

Metadados de assuntos de teses e dissertações em repositórios: estudo exploratório sobre controle de vocabulário

Mariângela Spotti Lopes Fujita

Universidade Estadual Paulista, Marília, SP, Brasil;
mariangela.fujita@unesp.br; <https://orcid.org/0000-0002-8239-7114>

Rosane Rodrigues de Barros Ribas

Universidade Estadual Paulista, São Paulo, SP, Brasil;
rosane.barros@unesp.br; <https://orcid.org/0000-0002-9550-8901>

Milena Maria Scarpino Rodrigues

Universidade Estadual Paulista, Araraquara, SP, Brasil;
milena.rodrigues@unesp.br; <https://orcid.org/0000-0002-8904-7686>

Telma Jaqueline Dias Silveira

Universidade Estadual Paulista, Marília, SP, Brasil;
telma.silveira@unesp.br; <https://orcid.org/0000-0002-2712-2011>

Resumo: Os metadados de teses e dissertações são inseridos nos repositórios automaticamente ou manualmente, em formato *Dublin Core*, diferente do utilizado pelas bibliotecas, gerando duplicidade de trabalho. O objetivo é comparar os procedimentos de tratamento temático de teses e dissertações existentes no catálogo *on-line* e no repositório, por meio da análise de metadados de assunto. Realizou-se estudo exploratório com estudo de caso sobre o uso de controle de vocabulário em metadados de teses e dissertações, com o repositório institucional da Universidade Estadual Paulista e o catálogo *on-line* Athena da mesma instituição. O estudo exploratório teve duas etapas: estudo documental da trajetória histórica de tratamento de teses e dissertações da universidade e análise de procedimentos de tratamento temático de teses e dissertações nas bases de dados, por autores e catalogadores. A análise de procedimentos de tratamento temático empregou observação de funcionamento e de rotinas e padrões, para avaliação de controle de vocabulário na análise de metadados de assuntos. Os resultados obtidos revelaram que o controle de vocabulário não é feito nas interfaces de busca para recuperação da informação e nem na lista alfabética de palavras-chave, o que significa falta de orientações aos autores, quanto à atribuição de descritores. As análises demonstraram dois tratamentos temáticos, cujos resultados são diferentes representações de assuntos, em linguagem natural, no repositório, e em linguagem controlada, no catálogo. Foram elaboradas duas propostas para o tratamento temático em metadados de assunto de teses e dissertações em repositórios e catálogos de

bibliotecas. Conclui-se que as propostas poderão ser auxiliadas pela elaboração de uma política de indexação do repositório que preveja o compartilhamento de dados resultantes do autoarquivamento de teses e dissertações com o catálogo e que possa aceitar a atribuição de descritores de vocabulário controlado, além da atribuição de palavras-chave.

Palavras-chave: metadados de assunto; repositórios institucionais; dissertações; teses; política de indexação

1 Introdução

Teses e dissertações são produções científicas inéditas e originais, concretizadas em decorrência do desenvolvimento de projetos de pesquisa por pesquisadores de Programas de Pós-Graduação certificados em universidades. Todos os sistemas de bibliotecas universitárias são depositários de teses e dissertações, desde o formato impresso. Quando o exemplar definitivo impresso era enviado à biblioteca, fazia-se o tratamento descritivo e temático para incluir a ficha catalográfica no catálogo físico, e o exemplar era armazenado em estante física. Com a evolução dos catálogos impressos para o *on-line*, os mesmos procedimentos eram realizados, com a vantagem de o catálogo *on-line* poder ser acessado por usuários remotos, além dos fisicamente próximos e, se fosse de interesse, solicitar o exemplar físico por empréstimo.

A *web* e a internet tornaram possível o acesso às teses e dissertações em formato digital e, com isso, surgiram as bibliotecas digitais, decorrentes do trabalho feito no catálogo *on-line*, cujos *softwares* não comportavam a inserção do documento digital junto aos registros bibliográficos. A transição para os repositórios com *software open access* rapidamente superou as bibliotecas digitais de teses e dissertações, nos moldes em que foram criadas, por proporcionarem interoperabilidade tecnológica e semântica.

Nos repositórios atuais, os metadados de teses e dissertações são inseridos automaticamente ou manualmente, em formato *Dublin Core*, diferente do formato usado pelas bibliotecas digitais e dos catálogos *on-line* de bibliotecas. Na prática, os sistemas de bibliotecas universitárias realizam operações diversas, em bases de dados diversas, as quais, em muitos casos, não são interoperáveis tecnologicamente e muito menos semanticamente. Essa

situação gera duplicidade de trabalho, quando é preciso maximizar esforços das reduzidas equipes de bibliotecários nos sistemas de bibliotecas universitárias.

Metadados de assuntos de teses e dissertações em repositórios institucionais poderiam ser norteados por política de indexação e controle de vocabulário de tesouros já aplicados por catalogadores, em catálogos *on-line*. Desse modo, a proposição é elaborar política de indexação para o tratamento temático em metadados de assunto de teses e dissertações em repositórios e catálogos de bibliotecas. Para isso, o objetivo desta investigação é realizar análise comparada dos procedimentos de tratamento de teses e dissertações existentes no catálogo *on-line* e no repositório, para verificação do controle de vocabulário, por meio da análise de metadados de assunto.

As principais contribuições deste estudo são dirigidas a uma proposta de metodologia de avaliação de controle de vocabulário em repositórios (FUJITA, 2022), que oriente a elaboração de política de indexação para teses e dissertações, durante autoarquivamento para autores e catalogadores, com o uso de vocabulário controlado.

2 Fundamentação teórica

O resultado do tratamento temático de teses e dissertações, em repositórios e catálogos, está contido em metadados de assunto que podem ser analisados de diferentes perspectivas – semântica, sintática, morfológica e pragmática. O tratamento temático refere-se à atribuição de assuntos, mediante linguagem livre ou vocabulário controlado. A atribuição de assuntos nos catálogos de bibliotecas sempre foi realizada por bibliotecários catalogadores, os quais são profissionais especialistas da informação, mas, em repositórios, as teses e dissertações são arquivadas pelos próprios autores, que, no processo de autoarquivamento, preenchem os metadados e atribuem os assuntos, através de palavras-chave em linguagem livre. Em casos de repositórios que fornecem orientação aos autores para uso de vocabulários controlados, essa atribuição poderá incluir termos com controle de vocabulário. Nessa perspectiva, estudos de avaliação de metadados de assuntos são implementados com a finalidade de compreender o processo de tratamento de teses e dissertações e, em específico, o processo de indexação e controle de vocabulário.

Park e Richard (2011) consideram que os repositórios institucionais são um novo canal de suporte, transformação e distribuição de conteúdo acadêmico, e os protocolos de metadados são uma ferramenta valiosa para divulgar esse conteúdo à comunidade acadêmica. Da mesma forma que os padrões para a produção de registros de catalogação tradicional evoluíram, ao longo de décadas de trabalho nacional e internacional, os mais recentes protocolos de metadados, como o *Metadata for Electronic Theses and Dissertations* (ETD-MS), continuam a desenvolver e oferecer aos usuários acesso aos recursos bibliotecários e universitários.

Essa avaliação de metadados descreve como os elementos de metadados de dissertações e teses eletrônicas são usados atualmente em repositórios institucionais canadenses e discute algumas das questões que afetam a padronização da prática, nessa área. Mais instituições-membros da *Canadian Association of Research Libraries* (CARL) adicionaram *Electronic Theses and Dissertations* (ETDs) a seus repositórios, um desenvolvimento que oferece uma oportunidade para pesquisas adicionais, usando um *pool* maior de amostras universitárias do que o examinado. Além de fornecer informações sobre as práticas de metadados, as quais existem na maioria das universidades canadenses, um estudo de acompanhamento também revelaria se houve algum progresso no sentido de padronização adicional, nas dez instituições originais mostradas.

Os resultados obtidos demonstram que os elementos de metadados atuais têm um nível significativo de inconsistência e variação. O resultado desse estudo pode ser benéfico para a obtenção de interoperabilidade entre repositórios institucionais e para o desenvolvimento de um perfil de aplicativo padronizado para repositórios institucionais canadenses.

A análise da literatura revela que pouca pesquisa até o momento foi conduzida, com o objetivo de avaliar especificamente os metadados de assunto em bibliotecas digitais. Estudos disponíveis de metadados de assunto em bibliotecas digitais focalizaram metadados em nível de coleção, que descrevem coleções inteiras de objetos de informação, em oposição a metadados em nível de item, os quais descrevem cada objeto de informação individual. Por exemplo, Zavalina (2011a, 2011b) examinou e comparou os metadados de assunto, em

nível de coleção de texto livre (ou seja, os valores de dados no campo de metadados da Descrição), em várias bibliotecas digitais.

O estudo de acompanhamento (ZAVALINA, 2012) comparou os valores dos dados na descrição detalhada e quatro campos de metadados da disciplina de vocabulário controlado – Assuntos, Cobertura Temporal, Cobertura Geográfica e Tipos de Objeto/Gêneros – em três bibliotecas digitais: *American Memory*, *História de Abertura* e *Biblioteca Europeia*.

Esses dois trabalhos usaram uma análise detalhada de conteúdo manual e se concentraram mais nas características qualitativas dos metadados das disciplinas do que nas quantitativas. Alguns indicadores quantitativos que foram medidos no estudo de Zavalina (2012) incluem o comprimento do valor de dados (medido como o número de caracteres) - intervalo, mediana, média, variância e desvio padrão - de cada um dos cinco campos de metadados do assunto, nos registros.

Na pesquisa de Zavalina (2014), objetivando analisar empiricamente a complementaridade de metadados de assunto no nível de coleção de texto livre, com vocabulário controlado, os resultados obtidos evidenciam um alto nível de complementaridade, que permite uma representação mais completa do conteúdo intelectual dos objetos de informação e, assim, melhora o acesso do assunto aos usuários.

A análise de metadados de assuntos é realizada quantitativamente, em investigação de Tarver *et al* (2015), com um grande conjunto de dados de mais de oito milhões de registros de metadados da *Digital Public of America* (DPLA), uma biblioteca com sistema de agregação de metadados. Adotaram, para isso, abordagem de *big data*, a fim de analisar todo o conjunto de dados, cujos procedimentos implicaram análise de registros de itens individuais com os metadados originais.

O estudo relatado nesse artigo é uma das primeiras tentativas de avaliar sistematicamente os metadados do assunto e o primeiro a usar um agregador muito grande, como a *Digital Public Library of America*, como seu alvo. Os resultados obtidos, após essa avaliação sistemática, revelam que existe alta variação de tipos de preenchimento de campos de assunto entre os registros, ou seja, de nenhum assunto a mais de mil.

Essa análise fornece uma estrutura para discussão geral sobre assuntos em coleções digitais e em grandes agregados. Uma constatação notável é a alta variabilidade do número de instâncias de campos de assunto entre os registros, variando de nenhum assunto a mais de mil.

Embora determinar a precisão e a “qualidade” dos metadados dos assuntos, nesses registros, seja essencialmente impossível, em grande escala, essa análise fornece dados relacionados à integridade, ou seja, se todos os registros têm ou não assunto(s), assumindo que cada registro deve incluir pelo menos um termo de assunto. Ele também destaca os registros de metadados que não se encaixam no modelo de um registro médio, em uma determinada biblioteca digital, e podem ser indicativos de registros de problemas ou metadados de baixa qualidade.

Hanrath e Radio (2017) investigaram o comportamento de pesquisa de usuários de repositórios institucionais (RI), com relação a assuntos, como um meio de estimar o impacto potencial da aplicação de um vocabulário de assuntos controlados a um RI.

A correspondência entre os termos de pesquisa do usuário e os valores do campo de assunto existentes foi relativamente baixa, em menos de cinco por cento das consultas do usuário, os quais correspondem aproximadamente aos valores do campo de assunto. Semelhante à comparação de termos de pesquisa com termos *Faceted Application of Subject Terminology* (FAST), quando as consultas do usuário são divididas manualmente em relação aos valores do campo Assunto, a porcentagem próxima aumenta para quase dez por cento. A taxa de correspondência para Assunto é maior do que Título, mas nenhum dos campos parece corresponder a uma porcentagem particularmente substancial das consultas de pesquisa do usuário examinadas.

A mecânica de realmente aplicar um vocabulário de assunto controlado ao conteúdo de RI pode representar desafios, mesmo quando os gestores de RI decidem que os benefícios esperados valem o esforço. Mas a aplicação retroativa de um vocabulário controlado, bem como o desenho dos mecanismos para sua implementação contínua, pode consumir tempo e recursos consideráveis. O benefício de aplicar um vocabulário controlado aos termos do assunto deve ser pesado em relação aos seus custos, incluindo o custo de

oportunidade de tempo e recursos que podem ser usados para melhorar outros metadados de RI.

3 Metodologia

A investigação sobre o tema controle de vocabulário na representação de teses e dissertações tem como objeto de análise os metadados de assunto em catálogos *on-line* e repositórios universitários administrados por bibliotecas. Esta pesquisa se caracteriza como qualitativa e seu desenvolvimento se dará como pesquisa exploratória.

A esse respeito, Gerhardt e Silveira (2009, p. 31) assinalam: “A pesquisa qualitativa não se preocupa com representatividade numérica, mas, sim, com o aprofundamento da compreensão de um grupo social, de uma organização, etc.” A pesquisa exploratória, no contexto da pesquisa qualitativa, tem o objetivo de aproximar o pesquisador da compreensão do problema de pesquisa, de modo a construir hipóteses. Para isso, a pesquisa exploratória utilizou análise descritiva e comparada de exemplos de procedimentos com metadados de assunto, em catálogos *on-line* e repositórios, para elucidar o problema investigado, a partir de estudo de caso. O estudo de caso é tomado como procedimento para conhecimento em profundidade de como e por que se tem uma determinada situação, de sorte a descobrir o que há nela de mais essencial e característico (FONSECA, 2002¹ *apud* GERHARDT; SILVEIRA, 2009, p. 39).

Nessa perspectiva, o estudo exploratório sobre o uso de controle de vocabulário em metadados de teses e dissertações foi desenvolvido com o Repositório Institucional Unesp e o catálogo *on-line* Athena, da rede de bibliotecas universitárias da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp). Considerando-se que são duas bases de dados com procedimentos e sistemas diferentes, quanto ao controle de vocabulário, foi analisada a possibilidade de interoperabilidade entre as duas bases de dados, de modo a aproveitar o uso de controle de vocabulário do Tesouro Unesp por catalogadores.

Com o objetivo de efetuar análise comparada dos procedimentos de tratamento de teses e dissertações existentes no catálogo *on-line* e no Repositório, para verificação do controle de vocabulário, por meio da análise de

metadados de assunto, o estudo exploratório foi dividido em duas etapas: estudo documental da trajetória histórica de tratamento de teses e dissertações em bases de dados na Unesp e análise de procedimentos de tratamento temático de teses e dissertações nas bases de dados da Unesp, Catálogo Athena e Repositório Institucional Unesp, por autores e catalogadores.

No estudo documental de tratamento de teses e dissertações em bases de dados na Unesp, foram analisados os principais marcos legais institucionais de criação e desenvolvimento. A pesquisa documental foi realizada através de dois tipos de fontes: levantamento, coleta e análise de toda a documentação histórica da implantação da biblioteca digital de teses e dissertações da Unesp, descrita em artigos e trabalhos apresentados em eventos como o Seminário Nacional de Bibliotecas Universitárias (SNBU) e Congresso Brasileiro de Bibliotecas e Documentação (CBBD), bem como de seus manuais, normas, portarias, resoluções e padrões anteriores e atuais.

A análise de procedimentos de tratamento temático de teses e dissertações por autores e catalogadores utilizou processos de observação de funcionamento e verificação de rotinas e padrões de elaboração de registros bibliográficos de teses e dissertações, para avaliação de controle de vocabulário na análise de metadados.

Para a observação de funcionamento do tratamento temático das teses e dissertações, foi feito estudo exploratório na interface de pesquisa integrada do Portal de bases de dados da Rede de Bibliotecas da Unesp, para análise de metadados de teses e dissertações, com base em Fujita e Tolare (2019), fundamentada no estudo de Mondoux e Shiri (2009). Os padrões de organização da informação foram alvo do estudo de observação exploratória, a partir da verificação de rotinas e padrões de elaboração de registros bibliográficos de teses e dissertações, tanto nas bases de dados da rede de bibliotecas da Unesp como no Repositório, com enfoque somente para o campo de assunto. Com os resultados obtidos, foi elaborada proposta de política de indexação para o tratamento temático em metadados de assunto de teses e dissertações, em Repositórios de teses e dissertações e catálogos de bibliotecas.

4 Da biblioteca digital de teses e dissertações eletrônicas ao Repositório: evolução histórica e procedimentos de tratamento temático

Além das diferenças históricas e de funcionamento das bases de dados, catálogo e repositório, cuja delimitação situacional será efetuada pelo estudo documental, é necessário o entendimento do contexto realizado pela análise de procedimentos de tratamento temático de teses e dissertações, condições pelas quais foi organizada a apresentação das duas etapas a seguir.

4.1 Estudo documental da trajetória histórica de tratamento de teses e dissertações em bases de dados na Unesp

Há cerca de 20 anos, a Unesp iniciou a automação de sua base de dados de dissertações e teses e, durante esse período, várias ações e procedimentos foram implantados, até o estabelecimento do Repositório Institucional da Universidade e, mais atualmente, da Política de Acesso Aberto à sua produção intelectual, ações cuja síntese e procedimentos estão apresentados no Quadro 1:

Quadro 1 - Síntese do estudo documental sobre a base de dados eletrônica de teses e dissertações da Unesp

Ano	Trajétória histórica
2000	Criação da Base de dados eletrônica de texto completo de dissertações e teses da Unesp, E-THESES.
2002	Regulamentação para entrega dos originais das dissertações e teses, em formato eletrônico, na biblioteca de cada unidade universitária.
2006	Mudança de nomenclatura da base de dados de dissertações e teses para C@thedra.
2009	Estabelecido procedimento para a autorização de publicação de teses e dissertações, através de um “Formulário de Autorização”, preenchido pelo autor, visando à disponibilização integral ou parcial das teses e dissertações defendidas na Unesp.
2013	Criação do Grupo Gestor da Política do Repositório Institucional Unesp, encarregado do desenvolvimento, implantação e manutenção do repositório institucional da Universidade.
2015	Estabelecidas normas para o autoarquivamento das dissertações e das teses defendidas nos Programas de Pós-Graduação da Unesp, no Repositório Institucional Unesp, tornando o autoarquivamento obrigatório.
2016	Aprovação do Regulamento Interno do Repositório.
2021	Instituição da Política de Acesso Aberto à Produção Intelectual da Unesp.

Fonte: Elaborado pelas autoras.

O Repositório Institucional Unesp tem, como propósito, armazenar, preservar, divulgar e dar acesso à produção científica, acadêmica e administrativa da Universidade. Também deve ter a capacidade de interoperar com sistemas nacionais e internacionais, observando-se o uso de padrões e protocolos de integração no modelo de *Open Archives* (Arquivos Abertos).

Entre os objetivos da Política de Acesso Aberto à Produção Intelectual da Unesp destacam-se o aumento da visibilidade da produção intelectual da Universidade e a consolidação do Repositório Institucional Unesp, como instrumento de gestão para os processos institucionais que necessitem de informações sobre a produção intelectual da Unesp.

O Repositório Institucional Unesp está organizado em comunidades que representam as Unidades universitárias da Unesp, as Unidades complementares e a Reitoria. A criação e a alteração de comunidades e de subcomunidades ocorre em consonância com a criação ou alteração de denominação de Unidades universitárias, Unidades complementares, Unidades auxiliares, Centros interdepartamentais, departamentos, programas de pós-graduação e seções técnico-administrativas, de modo que o Repositório Institucional Unesp reflita a organização da Universidade.

Para ser incluído no Repositório Institucional Unesp, o documento deve atender a vários critérios, entre os quais se ressalta que seja “[...] resultado de atividade científica, acadêmica, artística, técnica ou administrativa realizada nas dependências da Universidade ou em seu nome [...]” (GRUPO GESTOR DO REPOSITÓRIO INSTITUCIONAL UNESP, 2016, p. 2), e o acesso aos metadados e documentos arquivados é livre e gratuito a qualquer indivíduo, não sendo necessário qualquer cadastro ou pagamento, respeitando-se os embargos, quando for o caso.

Os atores envolvidos diretamente nos procedimentos de autoarquivamento são: o autor (pesquisador), a Seção Técnica de Pós-Graduação vinculada ao Programa de Pós-Graduação em que o autor esteve matriculado, a biblioteca da Unidade universitária à qual a Seção Técnica de Pós-Graduação está ligada, o Conselho do Programa de Pós-Graduação a que pertenceu o autor e a Coordenadoria Geral de Bibliotecas (CGB).

No primeiro momento da implantação, a CGB foi responsável pela verificação dos dados informados pelo autor, observando sua conformação ao padrão estabelecido para a descrição de documentos no Repositório Institucional Unesp e realizando, quando necessárias, alterações nos dados, para que ficassem de acordo com esse padrão. Às bibliotecas, nesse momento, competia catalogar a versão impressa no *software* de gerenciamento de bibliotecas, seguindo a

política de catalogação definida pela Rede de Bibliotecas, e incluí-la em seu acervo, em função de suas próprias normas.

A tarefa de catalogação era efetivada em sistema utilizado por toda a rede de bibliotecas, denominado Aleph. O endereço URL do recurso informacional deveria ser incluído no registro bibliográfico, no momento da catalogação (campo 856 da planilha de catalogação, formato MARC21).

Em setembro de 2016, foi implementada a integração entre o Aleph e o Repositório Institucional Unesp. Desde então, um arquivo em lote passou a ser executado diariamente, transferindo para o Aleph os registros coletados do Repositório e convertidos de *Dublin Core* para MARC21, formato adotado para catalogação no Aleph. O sistema também enviava, de forma automática, um *e-mail* informativo para a Seção Técnica de Aquisição e Tratamento da Informação (STATI) da biblioteca ligada à coleção na qual o item foi depositado. O *e-mail* continha uma planilha com a relação dos registros importados. Em seguida, os bibliotecários da STATI faziam a conferência dos registros e os ajustes necessários, seguindo os padrões de qualidade de registros bibliográficos da Unesp para teses e dissertações, de registros de autoridade, e a política de indexação. Não era mais necessário incluir o campo 856 da tese ou dissertação, bastando testar o *link* trazido do Repositório.

Em dezembro de 2017, foi publicado pela CGB o “*Tutorial para Verificação de Dissertações e Teses Submetidas no Repositório Unesp*” (COORDENADORIA GERAL DE BIBLIOTECA, 2017). Esse documento, juntamente com reuniões concretizadas por videoconferências com todas as Unidades, visava a instruir as Bibliotecas sobre como realizariam as verificações das submissões de suas próprias Unidades, a partir daquele momento.

Durante o ano de 2020, procedeu-se à alteração do Sistema Aleph para a versão *on-line* da mesma empresa (*ProQuest Ex Libris*), denominado Alma, e à migração de todas as informações, incluindo os registros bibliográficos, de um sistema para o outro. Considerada uma plataforma de serviços, o Alma permite, entre outras funcionalidades, o armazenamento em nuvem dos metadados de todo o acervo bibliográfico (impresso e eletrônico) e o acesso ao acervo digital.

Com relação às Teses e Dissertações, o Alma, assim como o Aleph, continua sendo uma base independente do Repositório Institucional Unesp, onde

será necessário continuar fazendo a migração dos registros de Teses e Dissertações de um sistema para o outro e as correções, entre outras, dos assuntos, de acordo com o Tesouro Unesp, realizadas pelos bibliotecários da Universidade.

4.2 Análise de procedimentos de tratamento temático de teses e dissertações nas bases de dados da Unesp

A análise de procedimentos de tratamento temático de teses e dissertações foi antecedida pela observação dos recursos de interfaces de busca do usuário e incorporação do Tesouro Unesp das duas bases de dados, cuja análise tomou como base a metodologia de Fujita e Tolare (2019), fundamentada no estudo de Mondoux e Shiri (2009).

A pesquisa integrada do Portal de bases de dados da Rede de Bibliotecas da Unesp, denominado Athena, e o Repositório Institucional Unesp dispõem de várias opções de navegação, conforme quadro abaixo:

Quadro 2 - Opções de navegação baseadas na metodologia de Fujita e Tolare (2019)

Opções de navegação	Catálogo Athena	Repositório Institucional Unesp
Opções de busca simples e avançada	X	X
Busca avançada com pesquisa em campos específicos dos metadados	X	X
Busca em listas de assuntos em ordem alfabética	X	X
Busca por comunidades e coleções	X ²	X
Busca facetada		
Lista de assuntos		X
Tesouro	X	

Fonte: Elaborado pelas autoras.

Ambas as bases de dados das bibliotecas da Rede Unesp possuem opções de busca simples e avançada; busca avançada com pesquisa em campos específicos dos metadados; busca em listas de assuntos em ordem alfabética. A opção de busca facetada, apresentada nas pesquisas das autoras e encontrada em outros repositórios institucionais, não existe no Repositório Institucional Unesp. A lista de assuntos em ordem alfabética é composta por palavras-chave em linguagem natural, sem controle de vocabulário, com problemas de variações terminológicas sintáticas e semânticas e de grafia dos termos. Por outro lado,

somente no catálogo Athena é disponibilizado *link* de acesso ao Tesouro Unesp, com possibilidade de consulta dos termos autorizados para indexação e busca.

A análise do tratamento temático de teses e dissertações teve enfoque no trabalho dos catalogadores com os metadados de assuntos, a partir do autoarquivamento, no Repositório, feito pelos autores, até a disponibilização ao usuário final, no catálogo Athena. As etapas de procedimentos de tratamento temático pelos catalogadores consistem em: (1) Autoarquivamento no Repositório, pelo autor; (2) Análise, formatação e aprovação pela biblioteca de origem, ainda no ambiente do Repositório; (3) Correção ortográfica dos metadados, pela biblioteca de origem, também no ambiente do Repositório; (4) Verificação da catalogação do registro bibliográfico e indexação dos assuntos, na Plataforma Alma, fora do ambiente do Repositório.

(1) Autoarquivamento no Repositório, pelo autor.

A partir da implantação do Repositório Institucional Unesp, os registros bibliográficos das teses e dissertações passaram a ser importados diariamente para o programa gerenciador dos serviços das bibliotecas, que hoje é a Plataforma Alma, na planilha de catalogação, agilizando o processo, visto que, com o registro bibliográfico criado previamente, os bibliotecários catalogadores de cada unidade universitária não precisam mais começar a catalogação descritiva do início.

Com o acesso aberto, foi possível o autoarquivamento pelo próprio autor (discente dos Programas de Pós-Graduação da Universidade), que passa a ser o responsável pela submissão da dissertação ou tese e, por sua vez, a biblioteca de origem, por realizar o recebimento, visualização, aprovação e correção dos metadados.

De acordo com Fujita e Tolare (2019), os repositórios digitais têm cumprido importante função na política do acesso aberto, principalmente em universidades que necessitam prover a comunicação científica, com preservação e acesso a essas publicações digitais:

Em universidades, os repositórios são institucionais e contém toda a produção científica das áreas do conhecimento abrangidas pelas atividades acadêmicas à semelhança de um catálogo bibliográfico online. Essa circunstância é própria do ambiente universitário onde a produção científica abrange diferentes áreas do conhecimento e são especializadas com inovação científica e tecnológica. (FUJITA; TOLARE, 2019, p. 96).

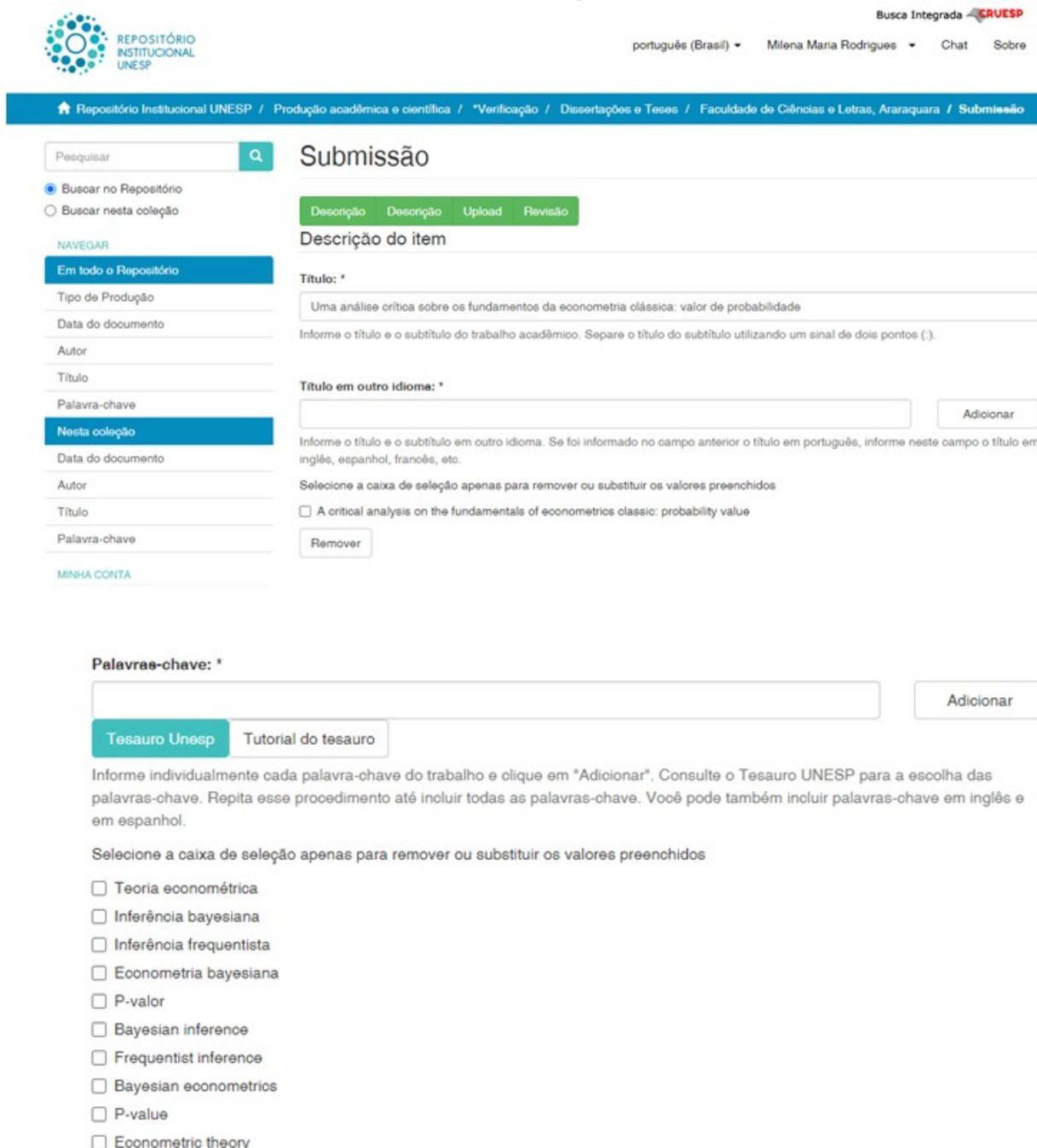
- (2) Análise, formatação e aprovação pela biblioteca de origem, ainda no ambiente do Repositório.

O bibliotecário responsável analisa a formatação e, se estiver conforme os padrões pré-estabelecidos, é providenciada a correção dos metadados e o processo de tratamento descritivo e temático da tese ou dissertação, realizado pelo autor, é validado e aprovado. Caso não esteja correto, o preenchimento de metadados da tese ou dissertação não é validado e é rejeitado pelo catalogador, o qual enviará ao autor as informações a serem corrigidas, devendo uma nova submissão ser feita por ele, para que o bibliotecário possa repetir o processo de aprovação.

- (3) Correção ortográfica dos metadados, pela biblioteca de origem, também no ambiente do Repositório.

A correção de metadados consiste na conferência das informações fornecidas pelo aluno, que ficarão disponibilizadas na página inicial do Repositório. Nessa etapa, os assuntos indicados pelo autor não são verificados pelo bibliotecário, permanecendo na linguagem natural. As Figuras 1 e 2 demonstram como o bibliotecário realiza a correção dos metadados, no Repositório. Observa-se que, nessa etapa, há a opção para consulta ao Tutorial do Tesouro Unesp e ao próprio Tesouro. No entanto, o bibliotecário não promove um controle dos termos indicados pelo autor, apenas suas correções ortográficas.

Figura 1 - Campos de metadados para correção de dissertação no Repositório Institucional Unesp



REPOSITÓRIO INSTITUCIONAL UNESP

Busca Integrada **CRUESP**

português (Brasil) Milena Maria Rodrigues Chat Sobre

Repositório Institucional UNESP / Produção acadêmica e científica / *Verificação / Dissertações e Teses / Faculdade de Ciências e Letras, Araraquara / **Submissão**

Pesquisar

Buscar no Repositório
 Buscar nesta coleção

NAVEGAR

Em todo o Repositório

Tipo de Produção

Data do documento

Autor

Título

Palavra-chave

Nesta coleção

Data do documento

Autor

Título

Palavra-chave

MINHA CONTA

Submissão

Descrição Descrição Upload Revisão

Descrição do item

Título: *

Uma análise crítica sobre os fundamentos da econometria clássica: valor de probabilidade

Informe o título e o subtítulo do trabalho acadêmico. Separe o título do subtítulo utilizando um sinal de dois pontos (:).

Título em outro idioma: *

Informe o título e o subtítulo em outro idioma. Se foi informado no campo anterior o título em português, informe neste campo o título em inglês, espanhol, francês, etc.

Selecione a caixa de seleção apenas para remover ou substituir os valores preenchidos

A critical analysis on the fundamentals of econometrics classic: probability value

Remover

Palavras-chave: *

Tesaurus Unesp Tutorial do tesaurus

Informe individualmente cada palavra-chave do trabalho e clique em "Adicionar". Consulte o Tesauro UNESP para a escolha das palavras-chave. Repita esse procedimento até incluir todas as palavras-chave. Você pode também incluir palavras-chave em inglês e em espanhol.

Selecione a caixa de seleção apenas para remover ou substituir os valores preenchidos

- Teoria econométrica
- Inferência bayesiana
- Inferência frequentista
- Econometria bayesiana
- P-valor
- Bayesian inference
- Frequentist inference
- Bayesian econometrics
- P-value
- Econometric theory

Fonte: Repositório Institucional Unesp (2022).

A Figura 2 demonstra como é visualizada uma dissertação pesquisada por qualquer usuário no Repositório Institucional Unesp, após a correção dos metadados.

Figura 2 - Visualização final da dissertação disponível no Repositório Institucional Unesp

Autor	2021-10-29	This dissertation aimed at critically investigating all econometric assumptions regarding the probability value (p-value) and questioning whether its use should be seen as an absolute methodology within statistical models in economics. To this end, a detailed analysis was made of all the literature on the debate about statistical inference: in particular, about estimations from frequentist and Bayesian statistics, and, at the same time, a parallel with econometric theory was developed. The research initially details the Fisher x Neyman-Pearson discussion, a fact that will be important to understand the main problems of the p-value and how all these methodological obstacles were transported t
Título		
Palavra-chave	Tipo Dissertação de mestrado	
MINHA CONTA		
Sair	Pós-graduação Economia - FCLAR 33004030080P0	
Perfil		
Submissões	Download do arquivo santos_lms_me_arafcl.pdf (700.0Kb)	(...)
CONTEXTO		
Editar coleção	Direito de acesso Acesso aberto	Como citar este documento Disponível em: <>
Mapear itens		Palavra-chave Teoria econométrica Inferência bayesiana Inferência frequentista Econometria bayesiana P-valor Bayesian inference Frequentist inference Bayesian econometrics P-value Econometric theory
Exportar coleção		Idioma Português
Exportar metadados		Financiador Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)
SOBRE O REPOSITÓRIO		
Estatísticas de acesso		Número do financiamento CAPES: Programa de Demanda Social (DS) - Código de Financiamento 001
Acompanhamento de submissões		

Fonte: Repositório Institucional Unesp (2022).

(4) Verificação da catalogação do registro bibliográfico e indexação dos assuntos, na Plataforma Alma, fora do ambiente do Repositório.

No ambiente da Plataforma Alma, os catalogadores seguem a Política de Indexação da rede de bibliotecas da Unesp, conforme Manual (UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO", 2017), e utilizam o Tesouro Unesp para controle de vocabulário da linguagem natural. Na Figura 3, visualizam-se os assuntos na linguagem natural, conforme o depósito original do autor e após a importação automática do Repositório para o Alma (campos 650 da planilha de catalogação, formato MARC21). Podemos observar também o campo 856 do formato MARC21, com o *link* para acesso à tese ou dissertação no Repositório.

Figura 3 - Registro no Alma em Formato MARC21

Uma análise crítica sobre os fundamentos da econometria clássica [recurso eletrônico] : valor de probabilidade / Lucas Mikael da Silva dos Santos. - (999582901306341)

Registro 01 MARC21 Bib

Pearson, fato que será importante para entender os principais problemas do p-valor e como todos esses obstáculos metodológicos foram transportados para a econometria. Após isso, é apresentado de maneira resumida todo o desenvolvimento da teoria econometria e como o p-valor se tornou uma pedra angular para a metodologia tradicional da mesma. Por fim, da mesma forma, também é destacado de maneira sucinta os principais fundamentos e dilemas da teoria bayesiana. Conclui-se que apesar da estatística bayesiana também ter diversos pontos a serem questionados referente a sua utilização, não existe uma justificava plausível para o domínio metodológico da inferência frequentista - fato que é ilustrado pelo uso do p-valor - dentro do mainstream econométrico.

536	\$\$a	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)	\$\$c	CAPES: Programa de Demanda Social (DS) - Código de Financiamento 001
538	\$\$a	Sistema requerido: Adobe Acrobat Reader		
546	\$\$a	Texto em português, resumos em português e inglês		
650	7	\$\$a	Economia	\$\$2 UNESP
650	7	\$\$a	Econometria	\$\$2 UNESP
650	7	\$\$a	Teoria bayesiana de decisão estatística	\$\$2 UNESP
650	7	\$\$a	Estatística matemática	\$\$2 UNESP
650	7	\$\$a	Econometria bayesiana	\$\$2 UNESP
650	0	\$\$a	Econometrics	
695	\$\$a	TM		
696	\$\$a	Mestre		
710	2	\$\$a	Universidade Estadual Paulista (Unesp),	\$\$b Faculdade de Ciências e Letras, Araraquara.
856	4	0	\$\$u	http://hdl.handle.net/11449/215686 \$\$y Disponível em: Repositório Institucional UNESP
940	\$\$a	Economia		
941	\$\$a	Mestre em Economia		
942	\$\$a	Economia	\$\$b	33004030080P0
943	\$\$a	2021/10/29		
945	\$\$a	Métodos quantitativos aplicados e modelo de simulação de agentes		

Fonte: Unesp. Primo.Ex Libris Group (ATHENA, 2022).

Com o registro na base, em formato MARC21, é realizada a verificação e correção dos assuntos descritos pelo autor, na linguagem natural, alterando-se para os termos disponíveis, de acordo com o Tesauro Unesp (campos 650, formato MARC21).

O Quadro 3 evidencia que todos os termos em linguagem natural, indicados pelo autor, sofreram alterações no processo de indexação pelo bibliotecário, seguindo o Tesauro Unesp:

Quadro 3 - Controle de vocabulário da Linguagem Natural com uso do Tesauro Unesp

Linguagem natural	Linguagem controlada
	Economia
Teoria econométrica	Econometria
Inferência bayesiana	Teoria bayesiana de decisão estatística
Inferência frequentista	Estatística matemática
Econometria bayesiana	<i>Econometrics</i>
P-valor	

Fonte: Elaborado pelas autoras.

O bibliotecário indexador, nesse caso, incluiu “Economia” como grande área, analisou os demais termos, descrevendo-os conforme o respectivo termo

autorizado e acrescentou também um termo em inglês, exigido pelo Padrão de Qualidade de Registros Bibliográficos da Unesp: teses e dissertações.

Um exemplo da análise de um termo na linguagem natural alterada, de acordo com o Tesouro Unesp, podemos observar na Figura 4. Ao consultar o termo “Inferência frequentista”, ele apresenta a opção de "Inferência estatística", que se adequa ao tema da tese ou dissertação, mas, no caso, é uma remissiva de "estatística matemática”, de sorte que o termo autorizado representado no registro passou a ser “Estatística matemática”.

Figura 4 - Tela de consulta à base de autoridades da Unesp³ no Alma



The screenshot shows a search interface for the Unesp authority database. The search term is "Inferência frequentista". The results list includes "Inferência estatística (Local)" and "Estatística matemática". A red arrow points to "Inferência estatística (Local)" in the search results, and another red arrow points to "Estatística matemática" in the list of related terms.

Fonte: Unesp.Primo. Ex Libris Group (ATHENA, 2022).

Concluído todo o processo de correção da catalogação e indexação, pelo bibliotecário, as Figuras 5 e 6 demonstram como o usuário visualiza o registro, ao realizar uma busca no Catálogo Athena, mostrando que se trata de um material digital, com o *link* para o Repositório Institucional da Unesp e *links* também para os assuntos, relacionando outros materiais com a mesma descrição disponíveis em todas as bibliotecas da Universidade.

Figura 5 - Visualização do trabalho disponível no Catálogo Athena após indexação dos assuntos



DISSERTAÇÃO
Uma análise crítica sobre os fundamentos da econometria clássica valor de probabilidade
 Santos, Lucas Mikael da Silva dos; Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Letras, Araraquara.
 2021

Disponível on-line >

VOLTAR AO TOPO

ENVIAR PARA

VFR ON-LINE

DETALHES

TALVEZ SE INTE...

LINKS

TAGS

COLEÇÃO

Enviar para



Ver On-line

Disponibilidade de Versões Digitais
 Mostrando 1 de 1 resultados

Fonte: Unesp. Primo. Ex Libris Group (ATHENA, 2022).

Figura 6 - Visualização do trabalho disponível no Catálogo Athena após indexação dos assuntos



Fonte: Unesp. Primo Ex Libris Group (ATHENA, 2022).

Assim, após a tese ou dissertação ser catalogada e indexada com os termos do tesauro Unesp, o usuário conseguirá localizá-la no Catálogo Athena, buscando por assunto, de acordo com o Tesauro Unesp. No entanto, a mesma tese ou dissertação, no Repositório Institucional Unesp, continua sendo disponibilizada somente com a linguagem natural autoarquivada, ficando assim o mesmo documento com diferentes formas de recuperação em sistemas da mesma Instituição.

5 Proposta de política de indexação de teses e dissertações no autoarquivamento no Repositório

As análises apontaram que são realizados dois tratamentos temáticos para teses e dissertações, cujos resultados são diferentes representações de assuntos, em linguagem natural, no repositório, e em linguagem controlada, no catálogo. No repositório, o autor é o catalogador e indexador de sua tese ou dissertação, quando preenche os metadados para a descrição física e temática, ao atribuir palavras-chave da linguagem natural para a representação do conteúdo. No catálogo, o bibliotecário é o catalogador e indexador da tese ou dissertação, quando preenche os campos do registro bibliográfico e atribui os termos do tesauro no campo de assunto, conforme a Política de Indexação da rede de bibliotecas da Unesp, a qual não permite o uso de linguagem natural para o catálogo.

Essa situação divergente e contraditória impossibilita que haja a convivência híbrida das linguagens natural e controlada, tanto no repositório quanto no catálogo. Certamente, propiciaria um enriquecimento terminológico benéfico na representação e na recuperação, haja vista a linguagem natural se beneficiar da precisão e exaustividade do controle de vocabulário, e a linguagem controlada se beneficiar do enriquecimento de vocabulário, com a variação terminológica da linguagem natural, pela inclusão de novos termos, segundo o avanço científico das áreas de conhecimento.

Enquanto o catálogo se beneficia de uma padronização proporcionada pelas políticas de catalogação e de indexação com a adoção do Tesauro Unesp, como vocabulário controlado para a representação e recuperação, o repositório não tem uma política de catalogação e de indexação, mas poderia implantá-la, a

partir da experiência existente junto ao catálogo e se beneficiar da padronização e constante atualização de um vocabulário controlado.

Entretanto, a elaboração de uma política de indexação para o repositório depende, inicialmente, de estudos de avaliação de controle de vocabulário e de compatibilização terminológica com o vocabulário controlado. Enquanto isso, seria recomendável que, na situação específica do controle de vocabulário das teses e dissertações, no repositório, a ser utilizado posteriormente, no catálogo, sejam adotados procedimentos para incluir o uso do Tesouro Unesp para a indexação, a partir da adoção de duas alternativas viáveis:

- a) após autoarquivamento e atribuição de palavras-chave da linguagem natural, pelo autor da tese ou dissertação, o catalogador entra no registro do repositório e faz a atribuição de descritores do Tesouro Unesp, no metadado de assunto, e, depois, exporta para o catálogo, deixando no repositório as palavras-chave e os descritores do Tesouro Unesp, em seus respectivos metadados; ou;
- b) durante o autoarquivamento, o autor atribui palavras-chave da linguagem natural e faz a representação obrigatória por termos do Tesouro Unesp, no metadado de assunto, dentro do Repositório. Depois, o catalogador faz a validação dos descritores do Tesouro Unesp e exporta apenas os descritores selecionados para o catálogo.

Cada uma das alternativas apresenta a solução de combinação híbrida para o repositório entre a linguagem natural, com base na atribuição de palavras-chave dos autores, e a linguagem controlada do Tesouro Unesp. Além disso, as alternativas envolvem a proposta de aproveitamento da indexação realizada no repositório, para uso no catálogo, sem que haja duplicação de esforços para o catalogador e o indexador.

A escolha de qualquer uma das alternativas dependerá da definição de uma política de indexação para o repositório, que inclui o uso do Tesouro Unesp como vocabulário controlado.

6 Conclusões

O estudo exploratório, com análise do Repositório Institucional da Unesp, tomado como estudo de caso, revelou que o controle de vocabulário não é efetuado nas interfaces de busca para recuperação da informação, nem a lista alfabética de palavras-chave atribuídas na produção científica arquivada e autoarquivada passa por procedimentos de controle de vocabulário, o que significa falta de orientações aos autores sobre o uso de vocabulário controlado para atribuição de descritores.

Após comparação dos procedimentos de tratamento de teses e dissertações existentes no catálogo *on-line* e no Repositório, para verificação do controle de vocabulário, por meio da análise de metadados de assunto, as principais discussões até aqui realizadas e a serem evidenciadas como resultados desta investigação são dois problemas enfrentados pelos gestores e bibliotecários responsáveis por repositórios: o primeiro diz respeito a diferentes formatos e procedimentos de tratamento descritivo e temático entre o repositório e o catálogo e consequente duplicação de esforços; o segundo é a falta de definição de uma política de catalogação e indexação para o repositório, capaz de determinar orientações de tratamento descritivo e temático ao autor que faz o autoarquivamento no repositório.

A existência do repositório é importante institucionalmente, pela necessidade de arquivamento e preservação da produção de uma instituição, porém, suas características estruturais para arquivamento ou autoarquivamento são diferentes dos tradicionais sistemas de informação, os quais não têm a possibilidade de, por exemplo, liberar autoarquivamento para um autor ou outra interface que possibilite interação externa.

Dessa forma, os processos de tratamento descritivo e temático, embora possam seguir padronizações vigentes, precisam ser ajustados às características estruturais do repositório – e isso é, sem dúvida, duplicação de trabalho para os bibliotecários, caso seja a mesma equipe. Entretanto, mesmo que exista outra equipe dedicada ao repositório, não seria coerente ignorar que as mesmas teses e dissertações autoarquivadas e processadas no repositório estarão no catálogo com tratamento diferente.

Essas duas situações poderão ser solucionadas pela elaboração de uma política de indexação do repositório que preveja o compartilhamento de dados resultantes do autoarquivamento de teses e dissertações, com o catálogo ou com outro sistema de informação, e que possa aceitar a atribuição de descritores de vocabulário controlado, além da atribuição de palavras-chave.

A política de indexação será determinante para a elaboração de orientações, a fim de os autores e catalogadores efetuarem a atribuição de palavras-chave e de descritores do vocabulário controlado, conforme definições de nível de especificidade e de exaustividade, além de consistência e correção, como, por exemplo, na quantidade mínima de três palavras-chave e descritores, uso de maiúsculas e minúsculas, singular e plural e nível de especificidade.

Como recomendação de providências de curto prazo, seria fundamental a criação de mais um metadado, no repositório, para a inclusão de descritores do Tesouro Unesp, além das palavras-chave. Desse modo, passariam a existir dois metadados: um de palavras-chave e outro de descritores do Tesouro Unesp. A recomendação de médio e longo prazo abrange a inclusão de metadados de assunto, com descritores do Tesouro Unesp para todas as teses e dissertações existentes no repositório.

Essas providências de inserção do controle de vocabulário, na representação de assuntos de teses e dissertações, certamente aumentarão a acessibilidade e recuperação por assuntos e, ao fim, a almejada visibilidade científica, finalidade ansiada pelos autores e gestores do repositório.

Financiamento

Resultado de parte do desenvolvimento de Projeto de Pesquisa financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com Bolsa de Produtividade em Pesquisa Nível 1B (Processo nº 304760/2017-0).

Referências

ATHENA. São Paulo, 2022. Disponível em: <http://athena.biblioteca.unesp.br/>. Acesso em: 09 set. 2022.

COORDENADORIA GERAL DE BIBLIOTECAS. **Tutorial para verificação de dissertações e teses submetidas no Repositório Institucional Unesp.**

Versão 4.4. São Paulo: Coordenadoria Geral de Bibliotecas - Unesp, 2017.

FUJITA, M. S. L. Sistematização de modelo de avaliação do controle de vocabulários em repositórios: relato de pesquisa com o Repositório Institucional Unesp. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, São Paulo, v. 20, n. 00, p. 1-22, 2022.

FUJITA, M. S. L.; TOLARE, J. B. Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. **Informação & Informação**, Londrina, v. 24, p. 93-125, 2019.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

GRUPO GESTOR DO REPOSITÓRIO INSTITUCIONAL UNESP. **Política de Gestão do Repositório Institucional da Unesp**: regulamento interno. São Paulo: Unesp, 2016.

HANRATH, S.; RADIO, E. User search terms and controlled subject vocabularies in an institutional repository. **Library Hi Tech**, Wagon Lane, v. 35, n. 3, p. 360-367, 2017.

MONDOUX, J.; SHIRI, A. Institutional repositories in Canadian post-secondary institutions: user interface features and knowledge organization systems. **Aslib Proceedings: New Information Perspectives**, Wagon Lane, v. 61, n. 5, p. 436-458, 2009.

PARK, E. G; RICHARD, M. Metadata assessment in e-theses and dissertations of Canadian institutional repositories. **The Electronic Library**, Wagon Lane, v. 29, n. 3, p. 394-407, 2011.

REPOSITÓRIO INSTITUCIONAL UNESP. São Paulo, 2022. Disponível em: <http://hdl.handle.net/11449/215686>. Acesso em: 09 set. 2022.

TARVER, H. *et al.* An exploratory analysis of subject metadata in the digital public library of America. *In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS*, 20., 2015, São Paulo. **Proceedings [...]**. São Paulo, 2015, p. 30-40.

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”. Coordenadoria Geral de Bibliotecas. **Manual de política de indexação para as bibliotecas universitárias da Unesp**. São Paulo: Unesp, 2017.

ZAVALINA, O. L. Contextual metadata in digital aggregations: application of collection-level subject metadata and its role in user interactions and information retrieval. **Journal of Library Metadata**, Philadelphia, v. 11, n. 3-4, p. 104-128, 2011a.

ZAVALINA, O. L. Free-text collection-level subject metadata in large-scale digital libraries: a comparative content analysis. *In: DCMI INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS*, 11., The Hague, 2011b. **Proceedings** [...]. The Hague, 2011b. p. 147-157.

ZAVALINA, O. L. Exploring the richness of collection-level subject metadata in three large-scale digital libraries. **International Journal of Metadata, Semantics, and Ontologies**, Denton, v. 7, n. 3, p. 209-221, 2012.

ZAVALINA, O. L. Complementarity in subject metadata in large-scale digital libraries: a comparative analysis. **Cataloging & Classification Quarterly**, Denton, v. 52, n. 1, p. 77-89, 2014.

Thesis and dissertation subject metadata in repositories: an exploratory study on vocabulary control

Abstract: The metadata of theses and dissertations are inserted in the repositories automatically or manually, in Dublin Core format, different from the one used by the libraries, generating duplicity of work. The objective is to compare the procedures of thematic treatment of theses and dissertations existing in the online catalogue and in the repository, through the analysis of subject metadata. An exploratory study was carried out with a case study on the use of vocabulary control in theses and dissertations metadata, with the institutional repository of São Paulo State University and the Athena online catalog of the same institution. The exploratory study had two stages: documentary study of the historical trajectory of treatment of theses and dissertations at university and analysis of procedures of thematic treatment of theses and dissertations in databases, by authors and catalogers. The analysis of thematic treatment procedures employed observation of functioning and of routines and patterns, for evaluation of vocabulary control in the analysis of subject metadata. The results obtained revealed that the vocabulary control is not done in the search interfaces for information retrieval and neither in the alphabetical list of keywords, which means lack of guidance to authors, regarding the assignment of descriptors. The analyses showed two thematic treatments, which results in different subject representations, in natural language, in the repository, and in controlled language, in the catalogue. Two proposals were elaborated for the thematic treatment in subject metadata of theses and dissertations in repositories and library catalogues. It is concluded that the proposals may be assisted by the elaboration of an indexing policy of the repository that foresees the sharing of data resulting from self-archiving of

theses and dissertations with the catalogue and that can accept the attribution of controlled vocabulary descriptors, besides the attribution of keywords.

Keywords: subject metadata; institutional repositories; dissertations; theses; indexing policy

Metadatos de materias de tesis y disertaciones en repositories: estudio exploratorio sobre el control del vocabulario

Resumen: Los metadatos de tesis y disertaciones se insertan en los repositórios de forma automática o manual en formato Dublin Core, diferente al que utilizan las bibliotecas, generando duplicidad de trabajos. El objetivo es realizar un análisis comparativo de los procedimientos de tratamiento temático de tesis y disertaciones existentes en el catálogo en línea y en el Repositorio a través del análisis de los metadatos de las materias. Se realizó un estudio exploratorio con estudio de caso sobre el uso del control de vocabulario en metadatos de tesis y disertaciones con el Repositorio Institucional de la UNESP y el catálogo en línea Athena de la UNESP. El estudio exploratorio tuvo dos etapas: estudio documental de la trayectoria histórica del tratamiento de tesis y disertaciones en la Unesp y análisis de los procedimientos de tratamiento temático de tesis y disertaciones en las bases de datos de la Unesp por autores y catalogadores. El análisis de los procedimientos de tratamiento temático utilizó la observación del funcionamiento y las rutinas y patrones para evaluar el control del vocabulario en el análisis de los metadatos de las materias. Los resultados obtenidos revelaron que el control del vocabulario no se realiza en las interfaces de búsqueda para la recuperación de información, ni en la lista alfabética de palabras clave, lo que significa falta de orientación a los autores en la atribución de descriptores. Los análisis mostraron dos tratamientos temáticos, cuyos resultados son diferentes representaciones de temas en lenguaje natural en el repositorio y en lenguaje controlado en el catálogo. Se realizaron dos propuestas para el tratamiento temático de metadatos de materias de tesis y disertaciones en Repositorios y catálogos de bibliotecas. Se concluye que las propuestas pueden ser ayudadas por la elaboración de una política de indexación del repositorio que prevea el intercambio de datos resultantes del autoarchivo de tesis y disertaciones con el catálogo y que pueda aceptar la atribución de descriptores de vocabulario controlado además de la atribución de palabras -clave.

Palabras clave: metadatos de las materias, repositórios institucionales, disertaciones, tesis, política de indexación

Recebido: 12/04/2022

Aceito: 24/06/2022

Declaração de autoria

Concepção e elaboração do estudo: Mariângela Spotti Lopes Fujita, Rosane Rodrigues de Barros Ribas, Milena Maria Scarpino Rodrigues, Telma Jaqueline Dias Silveira.

Coleta de dados: Rosane Rodrigues de Barros Ribas, Milena Maria Scarpino Rodrigues, Telma Jaqueline Dias Silveira.

Análise e interpretação de dados: Rosane Rodrigues de Barros Ribas, Milena Maria Scarpino Rodrigues, Telma Jaqueline Dias Silveira.

Redação: Rosane Rodrigues de Barros Ribas, Milena Maria Scarpino Rodrigues, Telma Jaqueline Dias Silveira.

Revisão crítica do manuscrito: Mariângela Spotti Lopes Fujita, Rosane Rodrigues de Barros Ribas, Milena Maria Scarpino Rodrigues, Telma Jaqueline Dias Silveira.

Como citar:

FUJITA, Mariângela Spotti Lopes *et al.* Metadados de assuntos de teses e dissertações em repositórios: estudo exploratório sobre controle de vocabulário. **Em Questão**, Porto Alegre, v. 28, n. 4, e-123710, out./dez. 2022. <https://doi.org/10.19132/1808-5245284.123710>



¹ FONSECA, J. J. S. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002. Apostila. *Apud* Gerhardt e Silveira (2009).

² A opção de busca por comunidades e coleções no Catálogo Athena é descrita somente como coleções.

³ A base de autoridades da Unesp no Alma é fonte de alimentação do Tesouro Unesp.