3>>

<sup>15</sup> Disponível em: <<http://dublincore.org/documents/usageguide/elements.shtml>>

OAI-PMH. Por contemplar a universidade dos provedores de dados, OAI\_DC é o esquema usado pelos provedores de serviço genéricos. No levantamento, foram colhidos metadados de 225 recursos, que correspondem a artigos, resumos de dissertações e teses, editorial e expediente, das 17 edições da Revista publicadas no período de 2003 a 2010 (Quadro 1).

Quadro 1 – Totais de Metadados colhidos por Elemento do DCMI

Elemento	Recursos descritos	Total de Valores	Média de valores por recurso descrito pelo elemento	Total de Valores Diferentes
Assunto	190	969	5,1	6.965,1
Cobertura	24	61	2,54	51
Colaborador	28	28	1	18
Criador	225	338	1,5	271
Data	225	225	1	12
Descrição	149	422	2,17	421
Direitos Autorais	225	225	1	1
Editor	225	225	1	1
Fonte	225	225	1	17
Formato	225	264	1.174	4
Identificador	225	225	1	255
Idioma	213	213	1	3
Relação	18	65	3,6	65
Tipo	141	155	1,09	17
Título	225	454	2,01	446

Fonte: dados da pesquisa

A qualidade de cada um dos quinze elementos de DC é avaliada a seguir.

**Título** (*title*) - representa o nome pelo qual o recurso é formalmente reconhecido. As melhores práticas recomendam que cada recurso deva possuir um elemento título. Com exceção dos títulos paralelos em obras de edição bilíngue, todos os demais títulos deverão ser especificados no refinamento (qualificação) do elemento, na forma de título alternativo (*dc.title.alternative*). Há **conformidade** deste elemento com relação às melhores práticas, pois cada recurso possui um elemento título original. Tanto o título original como suas traduções são representados através desse elemento, estando conforme o princípio *dump-down*, em que título alternativo, que é o elemento recomendado para representar traduções, deve ser fornecido como título. Exatidão, consistência e credibilidade são analisadas no Quadro 2.

Quadro 2 – Exatidão, Completude, Consistência e Credibilidade no Título

Exatidão	Títulos corretos. Não são verificados problemas de digitação ou erros de linguagem. Casos insignificantes de títulos com caracteres não alfanuméricos, que podem ser resultantes de algum processo de migração.
Completude	Título presente em todos os itens. Nem todos os itens possuem títulos traduzidos, o que não caracteriza incompletude. Média de 2,0 títulos por documento, indicando o título original e sua tradução.
Consistência	Títulos com traduções coerentes. Há falta de critério para os títulos de editoriais, apresentações e expedientes. Uns possuem o tipo como título (ex.: “Editorial”) e outros tem como título o tipo seguido da data, para evitar ambiguidade (ex.: “Editorial 2009/2”). Alguns itens não possuem tradução, enquanto outros possuem traduções para mais de um idioma.
Credibilidade	Títulos fiéis aos títulos dos documentos. Não cabe nesse estudo a verificação da credibilidade do título do pelo autor para o documento.

Fonte: dados da pesquisa

**Criador** (*creator*) - representa a entidade responsável em primeira instância pela existência do recurso. Em geral, trata-se de uma pessoa, uma organização ou um serviço. As melhores práticas recomendam: usar Colaborador no lugar de Criador, com exceção no caso em que há uma nítida e importante participação de uma pessoa ou organização e não exista um termo adequado de refinamento no *OLAC Role Vocabulary*<sup>16</sup> para utilizar com Colaborador. As recomendações dizem para colocar pelo menos um elemento para cada item. Nos metadados colhidos, a média resultou de dois valores para cada item descrito, evidenciando coautorias nos trabalhos. Há 117 nomes de autores repetidos, indicando autores que apresentaram mais de uma contribuição à revista. Há conformidade com as melhores práticas quanto à existência de no mínimo um autor para cada documento. Autores poderiam ser representados via elemento Criador, pois este termo é contemplado em *OLAC Role Vocabulary*. Exatidão completude, consistência e credibilidade são analisadas no Quadro 3.

Quadro 3 – Exatidão, Completude, Consistência e Credibilidade em Criador

Exatidão	Valores exatos, mas foram encontrados alguns poucos registros com falta de letras e caracteres alheios ao nome. Ocorrência de três registros com a formação do autor junto ao seu nome.
Completude	Todos os 225 itens da revista foram descritos através deste elemento.
Consistência	Há falta de consistência pela falta de padronização. Existem autores que estão representados em mais de uma forma.
Credibilidade	Autores correspondem aos descritos nos documentos.

Fonte: dados da pesquisa

<sup>16</sup> OLAC Role Vocabulary <<http://www.language-archives.org/REC/role-20060406.html>>

**Colaborador** (*contributor*) - colaborador representa aquele que contribuiu com o conteúdo do recurso. As melhores práticas recomendam usar esse elemento no lugar de Criador no caso da existência de mais de um autor sem nítida eminência de um deles. A prática não é observada na Revista, pois o elemento Criador é utilizado para representar todos os autores. O elemento está presente em 28 registros, cujos valores correspondem, principalmente, a entidades que aportam recursos financeiros para as pesquisas, o que está em conformidade com as melhores práticas. Exatidão, completude, consistência e credibilidade são analisadas no Quadro 4.

Quadro 4 – Exatidão, Completude, Consistência e Credibilidade em Colaborador

Exatidão	Em um registro foi utilizado hífen (-) para não adjudicar alguém como colaborador. Há um registro que refere a uma fundação assistencial previdenciária (FAPERS), mas o provável valor correto é FAPERGS.
Completude	O <b>elemento</b> é usado exclusivamente para representar financiador.
Consistência	Falta de padronização levou a representar, por exemplo, a entidade CNPq de cinco formas diferentes. Para identificar duas entidades de forma conjunta, utilizou-se o conectivo (y), hífen ou ponto e vírgula. Resultou na interpretação do registro como de uma nova entidade e não como duas.
Credibilidade	Patrocinadores indicados nos documentos.

Fonte: dados da pesquisa

**Assunto** (*subject*) - esse elemento é usado para representar um tópico do conteúdo do recurso. Tópicos espaciais e temporais devem ser representados por meio do elemento cobertura (*coverage*). As melhores práticas para assunto recomendam utilizar vocabulário controlado ou sistemas de classificação formais. De forma geral, os assuntos estão representados pelas palavras-chave dos recursos, estando assim em conformidade com as práticas, porém não exclusivamente. A existência de valores como ciência da informação e ciências da informação, ou análise de citação e análise de citações, e a elevada ocorrência de valores diferentes (72%) evidenciam que na Revista inexistem um vocabulário controlado, o que mostra uma não conformidade. Exatidão, completude, consistência e credibilidade são analisadas no Quadro 5.

Quadro 5 – Exatidão, Completude, Consistência e Credibilidade em Assunto

Exatidão	Assuntos corretos. Existem assuntos que expressam locais (como <i>America Latina</i> ) e tempo (como <i>Década de 30</i> ), deveriam constar em Cobertura.
Completude	Existem 190 itens com assunto, de um universo de 225. Dentre os 65 recursos sem assunto, 35 são Editorial, Expediente e Resumo, o que é considerado aceitável, restando ainda 30 itens (13%) sem motivo aparente.
Consistência	Comprometida pela falta de critérios para a definição dos assuntos
Credibilidade	Termos representam seus conteúdos pela perspectiva do autor.

Fonte: dados da pesquisa

**Descrição** (*description*) – visa descrever o conteúdo do recurso, incluindo resumo, sumário ou descrição em texto livre. As melhores práticas recomendam prover esses elementos sempre que possível, pois o campo é fonte para termos indexáveis. Os resultados mostram 194 recursos registrados com 421 valores diferentes. Esses valores são resumos, e suas traduções em inglês e/ou espanhol, o que demonstra conformidade com as melhores práticas. A média de 2,17 resumos por elemento descrito decorre de haver pelo menos uma tradução do resumo. A análise é apresentada no Quadro 6.

Quadro 6 – Exatidão, Completude, Consistência e Credibilidade em Descrição

Exatidão	Valores corretos. Em dois recursos, um do tipo Apresentação e outro do tipo Prefácio, o símbolo “.” foi usado para indicar ausência de descrição.
Completude	Ausente em 31 recursos, na maioria documentos dos tipos Editorial, Apresentação, Expediente e Prefácio. Cada recurso possui um ou mais valores para o elemento por constar o resumo em original e uma ou duas versões para outro idioma.
Consistência	Descrições em mais de um idioma consistentes entre si. Documentos com descrição em somente um idioma, documentos com descrições em dois idiomas e documentos com descrições em três idiomas.
Credibilidade	Descrições correspondem aos resumos (e traduções) presentes nos artigos.

Fonte: dados da pesquisa

**Cobertura** (*coverage*) - este elemento destina-se à identificação espacial ou temporal do conteúdo recurso. Inclui nomes de lugares, coordenadas geográficas, períodos de tempo ou jurisdição. As melhores práticas recomendam o uso de vocabulário controlado. O elemento consta em 24 recursos, a maioria traz lugares e datas, não usa vocabulários controlados. O Quadro 7 apresenta a análise do elemento.

Quadro 7 – Exatidão, Completude e Consistência em Cobertura

Exatidão	Uso de hífen (-) para indicar campo vazio. Valores que não atendem à definição do elemento, como “ <i>Resultados parciais da tese em andamento</i> ”, “revisão bibliográfica”. Valores carecendo de exatidão, como “ <i>Época moderna e atual</i> ”, “ <i>últimas décadas</i> ”.
Completude	Usado em apenas 24 recursos (9,3%). Outros itens poderiam ser descritos, como o título “ <i>Reconfiguração do mercado editorial brasileiro de livros didáticos no início do século XXI: história das principais editoras e suas práticas comerciais</i> ” (temporal).
Consistência	Falta de vocabulário controlado. Diversas maneiras e símbolos diferentes para identificar intervalos. Ex.: 1995-2007; 2000 – 2007; 2003 a 2005; 2006/2007; <i>Anos 1968 a 1978</i> ; <i>Anos 30-40</i> .
Credibilidade	Os poucos valores permitem ter noção sobre o período e local, compromete a credibilidade.

Fonte: dados da pesquisa

**Fonte** (*source*) – indica uma referência a um recurso do qual o recurso descrito é derivado, sendo que o recurso descrito pode ser derivado de sua fonte em parte ou no todo. As melhores práticas recomendam referenciar a fonte por um texto ou um número em conformidade com um sistema formal de identificação. Todos os 225 recursos possuem uma única fonte, que é a Revista estudada nesta pesquisa e há uma regra para a representação da edição. Há **conformidade**, pois o valor usado corresponde à edição do documento descrito, e tal valor é um texto em conformidade com um sistema formal de identificação. A análise do elemento encontra-se no Quadro 8.

Quadro 8 – Exatidão, Completude, Consistência e Credibilidade em Fonte

Exatidão	Valores corretos e que seguem uma regra para identificar a fonte.
Completude	Todos os itens são descritos.
Consistência	Valores identificam unicamente cada fonte.
Credibilidade	Valor corresponde à edição da qual o documento faz parte.

Fonte: dados da pesquisa

**Identificador** (*identifier*) – indica uma referência não ambígua a um recurso em um dado contexto, como, por exemplo, *Digital Object Identifier* (DOI), a *International Standard Book Number* (ISBN) ou localização do documento na *Web* (URI). Na Revista, todos os

recursos utilizam a URI do documento, estando em **conformidade** com *Dublin Core*. Exatidão, completude, consistência e credibilidade são analisadas no Quadro 9.

Quadro 9 – Exatidão, Completude, Consistência e Credibilidade em Identificador

Exatidão	URIs codificadas corretamente e representam documentos válidos.
Completude	Todos os documentos possuem o elemento.
Consistência	URIs identificam unicamente os documentos.
Credibilidade	URIs são válidas e representam os documentos descritos.

Fonte: dados da pesquisa

**Relação** (*relation*) – indica uma referência a um recurso relacionado. As melhores práticas recomendam que essa referência seja um texto ou um número em conformidade com um sistema formal de identificação, como URI. Na Revista, o elemento é usado em 18 recursos (8% do total). Todos os valores são URIs que referenciam documentos no domínio da revista, mas essas URIs não são válidas. A análise consta no Quadro 10, a seguir.

Quadro 10 – Exatidão, Completude, Consistência e Credibilidade em Relação

Exatidão	URIs codificadas corretamente, mas não representam os documentos.
Completude	O elemento relação não é exigido para todos os registros.
Consistência	Valores não representam documentos válidos.
Credibilidade	Valores não são verdadeiros, pois não representam documentos válidos.

Fonte: dados da pesquisa

**Formato** (*format*) – indica a manifestação física ou digital do recurso, como o tipo da mídia e dimensões de tamanho ou duração, inclusive pode indicar os equipamentos ou software adequado para operar o recurso. As melhores práticas recomendam o uso de um vocabulário controlado, como *Internet Media Types* [MIME]<sup>17</sup> para descrever os formatos de documentos digitais. Quatro valores diferentes são usados para descrever os formatos dos 225 recursos, os quais estão em conformidade com o vocabulário MIME. No Quadro 11 consta a análise do elemento.

<sup>17</sup> MIME *Media Types* <<http://www.iana.org/assignments/media-types/index.html>>

Quadro 11 – Exatidão, Completude, Consistência e Credibilidade em Formato

Exatidão	Nomes dos formatos seguem o vocabulário MIME.
Completude	Todos os itens são descritos. Existem 38 documentos que foram disponibilizados em dois formatos (html e PDF) e descritos por meio dos valores correspondentes do vocabulário MIME.
Consistência	Existem dois critérios para descrever documentos PDF, demonstrando incoerência, embora os valores usados sejam compatíveis com o formato do documento ( <i>application/pdf</i> e <i>application/force-download</i> ).
Credibilidade	Formatos descritos fieis aos formatos dos documentos.

Fonte: dados da pesquisa

**Tipo** (*type*) – representa a natureza ou o gênero do conteúdo do reuso. As melhores práticas orientam usar vocabulário controlado e, para fins de interoperabilidade, recomendam o uso de um tipo genérico de DCMI *Type Vocabulary* [DCMI-TYPE]<sup>18</sup>. O elemento consta em 141 itens. Em 139 recursos o valor é “avaliado por pares”. Há conformidade no uso da norma nos 139 recursos descritos como sendo “avaliado[s] por pares”, pois o valor indica que o recurso é do tipo artigo avaliado por pares. Entretanto, observa-se falta qualidade desse valor quando avaliado por outros critérios de qualidade (Quadro 12). Também não há o uso de pelo menos um valor de DCMI *Type*.

Quadro 12 – Exatidão, Completude, Consistência e Credibilidade em Tipo

Exatidão	Valor inadequado para indicar os artigos avaliados por pares. O valor poderia ser “artigo avaliado por pares”. Existência de valores errados, na sua maioria indicando a natureza da pesquisa abordada pelo artigo (ex.: pesquisa de fonte empírica).
Completude	Nem todos os recursos possuem tipo. A informação para indicar artigo avaliado por pares é incompleta.
Consistência	Coerência na descrição de recursos do tipo artigo, embora utilizando um valor não adequado (“avaliado por pares”).
Credibilidade	Indicação do tipo do artigo é fiel ao tipo do documento.

Fonte: dados da pesquisa

**Data** (*date*) – o elemento data se refere a alguma data do ciclo de vida do documento. As recomendações de melhores práticas indicam o uso para descrever a criação ou a disponibilização do recurso. Recomenda-se a codificação dos valores pela ISO 8601 e com o formato AAAA-MM-DD (ano-mês-dia). Se a data completa não for conhecida, recomendam-se os formatos AAAA (ano) ou AAAA-MM (ano-mês). O elemento está presente em todos os 225 recursos e segue o formato AAAA-MM-DD.

<sup>18</sup> DCMI *Type Vocabulary* <<http://dublincore.org/documents/dcmi-type-vocabulary/>>



As datas indicam as datas de publicação dos documentos (data das edições). Porém, para documentos publicados até 2005, o valor é “2006-12-9”. Acredita-se que indique o momento em que os registros anteriores a 2006 foram migrados para o sistema, que é uma data relacionada ao ciclo de vida do recurso. Por isso, considera-se o elemento em conformidade com as melhores práticas, embora haja falta de consistência (Quadro 13).

Quadro 13 – Exatidão, Completude, Consistência e Credibilidade em Data

Exatidão	Valores de acordo com as recomendações de codificação
Completude	Valores de acordo com as recomendações de codificação e presentes em todos os elementos.
Consistência	Consistência comprometida pelo uso de dois critérios para data. Data de expedição (para os recursos de 2006 em diante) e provavelmente data de migração (para recursos anteriores a 2006).
Credibilidade	Há exatidão. Valores fieis às datas do ciclo de vida dos documentos.

Fonte: dados da pesquisa

**Idioma** (*language*) – expressa o idioma do conteúdo intelectual do documento. As melhores práticas recomendam as especificações RFC3066 e ISO 639 para codificação dos valores para idiomas e que este elemento deva ser repetido caso o recurso esteja em mais de um idioma. O elemento está em **conformidade** com as melhores práticas. É usado em 213 registros com os valores compatíveis com as recomendações: “pt”, “en” e “es” para indicar, respectivamente, conteúdos em português, inglês e espanhol. O Quadro 14 apresenta Exatidão, Completude, Consistência e Credibilidade em Idioma.

Quadro 14 – Exatidão, Completude, Consistência e Credibilidade em Idioma

Exatidão	Há exatidão. Valores de acordo com as recomendações de codificação.
Completude	Há completude para os artigos. Entretanto, o idioma é usado em 213 dos 225 registros. Não está presente na maioria dos Editoriais e Expedientes.
Consistência	Há exatidão. Valores de acordo com as recomendações de codificação.
Credibilidade	Há fidedignidade. Informações de idioma compatíveis com os idiomas dos recursos. Ocorrência de um recurso em português descrito como sendo em inglês.

Fonte: dados da pesquisa

**Editor** (*publisher*) – Indica quem é responsável por tornar o recurso disponível. Pode ser uma pessoa, organização ou serviço, e as melhores práticas recomendam a não utilização desse elemento caso editor e criador sejam a mesma pessoa. O elemento está **conforme** às melhores práticas, pois a Revista é editada pela Faculdade de Biblioteconomia e Comunicação da UFRGS e utilizou-se como editor a instituição maior relacionada ao recurso, isto é, a Universidade Federal do Rio Grande do Sul. Seu valor é idêntico em todos os registros e atende aos critérios analisados (Quadro 15).

Quadro 15 – Exatidão, Completude, Consistência e Credibilidade em Editor

Exatidão	Texto escrito de forma correta.
Completude	Texto identifica a instituição. Não há uso de siglas ou abreviações que impeçam a identificação da instituição.
Consistência	Valores idênticos e em todos os campos.
Credibilidade	Indica a instituição à qual a revista pertence.

Fonte: dados da pesquisa

**Direitos Autorais** (*rights*) - indica informações de direitos sobre o recurso. As melhores práticas recomendam uma declaração de direitos para o recurso ou uma referência a um serviço que irá prover tais informações, que engloba direitos sobre propriedade intelectual, direito autoral e/ou direitos de propriedade. Advertem que a ausência desse elemento impede que seja feita qualquer suposição sobre os estados de direito relacionados ao recurso. O texto está **conforme** às recomendações de *Dublin Core*, pois dispõe que os direitos intelectuais são dos autores, que cedem para a Revista o seu ineditismo, e que os artigos podem ser utilizados para fins educacionais e não comerciais. Usa um texto idêntico nos 225 registros, não sendo aplicável a advertência das melhores práticas referente à ausência do mesmo. A análise consta no Quadro 16.

Quadro 16 – Exatidão, Completude, Consistência e Credibilidade em Direitos Autorais

Exatidão	Há exatidão. Texto correto.
Completude	Texto com informações necessárias para o propósito do elemento. Entretanto, texto somente em inglês pode limitar o seu campo de atuação.
Consistência	Os direitos autorais de todo o material veiculado na revista está sob o mesmo regime. Em todos os itens consta a mesma orientação para o uso da informação disponibilizada.
Credibilidade	Há credibilidade.

Fonte: dados da pesquisa

## 6 CONCLUSÕES

O estudo dos metadados da Revista Em Questão permitiu investigar e analisar sua adequação às recomendações do *Dublin Core Metadata Initiative* (DCMI). Os resultados apontaram deficiências na política de metadados da Revista, já que não há padronização na adoção do DCMI, nem quanto à definição do elemento nem quanto à instrução de uso, o que caracteriza inconformidades. Também foram notados metadados com registros faltantes, ambíguos e com incoerências no tratamento ao longo do tempo. A falta de qualidade dos dados corresponderam a diversas dimensões da qualidade. Foram identificadas quatro categorias com relação à qualidade dos elementos da Revista:

a) Elementos em conformidade com as melhores práticas e que atendem aos critérios de qualidade, com casos insignificantes de falta de qualidade: Identificador, Idioma, Editor, Fonte e Título;

b) Elementos em conformidade com as melhores práticas e que atendem aos critérios de qualidade, com alguns casos não comprometedores de falta de qualidade: Formato (inconsistência ao usar dois valores para indicar o formato PDF), Descrição (não completude ao estar ausente em documentos que não são artigos);

c) Elementos em conformidade com as melhores práticas e que atendem aos critérios de qualidade, com casos de falta de qualidade, que já comprometem a qualidade: Data (inconsistência devido a dois critérios: migração e edição da Revista);

d) Elementos com a qualidade comprometida: Colaborador (inconsistência devido à falta de controle para o valor, incompleto por representar somente o patrocinador), Cobertura (incompleto por ser usado de forma insuficiente, inconsistente por falta de controle para o valor e por apresentar valores errados), Criador (inconsistente por falta de controle para o valor), Tipo (exatidão por usar “avaliado por pares” para artigos, incompleto por não descrever documentos que não são artigos).

Recomenda-se o estabelecimento de uma política de metadados que esteja alinhada com a estratégia da Revista no sentido de tornar seu conteúdo visível, acessível e pronto para uso, pois as revistas são componentes importantes no fluxo de comunicação científica. Ressalta-se também a importância de instituir o treinamento dos envolvidos para a melhoria da qualidade dos metadados e, por consequência, da própria Revista. Esse tipo de ação pode ser pensada de forma isolada ou conjunta com outras revistas hospedadas no SEER. Assim, abre-se espaço para reforçar o trabalho em colaboração institucional.

## REFERÊNCIAS

- CETTO, Ana María; ALONSO GAMBOA, José Octavio (org.) **Calidad e Impacto de la Revista Iberoamericana**. México: LATINDEX, UNAM, 2011. Disponível em: <<http://www.latindex.unam.mx/librociri/>>
- KURAMOTO, Hélio. Informação Científica: proposta de um novo modelo para o Brasil. *Ciência da Informação*, Brasília, v. 35, n. 2, p. 91–102, maio./ago. 2006. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/831/678>> Acesso em: 20 mar. 2011.
- MELEROS, Remedios. La Revista Digital Vista desde un Entorno de Acceso Abierto: modelos y estrategias para favorecer su visibilidad. In.: CETTO, Ana María; ALONSO GAMBOA, José Octavio (org.) **Calidad e Impacto de la Revista Iberoamericana**. México: LATINDEX, UNAM, 2011. p. 442-456. Disponível em: <<http://www.latindex.unam.mx/librociri/>>
- PACKER, Abel; MENEZHINI, Rogério. Visibilidade da Produção Científica. In: POBLACIÓN, Dinah Aguiar; WITTER, Geraldina Porto; SILVA, José Fernando Modesto da. **Comunicação & Produção Científica**: contexto, indicadores e avaliação. São Paulo: Aditora, 2006. p. 235-259
- PAVAN, Cleusa. **Práticas Sociais na Comunicação Científica**: a avaliação pelos pares nas revistas brasileiras de Ciência da Informação. 2008. 171 f. Dissertação (Mestrado em Comunicação e Informação). Orientadora Prof. Ida Regina Chittó Stumpf. Universidade Federal do Rio Grande do Sul, (UFRGS). Porto Alegre, 2008. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/13798>>. Acesso em: 29 set. 2011.
- ROCHA, Rafael Port da. Metadados, *Web Semântica*, Categorização Automática: combinando esforços humanos e computacionais para a descoberta e uso dos recursos da *web*. **Em Questão**, Porto Alegre, v. 10, n. 1, p. 109-121, jan./ jun. 2004. Disponível em: <<http://seer.ufrgs.br/EmQuestao/article/view/86/1127>>. Acesso em: 23 jun. 2011.
- SENÉ, María Luisa. Los Metadatos y su Lugar en la Arena Internacional. In.: **CONGRESO INTERNACIONAL DE INFORMACIÓN (INFO)**, 5. abr. 2004, La Habana, 8 p. Disponível em: <<http://www.mendeley.com/research/los-metadatos-y-su-lugar-en-la-arena-internacional/>>. Acesso em: 22 out. 2011.
- SENSO, José A.; PIÑERO, Antonio de la Rosa. El Concepto de Metadato. Algo más que descripción de recursos electrónicos. **Ciência da Informação**. Brasília, v. 32, n. 2, p. 95-106, maio/ago. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n2/17038.pdf>>. Acesso em: 6 ago. 2011.
- WEITZEL, Simone. **Os repositórios de e-prints como nova forma de organização da produção científica**: o caso da área das Ciências da Comunicação no Brasil. Tese. (Doutorado em Ciências da Informação). Orientadora Prof. Dr. Sueli Mara dos Santos Pinto Ferreira. Universidade de São Paulo (USP). 2006. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-14052009-133509/pt-br.php>>. Acesso em: 25 set. 2011.