

Análise focada em metadados sob a luz do padrão MTD-BR

Jaqueline Costa Alves
Lígia Maria Arruda Café

RESUMO

A dificuldade de acesso e recuperação da informação ocasionada pelo aumento do volume de informações disponibilizadas sem uma forma estruturada de representação, é uma das conseqüências do uso cada vez mais freqüente da *world wide web* (www). Fazendo parte dessa realidade, as bibliotecas digitais procuram garantir a interoperabilidade por meio da descrição física e de conteúdo de seus itens, adotando padrões específicos de metadados. Nesse contexto, o presente artigo descreve a pesquisa desenvolvida no curso de Mestrado em Ciência da Informação da Universidade Federal de Santa Catarina, cujo objetivo foi analisar o conteúdo dos metadados das dissertações eletrônicas da Universidade do Estado de Santa Catarina - UDESC, disponibilizadas para a Biblioteca Digital de Teses e Dissertações- BDTD, a fim de verificar se os mesmos estão sendo informados de acordo com o padrão MTD-BR e, dessa forma, contribuir para a localização e recuperação das dissertações eletrônicas dessa instituição. Quanto aos procedimentos metodológicos, a abordagem geral da pesquisa foi exploratória, tendo sido os metadados analisados de forma qualitativa e quantitativa. A abordagem qualitativa baseou-se nas recomendações do padrão MTD-BR referentes ao conteúdo dos elementos de metadados e seus respectivos atributos e a abordagem quantitativa viabilizou a comparação da quantidade de metadados informados corretamente e de forma incorreta. Os resultados obtidos possibilitaram a apresentação de sugestões de melhoria do conteúdo dos metadados informados pela UDESC para a BDTD e permitiram concluir que a biblioteca digital da UDESC, embora tenha uma boa qualidade no preenchimento dos metadados, pode vir a aprimorar este procedimento em favor da interoperabilidade e da recuperação da informação.

PALAVRAS-CHAVE:Metadados. *Dublin Core*. ETD-MS. MTD-BR. BDTD. NDLTD.

1 Introdução

As pesquisas sobre tratamento e recuperação da informação (RI) evoluem constantemente e, dessa forma, viabilizam o desenvolvimento de novas tecnologias capazes de torná-los cada vez mais eficientes. Nesse sentido, a Ciência da Informação assume um papel estratégico, considerando a importância da organização e representação da informação para a sua recuperação.

ARI possibilita a localização de documentos por meio de consultas realizadas, normalmente, por palavras-chave. Mas, como diferentes fatores influenciam na RI, consultas desse tipo podem deixar de recuperar documentos relevantes ou, o documento recuperado e apresentado para o usuário pode não ser o que melhor atende sua necessidade. Nesse sentido, Baeza-Yates e Ribeiro Netto (1999) sinalizam que os documentos recuperados, freqüentemente, não são suficientes para satisfazer a necessidade de informação dos usuários. Dentre os fatores que podem prejudicar a RI, Chen (1999) chama a atenção para as variações nas estruturas e formatos das bases de dados, as diferentes formas de documentos disponibilizados (texto, áudio e vídeo) e a abundância de conteúdos multilingües nas aplicações da *web*.

Pode-se dizer também que a RI é prejudicada muitas vezes por falhas que ocorrem no processo de organização da informação (OI). Quando se fala em OI, dois princípios são fundamentais: especificidade¹ e exaustividade². Ao nosso entender, essas orientações têm influência direta nos níveis de precisão³ e revocação⁴ medidos na RI. Assim, enquanto a exaustividade e a especificidade estão relacionadas com a organização da informação, a revocação e a precisão se referem a recuperação, sendo que o processo de recuperação depende da qualidade do tratamento dado aos documentos. Essa qualidade está associada igualmente a adoção de padrões, principalmente quando se tratam de ambientes disponibilizados na *web* em que o processo de recuperação da informação está cada vez mais sendo executado de forma automática.

De acordo com Souza e Alvarenga (2004), a padronização da descrição da informação é essencial. Segundo esses autores:

[...] é necessária uma padronização de tecnologias, de linguagens e de metadados descritivos, de forma que todos os usuários da *web* obedeçam a determinadas regras comuns e compartilhadas sobre como armazenar dados e descrever a informação armazenada e que esta possa ser consumida por outros usuários humanos ou não, de maneira automática e não ambígua. (SOUZA ; ALVARENGA, 2004, p. 134)⁵

Diante do exposto, este artigo relata uma pesquisa de mestrado realizada no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, em que foi realizada uma análise do conteúdo dos metadados das dissertações eletrônicas da UDESC⁶, disponibilizados para a Biblioteca

¹ Especificidade é “[...] a extensão em que o sistema nos permite ser precisos ao especificarmos o assunto de um documento que estamos processando.” (FOSKETT, 1973, p. 12).

² Exaustividade é “[...] a extensão com que analisamos um determinado documento, a fim de estabelecer exatamente qual o conteúdo temático que temos de especificar.” (FOSKETT, 1973, p. 13).

³ Precisão “[...] relação entre os documentos relevantes recuperados e o número total de documentos recuperados.” (PIEADADE, 1977, p. 3).

⁴ Revocação é “[...] a relação entre os documentos relevantes recuperados e o número total de documentos relevantes existentes na coleção.” (PIEADADE, 1977, p. 3).

⁵ Documento eletrônico.

⁶ Universidade do Estado de Santa Catarina.

Digital de Teses e Dissertações (BDTD), a fim de verificar se os mesmos estão sendo informados de acordo com o padrão MTD-BR. Dessa forma, o estudo procura contribuir para a melhoria da qualidade da descrição das teses e dissertações contidas em repositórios digitais científicos, o que resulta no aprimoramento da circulação da informação nesses ambientes.

A estrutura do artigo é constituída pela introdução, exposição sobre bibliotecas digitais de teses e dissertações, metadados, padrões de metadados, descrição da metodologia utilizada na pesquisa, análise dos metadados e conclusão.

2 Bibliotecas digitais de teses e dissertações

Diante das recentes possibilidades oferecidas pelas TICs, as bibliotecas digitais estão tornando cada vez mais forte a idéia de preservação, conservação e, principalmente, acesso à informação. Para Borgman (1996), a definição de bibliotecas digitais pode ser feita por duas idéias:

1. Bibliotecas Digitais são um conjunto de fontes eletrônicas e serviços técnicos associados para a criação, pesquisa e uso da informação, que possibilitam uma extensão e um aumento do armazenamento da informação e dos sistemas de recuperação, manipulando dados digitais em qualquer formato (texto, imagens, sons; imagens dinâmicas e estáticas) em redes distribuídas de trabalho. O conteúdo das bibliotecas digitais inclui dados e metadados que descrevem vários aspectos do dado (representação, criador, dono, direitos de reprodução) e metadados que consistem em ligações ou relacionamentos com outros dados ou metadados, sejam esses externos ou internos à biblioteca digital.

2. Bibliotecas Digitais são construídas, coletadas e organizadas, por (e para) uma comunidade de usuários, e suas capacidades funcionais satisfazem as necessidades de informação e usos dessa comunidade. Nelas os indivíduos e grupos interagem uns com os outros, usando sistemas e fontes de dados, informação e conhecimento. Neste sentido, representa extensão, aumento e integração de uma variedade de instituições de informação, como locais físicos, onde recursos são selecionados, coletados, organizados, preservados, e acessados em função de uma comunidade de usuários. Estas instituições de informação incluem, entre outras, bibliotecas, museus, arquivos e escolas, mas as bibliotecas digitais também se estendem e servem outros ambientes de comunidades, incluindo salas de aulas, escritórios, laboratórios, casas e espaços públicos. (BORGMAN, 1996, p.6)⁷.

■
⁷ Documento eletrônico.

Sayão (2007) ressalta que:

Uma biblioteca digital no seu sentido pleno - não é meramente um repositório ou uma coleção de informações em formato digital; também não é somente uma tecnologia ou um conjunto de tecnologias que se pode avaliar isoladamente. Antes disso, é um sistema aberto, de múltiplas interligações e múltiplos subsistemas, envolvendo um ambiente organizacional, profissionais especializados provenientes de diversas áreas, recursos informacionais, usuários claramente definidos, tecnologia de informação, procedimentos, padrões e protocolos e, não menos importante, compromissos de longo prazo. (SAYÃO, 2007, p. 19)⁸.

■
⁸ Documento eletrônico.

Nesse contexto, o mesmo autor afirma que as normas, padrões, formatos e protocolos cumprem um papel de fundamental importância, já que estabelecem as regras pelas quais os objetos

são descritos, identificados e preservados, seus dados são armazenados, e os sistemas aos quais estão inseridos se comunicam.

Dias (2001 p. 4)⁹ defende que “a biblioteca digital parece estar se firmando como a expressão que significaria, no contexto digital, um conjunto de artefatos, conhecimento, práticas e uma comunidade, que engendra compromissos realísticos assumidos por profissionais da informação, analistas de sistemas e usuários.”

A principal diferença entre a biblioteca tradicional e a biblioteca digital é a ampliação das formas de tratamento, acesso e recuperação das informações. Cunha(1999, p.259) sinaliza que “a futura biblioteca coexistirá em um ambiente onde os usuários estarão conectados a uma ampla variedade de recursos informacionais que muitas bibliotecas não poderão prover, a menos que se conectem a tais recursos...”. Mas, Dias(2001 p. 10) ressalta que, “em ambos os contextos, o tratamento da informação permanece como uma atividade crucial, pois cumpre função fundamental no trabalho de facilitar o acesso à informação.”

No contexto do registro da produção científica, os cursos de pós-graduação *strictu sensu* têm uma responsabilidade fundamental, considerando a quantidade de teses e dissertações geradas por ano e a necessidade de organização adequada desta produção.

Uma tendência mundial verificada no tratamento dessa documentação é a construção de bibliotecas digitais de teses e dissertações de forma integrada, com o intuito de se tornarem verdadeiros sistemas de informação. Essas bibliotecas procuram registrar seu acervo em formato eletrônico de maneira a viabilizar o acesso ao texto integral do documento. No entanto, deve-se ressaltar a importância da correta organização e representação física e de conteúdo desses objetos informacionais, sem as quais seria impossível acessá-los e recuperá-los.

Uma das grandes vantagens oferecidas pelas teses e dissertações em formato eletrônico é a redução de tempo e de custo para a realização das buscas de documentos, possibilitada pela centralização e extensa divulgação dos trabalhos científicos. Essa vantagem torna possível a execução de consultas simultâneas dos mais diversos locais, facilitando o acesso à informação e contribuindo para o aumento da visibilidade das pesquisas científicas. Outro benefício que merece destaque refere-se à preservação, uma vez que se torna mais fácil e prático o armazenamento dos documentos, em meio digital.

Em nível internacional, citamos a *Networked Digital Library of Theses e Dissertations* (NDLTD), criada em 1996, cujo objetivo principal é o desenvolvimento de uma rede mundial de teses e dissertações digitais que se dedica a promover a adoção, criação, uso, disseminação e preservação das cópias das teses e dissertações publicadas. Para atingir seus objetivos, encarrega-se de patrocini-

■
⁹ Documento eletrônico.

nar e incentivar as instituições interessadas em contribuir para a evolução das teses e dissertações eletrônicas, realizando estudos e atividades sobre o assunto.

Os países que participam da NDLTD, visando contribuir para a ampliação do desenvolvimento das bibliotecas digitais, se propõem a compartilhar experiências, resultantes de pesquisas por eles realizadas, com os países em desenvolvimento, a fim de que as mesmas sirvam-lhes de base. (NETWORKED..., 2008)

O projeto NDLTD se propõe a disponibilizar coleções individuais de ETDs arquivadas em instituições distintas, para estudantes e pesquisadores que procuram teses e dissertações em uma biblioteca digital. Apóia as bibliotecas digitais e promove encontros anuais, possibilitando a integração das universidades por meio de uma cooperação entre as mesmas.

Para tornar as ETDs de uma instituição acessíveis, primeiramente, o responsável pelas mesmas precisa disponibilizar seus metadados. Sendo assim, as instituições que produzem e armazenam ETDs disponibilizam metadados, como título, autor, etc., para a NDLTD. Dessa forma, é possível apresentá-las para os pesquisadores como uma única coleção e, quando um original relevante é encontrado, o pesquisador é direcionado à instituição que armazena este documento em formato digital.

O formato de metadados utilizado na NDLTD é o ETD-MS, um conjunto de padrões internacional para os elementos de metadados usados para a descrição de teses ou dissertações eletrônicas. Essa possibilidade de tornar metadados disponíveis para auxiliar na pesquisa é sustentada pela iniciativa dos arquivos abertos (*Open Archives Initiative* - OAI). A coleta dos metadados é feita por meio do protocolo OAI-PMH que torna possível a coleta de metadados de seus locais de origem para disponibilizá-los na NDLTD.

No Brasil, as bibliotecas digitais de teses e dissertações ocupam cada vez mais o seu espaço com o propósito de disponibilizar acervos em formatos digitais, visando à conservação e a disponibilização de seus conteúdos e assegurando o compartilhamento da informação para os mais diversos locais do mundo. Nesse sentido, o Brasil, influenciado por esta tendência mundial, criou a Biblioteca Digital de Teses e Dissertações (BDTD) e participa, por meio dela, como provedor de dados da NDLTD. A BDTD, implantada sob a responsabilidade do Instituto Brasileiro de Informação e Tecnologia (IBICT), tem como principais objetivos a integração dos sistemas de informação de teses e dissertações das diversas Instituições de Ensino Superior (IES) brasileiras e o estímulo ao registro e a publicação de teses e dissertações em formato digital. O modelo da BDTD integra as iniciativas de registro bibliográfico e de publicação eletrônica de teses e dis-

sertações existentes nos acervos das IES brasileiras. Ao integrar essas duas iniciativas, o IBICT amplia a abrangência da BDTD e disponibiliza para os usuários um catálogo nacional de teses e dissertações em texto integral e referencial provenientes das IES, possibilitando uma forma única de busca e acesso a esses documentos. O desenvolvimento da BDTD ocorre no âmbito do programa da Biblioteca Digital Brasileira, que, portanto, vale-se das soluções tecnológicas que têm sido adotadas naquele programa. Segundo Baptista et al. (2007), para a implantação da BDTD, o IBICT elaborou e implantou o Padrão Brasileiro de Metadados para Teses e Dissertações (MTD-BR). A BDTD se apóia em um modelo distribuído e utiliza as tecnologias de arquivos abertos, sendo as IES provedores de dados e o IBICT um agregador responsável pela coleta de metadados de teses e dissertações destes provedores de dados e pelo fornecimento de serviços de informação sobre esses metadados, expondo-os para serem coletados por outros provedores de serviços. (BIBLIOTECA..., 2008)

Para as IES que ainda não implementaram suas bibliotecas digitais de teses e dissertações, o IBICT oferece gratuitamente o Sistema de Publicação Eletrônica de Teses e Dissertações – TEDE. Para as instituições que já implantaram suas bibliotecas digitais de teses e dissertações adotando tecnologias diferentes do sistema TEDE, o IBICT possibilita que se integrem à BDTD por meio da implementação do Protocolo OAI-PMH em seus repositórios de dados, expondo metadados no padrão MTD-BR. (INSTITUTO..., 2008)

A BDTD se operacionaliza por meio da geração de uma base centralizada de metadados. Utiliza o processo de coleta automática de metadados (*harvesting*), que é capaz de ir buscar os metadados diretamente nos servidores das instituições provedoras das informações. Assim como a BDTD, que utiliza a coleta de metadados nacionais para alimentar uma base centralizada, os sistemas internacionais adotam essa mesma base para realizarem a coleta automática. Em particular, o IBICT promoveu a integração do Consórcio Brasileiro de Teses e Dissertações à NDLTD. Desta forma, a produção nacional de teses e dissertações é também disponibilizada internacionalmente.

Em termos gerais, o projeto da BDTD promove não só maior visibilidade da produção científica e tecnológica brasileira, em nível nacional e internacional, mas também gera capacitação nacional nas tecnologias de informação e comunicação usadas para implantação de bibliotecas digitais. Do ponto de vista do usuário final, ao acessar a BDTD, esse pode realizar buscas de forma unificada (em uma única base de dados) às informações de teses e dissertações existentes nas diversas instituições consorciadas. As

teses em meio eletrônico podem ser recuperadas no servidor da instituição provedora da informação. (INSTITUTO..., 2008)

As IES, como a UDESC, cooperam com a BDTD atuando como provedores de dados enquanto o IBICT atua como agregador, coletando os metadados das teses e dissertações dos provedores de dados, provendo serviços de informação sobre esses metadados e expondo-os para serem coletados por outros provedores de serviços.

3 Metadados

Mesmo com o desenvolvimento de novas tecnologias para acesso e recuperação das informações, muitos Sistemas de Recuperação da Informação (SRIs), bastante utilizados atualmente, não estão aptos a recuperarem adequadamente todas as informações relevantes para aqueles que as buscam; retornando respostas, algumas vezes, não satisfatórias. O grande volume de informações disponibilizadas na *web* sem uma forma estruturada de representação contribui para dificultar a recuperação da informação. Cardoso (2006) destaca que a crescente complexidade dos objetos armazenados e o grande volume de dados exigem processos de recuperação cada vez mais sofisticados.

Como citado anteriormente, a organização da informação é essencial para que a recuperação relevante da mesma tenha êxito. Segundo Café e Bräscher (2008), a descrição da informação é fundamental na recuperação da informação, pois cumpre a função de possibilitar o acesso a itens específicos que se encontram armazenados. Nesse sentido, McGarry (1999) sinaliza que a informação deve ser ordenada, estruturada ou contida de alguma forma, senão permanecerá amorfa e inutilizável. Dentre as ferramentas disponíveis atualmente para a padronização da descrição das informações, em rede eletrônica, encontram-se os metadados. A padronização dos metadados facilita o acesso e a recuperação da informação, assegurando aos usuários a fácil movimentação de dados entre diferentes sistemas e plataformas computacionais.

Para muitos autores, metadados significa dados sobre dados. Da mesma forma, Takahashi (2000, p. 172) conceitua metadados como: “[...] dados a respeito de outros dados, ou seja, qualquer dado usado para auxiliar na identificação, descrição e localização de informações. Trata-se, em outras palavras, de dados estruturados que descrevem as características de um recurso de informação.”. Neste sentido, metadado refere-se a uma estrutura descritiva da informação sobre outro dado. Segundo Berners-Lee (1997), metadados são definidos como uma máquina para compreensão das informações sobre os recursos *web* ou outras coisas.

Os metadados têm como objetivo viabilizar a compreensão

da informação por seres humanos e sistemas computacionais, ajudando na identificação, descrição, localização e gerenciamento de recursos da *web*. Segundo Alves e Souza (2007),

Os elementos de metadados têm o propósito primário de descrever, identificar e definir um recurso de informação com o objetivo de modelar e filtrar o acesso. Os metadados são importantes na organização, gestão e recuperação da informação digital, principalmente. Nesse sentido, são adotados procedimentos técnicos de catalogação, indexação e categorização dos conteúdos informacionais, o que possibilita a integração de fontes diversificadas e heterogêneas de informação. (ALVES; SOUZA, 2007, p. 22)¹⁰

■
¹⁰ Documento eletrônico.

Trata-se de um assunto que sempre esteve presente em vários domínios de aplicação. Entretanto, sua utilização passou a aumentar significativamente com a evolução das plataformas computacionais, tornando-se popular com o surgimento dos sistemas de gerenciamento de banco de dados relacionais, que os adaptavam para descrever as estruturas desses sistemas.

Iannella e Waugh (1997) afirmam que os metadados assumem o importante papel de apoiar o uso de recursos eletrônicos e serviços. Os mesmos autores pontuam que, contudo, para um suporte efetivo e implantação de sistemas de metadados muitas questões necessitam ainda de ser abordadas. Dentre estas, os autores citam a integração de diferentes conjuntos de metadados, a construção de ferramentas de produção e gestão de metadados e a criação de formas de estender padrões de metadados.

4 Padrões de metadados

Para satisfazer as necessidades dos usuários, de acordo com o cenário de aplicação no qual estes se encontram inseridos, são desenvolvidos e utilizados diversos padrões de metadados. Alves (2005, p. 115) esclarece que:

[...] formatos de metadados, também chamados de padrões de metadados, são estruturas padronizadas para a representação do conteúdo informacional que será representado pelo conjunto de dados-atributos (metadados). Em outras palavras, os formatos ou padrões de metadados podem ser considerados como formas de representação de um item documentário. (ALVES, 2005, p.115).

Dentre os padrões de metadados utilizados, pode-se citar:

- a) *Dublin Core* (DC): catalogação de documentos eletrônicos na *web*;
- b) *An Interoperability Metadata Standard for Electronic Theses and Dissertations*(ETD-MS): padrão de metadados para descrição de teses e dissertações adotado pela NDLTD;
- c) *Machine Readable Catalogue* (MARC): catalogação bibliográfica;
- d) MTD-BR: padrão brasileiro de metadados para descrição de teses e dissertações adotado pela Biblioteca Digital Brasileira de Teses e Dissertações (BDTD).

Convém ressaltar que padrões DC, MTD-BR e ETD-MS estão relacionados direta ou indiretamente com a estrutura da BDTD, merecendo, portanto, maior ênfase neste trabalho.

O DC pode ser definido como um conjunto de elementos de metadados planejado para facilitar a descrição de recursos eletrônicos. É um formato menos estruturado e mais flexível, que adota a sintaxe do *Resource Description Framework*(RDF) para assegurar a interoperabilidade entre outros formatos.

O ETD-MS pode ser definido como um conjunto de padrões internacional para os elementos de metadados usados para descrever uma tese ou dissertação eletrônica.

Estes padrões de metadados procuram descrever o autor, o trabalho e o contexto no qual a pesquisa foi produzida, de tal forma que estas informações tornem-se úteis para aqueles que procuram pelas teses e/ou dissertações.

Sendo este o padrão internacional, todas as bibliotecas digitais de teses e dissertações que desejarem fazer parte da base cooperativa (*union catalog*) devem estar aderentes a ele.

O MTD-BR foi desenvolvido no âmbito do projeto BDTD, visando atender aos objetivos de geração de produtos e serviços de informação, com vistas à identificação e localização das teses e dissertações eletrônicas (TDEs), bem como permitir a coleta de informação para geração de indicadores e integração com outros repositórios nacionais de informação de ensino e pesquisa no país. (INSTITUTO..., 2008). Possui um conjunto de 19 elementos básicos. Os mesmos são descritos na tabela abaixo (BIBLIOTECA..., 2008).

Elementos Básicos	Descrição
Controle	Dados de controle do registro de Teses e Dissertações
BibliotecaDigital	Dados da Instituição responsável pela publicação digital da tese ou dissertação
BibliotecaDepositaria	Dados da Biblioteca onde o exemplar físico da tese ou dissertação está arquivado
Título	Título da tese ou dissertação
Arquivo	Endereço eletrônico dos arquivos da tese ou dissertação na Biblioteca Digital
Idioma	Idioma da tese ou dissertação
Grau	Grau acadêmico associado à tese ou dissertação
Titulação	Nome do grau acadêmico associado à tese ou dissertação
Resumo	Resumo da tese ou dissertação
Coertura	Escopo espacial ou temporal da tese ou dissertação
Assunto	Tópicos tratados na tese ou dissertação e a tabela de onde estes tópicos foram extraídos, quando for o caso
LocalDefesa	Local de defesa da tese ou dissertação
DataDefesa	Data em que foi defendida a tese ou dissertação
Autor	Dados do Autor da tese ou dissertação
Contribuidor	Dados do Contribuidor da tese ou dissertação e forma de participação (papel)
InstituiçãoDefesa	Dados da Instituição onde a tese foi defendida
AgênciaFomento	Dados da Agência de financiamento que apoiou financeiramente o autor
Direitos	Informa as condições de distribuição, reprodução e utilização da tese ou dissertação
Extensão	Reservado para a criação de sub-itens (subcampos) para uso específico

Quadro 1: Componentes básicos MTD-BR

Fonte: Biblioteca Digital de Teses e Dissertações (BDTD)

A grande maioria dos elementos básicos possui elementos vinculados, ou seja, é composto por outros elementos. O elemento Controle, por exemplo, se subdivide em 4 elementos. Os mesmos podem ser visualizados, juntamente com sua descrição, na tabela abaixo:

Elementos Vinculados	Descrição
Sigla	sigla da instituição consorciada
DataAtualização	data em que foi realizada a última atualização do registro da tese ou dissertação
IdentificaçãoDocumento	código que identifica a tese ou dissertação na base de dados da biblioteca
Tipo	tipo de informação

Quadro 2: Elementos vinculados ao elemento controle do padrão MTD-BR

O total de elementos do padrão MTD-BR, considerando os básicos e os vinculados, é 67.

Além dos elementos básicos e vinculados, o padrão MTD-BR também possui os atributos **Idioma**, **Formato**, **Esquema**, **Norma**, **Papel** e **Namespace**. Os atributos são adotados para os elementos **Idioma**, **URL**, **Legenda**, **Resumo**, **Cobertura**, **Assunto**, **Citação**, **Contribuidor** e **Extensão**. A finalidade dos atributos é especificar o padrão que deve ser seguido para o preenchimento. Por exemplo: o atributo **Idioma**, dos elementos **Legenda** (vinculado ao elemento Arquivo), **Título** (elemento básico), **Resumo** (elemento básico), **Cobertura** (elemento básico), **Assunto** (elemento básico), **Direitos** (elemento básico), segue a Norma ISO 639¹¹ para definição do nome da língua em dois ou três caracteres. Deve-se salientar que todos os atributos são de uso facultativo.

É importante ressaltar que o padrão MTD-BR é compatível com os padrões ETD-MS e DC, o que possibilita a integração dos registros de teses e dissertações com registros de outros repositórios brasileiros, como o Banco de Currículos da Plataforma Lattes.

¹¹ A ISO 639 - *Codes for the representation of names of languages*. É uma norma técnica da ISO para especificação de nomes de idiomas. Fonte: <http://www.oasis-open.org/cover/iso639a.html>.

5 Metodologia

A abordagem geral da pesquisa foi exploratória, uma vez que, de acordo com Gil (1991), estas pesquisas têm como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses.

A estratégia de pesquisa adotada foi a sintética de caso ou estudo de caso, que, segundo Gil (1991), é caracterizado pelo estudo profundo e exaustivo de um ou poucos objetos, de maneira que permita o seu detalhado conhecimento, tarefa praticamente impossível para outras estratégias de pesquisa.

A população da pesquisa foi constituída pelas dissertações eletrônicas da UDESC disponibilizadas na BDTD. Não foram incluídas teses porque embora a UDESC tenha cursos de pós-graduação com grau de mestre e de doutor, este último teve seu

início no ano de 2008, não tendo, portanto, teses defendidas no período desta pesquisa.

O *corpus* da pesquisa se restringiu as dissertações defendidas nesta instituição, no período de 1997 a 2007, totalizando 451 dissertações. Destas, 443 estão disponibilizadas na BDTD e oito, apesar de não terem sido processadas pelo *harvesting* e, portanto, constarem somente na BDTD da UDESC, também foram consideradas no *corpus* de análise.

O método de coleta de dados utilizado foi o de registros institucionais (ou análise documental).

Foram analisados, de forma qualitativa e quantitativa, um total de 46.887 metadados referentes as dissertações eletrônicas da UDESC disponibilizados para a BDTD. Estes metadados foram informados no sistema TEDE, disponibilizado pela BDTD, por servidores que trabalham na biblioteca da UDESC e que receberam um treinamento para utilizar o referido sistema, bem como manuais de operação.

A abordagem qualitativa foi realizada utilizando-se como parâmetro as recomendações do padrão MTD-BR no que se refere ao conteúdo dos elementos de metadados e seus respectivos atributos.

A abordagem quantitativa teve como objetivo comparar a quantidade de informações corretas e incorretas de metadados. Os resultados quantitativos foram apresentados de forma descritiva com o auxílio de tabelas. O exame destas tabelas resultou em recomendações para a melhoria do conteúdo dos metadados informados pela UDESC para a BDTD.

6 Análise dos metadados

Para que se obtenha a informação confiável, de fácil acesso, com um tempo de resposta reduzido e com um custo acessível, é essencial uma adequada gestão de informações, assim como de seus fluxos. Com o objetivo de garantir estas características, o tratamento físico e de conteúdo assumem um papel fundamental, pois, analisam, traduzem e representam a forma e o assunto dos documentos com a finalidade de auxiliar na recuperação de informações.

Segundo Kobashi (1994), a circulação de informações e documentos, tal como é compreendida no âmbito da Ciência da Informação, comporta as seguintes operações básicas:

- a) coleta de documentos;
- b) tratamento de documentos e
- c) difusão dos documentos.

O tratamento é a fase que se dedica a descrever o documento sob dois enfoques: forma e conteúdo. Neste sentido, Kobashi (1994) defende que:

O tratamento documentário do suporte material, objeto da Representação Descritiva ou Catalogação, visa elaborar a descrição normalizada de aspectos físicos do documento (nome da obra, do autor, local de publicação, ano de publicação, editora, entre outros dados). Tal modalidade de tratamento permite o acesso físico ao documento. (KOBASHI, 1994, p.20)

A mesma autora afirma que o tratamento documentário do conteúdo visa, por outro lado, elaborar representações condensadas daquilo que é dito em um dado texto, sendo o resumo e o índice, as representações documentárias típicas.

Para que tanto o tratamento físico como o de conteúdo seja feito de forma consistente, é necessária a utilização de padrões. Estes garantirão a interoperabilidade dos SRIs. No caso da BDTD, como já mencionado, este padrão é definido pelo MTD-BR. Ele se compõe de três partes:

- a) Lista de elementos do *Schema*: esta lista engloba todos os elementos e atributos do padrão MTD-BR, ou seja, os elementos básicos, os elementos vinculados e os atributos;
- b) Padrões adotados para o preenchimento de elementos específicos, baseando-se em tabelas elaboradas e utilizadas por organizações nacionais e/ou internacionais, como Normas ISO ou listas normalizadas de siglas de bibliotecas depositárias utilizadas no CCN/COMUT¹², por exemplo;
- c) Padrões adotados para os atributos com a finalidade de identificar alguns elementos. Convém lembrar que os atributos do MTD-BR são elementos facultativos adotados no *Schema*.

O uso adequado do MTD-BR garante a circulação das teses e dissertações no meio acadêmico. Tomando por base a distinção entre forma e conteúdo citada por Kobashi (1994), examinamos os elementos do padrão MTD-BR de metadados, a fim de identificar quais se destinam à descrição física e quais se destinam à descrição de conteúdo do documento. Constatou-se que, dos 67 elementos analisados, 61 destinam-se à descrição física, três destinam-se à descrição de conteúdo do documento e três não foram informados¹³ pela UDESC. Os elementos que se destinam à descrição de conteúdo são: **Título, Resumo e Assunto**, os que deixaram de ser informados são: **Cobertura**, que indica o escopo espacial ou temporal da dissertação, **Area**, que informa a área de concentração do programa de pós-graduação em que a dissertação foi defendida e extensão, reservado para a criação de subitens (sub-campos) para uso específico e os que se destinam à física, todos os demais. O exame das tabelas abaixo (4,5,6 e 7) mostra que os elementos que se destinam à descrição de conteúdo apresentam uma qualidade superior a dos elementos que se destinam à descrição física.

Tendo como base os padrões do MTD-BR adotados para

■
¹² O Catálogo Coletivo Nacional de Publicações Seriadas (CCN), coordenado pelo IBICT, é uma rede cooperativa de unidades de informação de instituições localizadas no Brasil. Possibilita o acesso a publicações periódicas científicas e técnicas e reúne informações de centenas de catálogos, produzidos pelas principais bibliotecas do país, em um único catálogo nacional de acesso público. (IBICT)

■
¹³ Elementos não informados são aqueles que não foram informados pelo bibliotecário no momento da digitação dos metadados.

os metadados e para seus respectivos atributos, foi realizada a análise do conteúdo dos metadados informados pela UDESC para a BDTD. Ressalta-se que para cada metadado analisado, foi apresentada uma tabela sintetizando os seguintes pontos:

- a) Elemento: informa o nome do elemento analisado;
- b) Descrição: informa a descrição do elemento analisado;
- c) Preenchimento: informa se o preenchimento do elemento analisado é obrigatório ou facultativo;
- d) Padrão: informa se foi adotado algum padrão para o preenchimento do elemento analisado;
- e) Total de elementos: informa o total de elementos analisados;
- f) Corretos: informa o número de elementos analisados informados corretamente;
- g) Incorretos: informa o número de elementos analisados que foram informados incorretamente;
- h) Não informados: informa o número de elementos analisados que deixaram de ser informados.

Para exemplificar, abaixo pode-se visualizar como foi apresentada na dissertação, a análise dos elementos que compõem o elemento **Controle**:

Elemento	Descrição	Preenchimento	Padrão	Total de elementos	Corretos	Incorretos	Não informados
1.1 Sigla	Indica a sigla da Instituição Consorciada	Obrigatório	Não adotado	451	451	0	0
1.2 DataAtualizacao	indica a data da última atualização do registro da dissertação	Obrigatório	Protocolo OAI	451	450	0	1
1.3 Identificação Documento	indica o código que identifica a dissertação na base de dados da biblioteca	Obrigatório	Não adotado	451	451	0	0
1.4 Tipo	indica o tipo da fonte de informação	Obrigatório	DCMI Type Vocabulary	895	894	1	0

Quadro 3 - Descrição dos elementos Sigla, Data Atualização, Identificação, Documento e Tipo do elemento Controle

É importante destacar que quando o preenchimento do elemento analisado foi feito corretamente, não foi apresentada nenhuma discussão a respeito. Entretanto, foram tecidos alguns comentários sobre a consequência do preenchimento incorreto dos elementos, assim como da falta de preenchimento dos mesmos.

A partir da análise de todos os elementos do padrão MTD-BR e seus respectivos atributos apuraram-se os seguintes totais:

- a) 25073 elementos de preenchimento facultativo informados corretamente;
- b) 4002 elementos de preenchimento facultativo informados incorretamente;
- c) 7002 elementos de preenchimento facultativo que deixaram de ser informados;
- d) 10798 elementos de preenchimento obrigatório informados corretamente;
- e) 4 elementos de preenchimento obrigatório informados incorretamente;
- f) 8 elementos de preenchimento obrigatório que deixaram de ser informados.

A soma destes totais resultou em 46887 metadados analisados.

Na tabela a seguir, podem-se observar todos os elementos de preenchimento facultativo que foram informados incorretamente. Convém mencionar que, neste artigo, foram inclusos os percentuais, nas tabelas 4, 5, 6 e 7.

Elemento	Descrição	Elementos analisados	Informados incorretamente	
2.4.5 CNPJ	indica o CNPJ da Instituição que abriga a biblioteca digital	1	1	100%
Atributo Idioma do elemento Legenda	indica o(s) idioma(s) no(s) qual(is) o arquivo vinculado a Legenda está disponível	1042	1042	100%
5.3 NívelAcesso	indica se o acesso a dissertação poderá ser público ou não	1042	3	0,29%
12.1 Cidade	indica a cidade onde foi defendida a dissertação	451	2	0,44%
12.2 UF	indica a UF onde foi defendida a dissertação	451	2	0,44%
14.3 Lattes	indica o endereço eletrônico do currículo do autor na Plataforma Lattes	451	26	5,76%
14.4 CPF	indica o CPF do autor da dissertação	451	437	96,89%
14.5 Afiliação	indica a instituição à qual o autor é afiliado	1045	1045	100%
14.5.5 CNPJ	indica o CNPJ da Instituição à qual o autor da dissertação é afiliado	451	241	53,44%
15.3 Lattes	indica o endereço eletrônico do currículo do contribuidor na Plataforma Lattes	451	22	4,88%
15.5.5 CNPJ	indica o CNPJ da Instituição à qual o contribuidor da dissertação é afiliado	451	279	61,86%
16.5 CNPJ	indica o CNPJ da Instituição onde a dissertação foi defendida	451	451	100%
17.5 CNPJ	indica o CNPJ da agência de fomento que apoiou financeiramente o autor da dissertação	451	451	100%

Quadro 4 - Elementos de preenchimento facultativo informados incorretamente

O preenchimento incorreto de um elemento facultativo não ocasiona a exclusão da respectiva tese ou dissertação do processo de *harvesting*. Entretanto, prejudicará a recuperação da informação, caso a busca por meio deste referido elemento seja disponibilizada por algum repositório. Como é o caso do elemento Nível Acesso, que teve 0,29% de informações incorretas e é utilizado como campo de busca pela ND LTD.

Na tabela seguinte, podem-se observar todos os elementos de preenchimento facultativo que deixaram de ser informados.

Elemento	Descrição	Elementos analisados	Não informados	
3.4 Numero Chamada	indica o código que identifica a dissertação no acervo da biblioteca depositária	451	18	3,99%
Atributo Idioma do elemento Titulo	indica o idioma no qual foi informado o título da dissertação	615	5	0,81%
5.1 URL	indica o endereço eletrônico do arquivo ou dos metadados da dissertação na biblioteca de origem	1042	1	0,09%
Atributo Formato do elemento URL	indica o formato eletrônico no qual a dissertação está disponível	1042	1	0,09%
Atributo Idioma do elemento URL	indica o(s) idioma(s) no(s) qual(is) o(s) arquivo(s) da dissertação está disponível	1042	1042	100%
5.2 Legenda	indica a legenda do arquivo da dissertação	1042	2	0,19%
Atributo Idioma do elemento Resumo	Indica o idioma do resumo da dissertação	832	99	11,90%
10 Cobertura	indica o escopo espacial ou temporal da dissertação	451	451	100%
Atributo Idioma do elemento Cobertura	indica o idioma no qual foi informada a cobertura	451	451	100%
11 Assunto	indica os tópicos tratados na dissertação e a tabela de onde os mesmos foram extraídos	3655	2	0,05%
Atributo Idioma do elemento Assunto	indica o(s) idioma(s) no(s) qual(is) o(s) assunto(s) da dissertação está disponível	3655	168	4,60%
12.2 UF	indica a UF onde foi defendida a dissertação	451	5	1,11%
Atributo Norma do elemento Citação	indica a(s) norma(s) utilizada(s) para informar a forma pela qual o autor será citado	451	451	100%
14.3 Lattes	indica o endereço eletrônico do currículo do autor na Plataforma Lattes	451	324	71,84%
14.5.1 Nome (Inst. Autor)	indica o nome da Instituição à qual o autor da dissertação é afiliado	451	210	46,56%
14.5.2 Sigla (Inst. Autor)	indica a sigla da Instituição à qual o autor da dissertação é afiliado	451	210	46,56%
14.5.3 Pais (Inst. Autor)	indica o país da Instituição à qual o autor da dissertação é afiliado	451	210	46,56%

14.5.4 UF (Inst. Autor)	indica a Unidade Federativa da Instituição à qual o autor da dissertação é afiliado	451	210	46,56%
14.5.5 CNPJ (Inst. Autor)	indica o CNPJ da Instituição à qual o autor da dissertação é afiliado	451	210	46,56%
14.5.6 URL (Inst. Autor)	indica o endereço eletrônico da Instituição à qual o autor da dissertação é afiliado	451	210	46,56%
Atributo Papel Contribuidor	indica o tipo de contribuição e forma de participação na geração/aprovação da dissertação	451	1	0,22%
15.2 Citacao	indica a forma pela qual a pessoa física do contribuidor será citada	451	2	0,44%
15.3 Lattes	indica o endereço eletrônico do currículo do contribuidor na Plataforma Lattes	451	332	73,61%
15.5.1 Nome (Inst. Contribuidor)	indica o nome da Instituição à qual o contribuidor da dissertação é afiliado	451	172	38,14%
15.5.2 Sigla (Inst. Contribuidor)	indica a sigla da Instituição à qual o contribuidor da dissertação é afiliado	451	172	38,14%
15.5.3 Pais (Inst. Contribuidor)	indica o país da Instituição à qual o contribuidor da dissertação é afiliado	451	172	38,14%
15.5.4 UF (Inst. Contribuidor)	indica a Unidade Federativa da Instituição à qual o contribuidor da dissertação é afiliado	451	172	38,14%
15.5.5 CNPJ (Inst. Contribuidor)	indica o CNPJ da Instituição à qual o contribuidor da dissertação é afiliado	451	172	38,14%
15.5.6 URL (Inst. Contribuidor)	indica o endereço eletrônico da Instituição à qual o contribuidor da dissertação é afiliado	451	172	38,14%
16.7.2 Area	indica a área de concentração do Programa de Pós-Graduação onde a dissertação foi defendida	451	451	100%
18 Direitos	indica as condições de distribuição, reprodução e utilização da dissertação	451	1	0,22%
Atributo Idioma do elemento Direitos	indica o idioma no qual os direitos foram informados	451	1	0,22%
19 Extensao	reservado para a criação de sub-itens (sub-campos) para uso específico	451	451	100%
Atributo Namespace do elemento Extensao	indica o idioma no qual a extensão foi informada	451	451	100%

Quadro 5 - Elementos de preenchimento facultativo que deixaram de ser informados

A falta de informação de um elemento de preenchimento facultativo não implica na exclusão da respectiva tese ou dissertação do processo de *harvesting*. Entretanto, a recuperação da informação será prejudicada, caso a BDTD e/ou a NDLTD disponibilize a busca por meio deste elemento que deixou de ser informado. Pode-se visualizar, na tabela acima, que o elemento Número Chamada teve 3,99% de não informados; por ser um elemento de preenchimento facultativo, as respectivas dissertações não foram excluídas do processo de *harvesting*. Entretanto, essa falta de informação pode prejudicar aqueles que procuram adotar uma classificação cooperativa em rede, uma vez que o número de chamada contém o número de classificação utilizado pela biblioteca cooperante.

Na próxima tabela, podem-se observar todos os elementos de preenchimento obrigatório que foram informados incorretamente:

Elemento	Descrição	Elementos analisados	Informados incorretamente	
1.4 Tipo	indica o tipo da fonte de informação	895	1	0,11%
4 Titulo	indica o título da dissertação	615	1	0,16%
7 Grau	indica o grau acadêmico associado a dissertação	451	1	0,22%
14.1 Nome (Autor)	indica o nome da pessoa física do autor	451	1	0,22%

Quadro 6 - Elementos de preenchimento obrigatório que foram informados incorretamente

A informação incorreta de um elemento de preenchimento obrigatório não implica na exclusão desta dissertação do processo de *harvesting*, desde que o preenchimento esteja de acordo com os tipos especificados no padrão para este elemento. Na tabela acima, pode-se visualizar que o elemento Tipo teve 0,11% de informações incorretas. Como o padrão especificado pelo MTD-BR foi levado em consideração (o elemento foi informado como som, que, mesmo constando na tabela de tipos do MTD-BR, não é a informação correta). O correto, no caso apontado, seria registrar o metadado com a informação Texto ou Tese ou Dissertação, por ser a fonte de informação um texto e não um som. Dessa forma, o preenchimento incorreto deste elemento não resultou na exclusão desta dissertação do processo de *harvesting*. E, como a BDTD não disponibiliza a busca por meio do Tipo, o preenchimento incorreto deste elemento não teve conseqüências na recuperação dessa dissertação na base da BDTD. Entretanto, se considerarmos a busca por outros repositórios, o preenchimento incorreto desse metadado ocasionou problema de interoperabilidade. Este é o caso da NDLTD, que disponibiliza a busca por meio do elemento Tipo.

Nesta última tabela, podem-se observar todos os elementos de preenchimento obrigatório que deixaram de ser informados, resultando na exclusão das respectivas dissertações do processo de *harvesting*:

Elemento	Descrição	Elementos analisados	Não informados	
1.2 DataAtualizacao	data em que foi realizada a última atualização do registro da dissertação	451	1	0,22%
4 Titulo	título da dissertação	615	1	0,16%
5.1 URL	endereço eletrônico do arquivo ou dos metadados da dissertação na biblioteca de origem	1042	1	0,09%
9 Resumo	resumo da dissertação	832	2	0,24%
13 DataDefesa	data da defesa da dissertação eletrônica	451	2	0,44%
15.1 Nome	nome da pessoa física do contribuidor	451	1	0,22%

Quadro 7 - Elementos de preenchimento obrigatório que deixaram de ser informados

A falta de informação de um elemento de preenchimento obrigatório resulta na exclusão da respectiva dissertação da base de dados da BDTD e, conseqüentemente, da base de dados da

NDLTD. Por exemplo, na tabela acima se pode observar que o elemento Título teve 0,16% de elementos não informados, o que ocasionou a exclusão da respectiva dissertação do processo de *harvesting*. Dito de outra forma, os metadados da respectiva dissertação deixaram de fazer parte da base de dados da BDTD e da NDLTD.

A análise dos metadados informados pela UDESC para as dissertações defendidas entre 1997 e 2007, nos permitiu observar que:

- a) Os atributos dos elementos, cujo preenchimento é facultativo, muitas vezes, deixaram de ser informados ou foram informados incorretamente. Como por exemplo, os atributos **Idioma**, do elemento **URL** e **Norma**, do elemento **Citação**, que resultaram em 100% de não informados. Ou ainda, o atributo **Idioma**, do elemento **Legenda**, que teve 100% de informações incorretas. Por não ser de preenchimento obrigatório, a falta de informação de um atributo ou a informação incorreta não ocasiona a exclusão das respectivas dissertações do processo de *harvesting*. Entretanto, pode deixar o usuário frustrado com o resultado da busca, tendo em vista que estes dados estão previstos na descrição;
- b) os elementos cujo preenchimento é facultativo, muitas vezes, deixaram de ser informados. Como por exemplo, o elemento **Cobertura**, que teve 100% de não informados, o elemento **Lattes** (do autor), que teve 71,84% de não informados e o elemento **Lattes** (do contribuidor), que teve 73,61% de não informados. Embora a falta de informação de elementos facultativos não ocasione exclusão do processo de *harvesting*, pode impossibilitar a recuperação das respectivas dissertações, na base de dados da BDTD e/ou da NDLTD, caso seja disponibilizada a busca pelo elemento que deixou de ser informado, em algum destes repositórios;
- c) o maior número de preenchimentos incorretos foram dos elementos **Lattes**, **CPF** e **CNPJ**. É importante destacar que o currículo **Lattes** possui *links* para outras bases como a Scielo. Desta forma, o preenchimento incorreto deste elemento impossibilita o acesso ao currículo Lattes do autor e, conseqüentemente, aos *links* para outros trabalhos dele, assim como a recuperação da respectiva dissertação por meio da busca pelo Lattes na base de dados da BDTD. O preenchimento incorreto do elemento **CPF** não inviabiliza a recuperação das respectivas dissertações na base de dados da BDTD nem da NDLTD, uma vez que, atualmente, a busca por este elemento não é dispo-

nibilizada por nenhuma das duas instituições. Entretanto, poderá tornar impossível sua utilização caso, futuramente, a BDTD venha a adotar este elemento de descrição única como item de ligação com outras bases de dados como o Lattes. O preenchimento incorreto do **CNPJ**, por tratar-se de um campo de preenchimento facultativo, não implica na exclusão das respectivas dissertações do processo de *harvesting*. Também não terá conseqüências na busca das respectivas dissertações, porque, atualmente, não é disponibilizada a busca por meio deste elemento na BDTD e nem na NDLTD. Entretanto, é importante refletir sobre a possibilidade de relacionamentos com outras bases por meio desse elemento futuramente;

d) o preenchimento dos elementos que devem ser obrigatoriamente informados, na grande maioria das vezes, foi feito corretamente. Entretanto, alguns destes deixaram de ser informados, implicando na exclusão das respectivas dissertações do processo de *harvesting*. Desta forma, as mesmas deixaram de fazer parte da base de dados da BDTD e da NDLTD.

7 Conclusões

A pesquisa bibliográfica sobre metadados realizada neste trabalho possibilitou a formação de um referencial teórico sobre o assunto, o que contribuiu para evidenciar a importância do uso de metadados nos SRIs para que a busca e acesso tornem-se mais eficientes. Nesse sentido, a literatura mostrou que o preenchimento correto de um padrão único de metadados, ao auxiliar a organização da informação, permite que o documento circule mais agilmente na comunidade a que se destina.

Foi com esse espírito que esta pesquisa procurou propor recomendações para a melhoria do conteúdo dos metadados informados pela UDESC para a BDTD, no sentido de apresentar uma reflexão mais aprofundada sobre como os metadados estão sendo informados. A proposição das melhorias partiu da análise de todos os metadados informados pela UDESC e disponibilizados para a BDTD, de acordo com o padrão de metadados MTD-BR.

Para alcançar o objetivo proposto por este trabalho foi necessário estabelecer alguns objetivos específicos que direcionassem as proposições necessárias para a melhoria do conteúdo dos metadados informados pela UDESC para a BDTD. O primeiro desses objetivos, levantamento dos metadados descritivos alimentados pela UDESC e disponibilizados para a BDTD, foi atingido por meio da análise da base de dados que armazena os

metadados das dissertações eletrônicas da UDESC. Essa análise possibilitou que fossem filtrados todos os metadados das dissertações defendidas entre 1997 e 2007. Por meio do segundo objetivo deste trabalho, organização desses metadados, a fim de identificar quais se destinam à descrição física e quais se destinam à descrição de conteúdo do documento, apurou-se que, dos 67 elementos analisados, 03 se destinam à descrição de conteúdo do documento, 61 se destinam à descrição física do documento e 03 não foram informados pela UDESC. O terceiro objetivo, exame do conteúdo desses metadados com base nas regras estabelecidas pelos modelos adotados para os elementos do padrão MTD-BR de metadados e para seus respectivos atributos, nos permitiu verificar que, na grande maioria das vezes, os metadados são informados de acordo com o padrão MTD-BR. Entretanto, observou-se que os elementos de preenchimento facultativo, assim como os atributos dos elementos, deixaram de ser informados muitas vezes. Também se observou que os atributos dos elementos foram informados incorretamente para algumas dissertações. O preenchimento incorreto dos atributos e dos elementos de preenchimento facultativo, bem como a falta de informação dos mesmos, não ocasiona a exclusão dos respectivos documentos do processo de *harvesting*. Mas, caso seja disponibilizada a busca por algum destes elementos, a recuperação das respectivas dissertações pode tornar-se impossível na BDTD, se esta disponibilizar este tipo de busca, e na NDLTD, se a mesma também disponibilizar a busca por esse elemento. Observou-se que o maior número de preenchimento incorreto dos elementos ocorreu para os elementos Lattes, CPF e CNPJ, o que resultou na impossibilidade de relacionamentos com outros sistemas de recuperação por meio destes elementos. O preenchimento dos elementos que devem ser obrigatoriamente informados, na grande maioria das vezes, foi feito corretamente. Embora, poucas vezes, alguns destes não tenham sido informados, resultando na exclusão das respectivas dissertações do processo de *harvesting*.

Sendo assim, com o objetivo de melhorar a qualidade dos metadados informados pela UDESC para a BDTD, foram feitas as seguintes sugestões relativas ao sistema TEDE:

- a) Para auxiliar na dificuldade relativa ao preenchimento do metadado CPF e CNPJ, sugerimos que seja incluído um texto explicativo sobre como informá-los, juntamente com um exemplo. Ou ainda, automatizar a consistência do preenchimento destes metadados, a fim de impedir que os mesmos sejam registrados de maneira incorreta;
- b) para auxiliar o preenchimento do metadado Lattes, sugerimos que seja incluído um texto explicativo sobre como informá-lo, juntamente com um exemplo. Além disso,

seria interessante ter uma explicação da importância de seu preenchimento e como proceder para localizar o endereço correto, que fica dentro do currículo e não na barra de endereços do site da plataforma Lattes;

- c) para auxiliar o preenchimento do metadado Cidade de Defesa, sugerimos que seja incluída uma caixa de seleção com as cidades, a fim de evitar que sejam registradas cidades com erro de grafia e também para que as mesmas sigam um padrão;
- d) para auxiliar no preenchimento do metadado Ano de Defesa, sugerimos que seja incluída uma caixa de seleção com, pelo menos, dez anos anteriores ao atual, juntamente com o mesmo. Ou ainda, automatizar a verificação do ano de defesa, a fim de evitar cadastros com datas incorretas, o que pode interferir no resultado das buscas por período;
- e) ressalta-se a conveniência de conscientizar os responsáveis pelo preenchimento dos metadados, sobre a importância de informá-los corretamente, mesmo quando o preenchimento é facultativo.

Sugere-se ainda que seja realizado um estudo para verificar a real necessidade de alguns metadados facultativos, uma vez que a sua falta ou incorreção de preenchimento pode ocorrer devido à baixa relevância do mesmo para a comunidade que participa da rede.

Finalmente, concluiu-se que a qualidade na representação de recursos tem grande influência no resultado das buscas e, conseqüentemente, contribui para garantir a interoperabilidade e a sociabilização da informação. Especificamente em relação a BDTD, é importante ressaltar que esta qualidade aumenta a visibilidade da produção acadêmica, otimiza o uso dos dados, maximiza a troca de informações e influencia os fatores de impacto, favorecendo o progresso científico.

Analysis focuses on metadata in the light of the standard MTD-BR

ABSTRACT

The difficulty of access and retrieval of information, caused by increasing volume of information available without a structured representation, is one of the consequences of increasingly frequent use of the world wide web (www). As part of this reality, digital libraries seek to ensure interoperability through the physical description and content of its items, adopting specific standards for metadata. In this context, this article describes the research developed in the Master of Information Science, Federal University of Santa Catarina, which analyzed the contents of the metadata of the dissertations of the University of Santa Catarina - UDESC, made available to the Library Digital Theses and Dissertations-BDTD in order to verify that they are being informed in accordance with standard MTD-BR and thus contribute to the location and recovery of dissertations that

institution. As the methodological procedures, the approach of the research was exploratory and was the metadata analyzed qualitatively and quantitatively. The qualitative approach was based on recommendations of the standard MTD-BR for the content of metadata elements and their attributes and quantitative approach enabled the comparison of the amount of metadata misinformed and incorrectly. The results enabled us to make suggestions to improve the content of the metadata for informed UDESC for BDTD and concluded that the digital library of UDESC, although a good quality for the completion of the metadata, could improve this procedure in favor of interoperability and information retrieval.

KEYWORDS: Metadata. Dublin Core. ETD-MS. MTD-BR. BDTD. NDLTD.

El Análisis se concentra en los metadatos a la luz de la norma MTD-BR

RESUMEN

La dificultad de acceso y recuperación de información, causado por el aumento del volumen de información disponible sin una representación estructurada, es una de las consecuencias del uso cada vez más frecuentes de la *World Wide Web* (www). Como parte de esta realidad, las bibliotecas digitales tratan de garantizar la interoperabilidad a través de la descripción física y el contenido de sus artículos, la adopción de normas específicas para los metadatos. En este contexto, el presente artículo se describe la investigación desarrollada en la Maestría en Ciencias de la Información de la Universidad Federal de Santa Catarina, que analizó el contenido de los metadatos de las tesis de la Universidad de Santa Catarina - UDESC, pondrá a disposición de las Tesis de Biblioteca Digital y Disertaciones-BDTD a fin de verificar que están siendo informados, de conformidad con la norma MTD-BR y así contribuir a la localización y recuperación de las tesis de la institución. Como los procedimientos metodológicos, el enfoque de la investigación fue exploratoria y se analizaron los metadatos cualitativa como cuantitativamente. El enfoque cualitativo, basado en las recomendaciones del MTD-BR estándar para el contenido de elementos de metadatos y sus atributos y enfoque cuantitativo permitió la comparación de la cantidad de metadatos y mal informado incorrectamente. Los resultados nos permitieron hacer sugerencias para mejorar el contenido de los metadatos para UDESC informó de BDTD y concluyó que la biblioteca digital de UDESC, a pesar de una buena calidad para la realización de los metadatos, podría mejorar este procedimiento en favor de la interoperabilidad y la recuperación de la información .

PALABRAS CLAVE: Metadatos. *Dublin Core*. ETD-MS. MTD-BR. BDTD. NDLTD.

Referências

ALVES, M. D. R.; SOUZA, I. F. Estudo de correspondência de elementos metadados: Dublin Core e MARC21. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 4, n. 2, 2007. Disponível em: <<http://server01.bc.unicamp.br/seer/ojs/include/getdoc.php?id=389&article=94&mode=pdf>>. Acesso em: maio 2008.

ALVES, R. C. V. **Web semântica**: uma análise focada no uso de metadados. 2005. 180f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern**

- Information Retrieval.** New York: Addison Wesley, 1999.
- BAPTISTA, A. A. et al. COMUNICAÇÃO CIENTÍFICA: o papel da Open Archives Initiative no contexto do Acesso Livre. **Encontros Bibli:** revista eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 13, n. esp., 1. sem. 2008. Disponível em: <http://www.encontros-bibli.ufsc.br/bibesp/esp_06/apresentacao_esp_20071.pdf>. Acesso em: jul. 2008.
- BERNERS-LEE, T. **Metadata architecture:** documents, metadata and link. 1997. Disponível em: <<http://www.w3.org/DesignIssues/Metadata.html>>. Acesso em: fev. 2008.
- BIBLIOTECA DIGITAL BRASILEIRA DE TESES E DISSERTAÇÕES.. Disponível em: <<http://btdt.ibict.br>>. Acesso em: maio 2008.
- BORGMAN, C. L. Social aspects of digital libraries. In: WORKSHOP ON SOCIAL ASPECTS OF DIGITAL LIBRARIES, 1996, Los Angeles. **Final Report** UCLA-NSF, 1996. Disponível em <<http://www.ifla.org/documents/libraries/net/ucladl01.pdf>>. Acesso em: abr. 2008.
- CAFÉ, L.; BRÄSCHER, M. Organização da informação e bibliometria. **Encontros Bibli:** revista eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 13, n. esp., 1. sem. 2008. Disponível em: <<http://www.periodicos.ufsc.br/index.php/index.php/eb/article/view/1334/1032>>. Acesso em: jul. 2008.
- CARDOSO, O. N. P. Recuperação de Informação. **INFOCOMP Journal of Computer Science**, Lavras, MG, dez. 2006. Disponível em: <www.dcc.ufla.br/infocomp/artigos/v2.1/olinda.pdf>. Acesso em: fev. 2008.
- CHEN, H. Semantic research for digital libraries. **D-Lib Magazine**, v.5, n. 10, 1999. Disponível em: <<http://www.dlib.org/dlib/october99/chen//10chen.html>>. Acesso em: out. 2008.
- CUNHA, Murilo Bastos da. Desafios na construção de uma biblioteca digital. **Ciência da Informação**, Brasília, DF, v. 28, n. 3, p. 255-266, set./dez. 1999.
- DIAS, E. W. Contexto Digital e Tratamento da Informação. **DataGramaZero:** revista de Ciência da Informação, Belo Horizonte, v.2, n.5, 2001. Disponível em: <http://www.dgz.org.br/out01/Art_01.htm>. Acesso em: ago. 2008.
- FOSKETT, A. C. **A Abordagem temática da informação.** São Paulo: Polígono; Brasília: Ed. Universidade de Brasília, 1973.
- GIL, A.C. **Como elaborar projetos de pesquisa.** São Paulo: Atlas, 1991.
- IANNELLA, R.; WAUGH, A. **Metadata:** enabling the Internet. 1997. Disponível em: <<http://www.ifla.org/documents/libraries/cataloging/metadata/ianr1.pdf>>. Acesso em: jul. 2008.
- INSTITUTO BRASILEIRO DE INFORMAÇÃO E TECNOLOGIA. Disponível em: <<http://btdt.ibict.br>>. Acesso em: abr. 2008.
- KOBASHI, N. Y. **A Elaboração de informações documentárias, em busca de uma metodologia.** 1994. Tese (Doutorado em Ciências da Comunicação) - Escola de Comunicações e Artes, Universidade de São Paulo, 1994.
- McGARRY, K. **O Contexto dinâmico da informação:** uma análise introdutória. Brasília, DF: Briquet de Lemos/Livros,

1999.

NETWORKED DIGITAL LIBRARY OF THESES AND DISSERTATIONS. Disponível em: <<http://www.ndltd.org>> Acesso em: maio 2008.

PIEIDADE, M. A. R. **Introdução à teoria da classificação**. Rio de Janeiro: Interciência, 1977.

SAYÃO, L. F. Padrões para bibliotecas digitais abertas e interoperáveis. **Encontros Bibli**: revista eletrônica de Biblioteconomia e Ciência da Informação. Florianópolis, n. esp., 1. sem. 2007. Disponível em: <http://www.encontrosbibli.ufsc.br/bibesp/esp_06/bibesp_esp_06_sayao_esp_20071.pdf>. Acesso em: jul. 2008.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, DF, v. 33, n. 1, p. 132-141, jan./abr. 2004. Disponível em: <<http://www.ibict.br/cionline/include/getdoc.php?id=356&article=71&mode=pdf>> Acesso em: maio 2007.

TAKAHASHI, T. (Org.). **Sociedade da informação no Brasil**: livro verde. Brasília: Ministério da Ciência e Tecnologia, 2000.

Jaqueline Costa Alves

*Mestre em Ciência da Informação pela
Universidade Federal de Santa Catarina (UFSC).
Técnico Universitário de Desenvolvimento da
Universidade do Estado de Santa Catarina
(UDESC).
E-mail: jaqueline@udesc.br*

Lígia Café

*Doutora em Linguística pela Université Laval
(Canadá).
Professora do Departamento de Ciência da
Informação e do Programa de Pós-graduação em
Ciência da Informação da Universidade Federal de
Santa Catarina (UFSC).
E-mail: ligia@cin.ufsc.br*

Recebido em: 21/04/2010

Aceito em: 04/11/2010