



UNIVERSIDADE
ESTADUAL de LONDRINA

MORGANA CARNEIRO DE ANDRADE

**A INTEROPERABILIDADE SEMÂNTICA NA PERSPECTIVA DA
ORGANIZAÇÃO DO CONHECIMENTO:
UMA PROPOSTA PARA O REPOSITÓRIO INSTITUCIONAL DA
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO**

MORGANA CARNEIRO DE ANDRADE

**A INTEROPERABILIDADE SEMÂNTICA NA PERSPECTIVA DA
ORGANIZAÇÃO DO CONHECIMENTO:
UMA PROPOSTA PARA O REPOSITÓRIO INSTITUCIONAL DA
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão da Informação (MP) do Centro de Educação, Comunicação e Artes, da Universidade Estadual de Londrina, como requisito parcial para obtenção do título de Mestre em Gestão da Informação.

Área de Concentração: Gestão e Organização do Conhecimento.

Linha de Pesquisa: Organização e Representação da Informação e do Conhecimento

Orientação: Dra. Brígida Maria Nogueira Cervantes

Londrina
2012

Dados Internacionais de Catalogação na Publicação

A553i Andrade, Morgana Carneiro de Andrade

A interoperabilidade semântica na perspectiva da organização do conhecimento: uma proposta para o Repositório Institucional da Universidade Federal do Espírito Santo / Morgana Carneiro de Andrade. – Londrina, 2012.
212 f. il. ; 30 cm.

Orientação: Brígida Maria Nogueira Cervantes.
Dissertação (Mestrado em Gestão da Informação) –
Universidade Estadual de Londrina, Programa de Pós-Graduação em Gestão da Informação (Mestrado Profissional), 2012.
Bibliografia: f. 177-194

1. Interoperabilidade semântica. 2. Organização do conhecimento. 3. Repositórios institucionais. 4. Universidade Federal do Espírito Santo. I. Autor. II. Cervantes, Brígida Maria Nogueira. III. Universidade Estadual de Londrina, Programa de Pós-Graduação em Gestão da Informação da Universidade Estadual de Londrina. Programa de Pós-Graduação em Gestão da Informação (Mestrado Profissional). IV. Título.

CDD 025.4

MORGANA CARNEIRO DE ANDRADE

**A INTEROPERABILIDADE SEMÂNTICA NA PERSPECTIVA DA
ORGANIZAÇÃO DO CONHECIMENTO:
UMA PROPOSTA PARA O REPOSITÓRIO INSTITUCIONAL DA
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão da Informação (MP) do Centro de Educação, Comunicação e Artes, da Universidade Estadual de Londrina, como requisito parcial para obtenção do título de Mestre em Gestão da Informação.

BANCA EXAMINADORA

Profa. Dra. Brígida Maria Nogueira
Cervantes
UEL – Londrina - PR

Profa. Dra. Maria Elizabete Catarino
UEL – Londrina - PR

Prof. Dr. Ricardo César Gonçalves Sant’Ana
UNESP – São Paulo - SP

Londrina, ____ de ____ de ____.

AOS

IRMÃOS

E

U

AMIGOS

PARENTES

A

I

CÚMPLICES

Algumas pessoas me fizeram pensar, produzir e me divertir.
Algumas pessoas sempre me fizeram sorrir; outras, sempre chorar; outras, sorrir e
chorar.
Agradeço a todas, pois, se me tornei o que sou, devo a elas.
As que tiverem acesso a este texto saberão o quanto e como contribuíram para me
tornar o que sou, estar onde estou!

Obrigada a todos vocês que um dia fizeram ou fazem parte da minha vida.

Agradeço

À Universidade Federal do Espírito Santo, em especial à Diretora da Biblioteca Central, Rachel Rocha Madeira, por ter possibilitado a realização deste mestrado, por tantos anos adiado.

Ao Programa de Pós-Graduação em Gestão da Informação, modalidade Mestrado Profissional, da Universidade Estadual de Londrina, especialmente à Profa. Dra. Brígida Maria Nogueira Cervantes, por ter me acolhido.

ANDRADE, M. C. **A interoperabilidade semântica na perspectiva da organização do conhecimento**: uma proposta para o Repositório Institucional da Universidade Federal do Espírito Santo. 2012. 212f. Dissertação (Mestrado Profissional em Gestão da Informação) – Programa de Pós-Graduação em Gestão da Informação, Centro de Educação, Comunicação e Artes, Universidade Estadual de Londrina, Londrina, 2012.

RESUMO

Os repositórios institucionais têm-se tornado uma alternativa relevante para as instituições de pesquisa e ensino reunirem o conhecimento produzido em um único local, facilitando a gestão da produção intelectual e a sua socialização. Um dos pré-requisitos para que os repositórios cumpram esse papel é que eles possibilitem a interoperabilidade semântica, não só entre os repositórios como em relação aos outros serviços de informação. O objetivo deste estudo é analisar o elemento de metadado *Subject* do Dublin Core (dc.subject), sob a perspectiva da organização do conhecimento, a fim de fornecer subsídios para que o Repositório Institucional da Universidade Federal do Espírito Santo (UFES) alcance a interoperabilidade semântica, especificamente no nível dos dados categorizados. Como recursos metodológicos, foram adotadas as pesquisas documental e descritiva, com abordagem quali-quantitativa, utilizando o método de estudos de casos múltiplos. A amostra constitui-se de 38 repositórios institucionais contemplados pelos Projeto Piloto e Editais 01/2009 e 02/2010 IBICT/FINEP, dos quais 26 participaram da pesquisa. Dois dos repositórios institucionais não foram localizados; seis não responderam; e quatro justificaram sua não participação em razão de os repositórios se encontrarem em fase de implantação ou mudança de coordenação. Para a coleta de dados, foram utilizados um roteiro de pesquisa, uma planilha, que orientou a identificação das informações registradas nos *sites* dos repositórios selecionados, e um questionário eletrônico EncuestaFacil. Identificou-se que o padrão de metadados usado pela totalidade desses repositórios é o Dublin Core, e que alguns utilizam mais de um. A pesquisa apontou alguns fatores – como a falta de controle de vocabulário na maioria dos repositórios – que poderiam ser evitados com a disponibilização de instrumentos que facilitam a identificação de descritores pelos depositantes –; baixo índice de descritores traduzidos para o inglês; ausência de exposição de vocabulários controlados e de esquemas de classificação. Conclui-se que o panorama atual apresentado pelos repositórios institucionais nacionais necessita de um alinhamento com iniciativas internacionais em prol da interoperabilidade semântica. É necessário que os procedimentos adotados por gestores e pelo pessoal envolvido com os RIs sigam boas práticas em relação aos metadados e à Organização do Conhecimento.

Palavras-chave: Interoperabilidade semântica. Metadados. Repositório institucional. Organização do conhecimento.

ANDRADE, M. C. **The semantic interoperability in the context of knowledge organization**: a proposal for the Federal University of Espírito Santo Institutional Repository. 2012. 212f. Dissertation (Professional Máster in Information Management) – Programa de Pós-Graduação em Gestão da Informação, Centro de Educação, Comunicação e Artes, Universidade Estadual de Londrina, Londrina, 2012.

ABSTRACT

The institutional repositories have become a most considerable alternative for the research and education institutions. They optimize the scientific and technological advance and assemble the generated knowledge in a single spot which facilitates the intellectual production management and its socialization. One of the requirements for the repositories to fulfill their role is to enable the semantic interoperability, not only among repositories, but also concerning other information services. This study aims to analyse the Dublin Core metadata element set – Subject (dc.subject) from the perspective of the knowledge organization in order to provide support for the UFES institutional repository to achieve the semantic interoperability, specifically at the categorical data. The methodology used here was the descriptive and documentary research. The approach was quali-quantitative analysis and the method was the multiple case study. All the 38 institutional repository included in the Projeto Piloto and Editais 01/2009 and 02/2010 IBICT/FINEP were contacted (two of the IR were not located, six did not return the questionnaire and four apologized for not contributing due to administrative matters), so therefore only 26 participated of the research. For the data collection were used a research script which contains aspects that identify the repositories and a spreadsheet that guided the identification of the information recorded on the sites of the selected repositories and EncuestaFacil eletronic query. The pattern of metadata adopted by all those repositories is Dublin Core, and some make use of more then one (pattern). The research point out to some problems such as: the lack of vocabulary control in most repositories that could be avoided by making the instruments to identify the descriptors acessible to people in putting data; small rate of descriptors translated into English; absence of controlled vocabulary or classification schemas. This study demonstrates that current situation presented by the national institutional repositories urgs alignment with international initiatives in favor of semantic interoperability. It is necessary that managers and staff who are involved in the IRs adopt good practices and standard procedures regarding the metadata and the knowledge organization.

Keywords: Semantic interoperability. Metadata. Institutional repository. Knowledge organization.

LISTA DE ABREVIATURAS E SIGLAS

AACR2	–	Anglo-American Cataloguing Rules
AAT	–	Art & Architecture Thesaurus
ANSI	–	American National Standards Institute
APA	–	American Psychological Association
ARROW	–	Australian Research Repositories <i>Online</i> to the World
ASP	–	Academic Search Premier
ASP	–	Academic Search Premier
ASRC	–	American Subject Rules Classification
BASE	–	Bielefeld Academic Search Engine
BC	–	Bibliographic Classification
BD	–	Biblioteca Digital
BDs	–	Bibliotecas Digitais
BDTDs	–	Bibliotecas de Digitais de Teses e Dissertações
BID	–	Banco Interamericano de Desenvolvimento
BIOE	–	Banco Internacional de Objetos Educacionais
CABI	–	CAB Abstract Thesaurus
CALCO	–	Catálogo Legível por Computador
CAPES	–	Coordenação de Aperfeiçoamento Pessoal de Ensino Superior
CC	–	Colon Classification
CC:DA	–	Committee on Cataloging: Description and Access
CCA	–	Centro de Ciências Agrárias
CCO	–	Cataloging Cultural Objects
CCS	–	Centro de Ciências da Saúde
CDD	–	Classificação Decimal de Dewey
CDU	–	Classificação Decimal Universal
CEUNES	–	Centro Universitário Norte do Espírito Santo
CI	–	Ciência da Informação
CIDOC-CRM	–	CIDOC Conceptual Reference Model
CIP	–	Catálogo na Publicação

CLARA	–	Cooperação Latino-Americana de Redes Avançadas
CNRI	–	Corporation for National Research Initiatives
COMUT	–	Programa de Comutação Bibliográfica
CORBA	–	Sistemas Distribuídos Orientados a Objetos
CSA	–	Library and Information Science Abstracts
CSH	–	Canadian Subject Heading (Lista de cabeçalho canadense)
CTI	–	Centro Renato Archer
DACS	–	Describing archives: a content standard
DAI	–	Digital Author Identifier (Identificador Único de Autoria)
DARE	–	Digital Academic Repository
DC	–	Dublin Core
DCAP	–	Dublin Core Application Profile
DCMES	–	Dublin Core Metadata Element Set
DCMI	–	Dublin Core Metadata Initiative
DECS	–	Descritores em Ciências da Saúde
DINI	–	German Initiative for Networked Information
DOI	–	Digital Object Identifier
DRIVER	–	Digital Repository Infrastructure Vision for European Research
ETD	–	Digital Library for Thesis and Dissertations
ETD-MS	–	Electronic Theses and Dissertations Metadata Standard
ETHOS	–	Electronic Theses <i>Online</i> Service
FAPESP	–	Fundação de Amparo à Pesquisa do Estado de São Paulo
FCCN	–	Fundação para a Computação Científica Nacional
FINEP	–	Financiadora de Estudos e Projetos
FRBR	–	Requisitos Funcionais para Registros Bibliográficos
GEMET	–	General Multilingua Environmental Thesaurus
HTML	–	Hyper Text Markup Language
HUCAM	–	Hospital Universitário Antonio Cassiano de Moraes
IBICT	–	Instituto Brasileiro de Informação em Ciência e Tecnologia
IFLA	–	International Federation of Library Association and Institutions
INMETRO	–	Instituto Nacional de Metrologia, Qualidade e Tecnologia
INT	–	Instituto Nacional de Tecnologia

ISAD	– General International Standard Archival Description
ISBD	– International Standard Bibliographic Description
ISBN	– International Standard Book Number
ISKO	– International Society for knowledge Organization
ISNI	– International Standard for Name Identification
ISO	– International Standart Organization
ISO-ILL	– International Standard for Interlibrary Loan
ISSN	– International Standard Serial Number
ISTA	– Information Science & Technology Abstracts
JISC	– Joint Information Science Committee
KOS	– Knowledge Organization System
LC	– Linguagem Controlada
LCC	– Library of Congress Classification (Classificação da Biblioteca do Congresso Americano)
LCSH	– Library Congress Subject Headings
LILACS	– Literatura Latino-Americana e do Caribe em Ciências da Saúde
LISA	– Library and Information Science Abstracts
LOM	– Learning Object Metadata
LSP	– Linked Systems Project
MARC	– Machine Readable Cataloging
MDA	– MDA Archaeological Objects
MeSH	– Medical Subject Headings
METS	– Metadata Encoding and Transmission Standard
MIT	– Massachusetts Institute of Technology
ML	– catalogação digital
MODS	– Metadata Object Description Schema
MPEG	– Moving Picture Experts Group
MSC	– Mathematics Subject Classification
NEDTEC	– Núcleo de Estudos e de Difusão de Tecnologia em Floresta, Recursos Hídricos e Agricultura Sustentável
NIH	– National Institutes of Health
NISO	– National Information Standards Organizations

NKOS	–	Networked Knowledge Organization Systems/Services
NT	–	Termo próximo
OAI	–	Open Archive Initiative
OAI-DC	–	Open Archives Initiative Dublin Core
OAI-ORE	–	Open Archives Initiative Object Reuse and Exchange
OAI-PMH	–	Open Archives Initiative Protocol for Metadata Harvesting
OC	–	Organização do Conhecimento
OCLC	–	On-line Computer Library Center
OCRIS	–	Online Catalogue and Repository Interoperability Study
OI	–	Organização da Informação
OPAC	–	Online Public Access Catalog
OWL	–	Web Ontology Language
PIRUS	–	Publisher and Institutional Repository Usage Statistics
PRIMO_DC	–	Perfil de aplicação do Dublin core utilizado pelo PRIMO
QDC	–	Qualified Dublin Core
RCAAP	–	Repositório Científico de Acesso Aberto de Portugal
RDA	–	Resource Description and Access
RDF	–	Resource Description Framework
RDFS	–	Resource Description Framework Schema
RI	–	Repositório Institucional
RIIs	–	Repositórios Institucionais
RIUFES	–	Repositório Institucional da Universidade Federal do Espírito Santo
RiUNB	–	Repositório Institucional da Universidade de Brasília
RLIN	–	Research Libraries Information Network
ROAMap	–	Registry of Open access Repositories Mandatory Archiving Policies
RVM	–	Repertório de Vedettes-Matière
SARI	–	Serviço de Alojamento de Repositórios Institucionais
SES	–	Syntax Encoding Scheme
SIB	–	Sistema Integrado de Bibliotecas
SIRIS	–	Subject and Institutional Repositories Interactions Study
SKOS	–	Simple Knowledge Organization Schemes
SOC	–	Sistema de Organização do Conhecimento

SPARC	– Scholarly Publishing for Academic Resources Coalition
SRW	– Search Retrieval Web Service
SVEP	– Coordination of e-Publishing within Swedish Higher Education
SWAP	– Scholarly Work Application Profile
SWD	– Keyword Authority Control (Classificação Schlagwortnormdatei)
SWR	– Search Retrieval WebWeb Service
TCCs	– Trabalhos de Conclusão de Curso
TE	– Termo Específico
TEDE	– Sistema de Teses e Dissertações Eletrônicas
TG	– Termo Geral
TI	– Tecnologia da Informação
TR	– Termo Relacionado
UAB	– Universidade Aberta do Brasil
UFES	– Universidade Federal do Espírito Santo
UK-LOM	– United Kingdom Learning Object Metadata
UKOLN	– United Kingdom Office for Library and Information Networking
UNESCO	– Organização das Nações Unidas para a educação, a ciência e a cultura
UP	– Usado Para
URI	– Identificador Único de Recurso
VES	– Vocabulary Encoding Scheme
VR4A	– Padrão de metadado utilizado para descrição de imagens e trabalhos de artes e cultura.
W3C	– World Wide WebWeb Consortium
UK	– United Kingdom
XML	– Extensible Markup Language (Linguagem de Descrição Específica)
XMLS	– Extensible Markup Language Schema

LISTA DE FIGURAS

Figura 1 – Estrutura do repositório institucional sob uma visão cognitiva	47
Figura 2 – Sistemas de Organização do Conhecimento	63
Figura 3 – SOC compartilhados por Hodge (2000), Broughton et al. (2005), Hjørland (2008), e Vickery (2011).....	64
Figura 4 – Estruturas descritivas de metadados.....	100
Figura 5 – Metadados DC e qualificadores utilizados pelo Dspace@MIT	167
Figura 6 – Pesquisa por sistema de classificação	167
Figura 7 – A utilização da LCSH como instrumento de recuperação da informação.....	169
Figura 8 – Formulário para pesquisa avançada.	170

LISTA DE QUADROS

Quadro 1	– Premissas básicas para a Organização do Conhecimento.....	51
Quadro 2	– Abordagens em relação à Organização do Conhecimento.....	58
Quadro 3	– Tipos de ontologia.....	79
Quadro 4	– Síntese comparativa dos SOC	81
Quadro 5	– Princípios que auxiliam na formação de uma coleção digital de boa qualidade.	114
Quadro 6	– Objeto digital adequado	115
Quadro 7	– Qualidade dos metadados	115
Quadro 8	– Exemplos de boas práticas para obtenção da interoperabilidade semântica	132
Quadro 9	– Repositórios institucionais contemplados pelos editais IBICT/FINEP	149
Quadro 10	– Elementos metadados Dublin Core.....	152
Quadro 11	– Esquemas de codificação para descrição temática	153

LISTA DE TABELAS

Tabela 1 – Esquema de metadados utilizado.....	157
Tabela 2 – Quem é o responsável pela administração do RI?	158
Tabela 3 – Quem realiza o depósito?.....	159
Tabela 4 – Que tipo de material pode ser inserido no RI?	161
Tabela 5 – Os metadados são mediados ou controlados pela administração do RI?.....	162
Tabela 6 – Que padrão de conteúdo de metadados é utilizado?	162
Tabela 7 – Qual das afirmações melhor descreve a indexação de assuntos?.....	163
Tabela 8 – A atribuição de palavras-chave	165
Tabela 9 – Que controle de autoridade de assunto é utilizado?	166
Tabela 10 – Há algum instrumento ou procedimento para a descrição dos recursos?	168
Tabela 11 – Os metadados são compartilhados com a biblioteca OPAC?.....	170
Tabela 12 – O RI exporta metadado em OAI-PMH?	171
Tabela 13 – O seu repositório interoperar com algum outro sistema interno?	172
Tabela 14 – Opções de pesquisa no repositório	173

SUMÁRIO

1	INTRODUÇÃO	19
1.1	PROBLEMA	26
1.2	JUSTIFICATIVA	27
1.3	OBJETIVO GERAL	29
1.3.1	Objetivos específicos	30
2	DA UNIVERSIDADE COMO GERADORA E GESTORA DO CONHECIMENTO AOS REPOSITÓRIOS INSTITUCIONAIS	32
2.1	UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO	32
2.2	A PRODUÇÃO DO CONHECIMENTO E O MOVIMENTO DE ACESSO LIVRE	34
3	ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO	49
3.1	SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO	52
3.2	OS SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO E SEUS USOS	84
4	METADADOS DESCRITIVOS	94
4.1	PADRÃO DUBLIN CORE	102
5	A QUESTÃO DA INTEROPERABILIDADE	119
5.1	INTEROPERABILIDADE SEMÂNTICA	124
5.2	OS SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO E A INTEROPERABILIDADE SEMÂNTICA	132
6	METODOLOGIA	143
6.1	PROCEDIMENTO METODOLÓGICO	146
6.1.1	Definição da Amostra	147
6.1.2	Instrumento de Coleta de Dados	150
6.1.2.1	Elaboração do protocolo	150
6.1.2.2	Questionário	154
7	ANÁLISE E DISCUSSÃO DOS RESULTADOS	156
7.1	LIMITAÇÕES DA PESQUISA	174

8	CONSIDERAÇÕES FINAIS	175
8.1	FUTUROS TRABALHOS	178
	REFERÊNCIAS.....	180
	GLOSSÁRIO.....	198
	APÊNDICE.....	202
	APÊNDICE A – Roteiro para pesquisa.....	203
	APÊNDICE B – Planilha para identificação dos instrumentos e elementos de descrição temática dos recursos utilizados pelos repositórios institucionais	204
	APÊNDICE C – Questionário para identificação de serviços dos repositórios institucionais.....	205
	APÊNDICE D – Termo de consentimento livre e esclarecido.....	209
	ANEXO	211
	ANEXO A – Aprovação Comitê de Ética.....	212

Mudam-se os meios, sofisticam-se os instrumentos e surgem nomes novos para designar coisas velhas. Entretanto, a essência das coisas permanece. Os tempos digitais acenam para a concretização de sonhos: da grande biblioteca universal, do conhecimento sem fronteiras, da extensão infinita dos limites da memória, do processamento eletrônico de dados complexos, coisas que, embora profetizadas há anos, décadas, séculos, somente agora encontram as condições para se tornarem realidade (ALVARENGA, 2001, p. 13).

1 INTRODUÇÃO

A influência internacional crescente em todos os países, a complexidade existente no mundo em relação às diversas atividades desenvolvidas, assim como o aumento da quantidade de conhecimento necessário para enfrentar os constantes desafios científicos e tecnológicos levam à parceria internacional, em que o compartilhamento de informações e documentos de forma rápida e eficaz se torna essencial. Os repositórios digitais, na concepção de vários autores, são vistos como uma possibilidade de otimizar esse compartilhamento (COSTA; LEITE, 2009; CROW, 2002; LYNCH, 2003; MARCONDES; SAYÃO, 2009; SWAN, 2011).

Os repositórios digitais são desenvolvidos para desempenhar um papel estratégico, nos ambientes cultural e técnico, de distribuidor de conhecimento em larga escala, que aumenta o impacto de tomada de decisões, criação e transferência de conhecimento. Podem ser vistos, ainda, como uma ferramenta que estimula o talento, como incubador de inovações e como veículos de colaboração. Darby et al. (2008) ilustraram essas potencialidades ao citar o papel que os repositórios institucionais (RIs) representam para órgãos como Research Councils United Kingdom, ao possibilitar a emissão de relatórios de pesquisas desenvolvidas no Reino Unido; para as agências de financiamento como a UCAS Joint Academic Coding System e a Higher Education Statistics Agency, ao permitir mapear seus investimentos em pesquisas, e para o Governments Economic Impact Reporting Frameworks, ao identificar e monitorar repasses do dinheiro público para a pesquisa.

Hoje existe uma tendência para que a informação seja flexível, podendo ser utilizada para diferentes finalidades e baseada na necessidade do usuário; granular, em que o nível de detalhe da informação pode variar conforme a consulta, e fácil de analisar e visualizar. Nesse sentido, o repositório institucional (RI), um dos tipos de repositórios digitais, “[...] torna-se parte de um modelo holístico que contribui para a gestão do conhecimento institucional” (WHITE, 2009, p. 2), um serviço que pode ser compartilhado por todo o mundo, em consonância com o movimento de acesso aberto. Tal possibilidade se reflete na ampliação do acesso à produção científica, gerando uma diminuição dos custos com a informação. Com a implantação dos RIs, ainda é possível realizar atividades que auxiliam em gestão da produção científica institucional; desenvolvimento de um “CitationRank” semelhante

ao algoritmo “PageRank” do Google; registro¹ e seguimento de *downloads*, citações e padrões de uso; avaliação do grau de relacionamento entre investigadores de uma mesma unidade de investigação ou de diferentes unidades; detecção de autores/trabalho não citados/ignorados ou de plágio por meio de análises semânticas e armazenamento de determinados materiais por meios analógicos (*sites*, modelos 3D) (OS INVESTIGADORES..., 2011).

Com esses novos paradigmas, a concepção do RI cumpre o seu papel de reunir, disseminar a informação e socializar o conhecimento. O RI foi definido por Lynch (2003, p. 1) como “[...] um conjunto de serviços que uma universidade oferece aos membros de sua comunidade² para o gerenciamento e disseminação dos materiais digitais gerados pela instituição e por seus membros” (Tradução nossa). Salienta, ainda, que deve existir um compromisso com a organização, a preservação e o acesso e/ou a distribuição desse material.

Pode-se perceber, com base na literatura, que a atividade de pesquisa e seu produto transpõem os muros da universidade, que tem o papel de interagir não apenas com a comunidade científica, como também com a sociedade em geral. A Internet disponibilizada por bibliotecas e serviços de informação pode proporcionar à comunidade o livre acesso à informação, que, em geral, se reflete em liberdade, prosperidade e desenvolvimento dos indivíduos. Esse tipo de acesso é incentivado pela International Federation of Library Association and Institutions – IFLA (2002), quando afirma que não deve haver barreiras que impeçam a circulação das informações, principalmente aquelas que proporcionam igualdade e melhoria de qualidade de vida. Nessa perspectiva, é possível identificar que instituições de ensino que não tinham opção de utilizar o Portal de Periódicos CAPES possuem acesso a publicações científicas que, antes do movimento *Open Access*, só conseguiam com o pagamento às editoras e a outras instituições por serviços como o Programa de Comutação Bibliográfica (Comut). Nesse sentido, ações

¹ São metadados refletidos num formato único. Um registro é recuperado num XML como resposta a um pedido OAI-PMH de metadados de um item. (DRIVER, 2009)

² Ponto de entrada ao material depositado no repositório. Consiste numa *homepage* (ou página inicial) especificamente configurada para a comunidade, com o conjunto das coleções a ela pertencentes e o grupo de usuários a ela integrados, assim como suas informações, notícias e *links*, refletindo os interesses de tal comunidade. Em cada comunidade, podem existir inúmeras coleções, e em cada coleção podem existir inúmeros itens. Essa organização dá ao repositório a flexibilidade necessária para acomodar diferentes necessidades de cada comunidade, permitindo decidir sobre políticas (quem contribui com conteúdo, permissões de acesso, *workflow* e gerenciamento de coleções) (IBICT, 2011).

governamentais vêm possibilitando o acesso à Internet por alunos e professores de escolas públicas em várias regiões do Brasil, tornando o ensino mais dinâmico e auxiliando na interação com outras regiões de realidades distintas, como ocorre com o Projeto de Inclusão Digital desenvolvido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) com comunidades indígenas e catadores de lixo.

Entretanto, no cenário atual, os usuários se deparam, no momento da busca por informações e documentos no ambiente digital, com dois fatores: a heterogeneidade de sistemas e a abundância de informações. Autores como Patel et al. (2005), Shintaku e Brascher (2007) e Boteram (2010) defendem que o acesso a essas informações por única opção de busca proporciona a otimização da recuperação pelos usuários, e é nesse sentido que a interoperabilidade assume um papel de extrema relevância.

Para Crow (2002), a interoperabilidade permite ao pesquisador o acesso a diversos tipos de repositórios, o que facilita a pesquisa interdisciplinar, contribuindo para as abordagens multidisciplinares. A interoperabilidade compreende a identificação persistente de esquemas de metadados padronizados e um protocolo de coleta de metadados – esses descrevem a natureza dos dados armazenados, que incluem conteúdo, estrutura e administração de direitos.

Ainda podem ser incluídos metadados que informam uma área específica, por exemplo, a instituição que financiou a pesquisa. O modelo adotado pelos repositórios visa reunir publicações heterogêneas, que inclui, além dos *pre-prints*, dados de pesquisa, monografias, teses, dissertações, *papers*, literatura cinzenta e patentes. Nesse sistema, onde não será adotado somente um tipo de publicação, os repositórios interoperáveis coexistirão e se complementarão.

Portanto, pode-se deduzir, com base em alguns trabalhos, que a interoperabilidade permite a integração de processos bibliográficos e documentais heterogêneos entre diferentes sistemas de informação – ou seja, que serviços de busca e recuperação de informação direcionados aos usuários são disponibilizados a partir de sistemas que integram *hardwares*, *softwares*, estruturas de dados e interfaces diferentes, de forma mais eficaz (NATIONAL INFORMATION STANDARDS ORGANIZATIONS - NISO, 2004; SAYÃO; MARCONDES, 2008).

A possibilidade de usuários pesquisarem por uma única interface, utilizando uma linguagem comum ao acessar várias bases de dados e coleções ao mesmo tempo, apoiados na interoperabilidade, não é uma questão apenas de

economia, representa uma chance de aumentar o intercâmbio de ideias entre domínios, obtendo resultados de buscas mais exaustivas (MAI, 2003).

Uma organização verdadeiramente interoperável é capaz de intercambiar efetivamente [...] informações com outras organizações igualmente interoperáveis, permitindo que novos conhecimentos possam ser gerados a partir da identificação de relacionamentos entre conjuntos de dados previamente não relacionados. Na perspectiva do usuário, as interfaces devem apresentar para o usuário uma visão unificada em termos semânticos de diferentes recursos informacionais heterogêneos, ou seja: como nomeá-los, como referenciá-los, como utilizá-los em buscas, como acessá-los, como apresentá-los para o usuário (SAYÃO; MARCONDES, 2008, p. 136).

De acordo com Moen (2001), alguns fatores afetam a interoperabilidade:

- a) múltiplos protocolos (Archive Initiative Protocol for Metadata Haversting - OAI-PMH; Z39.50; Search Retrieval Web Service - SWR);
- b) múltiplas e diferentes operações e sistemas de recuperação da informação;
- c) múltiplos esquemas de metadados (semântica dos elementos);
- d) múltiplos formatos de dados;
- e) múltiplos idiomas e padrões de caracteres; e
- f) múltiplos vocabulários, ontologias e disciplinas (semântica dos valores).

Sayão e Marcondes (2008, p. 137-138), ao se referir à interoperabilidade, com base em Miller (2000) e na American National Standards Institute - UKOLN (UKOLN, 2005), acrescentam que a interoperabilidade pode ser dividida nos seguintes níveis:

- a) interoperabilidade técnica – tem por objetivo assegurar o envolvimento de um conjunto de organizações no desenvolvimento de padrões de comunicação, transporte, armazenamento e representação de informações, como Z39.50, Search Retrieval Web Service (SRW), International Standard for Interlibrary Loan

(ISO-ILL) e o Extensible Markup Language (XML). Envolve ainda os padrões individuais voltados para a comunidade;

- b) interoperabilidade semântica – corresponde ao “[...] significado ou semântica das informações originadas de diferentes recursos³ e é solucionada pela adoção de ferramentas comuns ou/e mapeáveis de representação da informação, como esquemas de metadados, classificações, tesouros e, mais recentemente, ontologias”;
- c) interoperabilidade política/humana – está ligada às políticas de informação adotadas, principalmente pelos governos, em relação a acesso livre, exclusão digital e ensino;
- d) interoperabilidade intercomunitária – refere-se à interdisciplinaridade, principalmente nas áreas de pesquisa e de acesso a informações. Utiliza fóruns para discussão e consenso em torno de práticas e procedimentos comuns;
- e) interoperabilidade legal – considera os aspectos legais ao acesso à informação; e
- f) interoperabilidade internacional – está associada a diversidade de padrões e normas internacionais, problemas de comunicação, barreiras linguísticas, “[...] diferenças no estilo de comunicação e falta de uma fundamentação comum”.

Entre os níveis de interoperabilidade mencionados, o presente estudo abordará, especificamente, a interoperabilidade semântica, no que concerne à semântica dos valores, por razões a serem apresentadas no decorrer deste trabalho.

Neste estudo, entende-se como interoperabilidade semântica

³ Qualquer sequência de bits, registrada como produção científica por uma pessoa autorizada (DRIVER, 2009).

[...] la capacidad de los sistemas de información (Bibliotecas Digitales y Repositorios Institucionales), para intercambiar información basándose en un significado común de los **términos y expresiones contenidos en los metadatos y documentos**, con el fin de asegurar la consistencia, representación y recuperación de los contenidos. Esto involucra el uso de esquemas semánticos que incluyen vocabularios controlados (tesauros, listados de encabezamiento de materias, anillos de sinónimos, taxonomías, ontologías, etc.) (GÓMEZ DUEÑAS, 2012, p. 3).⁴

A interoperabilidade semântica apresenta, além do significado que o dado pode ter de forma inequívoca por humanos e programas de computador, a diversidade de idioma; a sintática (diferenças na codificação e na representação da informação); e os esquemas de metadados, que, segundo definição de Sayão (2007, p. 34), são conjuntos de “[...] elementos de metadados projetados para um propósito específico como, por exemplo, descrever um recurso informacional [...]”. Entende-se, então, que os elementos de metadados representam as propriedades dos recursos⁵, e que recurso é o objeto⁶/item⁷. Esses elementos podem ter descrições mais específicas, atribuídas por meio de qualificadores⁸ que adicionam valores. Por sua vez, os qualificadores são termos que, associados ao elemento, fornecem informações adicionais mais precisas em relação ao significado do elemento (BAPTISTA; MACHADO, 2001; SOUZA; ALVES, 2009). Esse nível de interoperabilidade contém as linguagens documentárias utilizadas nas áreas de Ciência da Informação (CI).

De acordo com McCulloch (2004), a prevalência de diferentes terminologias e as limitações causadas aos usuários na recuperação das informações têm levado instituições e pesquisadores a procurar uma solução para esse problema. O uso de diversas terminologias compromete a interoperabilidade semântica entre os diferentes sistemas, o que não é desejável em um ambiente

⁴ “[...] é a capacidade dos sistemas de informação (bibliotecas digitais e repositórios institucionais), para intercambiar informações baseando-se em um significado comum dos termos e expressões contidas nos metadados e documentos, com o objetivo de assegurar a consistência, a representação e a recuperação dos conteúdos. Isto envolve o uso de esquemas semânticos que incluem vocabulários controlados [...]” (Tradução nossa).

⁵ Qualquer sequência de bits, registrada como produção científica por uma pessoa autorizada (DRIVER, 2009).

⁶ Sequência de *bits* digital (DRIVER, 2009).

⁷ É o documento individualmente considerado que compõe uma coleção. Para cada item, deve haver um metadado correspondente (IBICT, 2011).

⁸ Informação que ajuda a definir o conteúdo do elemento do metadado. Os qualificadores identificam o atributo do esquema de valor do recurso (tesauro, sistemas de classificação), fornecendo um elemento de conteúdo de metadado.

onde os usuários buscam obter resultados consultando várias fontes simultaneamente.

McCulloch (2004, p. 297) exemplifica a complexidade do assunto ao citar Hammond (2001), que afirma: “[...] um pesquisador especialista leva um ano para se familiarizar com um vocabulário novo e seu uso” (tradução nossa). O autor, então, questiona: “se isso for verdade”, quais as possibilidades dos usuários diante da diversidade terminológica, dos vários sistemas de organização do conhecimento (SOC), específicos e gerais, e, pior ainda, da utilização de sistemas caseiros ou criados por conta própria, específicos para um propósito, além da utilização da linguagem natural empregada por autores e provedores de conteúdo?

Para Boteram (2010), não apenas a abundância de informação e a heterogeneidade de sistemas utilizados para o acesso aos documentos comprometem a recuperação de informações pelos usuários. Eles também se deparam com questões relacionadas com os diferentes idiomas, vocabulários e esquemas de organização da informação (OI). Segundo Patel et al. (2005), McCulloch (2004), a adoção de princípios da organização do conhecimento (OC) pode proporcionar a obtenção da interoperabilidade semântica, permitindo o gerenciamento eficaz dessas informações. Esclarecendo que, embora seja encontrado na literatura o mesmo sentido para os termos Organização da Informação e Organização do Conhecimento, neste estudo entende-se que a OC está relacionada a

[...] atividades como descrição de documentos, indexação e classificação realizadas em bibliotecas, bases de dados bibliográficas, arquivos e outros tipos de ‘memórias institucionais’ por bibliotecários, arquivistas, especialistas em informação, especialistas em assuntos, assim como por algoritmos e leigos (HJØRLAND, 2008, p. 86).

Mas, quando considerada um campo de estudo, a OC se preocupa com a natureza e a qualidade dos processos e sistemas de organização do conhecimento. Os SOC são adotados na organização e na representação dos documentos (HJØRLAND, 2008).

A necessidade de estabelecer um padrão de interoperabilidade semântica para o RI da Universidade Federal do Espírito Santo (RIUFES), especificamente em relação aos valores do elemento de metadado *Subject*

(Assunto) do Dublin Core (*dc.subject*), norteou a escolha do tema deste estudo, que, inicialmente, utilizou a pesquisa bibliográfica para apresentar e discutir a interoperabilidade semântica, tendo como aporte teórico principal os seguintes autores: Miller (2000a, 2000b), Marcondes e Sayão (2001, 2002, 2009), McCulloch (2004), Patel et al. (2005), Heery et al. (2006a, b), DIGITAL REPOSITORY INFRASTRUCTURE VISION FOR EUROPEAN RESEARCH - DRIVER (2009) e a Organização do Conhecimento (OC), Hodge (2000), Hjørland (2003, 2007, 2008) e Broughton et al. (2005). Em um segundo momento, foi desenvolvido um estudo de casos múltiplos envolvendo RIs contemplados pelo projeto-piloto do IBICT/FINEP e pelos Editais IBICT/FINEP/PCAL/XBDB n. 02/2009 e 03/2009, que totalizam 38 repositórios de instituições de ensino/pesquisa.

1.1 PROBLEMA

Desde a implantação dos primeiros RIs, pressupõe-se que nem todas as instituições adotam os mesmos princípios para a sua concepção. Daí a preocupação com políticas que possibilitem a interoperabilidade semântica, o que permitiria aos RIs a comunicação entre eles, assim como com outros sistemas de informação, bases de dados e mecanismos de busca, como o Google e o Yahoo.

A interoperabilidade semântica é um dos tipos de interoperabilidade mais pesquisado, em vista da sua complexidade. Embora vários projetos, como o Project n.º 507618 DELOS (PATEL et al., 2005), *Online Catalogue and Repository Interoperability Study* - OCRIS project (BIRREL; DUNSIRE; MENZIES, 2009) e o Projeto DRIVER 2.0 (DRIVER, 2009), estejam desenvolvendo estudos e diretrizes para o estabelecimento desse tipo de interoperabilidade, muitos repositórios ainda não utilizam essas recomendações por problemas técnicos, estruturais, culturais ou de adequação. O OAI-PMH, desenvolvido pela Open Archives Initiative (OAI), possibilita o compartilhamento de metadados adotados por repositórios digitais, baseado em duas propriedades:

- a) interoperabilidade – que consiste na obrigatoriedade de uso do padrão Dublin Core; e
- b) extensibilidade – que permite a criação ou a utilização de padrões de metadados diferentes do Dublin Core.

Esse protocolo propicia, ainda, o compartilhamento de dados entre provedores de dados e de serviços (CATARINO, 2009).

Embora o protocolo OAI-PMH proporcione a interoperabilidade técnica, não assegura a interoperabilidade semântica. A ausência dessa interoperabilidade, ou a sua falha, compromete a comunicação entre os sistemas de informação. Até mesmo o surgimento de iniciativas isoladas reflete no que é denominado “ilhas de interoperabilidade”, ou seja, uma comunicação limitada, não permitindo que se relacione a informação de forma globalizada. Desse modo, o propósito da socialização do conhecimento, em âmbitos regional e mundial, fica comprometido (BAPTISTA, 2010).

O crescimento do número de RIs e, conseqüentemente, de depósitos baseados no autoarquivamento⁹, torna-se um fator preocupant, principalmente em relação ao elemento de metadado *Subject* (assunto) e seus valores, quando não ocorre um controle da padronização de terminologias ou vocabulários. Essa falta de padronização tende a prejudicar a consecução da interoperabilidade semântica, não apenas entre os repositórios, mas também com outros sistemas de informação, em razão das inconsistências semânticas – diferentes terminologias e significado – e dos esquemas de metadados. Essa perspectiva levou-nos a um questionamento, o qual este estudo pretende responder: Qual a contribuição da organização do conhecimento para a otimização da interoperabilidade semântica entre os repositórios institucionais?

1.2 JUSTIFICATIVA

O RI apresenta-se como uma perspectiva de reunir, organizar, preservar, divulgar e garantir o acesso confiável e permanente da produção intelectual gerada na universidade em um único local. Assim, implantá-lo constitui uma oportunidade única de nortear uma política formal para a cultura da produção e da disseminação das informações da instituição (LEITE, 2009). Posição respaldada por Crow (2002, p. 4) ao afirmar que os RIs também contribuem para mudanças no sistema de comunicação científica: ao expandirem exponencialmente o acesso à pesquisa; reafirmarem o controle sobre o saber pela academia; servirem como

⁹ Depósito de um documento digital pelo seu autor em Website publicamente acessível (IBICT, 2011).

indicadores tangíveis da qualidade de uma universidade; demonstrarem as relevâncias científica, social e econômica das atividades de pesquisa da instituição; e aumentarem a visibilidade, o *status* e o valor público da instituição.

Contudo, para que a preservação, a divulgação e o acesso a esses documentos sejam possíveis, é necessário que os repositórios possam “conversar” entre os diversos sistemas de informação existentes na Rede. O estabelecimento desse diálogo é viabilizado por vários tipos de interoperabilidade que buscam o desenvolvimento de regras e padrões (BOTERAM, 2010; MAI, 2003; MILLER, 2000a, 2000b; PATEL et al., 2005; ZENG, 2010). Para a realização deste estudo, abordou-se a interoperabilidade semântica em relação à perspectiva da OC, especificamente a gestão do conteúdo viabilizada pela adoção de elementos e instrumentos de representação do conhecimento, como esquemas de metadados e vocabulários controlados (esquemas de classificação, tesouros, ontologias). Eles possibilitam o estabelecimento de sistemas de informação e serviços normalizados, diminuindo o impacto causado pelo uso da linguagem natural e pela ausência de padronização de terminologia (CROW, 2002).

Os primeiros contatos dos serviços de informação com a interoperabilidade semântica no contexto da CI ocorreram por meio dos serviços cooperativos (*On-line Computer Library Center/Catálogo Legível por Computador - OCLC/CALCO*). Com o advento das bibliotecas digitais¹⁰ (BDs), a busca por essa interoperabilidade persiste em praticamente todos os serviços de informação disponibilizados na Internet, como é o caso dos RIs, e se reflete nos serviços de inclusão e recuperação da informação *on-line* (ALIPOUR-HAFEZI et al., 2010).

A necessidade de estabelecer a interoperabilidade semântica é uma preocupação conhecida no ambiente das bibliotecas (HILL et al., 1999; LYNCH; GARCIA-MOLINA, 1995; MARCONDES; SAYÃO, 2001; PAEPCKE et al., 2010), no entanto só se tornou evidente para os gestores de repositórios digitais recentemente (ALIPOUR-HAFEZI et al., 2010; PATEL et al., 2005; SAYÃO; MARCONDES, 2008).

¹⁰ Para melhor compreensão, biblioteca digital neste trabalho é definida como “[...] conjunto de objetos digitais construídos a partir do uso de instrumentos eletrônicos, concebidos com o objetivo de registrar e comunicar pensamentos, ideias, imagens e sons, disponíveis a um contingente ilimitado de pessoas, dispersas onde quer que a plataforma WWW alcance” (ALVARENGA, 2001, p. 3). Considerações em relação a esse tipo de biblioteca são extensivas aos RIs.

Patel et al. (2005) defende que essa interoperabilidade semântica possui três níveis de abstração: estruturas de dados (descrição dos dados); dados categorizados (dados que se referem ao universo, como classificações, tipologias e assuntos gerais) e dados com base em fatos (dados que se referem a particularidades, como autoria e local). A qualidade da interoperabilidade semântica nos níveis defendidos por Patel et al. (2005) proporciona a eficácia dos serviços oferecidos pelos repositórios. As ações para a promoção dessa interoperabilidade em termos globais devem partir de procedimentos locais. É necessária a utilização de regras e padrões definidos, pois, do contrário, os dados agregados nos repositórios estarão isolados (BAPTISTA, 2010; BOTERAM, 2010; CHAN; ZENG, 2002; DRIVER, 2009; FOULONNEAU; ANDRÉ, 2008; GUY, 2006).

No caso específico deste estudo, será analisada a interoperabilidade semântica com foco no nível de dados categorizados, que está relacionado ao significado dos dados representados pelos valores do elemento de metadado¹¹ *Subject* do padrão de metadados Dublin Core. Esse elemento representa o assunto do recurso, expresso por palavras-chave, frases-chave ou códigos de classificação que descrevem o conteúdo intelectual do recurso.

A busca de práticas que potencializem a interoperabilidade semântica em RIs e, especificamente, no RIUFES, é justificável pelo fato de essa ação permitir que os repositórios atinjam um maior fator de impacto dos resultados de pesquisa. Também propicia o reúso da informação e confere maior precisão e relevância aos processos de recuperação da informação, proporcionando a socialização do conhecimento em nível global.

1.3 OBJETIVO GERAL

Analisar o elemento de metadado do Dublin Core *Subject* (*dc.subject*) e os qualificadores utilizados para a organização do conhecimento, visando fornecer subsídios para a potencialização da interoperabilidade semântica de repositórios institucionais.

¹¹ Propriedade de um objeto que é definida em uma especificação. “Autor/*creator*”, Título/*title*” e Autor/*subject*” são propriedades definidas como elementos de metadados.

1.3.1 Objetivos Específicos

Foram estabelecidos os seguintes objetivos:

- a) identificar metadados descritivos para a organização do conhecimento nos repositórios institucionais;
- b) detectar elementos da representação temática que favoreçam a interoperabilidade semântica;
- c) mapear instrumentos da representação temática que propiciem a interoperabilidade semântica; e
- d) fornecer subsídios para estabelecer uma conexão entre os sistemas de organização do conhecimento e a interoperabilidade semântica.

Esta dissertação foi estruturada da seguinte forma: a introdução, que busca situar o leitor sobre os temas tratados no decorrer do trabalho, a identificação do problema e a justificativa para a realização do estudo. O objetivo geral também é apresentado, bem como os objetivos específicos. Em seguida, a segunda Seção contextualiza o papel da universidade como geradora de conhecimento e as possibilidades atuais de disponibilizar a sua produção intelectual por meio dos RIs, sistemas que estão integrados ao movimento de acesso aberto, o qual impulsiona a democratização desse saber. A Seção 3 contempla definições e utilização da organização do conhecimento, por meio dos elementos e instrumentos, para a representação temática. Na quarta Seção, são tratadas questões relativas aos metadados, com foco nos descritivos, procurando estabelecer a conexão entre os sistemas de organização do conhecimento e a interoperabilidade semântica. A Seção 5 aborda aspectos referentes a interoperabilidade semântica, identificação de projetos e ações que propiciem um melhor entendimento do assunto. Na Seção 6, são apresentados os procedimentos metodológicos adotados para o desenvolvimento da pesquisa, que envolveu o estudo de casos múltiplos de repositórios institucionais contemplados pelo Projeto IBICT/FINEP. Os resultados são analisados e discutidos, na Seção 7, buscando uma compreensão em relação ao estágio atual dos RIs nacionais e aos indicativos apontados na literatura para a

otimização da interoperabilidade. Finalizando, são apresentadas as considerações finais que mostram a relevância da utilização dos SOC para a obtenção da interoperabilidade semântica entre RIs.

Nas referências, são elencadas as fontes que possibilitaram uma identificação das relações entre os sistemas de organização do conhecimento, os metadados e a interoperabilidade semântica no contexto dos RIs, assim como subsidiaram a base teórica para o desenvolvimento e a conclusão do estudo. Acrescenta-se, ainda, um glossário para melhor entendimento dos termos utilizados no trabalho, seguido da apresentação de apêndices que ilustram alguns aspectos dos trabalhos, como a metodologia adotada.

2 DA UNIVERSIDADE COMO GERADORA E GESTORA DO CONHECIMENTO AOS REPOSITÓRIOS INSTITUCIONAIS

Essa seção apresenta as universidades como grande geradora da produção científica no contexto do movimento de acesso livre, assim como definições que permitirão a compreensão do trabalho no que tange ao papel dos repositórios, à organização do conhecimento e ao interesse da Universidade Federal do Espírito Santo em participar desse contexto.

2.1 UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

A história da Universidade Federal do Espírito Santo (UFES) está associada ao início do ensino superior no Espírito Santo, na década de 1930, com a criação, pela iniciativa privada, de alguns cursos, como Odontologia, Direito e Educação Física. Durante o governo de Jones dos Santos Neves (1951-1955), esses cursos isolados foram agrupados, e nasceu, no dia 5 de maio de 1954, a Universidade do Espírito Santo, mantida e administrada pelo Governo Estadual (PANTOJA, [1971]; UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2010).

Em 30 de janeiro de 1961, o presidente Juscelino Kubitschek, em um ato administrativo, transformou a instituição em uma universidade federal. Surgia, então, a Universidade Federal do Espírito Santo.

Aos poucos, a UFES foi se expandindo geograficamente e criando outras formas de acesso ao conhecimento (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2010). Na década de 1970, ainda na Capital, construiu um outro *campus*, onde foram instalados o atual Centro de Ciências da Saúde (CCS) e o Hospital Universitário Antonio Cassiano de Moraes (HUCAM), na região de Maruípe. No sul do Estado, na cidade de Alegre, criou o atual Centro de Ciências Agrárias (CCA), no início dos anos 70. Ao norte, em São Mateus, fundou, em 2006, o Centro Universitário Norte do Espírito Santo (CEUNES).

A UFES está organizada administrativamente da seguinte forma: Reitoria, seis Pró-Reitorias, quatro Secretarias, onze Centros de Ensino, três Conselhos Superiores e onze Órgãos Suplementares.

O gerenciamento administrativo da UFES é executado pela Reitoria e operacionalizado por seis Pró-Reitorias (Graduação, Pesquisa e Pós-Graduação,

Administração, Extensão e Planejamento e Desenvolvimento Institucional, Gestão de Pessoas e Assistência Estudantil), além de quatro Secretarias (Assuntos Comunitários, Comunicação, Cultura e Inclusão Social).

A Universidade constitui-se das seguintes unidades de ensino: Centro de Artes; Centro de Ciências Agrárias; Centro de Ciências Exatas; Centro de Ciências Humanas e Naturais; Centro de Ciências Jurídicas e Econômicas; Centro de Ciências da Saúde; Centro de Educação; Centro de Educação Física e Desportos; Centro Tecnológico; Centro Universitário Norte do Espírito Santo e Polo Universitário de São Mateus.

As atividades da Universidade são complementadas ainda pelos seguintes órgãos suplementares: Biblioteca Central; Coordenação de Interiorização; Hospital Universitário Cassiano Antonio Moraes; Instituto de Odontologia; Instituto Tecnológico; Núcleo de Processamento de Dados; Prefeitura Universitária; Rádio e Televisão Universitária; Instituto de Estudos e Educação Ambientais; Instituto de Inovação Tecnológica.

São 74 cursos de graduação, 60 programas de pós-graduação, com 39 cursos em nível de mestrado acadêmico e 7 de mestrado profissional e 14 de doutorado. Além disso, a UFES está integrada ao programa Universidade Aberta do Brasil (UAB) e ao Projeto Pró-Licenciatura, idealizados e executados pelo Ministério da Educação. Atualmente oferece quatro cursos de licenciatura nas áreas de Física, Química, Artes Visuais e Educação Física, e um bacharelado em Ciências Contábeis, totalizando três mil alunos.

Na UFES, a motivação pela criação do RI teve início no ambiente da Biblioteca Central, que, junto a outros órgãos administrativos (Pró-Reitoria de Pesquisa e Pós-Graduação e Núcleo de Processamento de Dados), conjugou esforços para sua implantação.

A Biblioteca Central é órgão suplementar vinculado diretamente à Reitoria e coordena todos os procedimentos técnicos do Sistema Integrado de Bibliotecas (SIB). Administrativamente, as setoriais estão vinculadas aos seus centros de ensino. A Biblioteca Central coordena o Sistema Integrado de Bibliotecas da Universidade Federal do Espírito Santo (SIB/UFES), que conta com sete unidades: Biblioteca Central, Biblioteca Setorial Tecnológica, Biblioteca Setorial Ciências da Saúde, Biblioteca Setorial CEUNES, Biblioteca Setorial Ciências Agrárias, Biblioteca

de Educação e Biblioteca Setorial Nedtec. O acesso ao acervo do SIB é permitido tanto à comunidade universitária quanto à externa.

O SIB/UFES procura se manter em constante processo de atualização para que sejam providas informações atualizadas e adequadas às necessidades das atividades de Ensino, Pesquisa, Extensão e Administração da UFES. Dessa forma, busca oferecer recursos em consonância com o contexto atual de acesso aberto, como é o caso da Biblioteca de Digital de Teses e Dissertações (BDTD) e, mais recentemente, o Repositório Institucional da Universidade Federal do Espírito Santo (RIUFES) aprovado pelo Edital n.º 003/2009 FINEP/IBICT, em novembro de 2010.

O RIUFES utiliza, como os outros 38 repositórios sob auspício do IBICT, o DSpace e o protocolo OAI-PMH. O RIUFES foi implantado em dezembro de 2010, adotando os 15 elementos Dublin Core, assim como os outros repositórios contemplados pelos editais da FINEP/IBICT. Passada a fase de implantação, outros aspectos foram levantados em relação à qualidade do RIUFES e aos serviços oferecidos. Um dos fatores que chamaram a atenção foi a preocupação com a interoperabilidade semântica, especificamente com a questão do controle de vocabulário para a recuperação dos recursos. O interesse em identificar o que poderia ser feito para que o RIUFES alcance essa interoperabilidade, objeto de discussão em várias comunidades, como a europeia, sustenta a realização desta pesquisa.

2.2 A PRODUÇÃO DO CONHECIMENTO E O MOVIMENTO DE ACESSO LIVRE

Os autores Godin e Gingras (2000, p. 277), que visualizam a universidade como protagonista na produção de conhecimento no cenário mundial, demonstram, em seu estudo, que atualmente existe uma diversificação de locais de produção de documentos científicos, mas que a universidade não diminuiu sua participação nesse contexto. Ao contrário, os novos “atores” dos sistemas de produção científica produzem em colaboração com as universidades. De acordo com os autores, “[...] as universidades estão, portanto, mais do que nunca no coração do sistema de produção de conhecimento”.

Para White (2009, p. 1-2),

[...] as universidades são claramente instituições intensas de conhecimento e os pesquisadores são trabalhadores essenciais com altos níveis de conhecimento teórico e do capital social das redes de pesquisa. [...] A tensão entre o pequeno grupo que inova para criar novo conhecimento e a rede de grande escala que facilita a troca mais ampla de conhecimento e o conteúdo do conhecimento está no coração da atividade de pesquisa (Tradução nossa).

Conforme Leite (2007, p. 144, 146), “O ambiente acadêmico, especialmente o das universidades, possui uma estrutura cultural, social e tecnológica constituída bastante favorável às práticas da gestão do conhecimento”, assim como para a comunicação científica, que é responsável por compartilhar o conhecimento científico baseado nas experiências e habilidades dos cientistas.

Leite e Costa (2006) também confirmam que a produção do conhecimento científico ocorre principalmente nas universidades, e que a pesquisa deve ser facilmente acessível e livre, para que seja voltada para o benefício de todos e para o crescimento científico, complementa Alberts (2002, apud FACHIN et al., 2009).

O movimento internacional de acesso livre veio contribuir para a consolidação das instituições de ensino e pesquisa como produtoras e disseminadoras desse conhecimento.

A insatisfação dos investigadores e dos bibliotecários em relação ao controle das editoras sobre a produção científica, principalmente com referência aos custos das assinaturas, e as possibilidades que a Internet disponibilizava contribuíram para a iniciativa dos arquivos de acesso livre.

O surgimento da iniciativa dos arquivos de acesso livre inicialmente se deu com o ArXiv, na década de 90, nos Estados Unidos. Em 2005, o ArXiv possuía em torno de 300.000 documentos, com aproximadamente 42.000 inseridos a cada ano (SWAN, 2005). Esse novo modelo de disponibilização de trabalhos diminui as dificuldades que os autores enfrentam ao divulgar o resultado das suas pesquisas internacionalmente e, com isso, aumenta a sua visibilidade e a estimativa de impacto na comunidade (MARCONDES; SAYÃO, 2009; SUBER, 2007; TOMAÉL; SILVA, 2007).

Vários fatos contribuíram para a consolidação do movimento de acesso livre: a Convenção de Santa Fé, em 1999, a Declaração de Budapest, a adoção do acesso aberto pela Universidade de Southampton, a Declaração de Berlin, a decisão dos pesquisadores de Harvard a favor do livre acesso e a

aprovação pelo governo americano do mandato do acesso livre para o National Institutes of Health (NIH), em 2007 (MARCONDES; SAYÃO, 2009; SUBER, 2007, 2010). Dez anos após esse movimento, estima-se que 15% da literatura científica estejam disponíveis em acesso livre, e o objetivo é que sejam contemplados, anualmente, 2,5 milhões de artigos científicos (KURAMOTO, 2011).

O acesso livre tem como foco principal os artigos de revistas, por serem entregues às editoras (*Copyright*) de forma gratuita, com o objetivo de maximizar o seu uso, entendimento e aplicação de impacto. Quando publicados, seu uso torna-se restritivo e oneroso, não apenas para quem os produz como também para as instituições que se beneficiaram e incentivaram a sua produção (HARNARD, 2007).

A base jurídica do acesso livre é o consentimento do titular dos direitos autorais (para a literatura mais recente) ou a expiração dos direitos de autor (publicações mais antigas, no caso do Brasil, após 70 anos da morte do autor). Uma maneira fácil, eficaz e cada vez mais comum para os titulares dos direitos autorais é a utilização das licenças *Creative Commons*. A licença *Creative Commons* tem por objetivo promover o conteúdo aberto, com base em licenças flexíveis para obras científicas, artísticas e literárias, possibilitando que os autores ou titulares do direito autoral tenham liberdade para liberar a reprodução e até a modificação de suas obras de acordo com os critérios definidos pela licença escolhida. Existem outras licenças de conteúdo aberto que também poderão ser empregadas. Os detentores dos direitos autorais podem, ainda, compor as suas próprias licenças ou declarações de permissão e anexá-las às suas obras, mas sempre assessorados juridicamente (SUBER, 2010; SWAN, 2011).

A proposta de desagregar o atual modelo de publicação não é destruir os sistemas das editoras, mas enfraquecer o seu monopólio nas instituições acadêmicas e nas bibliotecas. O aumento das iniciativas de acesso livre, como os repositórios, reflete no crescimento de competitividade que leva as editoras a manterem o preço das assinaturas mais em conta e participantes de um mercado competitivo (CROW, 2002). Acrescentam-se ainda outros aspectos, como mudanças tecnológicas, aumento significativo no volume de pesquisas, incertezas sobre a preservação e arquivamento dos conteúdos digitais e materiais de pesquisa.

O movimento de acesso livre propõe dois mecanismos para a disponibilização da produção científica, chamados de vias: a via dourada, que é a

disponibilização de artigos em periódicos científicos que adotam o conceito de livre acesso; e a via verde, que consiste no depósito desses trabalhos em RIs por autoarquivamento (HARNARD, 2007; KURAMOTO, 2011; MARCONDES; SAYÃO, 2009). O autoarquivamento é uma atividade complementar em que o autor, após publicar um artigo em periódico e anais de congresso ou defender uma dissertação ou tese, autoarquiva uma cópia no RI. O autor pode depositar um arquivo que usualmente é uma versão final (*post-print*), depois da revisão por pares, ou o texto antes dessa revisão ser completada (*pre-print*). Atualmente, muitas revistas permitem que os autores depositem a versão final do trabalho imediatamente, sem ressalvas, ou após algum tempo de sua publicação (SWAN, 2005).

Na perspectiva da via verde, Bailey Jr. (2005) afirmou que os RIs, além de preservar e possibilitar o acesso a esse material, transmitem confiabilidade ao processo, por ser um sistema reconhecido e com custo reduzido. O aumento da visibilidade dos resultados de pesquisa, do pesquisador e da instituição é apontado por Swan (2011) como a principal vantagem dos RIs.

Torino, Torino e Silva (2009) destacaram a utilização dos protocolos de coleta de metadados OAI-PMH, baseados em padrões nacionais e internacionais de interoperabilidade, como um dos fatores essenciais para a consolidação dessa filosofia. Após o estabelecimento desse protocolo, vieram as tecnologias de *softwares* livres para criar arquivos em conformidade com os padrões dos repositórios (HARNARD, 2007).

O movimento de acesso aberto proporcionou um aumento no número de repositórios digitais no mundo (2.205, em agosto de 2012), principalmente dos institucionais (1.811) (OPEN DOAR, 2012), como pode ser observado no Gráfico 1.

Gráfico 1 – Crescimento dos repositórios no mundo

Fonte: OpenDoar (2012).

Cada tipo de repositório digital¹² contempla um sistema de comunicação científica ou de produção intelectual:

- repositórios disciplinares ou temáticos – voltados à produção intelectual de áreas de conhecimento específicas;
- repositórios de teses e dissertações – destinados a bibliotecas digitais de teses e dissertações (BDTDs); e
- repositórios institucionais – direcionados à produção intelectual, geralmente, de uma instituição de ensino e/ou de pesquisa (COSTA; LEITE, 2009).

Independentemente do tipo de repositório, é imprescindível a participação de pesquisadores no processo de autoarquivamento (LYNCH; LIPPINCOTT, 2005; SWAN, 2005). Das 999 editoras registradas no *site* SHERPA/ROMEO, 64% permitem algum tipo de autoarquivamento – políticas de *Copyright* e de autoarquivo de editores (SHERPA ROMEO, 2010). De acordo com Kuramoto (2011), a América do Norte e a Europa são responsáveis por 81% dos mandatos registrados no Registry of *Open access* Repositories Mandatory Archiving Policies (ROARmap), enquanto a América do Sul e a África possuem apenas 3% desses mandatos, o que o autor identifica como contraditório, em razão do custo do

¹² É um sistema de computador usado para armazenar coleções de uma biblioteca digital e disseminá-las aos usuários (ARMS apud KURAMOTO, 2009).

acesso à informação científica que, mesmo nesses países mais desenvolvidos, ainda é muito alto.

Em relação ao RI, o termo foi cunhado pela Scholarly Publishing for Academic Resorces Coalition (SPARC) como “[...] coleções digitais de captura e preservação da produção intelectual de uma única comunidade ou multi-universidade” (CROW, 2002, p. 2, tradução nossa).

De acordo com o SPARC, o principal foco dos RIs é o armazenamento da produção científica de uma instituição. Entretanto, para Crow (2002), o RI pode armazenar outros tipos de conteúdos produzidos, incluindo material didático, relatórios institucionais, vídeos, programas de computador, dados, fotografias e materiais digitais que as instituições gostariam de preservar.

De acordo com Tomaél e Silva (2007, p. 3), “[...] o que caracteriza os RIs é o fato de serem orientados para a informação produzida no ambiente das instituições, sendo desenvolvidos, implementados e mantidos por elas”. Posição ratificada pela definição de Crow (2002, p. 2),

Repositórios institucionais são coleções digitais que captam e preservam o resultado intelectual das comunidades universitárias – respondem a duas questões estratégicas frente às instituições acadêmicas: 1) elas proporcionam um componente central na reforma da comunicação erudita por estimular a inovação numa estrutura de publicação desagregada; e 2) elas servem como indicadores tangíveis da qualidade de uma instituição, aumentando assim sua visibilidade, seu prestígio e o valor público (Tradução nossa).

Verdadeiros tesouros históricos pertencentes a universidades, museus, bibliotecas e organizações, antes indisponíveis para pesquisadores e para o público, atualmente podem ser disponibilizados pelos repositórios. Os RIs “[...] estão sendo criados para armazenar, preservar e manter os recursos digitais, permitindo uma recuperação eficaz; da mesma forma que serve de arquivo para a história das instituições”. Pelo fato de o “[...] conhecimento da organização ser um recurso imprescindível nas organizações, gerenciar esse recurso é assegurar o conhecimento do que já foi produzido e as novas aplicações que se criam” (DRAKE, 2004, p. 1, tradução nossa).

Dodebei (2009, p. 104) enriquece a fala de Drake (2004), ao afirmar que os RIs, pela sua condição de facilitar o acesso e disseminar a informação, assim

como a preservação de coleções, são vistos como “[...] modelos híbridos de patrimônio da humanidade que não disputam espaço com outros arquivos e bibliotecas digitais, apenas somam em qualidade de oferta”.

O fato dos repositórios serem “[...] institucionalmente definido; científica e academicamente orientado; cumulativo e perpétuo; aberto e interoperável” (CROW, 2002, p. 2, tradução nossa) possibilitam: alternativas de gestão de informação científica; preservação da produção intelectual; garantia de acesso amplo e irrestrito e interoperabilidade, caracteriza-os como uma ferramenta capaz de tornar acessível e de disseminar o capital intelectual de uma instituição (COSTA; LEITE, 2009, p. 163).

Esse contexto propicia ao usuário melhor acesso às pesquisas, às tendências históricas e conceituais entre artigos podem ser reveladas e a avaliação da qualidade da pesquisa e da produtividade poderá ser mais bem aferida (CROW, 2002; DARBY et al., 2008). Novos padrões e técnicas de avaliação serão desenvolvidos para analisar mais profundamente o *corpus* de pesquisa disponibilizada em acesso aberto, como relatado na Introdução deste trabalho. Crow (2002) ainda visualiza que *links* de citações abertas e sofisticadas análises retrospectivas permitiriam a criação de sumários que poderiam identificar mais eficientemente a trajetória de um determinado conceito ou sujeito da pesquisa, ou até mapear tendências de novas pesquisas.

Nesse processo, os bibliotecários possuem o papel relevante de contribuir ativamente para uma mudança no processo de comunicação científica. As bibliotecas que armazenam, organizam e disponibilizam a produção científica aumentarão a interação com os pesquisadores, à medida que devem auxiliá-los no controle dos formatos e padrões de arquivamento. Devem contribuir ainda para definição dos metadados, controle de autoridades e de conteúdo, requisitos necessários para o aumento da interoperabilidade.

O Projeto MILAGRE (fazendo dos Repositórios Institucionais um ambiente de aprendizagem colaborativo) investigou as questões e as tendências relacionadas com o planejamento e a implementação de RI, como as suas finalidades, os papéis e também os critérios de seleção do sistema, os recursos do sistema, as políticas, serviços e sustentabilidade. Os dados foram coletados em entrevistas por telefone com 36 funcionários dos RIs e as pessoas que desempenharam papéis importantes no desenvolvimento de RIs em suas

respectivas instituições. Os entrevistados não foram apenas de instituições acadêmicas onde um RI foi implementado, mas também participaram aqueles que planejaram, fizeram teste piloto, ou aqueles que têm planos para lançar um RI. Independentemente da fase do desenvolvimento dos RIs, os entrevistados manifestaram o seu entusiasmo de forma consistente para a implantação de repositórios e preparo para o acesso e preservação do material digital. Os resultados indicam que a visão pessoal de RI, como infraestrutura para a participação da universidade no movimento de acesso aberto, está confiante sobre a sustentabilidade em longo prazo. As possibilidades dos RIs têm evoluído ao longo dos anos. Os RIs estão cada vez mais percebidos em termos de objetivos globais, e não como um conjunto de funções, recursos ou serviços específicos. Com base nesses resultados, sugere-se que os RIs busquem modelos de serviços que possam permitir maior possibilidade de servir a comunidade, contribuindo com a aprendizagem dos indivíduos (RIEH et al., 2008).

Melero et al. (2009) elaboraram um relatório sobre a situação dos RIs na Espanha, nos anos de 2008 e 2009. Foi realizada uma pesquisa, na Internet, com 104 instituições; 25 instituições que já possuíam repositórios e treze que não o tinham responderam aos questionários. O resultado mostra que os repositórios contêm, em sua maioria, metadados e texto completo de artigos e teses. As políticas institucionais sobre autoarquivo e serviços criados em repositórios superiores foram relatadas pelos gestores dos repositórios como as mais relevantes, embora a maioria dos materiais depositados tenha sido feito pelo pessoal administrativo. O aumento da visibilidade e citação, o interesse pela administração da instituição e a simplicidade do uso e dos serviços de pesquisa foram identificados como estimuladores para o depósito em repositório. Em contrapartida, foram apontados, como os aspectos mais inibidores, a ausência de políticas, a falta de integração com outros sistemas nacionais e internacionais e a falta de esforços de conscientização entre a comunidade acadêmica.

Darby et al. (2008), membros do Information Systems Committee (JISC), realizaram o Subject and Institutional Repositories Interactions Study (SIRIS), um estudo em relação aos sujeitos – pesquisadores, organizações financiadoras, instituições, produtores – e os RIs. O estudo buscou estabelecer recomendações para melhorar as interações entre os RIs no Reino Unido, já que, para os autores, o conceito de repositórios representa uma rede de relações com outros repositórios e

fontes de informação, pois os autores dos trabalhos podem estar relacionados com outras instituições por meio de grupos de pesquisas e publicações.

Uma pesquisa *on-line* foi realizada com 125 administradores de repositórios, utilizando questionário, que contabilizou um retorno de 25% (32 responderam). Foram realizadas ainda 30 entrevistas com participantes de órgãos financiadores, gestores de repositórios e peritos. Os resultados das entrevistas e do levantamento foram analisados com relação à literatura científica publicada nos repositórios, aos objetivos desejados e opiniões das organizações das partes interessadas, e aos projetos de relevância para as interações dos RIs.

Foram identificados três tipos de repositórios: institucional, repositório temático e o financiador. O RI é tido como o mais prolífico, menos maduro e menos bem-equipado. Com base nas respostas dos questionários, constatou-se uma preocupação, por parte dos gestores, em relação à qualidade, constatando que 83% dos entrevistados realizam controle de qualidade pós-depósito. O percentual de depósito feito pelo autor é de 25%, 46% mediados por equipe do repositório e 18% por transferência em massa de outras fontes.

Como o foco da pesquisa era a interação entre RIs, foram identificados alguns fatores que influenciam essa interação:

- a) desejo de um padrão mínimo de elementos de metadados. Metadados que não possuem identificadores fundamentais, como afiliação institucional clara, identificação explícita da versão e informações sobre licenciamento, têm valor limitante. Um dos problemas identificados na interação entre os repositórios diz respeito à autoridade e à coerência da identificação dos autores. Problema que compromete também a função administrativa do repositório. Essa função pode ser vista como um importante condutor na padronização e controle de qualidade dos metadados, dada a importância de identificação inequívoca de autor e classificação de código que mapeia as comunidades;
- b) os metadados descritivos – classificação de assunto e declaração de versão adicionada ao conjunto básico necessário para a identificação elementar –, tornam o conteúdo do

repositório pesquisável e utilizável. Existe uma tendência de ser, no mínimo, funcional, levando a crer que a probabilidade de riqueza de especificações de metadados mais sofisticados, tais como o Scholarly Work Application Profile (SWAP), não incorra no mesmo problema do Machine Readable Cataloging (MARC) – pouco explorado ou utilizado de maneira ineficaz.

O SWAP é um perfil de aplicação desenvolvido para solucionar problemas de metadados identificados no Eprints utilizado por repositórios do Reino Unido. Um perfil de aplicação representa as especificações de como um esquema de metadados são combinados e adaptados para descrever um determinado conjunto de recursos, explicitando quais e como os elementos são utilizados na descrição (NISO, 2007). Enfim, o SWAP busca enriquecer os metadados em função do serviço de busca para agregar conteúdo aos repositórios digitais ao seu modelo de domínio e tem como base os Requisitos Funcionais para Registros Bibliográficos (FRBR), que tem por objetivo produzir descrição detalhada dos registros bibliográficos com o intuito de suprir as necessidades dos usuários.

Darby et al. (2008) esclarecem que é importante a mudança de percepção de uma visão centrada no repositório-centralizador, baseada na propriedade da informação, para uma visão sistêmica, fundamentada na criação de mapas ou em caminhos para a informação. Iniciativas como a *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE) propõem um modelo que, em vez de duplicar e transportar dados, cria visualizações agregadas das informações mantidas em locais diferentes da *Web* (identificador único de recursos - URIs). A ORE trata a *Web*

[...] como um único e enorme banco de dados de informações, capaz de sustentar uma variedade e complexidade de inter-relações. Alguns desenvolvedores de *software* afirmam que o esforço deve ser direcionado não na criação de soluções para o transporte de conteúdos entre repositórios, mas a criação de serviços de sobreposição que irão fornecer *gateways* semanticamente ricos pra o conteúdo distribuído através da *Web* (DARBY et al., 2008, p. 128, tradução nossa).

Os autores estabelecem quatro diretivas principais para a interação entre repositórios:

- a) população dos repositórios – processo de depósito simples, integrado aos fluxos de trabalho do autor; avaliação da pesquisa nacional e relatório de demandas; maior uso do protocolo OAI-PMH para permitir a transferência de conteúdo; utilização do SWAP (referência descritiva de trabalhos acadêmicos) para fornecer a equivalência semântica para a troca de dados; integração do depósito com fluxos de trabalho dos editores;
- b) coleta e análise de medidas – participação em conselhos de pesquisa para atender às exigências dos indicadores de impacto econômico; participação do projeto Publisher and Institutional Repository Usage Statistics (PIRUS);
- c) preservação – garantia de utilização continuada do conteúdo; aplicação da International Organization for Standardization (ISO) em Auditoria; e certificação de repositório digital; e
- d) agregação dos resultados de pesquisa – implementação de repositórios de dados e tecnologia de vinculação (projeto Citation, entre outros); modelo comum para a descrição; e intercâmbio de totalizações de fontes da Web.

Por fim, os autores recomendam que sejam aplicadas normas para a identificação clara dos autores, financiadores e instituições de ensino superior, adoção de padrões de intercâmbio de informações e confiança em outros repositórios.

Carvalho (2009) desenvolveu uma pesquisa para verificar a percepção das bibliotecas universitárias em relação à implantação de repositórios no Brasil. Os aspectos analisados foram: liderança, depositantes, conteúdos, sistemas e financiamento dos RIs. O estudo foi exploratório, com abordagem quali-quantitativa, em que se empregou um questionário com 17 questões, fechadas e abertas. A autora salientou alguns problemas que ocorreram na pesquisa, como o tempo limitado para a investigação e análise de dados por extrato. Houve um número reduzido de questionários respondidos (amostra 141 e 20 respondidos) e grande número de questões sem repostas. A autora identificou como causa o provável desconhecimento do assunto pelos profissionais, já que apenas 11 repositórios tinham sido implantados e 3 encontravam-se em fase de implantação.

Em relação à liderança, identificou-se que a biblioteca universitária tem participação na coordenação das iniciativas de RI. Às vezes essa coordenação é compartilhada com outros setores. Os docentes, pesquisadores e alunos de mestrado e doutorado são os principais depositantes. A indicação de funcionários, segundo a autora, como depositante, pode representar que o arquivamento tenha sido feito por pessoal técnico. O número de depósito ainda se encontra muito baixo. Inclui dissertações/teses, anais, trabalhos de congressos, trabalhos de conclusão de curso (TCCs) e fotografias. Os dois sistemas mais utilizados foram o DSpace e o sistema de Teses e Dissertações Eletrônicas (TEDE). Carvalho (2009) relaciona esse resultado com os de outros autores, em que o *software* Virginia Tech Digital Library for Thesis and Dissertations (ETD), que originou o TEDE, é o menos utilizado, o que sugere que as dissertações e teses estão sendo depositadas em repositórios de conteúdos mais abrangentes e não em bibliotecas digitais.

Dos sistemas utilizados para gerenciar a biblioteca, Pergamun, Aleph e sistemas desenvolvidos pelas próprias universidades, apenas o Aleph possui protocolo OAI-PMH. Os demais não fazem coleta automática dos metadados. Já o Nou-Rou (plataforma para bibliotecas digitais - BD), desenvolvido para o arquivamento de vários tipos de publicação, utiliza o OAI-PMH, proporcionando a interoperabilidade com outros sistemas. As fontes de financiamento são recursos provenientes de fontes externas, como agências de financiamento (Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP, Financiadora de Estudos e Projetos – FINEP e IBICT).

Carvalho (2009, p. 9) concluiu que a implantação dos RIs no Brasil deve evitar o retrabalho, utilizando uma customização padrão com metadados que contemplem os diversos tipos de documentos, e que a busca da interoperabilidade deve ser perseguida a fim de aproveitar os “[...] esforços e recursos na organização da produção científica nacional de acesso aberto”.

Uma outra vertente em relação aos RIs foi explorada pelas autoras Rosa e Gomes (2010), que desenvolveram um estudo com o RepositóriUM, da Universidade do Minho, em que buscaram identificar aspectos em relação aos depositantes e aos públicos usuários dos RIs. Foram identificados como usuários do RI três grupos principais: gestores, autores e usuários. Este último foi pouquíssimo abordado pela literatura e, como as próprias autoras identificaram no trabalho de

McKay (2007), chega a ser, se não curioso, um contra-senso, já que o objetivo final dos repositórios são os usuários finais.

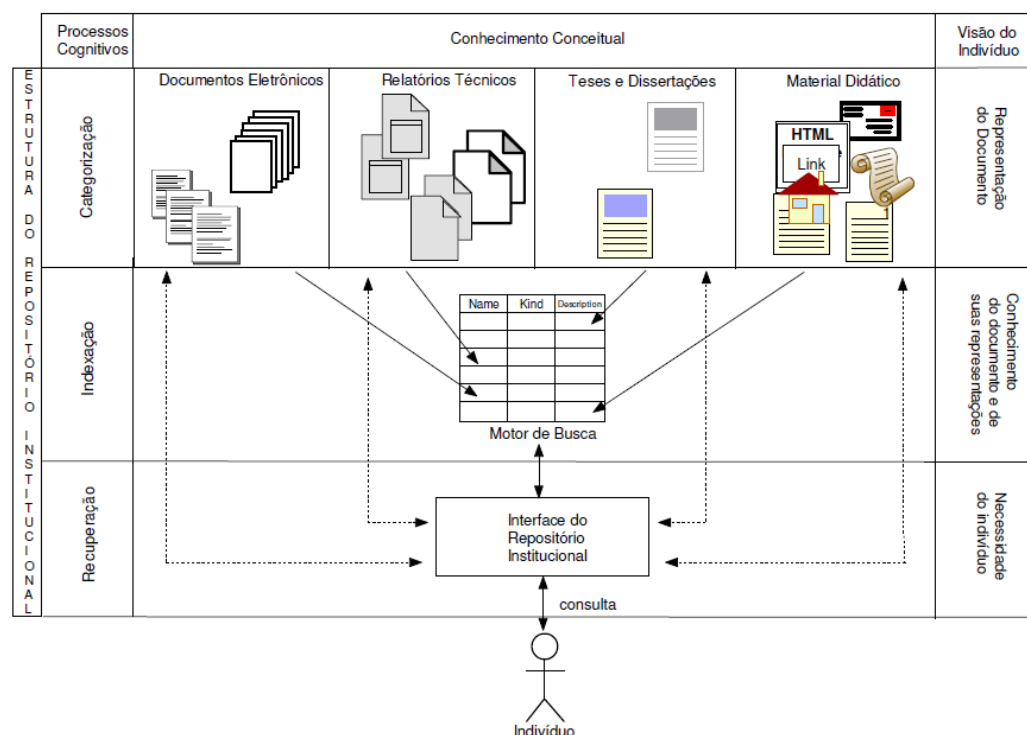
Para cada grupo, utilizou-se um instrumento de coleta de dados diferente, disponibilizado *on-line* pelo sistema SurveyMonkey. Para os usuários finais, internos ou externos à Universidade, foram analisados aspectos biográficos (idade, escolaridade e vínculo institucional) e os aspectos relativos ao uso do RI (frequência de acesso, motivo do uso, apreciação da qualidade dos recursos e conhecimento sobre o movimento de acesso aberto).

Identificou-se que, do total de 170 usuários respondentes, 160 são portugueses e 10 são brasileiros, com idade entre 20 e 50 anos; 50 são estudantes de pós-graduação e mestrado, vinculados à Universidade, 41 são estudantes de graduação e 54 não possuem vínculo. A maioria não sente dificuldade para realizar as pesquisas (138), 87 indicam um bom índice de relevância e qualidade de recursos; e 105 desconhecem o movimento de acesso livre. Rosa e Gomes (2010), com base no índice de satisfação sobre a obtenção de recursos no repositório, apontaram que as questões de pesquisa atendem às sugestões feitas por Fachin et al. (2009) em relação à visão cognitiva na estruturação de repositórios.

O estudo realizado por Fachin et al. (2009, p. 230) analisou a relação da gestão do conhecimento com a visão cognitiva na construção dos RIs. As autoras identificaram que, como em qualquer sistema de informação, nos RIs, não é apenas a recuperação que envolve aspectos cognitivos, mas também outros processos, como o da indexação, que é a forma como esse documento é compreendido e representado. Segundo essas autoras, os processos de indexação está “[...] ligado à abordagem do processamento de informação na psicologia cognitiva”.

Justifica-se, dessa forma, a utilização da visão cognitiva para a estruturação desses RIs, pela sua heterogeneidade de informações e de usuários. As autoras ratificam com a citação de Neves (2006, p. 41), “[...] a compreensão textual, a geração de texto (tradução para a linguagem de indexação e/ou resumo) e a representação do conteúdo (criação de linguagens de indexação) devem ser evidenciadas”.

Com base em alguns estudos, Fachin et al. (2009) desenvolveram o esquema de estruturação de RI, visando atender aos processos cognitivos dos usuários no uso desse sistema (Figura 1).

Figura 1 – Estrutura do repositório institucional sob uma visão cognitiva

Fonte: Fachin et al. (2009).

As buscas realizadas pelos usuários nesses repositórios sofrem interferências de vários fatores: conceitos dos usuários, conceitos existentes nos repositórios, como linguagens documentárias e políticas adotadas, e representação do documento. Tudo isso justifica, para os autores, a utilização da visão cognitiva na concepção de uma estrutura padronizada que permite a recuperação das informações de forma eficaz pelos seus usuários.

Aquisição, aplicação, armazenamento e recuperação do conhecimento sempre foram componentes para as conquistas e para a evolução dos seres humanos (FACHIN et al., 2009). Além disso, esses repositórios intervêm em duas questões estratégicas: favorecem o aumento da visibilidade e do valor público das instituições, servindo como indicador tangível da sua qualidade, e contribuem para a reforma do sistema de comunicação científica, expandindo o acesso aos resultados da investigação e reassumindo o controle acadêmico sobre a publicação científica. Os repositórios constituem uma manifestação clara da importância emergente da gestão do conhecimento no contexto da educação superior (CROW, 2002; LEITE, 2007).

Um dos desafios identificados na literatura para a implementação dos RIs é em relação à interoperabilidade, principalmente a semântica. Esse tipo de

interoperabilidade está estreitamente relacionado com a OC e os metadados utilizados pelos RIs (BHAT, 2010; BOTHERAM, 2010).

A recuperação da informação em sistemas diversos e no ambiente heterogêneo da *Web* identifica, na interoperabilidade semântica, um fator essencial para a recuperação dessas informações, com maior eficiência e eficácia.

A *Web* tem sido comparada a uma grande sala cheia de livros espalhados pelo chão. É um local onde se reúnem diversas comunidades e que tem sido chamado de "*Internet Commons*". Exige, entretanto, que haja comunicação entre as comunidades, com o objetivo de alcançar uma "economia" que possibilite transações comerciais, culturais, sociais e educacionais (WEIBEL, 1999, apud HODGE, 2000). Gilliland-Swetland (2000) também afirma que os metadados representam um conceito de economia informacional. Já Hodge (2000, p. 17) acredita que os SOC, que, de acordo com Brasher e Café (2008, p. 7), "[...] são sistemas conceituais que representam determinado domínio por meio da sistematização dos conceitos e das relações semânticas que se estabelecem entre eles [...]", também participam desse contexto ao fazer parte de um dos fatores que permitem a comunicação entre distintas comunidades. Para Hodge, os SOC podem ser utilizados para: a) promover alternativas de acesso ao assunto; b) adicionar significado aos recursos da biblioteca digital; c) apoiar o acesso multilíngue; e d) fornecer termos que possam expandir a pesquisa em textos livres nos domínios que são relativamente desconhecidos pelos usuários.

3 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

Disserta-se nessa seção a respeito da organização e representação do conhecimento, assim como os sistemas que possibilitam estruturação, representação e organização do conhecimento. São abordadas definições, contextualizações e considerações sobre relevância para o acesso e a recuperação da informação. Com isso, o leitor poderá compreender as relações entre organização do conhecimento, contexto digital, explosão informacional e obtenção da informação em um contexto onde rapidez, exatidão e reuso do conteúdo tornam-se imprescindíveis.

A organização faz parte do cotidiano das pessoas desde os primórdios. Há uma busca por ordenamento e agrupamento, assim como pela identificação de padrões por vários profissionais e organizações. As formas como determinados objetos ou itens são organizados e o motivo que levaram a essa organização provocaram análises, estudos e teorias. No caso específico da OC, é importante definir organização e conhecimento, para que isso possa auxiliar a compreensão da extensão do seu significado.

Organização pode ser vista como um arranjo lógico ou compreensível de elementos que se encontram separados. No contexto informacional, essa organização tem por objetivo facilitar o uso de documentos ou registros do conhecimento (HJØRLAND, 2007). E conhecimento pode ser definido como “[...] processo intelectual ou emocional que um indivíduo realiza para compreender um fenômeno do mundo exterior, e incorporar seu resultado como reafirmador ou removedor da sua concepção do mundo” (BARITÉ, 2001, p. 44, tradução nossa).

Esforços vêm sendo empreendidos com o objetivo de organizar esse conhecimento em grupos que fazem parte de nossa linguagem, educação e ciência, desde Aristóteles, passando pelo período em que a guarda do conhecimento era feita pelos monges, culminando com o desenvolvimento das bibliotecas públicas, quando houve grande preocupação em organizar o material ali armazenado, seguindo com as classificações das ciências na época moderna, que surgiram a partir do Renascimento e com as ideias de Francis Bacon, até os atuais SOC (HODGE, 2000; SAN SEGUNDO MANUEL, 1996).

De acordo com Barité (2001), a OC busca oferecer subsídios para o tratamento da informação, em especial, o tratamento temático da informação, assim como a gestão de seu uso social.

O termo OC foi estabelecido por volta de 1900, por Charles Cutter, W.C. Berwick Sayers e Ernest Cushingon Richardson, e o fator preponderante para o estabelecimento desse campo de estudo foi a contribuição intelectual de Henry Bliss (HJØRLAND, 2007).

Tennis (2008, p. 103) define a OC como a “[...] área do conhecimento interessada no estudo e crítica dos processos de organização e representação documental que a sociedade vê como digno de preservação”.

Para Hjørland (2007, p. 2), atualmente, o conhecimento é considerado mais em relação a teorias e perspectivas, o que poderia ser entendido como “reivindicações de conhecimento”. Formas de ordenação do conhecimento são vistas como pragmáticas, como suportes de diferentes atividades humanas, já que a organização não é nem objetiva nem neutra; ela possui um propósito em relação aos diferentes objetivos e atividades. Identifica-se assim uma tendência de o conhecimento ser uma epistemologia mais pragmática.

Barité (2001, p. 41) conceitua como objeto de estudo da OC:

O conhecimento socializado, e como disciplina propicia o desenvolvimento de técnicas para a construção, a gestão, o uso e a avaliação de classificações científicas, taxonomias, nomenclaturas e linguagens documentárias. Por outro lado, aborda metodologias de uso e recuperação pela linguagem natural (tradução nossa).

Broughton et al. (2005) também entendem que a OC representa os sistemas que organizam o conhecimento (SOC), tais como registros bibliográficos, sistemas de classificação, tesauro, redes semânticas e processos de organização do conhecimento, como classificação e descrição de documentos, catalogação descritiva, indexação e análise de assunto. A OC é adotada em sistemas de informação tradicionais, como museus e bibliotecas, bancos de dados *on-line*, no ambiente da Internet ou em ambientes externos aos sistemas de informações tradicionais, como os sistemas denominados de gestão de informações pessoais, ou na indexação de contracapa de livros. Os SOC ainda podem ser universais, quando cobrem todos os campos de conhecimento ou são limitados a determinados domínios ou tipos de documentos.

Dois aspectos são fundamentais para a compreensão do aumento das discussões e pesquisas sobre os SOC. Um deles está relacionado à seguinte questão: Por que organizar o conhecimento? Talvez essa pergunta possa ser respondida pelas premissas defendidas por Barité (2001) (Quadro 1):

Quadro 1 – Premissas básicas para a Organização do Conhecimento

PREMISSAS	CONSIDERAÇÕES
O conhecimento é um produto social, uma necessidade social e um dínamo social	O conhecimento é visto como um saber acumulativo que pode suprir necessidades do ser humano com possibilidades de geração de novos conhecimentos
O conhecimento se realiza a partir da informação e, ao socializar-se, transforma-se em informação	O conhecimento é construído a partir da informação enriquecida pela experiência de cada um. Por ser uma conquista subjetiva, o conhecimento só se socializa quando é aplicado ou quando é difundido. Com a socialização do conhecimento, é possível que os organizadores do conhecimento utilizem formas apropriadas de organização, para que esse conhecimento seja transformado em informação. Para tanto, requer que os sistemas de representação do conhecimento acompanhem a dinâmica da sua geração
A estrutura e a comunicação do conhecimento formam um sistema aberto	A distribuição social e a geração de novo conhecimento estão sujeitas a fatores políticos, econômicos e sociais. Podem ser estudadas em relação a vários prismas: aquisição, organização, transmissão, utilização, tipos e transformações
O conhecimento deve ser organizado para seu melhor aproveitamento individual e social	Independentemente da informação requerida (especializada ou leiga), o sistema adotado para organizar o conhecimento deve permitir que os conceitos ali utilizados sejam compreendidos pela maior quantidade possível de usuários
Existem “N” formas possíveis de organizar o conhecimento.	Os sistemas de organização do conhecimento existentes e as novas perspectivas com o uso das novas tecnologias ampliam as possibilidades de organização do conhecimento
Toda organização do conhecimento é artificial, provisória e determinista	É artificial, pois se baseia em sistemas de conceitos. Determinista, por se utilizar de conceitos unilaterais; provisória, já que depende do seu uso – o tratamento da informação sofre transformações de acordo com as mudanças que ocorrem na sociedade. Também é científica, por estar sujeita ao método científico, e social, em razão do seu caráter social (MIRANDA, 1999)
O conhecimento é registrado sempre em documentos, como conjunto organizado de dados disponíveis, e admite usos indiscriminados	Esse conhecimento registrado transforma-se em informação, que poderá ser utilizada para inúmeras finalidades e por pessoas com diferentes formações, possibilitando a geração de um novo conhecimento
O conhecimento se expressa em conceitos e se organiza mediante sistemas e conceitos	As classificações científicas, as taxonomias e as nomenclaturas são os sistemas de conceitos mais adotados por diferentes especialidades e os que servem de modelo de organização para os sistemas de classificação e dos tesouros
Os sistemas de conceitos se organizam para fins científicos, funcionais ou de documentação	Os tesouros podem ser entendidos como um exemplo de como os sistemas de conceitos atendem a várias demandas da representação do conhecimento
As leis que regem a organização dos sistemas de conceitos são uniformes e previsíveis e se aplicam por igual a qualquer área	A Teoria do Conceito defendida por Dalhberg e a Teoria da Classificação Facetada de Ranganathan representam teorias que buscam estabelecer a teoria e a metodologia para o desenvolvimento da representação do conhecimento

Fonte: Barité (2001)

Os RIs podem encontrar nessas premissas as bases de sua implementação e sua organização. Por tratar-se de um sistema aberto e pelo fato das informações ali depositadas poderem ser organizadas utilizando métodos e sistemas de organização do conhecimento (SOC), fica visível a estreita relação entre a concepção dos RIs e a utilização dos princípios da Organização do Conhecimento como norteadores para gerenciamento dos seus recursos. Pode-se afirmar ainda que as ideias de Barité (2001) contribuem para o papel a ser desempenhado pelos RIs no pensamento de Crown (2002), Lynch (2003) e Bailey Jr. (2005): o de socialização do conhecimento.

O outro aspecto que está relacionado com o contínuo interesse dos pesquisadores no desenvolvimento dos SOC é justificado pelas funções atribuídas à OC: a) facilitar a pesquisa entre catálogos, bibliografias e sistemas de recuperação; b) prover informações sobre documentos relevantes para os usuários; e c) organizar documentos na estante ou usar outros tipos de disposição (BROUGHTON et al., 2005; HJØRLAND, 2007).

As funções da OC estão voltadas para auxiliar o usuário a navegar em espaços informacionais, a recuperar documentos, a tomar decisões sobre futuras atividades de pesquisa e a proporcionar uma visão ampla dos recursos de informação.

Esses sistemas não se limitam aos espaços tradicionais das bibliotecas; devem ser vistos como uma forma de participação ativa no contexto digital, por meio dos esquemas de classificação, vocabulários controlados e taxonomias (BROUGHTON et al., 2005). Gomes (2009) reforça esse posicionamento ao argumentar que o entrosamento entre os ramos da Tecnologia da Informação (TI) e da CI, em virtude da necessidade do desenvolvimento de atividades como ontologias, taxonomias e inteligência artificial, contribui para a organização do conhecimento.

3.1 SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

A Internet, a Web e as bibliotecas digitais contribuem com o número de recursos que podem ser disponibilizados para diferentes comunidades. A tecnologia facilita a comunicação entre comunidades distintas, entretanto a descoberta dos recursos e a acessibilidade são necessárias para que os conteúdos

e a organização sejam entendidos por essas diferentes comunidades. Para promover esse acesso, os SOC possuem um papel fundamental ao possibilitar acesso alternado de assunto, modos de compreensão, acesso multilíngue e fornecimento de termos para a expansão de busca em textos livres.

Os SOC incluem classificação e esquemas de categorização que organizam materiais em um nível geral, como cabeçalhos de assuntos, que permitem acessos mais detalhados, e também listas de autoridades, que controlam informações como nomes geográficos e nomes pessoais. Os SOC também incluem vocabulários altamente estruturados, como tesouro, e esquemas menos tradicionais, como redes semânticas e ontologias. Os SOC podem ser aplicados em registros de metadados¹³ para cada recurso, incorporado em *metatags* ou separadamente, de uma biblioteca digital como parte de um mecanismo de acesso. Independente da localização dos SOC em relação a recursos, sua origem ou seu tipo, eles têm o papel de organizar o conteúdo para possibilitar a recuperação de itens relevantes de uma coleção¹⁴ digital (HILL et al., 1999).

Os SOC são vistos como uma forma de melhorar o acesso aos recursos digitais, utilizando o controle de vocabulário e a organização do conhecimento. O controle de vocabulário visa reduzir a ambiguidade da linguagem natural para descrever e recuperar itens para fins de busca de informação. Os vocabulários controlados consistem de termos, palavras de linguagem natural selecionadas para recuperação da informação (TUDHOPE; NIELSEN, 2006). Dois aspectos da linguagem natural comprometem a recuperação da informação: termos diferentes podem representar o mesmo conceito, enquanto o mesmo termo pode representar conceitos diferentes.

Apesar das diferenças inerentes aos SOC, Hodge (2000) descreve algumas características em comum:

- a) o SOC impõe uma visão particular do mundo em uma coleção e em itens;

¹³ Conjunto de metadado de uma estrutura relevante, compreendendo todos os elementos relevantes, descrevendo um objeto.

¹⁴ Listagem de itens de alguma forma relacionados ou similares. Uma coleção pode ser caracterizada e classificada de acordo com algumas qualidades, como a extensão do arquivo, a natureza do documento, entre outros (em torno de um tópico ou tipo de informação). Para cada critério previamente estabelecido, poder-se-á dar origem a uma coleção específica. Não se pode criar uma coleção sem uma comunidade (IBICT, 2011).

- b) o SOC, ao ser utilizado, influencia a caracterização de uma mesma entidade; e
- c) o SOC deve intermediar o conceito adotado para representar objeto e o conceito a ser empregado pelo usuário.

E sugeriu uma taxonomia para os SOC, no contexto da CI:

Lista de termos

- Lista de autoridades. Lista que tem por objetivo controlar variações de nomes (indivíduos, países ou organizações) para uma entidade ou domínio em um campo específico.
- Glossários. Lista de termos com definições. O termo pode ser de um campo específico de assunto ou de uma atividade.
- Dicionários. Lista alfabética de palavras e suas definições. O dicionário pode atribuir sinônimos e, por meio das definições, relacionar palavras. Não existe uma estrutura hierárquica ou uma intenção de agrupar os conceitos.
- *Gazetteers*. A *gazetteer* é uma lista de nomes de locais. Esses *gazetteers* sempre organizam nomes de locais utilizando um esquema de classificação ou categorias de assuntos.

Classificação e categorias

- Cabeçalho de assunto. Esse esquema fornece um conjunto de termos controlados para representar os assuntos dos itens na coleção.
- Esquemas de classificação, taxonomias e esquemas de categorização.

Esses esquemas são adotados para agrupar entidades em classes mais gerais. Utilizam-se taxonomias em modelos orientados a objeto e em sistemas de gestão do conhecimento, para indicar grupos de objetos baseados em características particulares. Os esquemas de categorização podem ser empregados para reunir os termos de tesouros em tópicos.

Lista de relacionamentos

- Tesouro. É baseado em conceitos e mostra relações entre os termos. As relações apresentadas nos tesouros incluem hierarquia, equivalência e associação ou semelhanças.
- Redes semânticas. O processamento da linguagem natural tem incrementado o desenvolvimento de redes semânticas, em que as estruturas de conceitos e termos não são hierárquicas; apresentam-se como uma rede ou teia.
- Ontologia. É o SOC mais atual. Descreve o conhecimento em uma área específica, sempre conectada por sistemas, como *data mining*, ou de gestão do conhecimento.

Nem todos os itens apresentados por Hodge (2000) são compartilhados por Broughton et al. (2005), que elencam os seguintes instrumentos:

- a) Classificação Decimal de Dewey (CDD), Classificação Decimal Universal (CDU) e Library of Congress Classification (LCC) – são sistemas de classificação tradicionais utilizados por bibliotecas e bases de dados, que ainda possuem papel relevante nas atividades práticas das bibliotecas, assim como influenciam o ensino da OC;
- b) sistemas de análise facetada de Bliss – concebidos por Ranghanatan e desenvolvidos pelo British Classification Research Group e pelos editores do Sistema de Classificação Bliss, são identificados como um sistema de classificação que pode ser empregado no ambiente digital, que segue a tradição da pesquisa em análise facetada, vista como a mais pura das abordagens da OC;
- c) vocabulário de assunto (Library Congress Subject Headings - LCSH) – de acordo com Hjørland (2008), é uma forma padronizada de controlar termos sinônimos e homônimos;
- d) sistemas baseados em busca de textos – ferramentas de busca e motores de busca;

- e) tesouros baseados em sistemas – sistemas artificiais de signos normalizados que permitem representação mais fácil e efetiva do conteúdo documental, com o objetivo de recuperar, manual ou automaticamente, a informação que o usuário solicita (TRISTÃO; FACHIN; ALARCON, 2004);
- f) mapas bibliométricos – são técnicas bibliométricas usadas, por exemplo, para fazer a análise de citação ou métodos de construção e manutenção semiautomática de tesouros;
- g) algoritmos em ferramentas de busca – na nova geração de ferramentas de busca, o posicionamento passou a depender da relevância, que é definida por algoritmos – cálculos que servem para definir o quanto uma página é importante. Um dos mais conhecidos algoritmos de busca é o Google PageRank (TEIXEIRA, 2011);
- h) sistemas de arquivo (baseado no princípio da proveniência);
- i) ontologias – “[...] especificação formal e explícita de uma conceitualização compartilhada” (GRUBBER, 1996, p. 1);
- j) *Web* semântica – é uma extensão da *Web*, corrente em que se dá um significado bem preciso à informação, aumentando a possibilidade de trabalho cooperativo entre computadores e entre pessoas (BERNERS-LEE, 2001). Segundo a visão de Tim Berners-Lee, a *Web* seria como uma base de dados global em que a infraestrutura da *Web* Semântica permitiria às máquinas fazer deduções e organizar a informação como os humanos;
- k) mapas tópicos – é um padrão International Organization for Standardization (ISO) para descrever estruturas de conhecimento associadas a fontes de informação. Podem representar a informação utilizando tópicos que expressam conceitos (pessoas, países, arquivos, eventos etc.), associações entre conceitos e ocorrências (relações entre tópicos).

Hjørland (2007) acrescenta à lista de Broughton et al. (2005) mapas conceituais; hipertextos; anéis de sinônimos, considerados um tipo de vocabulário

controlado¹⁵, mas utilizado apenas durante a recuperação; tipologias, *Webpacs* e folksonomia. O Quadro 2 apresenta as abordagens de Broughton et al. (2005) e de Hjørland (2008) em relação à OC.

¹⁵ É um conjunto limitado de termos autorizados para uso na indexação e busca de documentos (LOPES, 2002).

Quadro 2 – Abordagens em relação à Organização do Conhecimento

ABORDAGENS	BROUGHTON	HJØRLAND
Abordagem tradicional Sistemas de classificação usados nas bibliotecas e bases de dados	Documentos de unidades organizadas, mas o termo OC implica uma abstração mais ambiciosa para basear a classificação no conhecimento científico e acadêmico	Os princípios atribuídos a essa abordagem são: a) Controle de vocabulário b) Regra de Cutter sobre especificidade c) Princípio de Hulme de garantia literária (1911) d) Princípio de organização do geral para o específico CDD, CDU. A CDD foi publicada pela primeira vez em 1876. Tem a vantagem de ser um padrão e não um sistema direcionado a uma coleção particular, um domínio ou um grupo de usuário
Análise facetada	Ideias. Essa abordagem afasta a base empírica dos documentos e introduz princípios lógicos para a OC que se baseiam na intuição racional	Foi estabelecida por Ranganathan em 1933 e desenvolvida pela British Classification Research Group e os editores da Bliss Classification Systems. Essa abordagem continua sendo adotada, inclusive, em ambientes digitais. É a abordagem teórica mais diferente e pura na OC, não necessariamente a mais importante. Influenciou outros sistemas de classificação, como a CDD. Seus pontos fortes são os princípios lógicos e a forma como fornece estruturas no SOC (classificações, tesauros)
A tradição da recuperação da informação	São palavras, relação copalavras e relações palavra-documento das unidades. Entretanto, a informação é a unidade reivindicada	A RI foi incrementada em 1950. Assim como os tradicionais sistemas de classificação e o método de análise facetada, foi seriamente criticada em relação à eficiência na recuperação da informação
Visão orientada ao usuário	Individual, visão cognitiva	Tem influenciado a CI nas últimas décadas
Abordagem bibliométrica	Documentos e padrões de citações entre documentos	Teve início com Garfield, ao elaborar o Science Citation Index, em 1963. Esforços têm sido empreendidos para relacionar a bibliometria com a abordagem tradicional da OC e a recuperação da informação. Algumas pesquisas têm sido desenvolvidas para investigar a possibilidade da construção de tesauros baseados em citações relacionadas e em documentos
Abordagem na análise de domínio	Conhecimento recolocado com reivindicação do conhecimento (reivindicação do conhecimento documentado) ou atividades. O que é organizado não são verdades eternas, mas funciona como reivindicações justificadas a partir de alguma perspectiva epistemológica.	Formulada a partir de 1994

Fonte: Adaptado de Broughton et al. (2005) e Hjørland (2008).

Vickery (2011) afirma que os SOC são baseados em reunião e exposição de palavras ou frases que possuem uma relação semântica entre elas. Essa definição compreende:

- a) cabeçalhos de assunto – uma lista que relaciona um termo com outro, tratados como sinônimos, e que pode ter uma referência indicando uma relação semântica indeterminada entre os termos;
- b) classificação enumerativa pré-coordenada – relações herárquicas que representam uma relação genérica entre a classe e a subclasse, mas que, na prática, podem ser utilizadas como relações de associação;
- c) classificação facetada – subdivide um campo de assunto ou de domínio em categorias facetadas antes de introduzir a hierarquia;
- d) tesouro alfabético – expressa quaisquer relações de classes genéricas no todo ou em parte, e é utilizado para relacionar termos tratados como os sinônimos nos SOC;
- e) sistemas de indexação pós-coordenada – indicadores de funções, quando combinados em uma composição;
- f) sistemas de pesquisa *on-line* – esses sempre usam um tesouro alfabético ou uma parte de um sistema global, empregando termos extraídos de um deles, como os descritores;
- g) lista de termos – usada nas interfaces de pesquisa *on-line*;
- h) ontologia – formalização sistemática de conceitos, definições, relacionamentos e regras que identificam o conteúdo semântico de um domínio em uma linguagem de máquina legível.

Enquanto, para Vickery (2011), os SOC são instrumentos complementares que auxiliam o usuário a encontrar sua informação, para Broughton et al. (2005), esses sistemas podem ser considerados como ferramentas semânticas que possibilitam a seleção de conceitos e informações baseados nas suas relações semânticas. Já os sistemas apresentados por Hodge (2000), utilizados por outros autores, como Hjørland (2008), Zeng e Chan (2006), são adotados para organizar a

informação e constituem o coração de todos os sistemas de recuperação de informação.

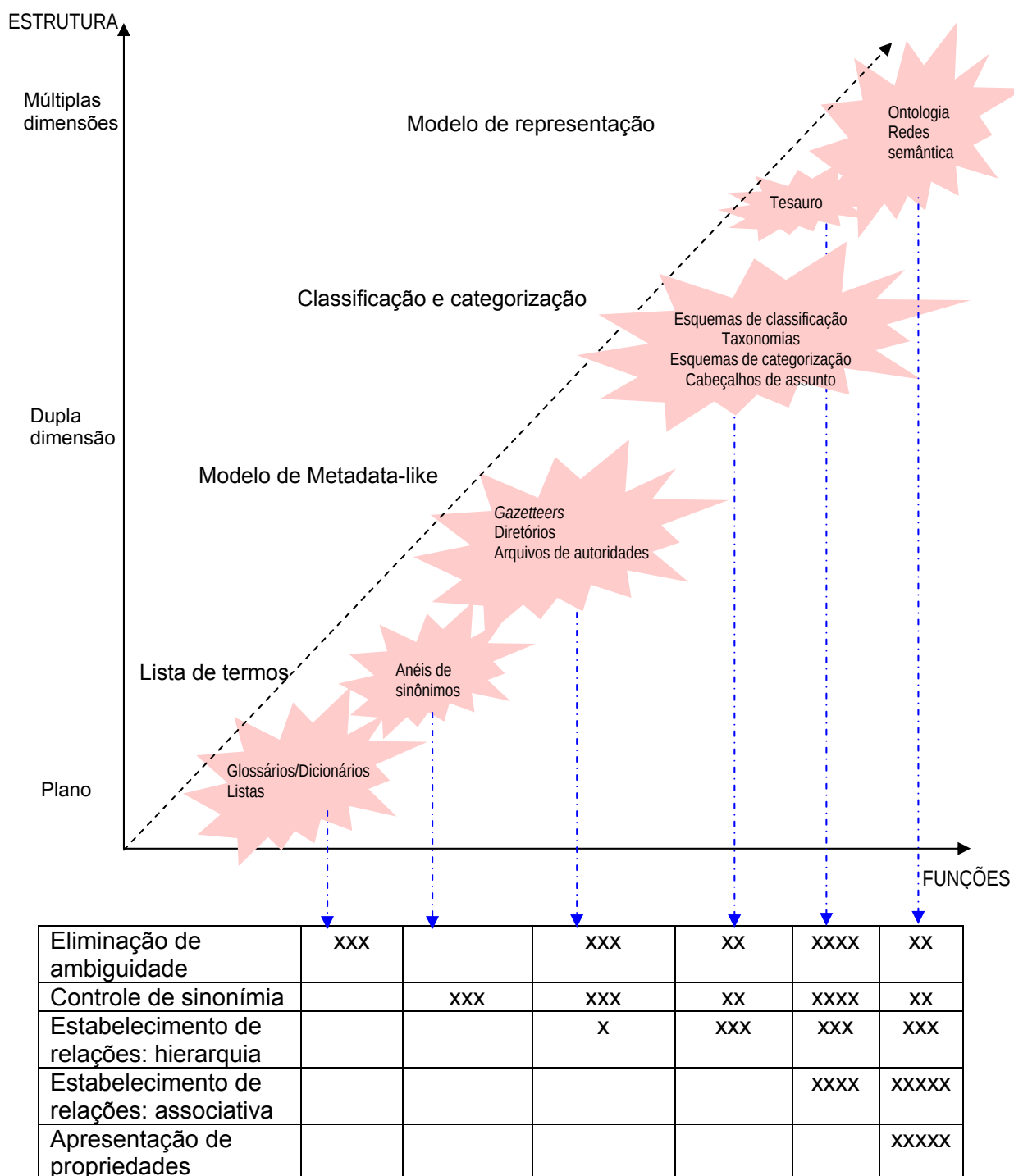
A base teórica dos SOC pode explicar como os termos e conceitos devem ser selecionados e definidos e como as suas relações semânticas devem ser definidas e selecionadas.

É preconizado, por vários autores, que os conceitos e semânticas devem relacionar o conceito de garantia literária ou de outros tipos de garantias (SAN SEGUNDO MANUEL, 1996). O princípio da garantia literária implica a decisão de adotar um SOC. Um exemplo é o vocabulário controlado, que tem por objetivo evitar sinonímias e homonímia, o que se torna possível ao se utilizar a indexação de termos empregando vocabulários padronizados (HJØRLAND, 2005). Outros possuem funções de eliminar a ambiguidade, estabelecer relacionamentos (hierárquico e associativo) e apresentar propriedades (ZENG, 2008).

Para Zeng (2008, p. 180), os SOC no ambiente de rede

[...] herdaram a maioria das estruturas que o mundo tem testemunhado por uma centena de anos, mas sistemas de organização do conhecimento em rede não são simples repetições do passado. Eles estão formando novas estruturas semânticas que funcionarão com grande impacto, muito mais extenso do que se imaginava (ZENG, 2008, p. 180, tradução nossa).

Zeng (2008), que elaborou um estudo exaustivo da literatura sobre SOC, oferece uma taxonomia desses sistemas agrupados nas seguintes categorias: Listas de termos, Modelo *Metadata-like*, Classificação e Categorização, e Modelo de Representação (Gráfico 2). A autora mostra as funções associadas a diferentes SOC e suas relações semânticas.

Gráfico 2 – Taxonomia dos sistemas de organização do conhecimento

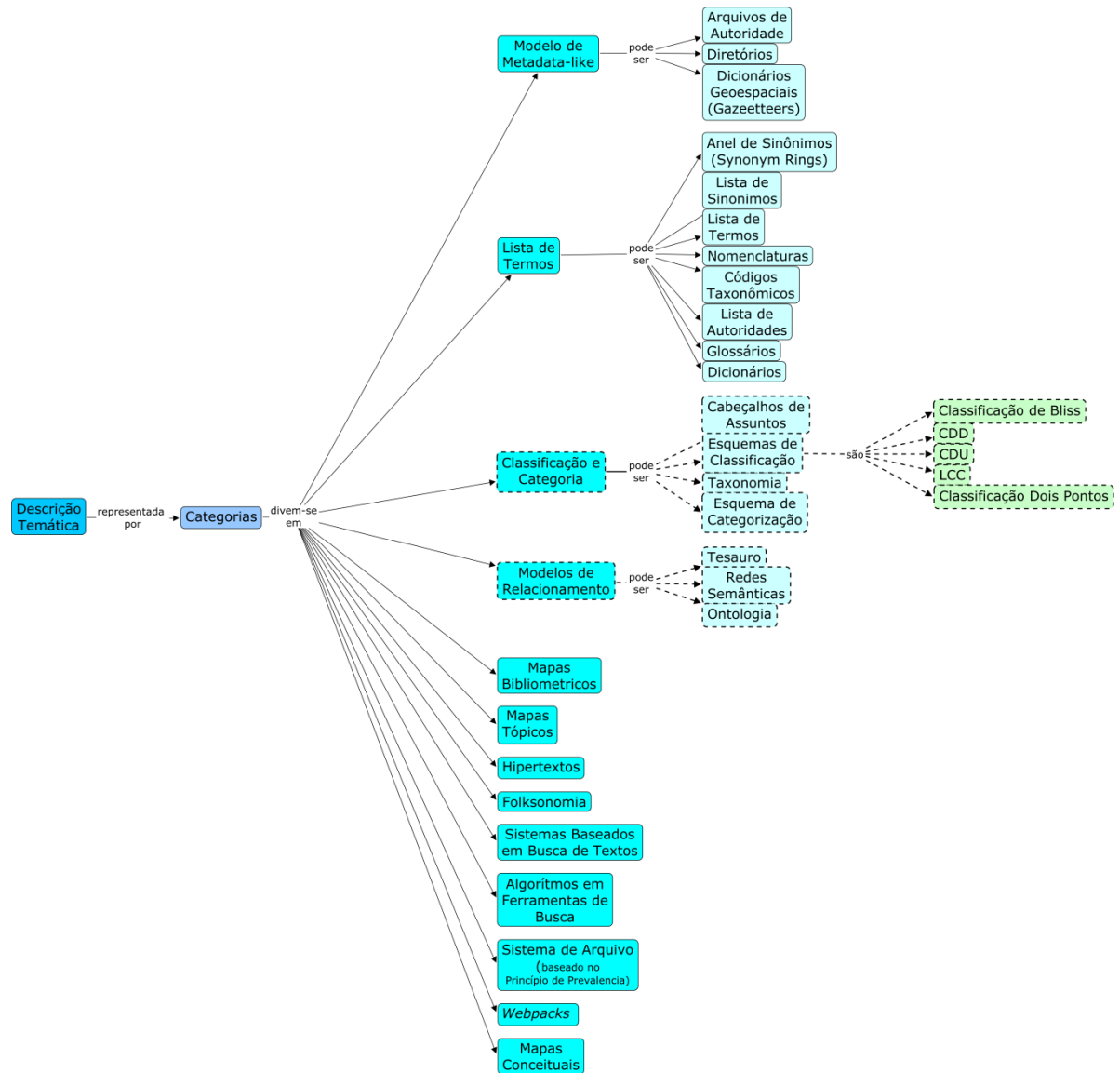
Fonte: Adaptado de Zeng (2008)

Embora alguns tipos de SOC sejam mais flexíveis e fáceis de se adaptarem às diferentes situações, há restrições quanto a um ou a outro sistema. Portanto, sempre devem ser considerados os interesses e as preocupações em relação às coleções e aos usuários como prioridade.

Lesk (1997, apud MIRANDA, 2005, p. 119) afirma que não existe um único SOC com o qual todos concordem. Especula-se que um único seria vantajoso, porém improvável de ser desenvolvido. A questão cultural pode limitar um sistema de organização do conhecimento, pois, o que é significativo para uma cultura pode não ser para outra. Entretanto, para efeito desta pesquisa, serão analisados os SOC compartilhados por Hodge (2000), Broughton et al. (2005) e Vickery (2011), por se entender que esses autores aparentam ter ideias comuns em relação à OC.

A Figura 2 apresenta os SOC acatados por Hodge (2000), Broughton et al. (2005), Hjørland (2008), Vickery (2011).

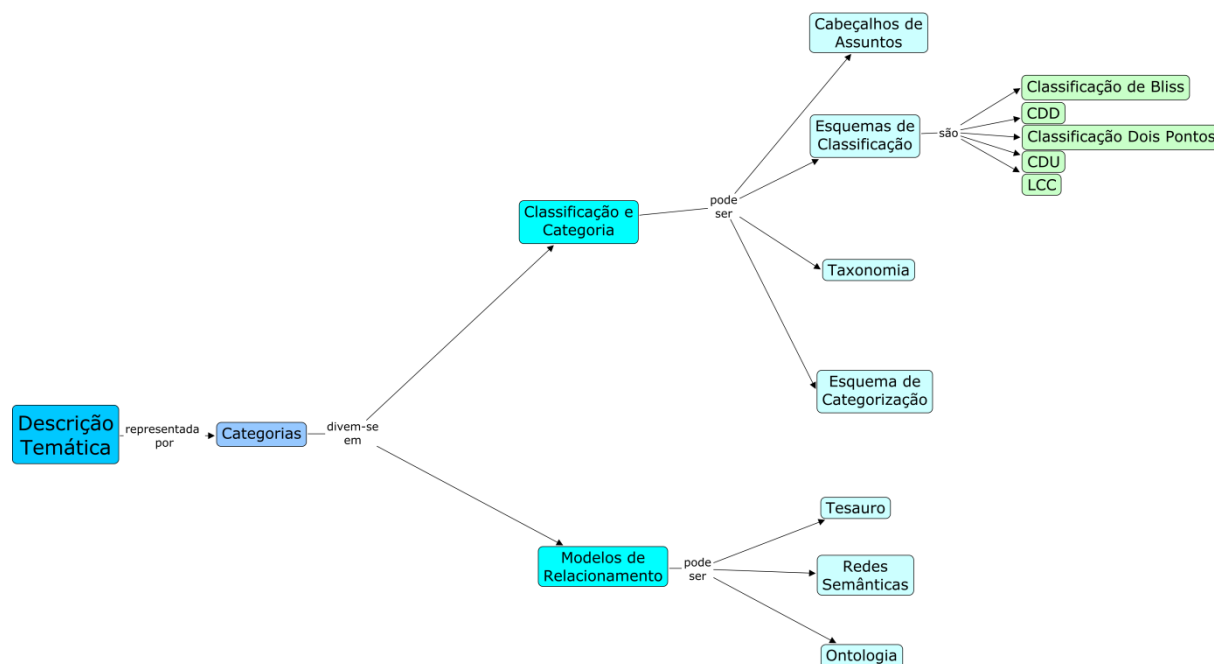
Figura 2 – Sistemas de Organização do Conhecimento



Fonte: Hodge (2000), Broughton et al. (2005), Hjørland (2008), Vickery (2011).

Serão definidos e exemplificados neste estudo apenas os SOC que são consenso entre os autores Hodge (2000), Broughton et al. (2005), Hjørland (2008) e Vickery (2011) (Figura 3).

Figura 3 – SOC compartilhados por Hodge (2000), Broughton et al. (2005), Hjørland (2008), e Vickery (2011).



Fonte: Hodge (2000), Broughton et al. (2005), Hjørland (2008), Vickery (2011).

Dessa forma, serão explicitadas definições e particularidades referentes a esses sistemas de descrição temática:

Classificação e Categoria

a) cabeçalhos de assunto

A lista de cabeçalhos de assunto contém termos controlados com o objetivo de representar os assuntos de um item de uma coleção. Ela é de ampla cobertura, apresenta uma estrutura superficial com uma estrutura hierárquica limitada (HODGE, 2000). Trata-se de uma linguagem pré-coordenada, ou seja, documentária¹⁶, nas quais os termos que a integram se coordenam num processo que precede sua utilização (CURRÁS, 2010).

De acordo com Vickey (2011, p. 3),

¹⁶ “[...] são linguagens construídas para indexação, armazenamento e recuperação da informação e correspondem a sistemas de símbolos destinados a traduzir os conteúdos dos documentos” (CINTRA, 2002, p. 39).

Trata-se de uma lista que relaciona termos a outros que são tratados como sinônimos, podendo ter referências indicando relações semânticas indeterminadas com outros termos. Quando os termos na lista são combinados para formar uma combinação de cabeçalhos, as relações entre os termos combinados não são explícitas, embora eles possam ser entendidos pelo usuário bem informado (Tradução nossa).

Essa lista é composta por instruções de uso, pelos cabeçalhos de assunto, pela remissiva “ver”, que serve para o controle da terminologia, e pela estrutura sindética (“ver também”), que remete o usuário a assuntos que se relacionam aos termos utilizados no cabeçalho de assunto (TEIXEIRA, 1979, apud MIRANDA, 2005).

Como exemplos de lista de cabeçalhos de assuntos, há o cabeçalho de assunto da Biblioteca do Congresso (LCSH) e o cabeçalho de assunto da área da saúde Medical Subject Headings (MeSH).

b) esquemas de classificação

A evolução da fase do saber único para as especialidades fez com que fossem elaboradas classificações no decorrer da história (CURRÁS, 2010). Como ponto de partida, há os registros da Biblioteca de Assurbanipal, em Nínive (825 a.C), onde seus materiais foram divididos em História, Direito, Ciência, Magia e Dogma (SAN SEGUNDO MANUEL, 1996). O maior avanço aconteceu na Biblioteca de Alexandria, onde Clímaco estabeleceu a classificação por matérias e autores. Outro marco foi a Classificação de August Comte, em que distinguia diferentes especialidades do saber (CURRÁS, 2010).

Gomes (2009) relata que a classificação bibliográfica é vista por Dahlberg (1993), Gnoli (2004a) e Needham (1974) como a primeira iniciativa de organização do conhecimento. O objetivo prioritário dessa classificação é a ordenação por grupos temáticos e a elaboração de catálogos e bibliografias sistemáticos que facilitem o acesso do usuário à informação e aos documentos (SAN SEGUNDO MANUEL, 1996).

Alguns sistemas de classificações atuais têm como base a classificação denominada árvore de Porfírio (233-305), que é um sistema de linguagem pré-coordenada, com cada assunto representado por um código ou notação. Esse tipo de classificação tem por finalidade principal localizar os assuntos e as relações mantidas entre si dentro de uma estrutura. Utiliza, também, notações constituídas de letras, símbolos e/ou números para representar o conteúdo temático dos documentos.

Os sistemas de classificação dividem-se em gerais, universais e específicos, baseados em unidades conceituais isoladas, termos ou suas partes, estruturados conforme métodos estabelecidos (CURRÁS, 2010, p. 57). Ainda possuem como uma das vantagens a possibilidade de ser utilizada para a organização conceitual que auxilia a busca por palavras no ambiente Web.

Em resumo, os sistemas de classificação dão ênfase nas subdivisões para baixo, dentro de classes mais específicas e menores, não fornecem relações sintáticas e podem limitar combinação entre classe ou expor conteúdos semânticos complexos, causando dificuldades para pesquisas em ambientes multidisciplinares (TRISTÃO; FACHIN; ALARCON, 2004).

Os sistemas de classificação analisados foram:

– Classificação Decimal de Dewey (CDD)

Esse sistema de classificação foi desenvolvido por Melville Louis Kossuth Dewey (Melvil Dewey) no período de 1873, ao tentar responder às necessidades práticas identificadas ainda como bibliotecário do Amherst College. Dewey não teve a intenção de elaborar uma classificação científica.

Segundo Mai (2003), essa classificação foi um dos primeiros projetos de interoperabilidade documental, ao ter como objetivo principal torná-la uma ferramenta de organização bibliográfica universal. Dessa forma, todas as bibliotecas teriam a mesma classificação para a organização das coleções, favorecendo a busca de informação pelos usuários.

Esse sistema de classificação possui uma linguagem pré-facetada e estrutura pós-controlada. As contribuições desse sistema são: possibilitar uma classificação temática que indique um lugar fixo para os livros nas estantes; permitir uma subordinação de classes e subclasses até o infinito com numeração seriada e propiciar a inclusão de um índice relativo às tabelas (SAN SEGUNDO MANUEL, 1996).

A CDD é o sistema de classificação bibliográfica mais utilizado em todo o mundo. Até 2006, era empregado em mais de 135 países e tinha sido traduzido para mais de 30 línguas. Também é adotado nos Estados Unidos por 95% das bibliotecas públicas e escolares, 25% das bibliotecas das faculdades e universidades e 20% das bibliotecas especiais (DEWEY, 2003, apud MIRANDA, 2005).

Desenvolvida, mantida e aplicada pela Divisão de Classificação Decimal da Biblioteca do Congresso (Library of Congress - LC), a CDD possui os números incorporados em registros bibliográficos de catalogação legíveis por computador (MARC) e distribuídos às bibliotecas por meio de comunicação computadorizada (Catalogação-na-Publicação - CIP e fichas da LC).

Na CDD, o conhecimento está organizado em 10 classes subdivididas em mais 10, que, por sua vez, se subdividem em mais 10, e assim sucessivamente. Essas classes correspondem a disciplinas tradicionais ou áreas de estudo, assim descritas:

000	Obras gerais
100	Filosofia
200	Religião
300	Ciências sociais
400	Filologia
500	Ciências puras
600	Ciências aplicadas
700	Arte, Esporte, Lazer
800	Literatura
900	História, Geografia e Biografias.

– Classificação de Bliss

A Bibliographic Classification (BC), ou Classificação de Bliss, foi elaborada por Henry Evely Bliss em 1910. O sistema é baseado na ordenação de classes que combinam a totalidade do conhecimento, divididas nas seguintes disciplinas: Ciências, Filosofia, História e Tecnologia e Arte.

Essa classificação foi desenvolvida com base nos seguintes princípios: consenso; arranjo de assuntos correlatos; gradação em especificidade – do geral para o específico; localização alternativa – flexibilidade do sistema;

brevidade da notação – notação sintética, expressiva e sistemática (CARLAN, 2010). As classes principais são representadas pelas letras do alfabeto (SOUZA, 2004).

– Classificação de Dois Pontos

A Classificação de Dois Pontos, Colon Classification (CC) ou Classificação de Ranganathan, foi desenvolvida por Shiyam Ramarita Ranganathan. Tem como característica a subdivisão dos assuntos em facetas e focos. Ranganathan propôs cinco áreas básicas de classificação ou facetas: personalidade, assunto, energia, espaço e tempo. Para cada uma, apresentou uma lista de valores possíveis, que chamou de isolados (SOUZA, 2004). A CC é o único esquema de classificação que adota um sistema completamente facetado, em que cada conceito é identificado por um termo linguístico e por um código alfanumérico (CARLAN, 2010).

Esse sistema inclui diversos graus de especialização, podendo representar todas as ciências através de conceitos universais, acompanhando as modificações e a evolução do conhecimento humano. A notação adotada por esse sistema é mista, com algarismos arábicos, letras minúsculas e maiúsculas, letras gregas e sinais gráficos, totalizando 70 caracteres (PIEADADE, 1977).

– Classificação Decimal Universal (CDU)

A Classificação de Dewey (CDD) foi transformada em Classificação Decimal Universal (CDU) pelos juristas Paul Otlet e Henry La Fontaine com o objetivo de organizar um repertório bibliográfico universal.

A CDU está dividida em 10 classes, em que a notação de cada classe é feita por apenas um algarismo. Cada classe se subdivide em classes de dois algarismos, colocando-se um ponto de três em três dígitos (SOUZA, 2004). É uma linguagem pré-facetada, pós-controlada.

A CDU é dotada de uma estrutura que possui a habilidade de representar por símbolos, não apenas os assuntos simples, mas também as várias relações entre os diversos assuntos (MIRANDA, 2005). Essa estrutura consta de tabelas principais, tabelas auxiliares e índice alfabético (SOUZA, 2004).

– Classificação da Biblioteca do Congresso (LCC)

A Classificação da Biblioteca do Congresso Americano (Library Congress Classification) foi criada em 1901 com base na Classificação de Cutter com modificações principalmente no sistema de notação. Trata-se de um sistema de classificação enumerativo que recorre à síntese quando utiliza as tabelas auxiliares (PIEADADE, 1977).

A LCC possui uma notação mista, constituída de letras maiúsculas e algarismos arábicos. Os assuntos são distribuídos em 21 classes principais, representadas por letras maiúsculas de A a Z. Cada uma das classes traz seu próprio índice alfabético e relativo, não existindo índice geral. A LCSH é utilizada para suprir essa ausência, pois essa lista traz a notação correspondente aos esquemas de classificação entre colchetes, ao lado dos cabeçalhos de assunto (MIRANDA, 2005).

c) taxonomia

A palavra taxonomia foi introduzida por A. de Candolle, em 1813, para designar as normas ou leis que se utilizavam na sistemática. A Biologia foi a área que mais utilizou os princípios da taxonomia e, em sentido mais amplo, a área da Lógica (CURRÁS, 2010).

Currás (2010, p. 68), em sua obra, apresenta a definição de taxonomia disponibilizada no “Tema Três”,

[...] é uma aplicação Web para gestão de linguagens documentárias, orientada especificamente para o desenvolvimento de Tesouros hierárquicos. Pode-se utilizar, também, no desenvolvimento de estruturas de navegação na Web ou como complemento articulador num gestor de conteúdos.

É uma ferramenta que tem grande aproximação com o conceito da Web Semântica, com possibilidade de padronizar conteúdos semânticos em *sites* e possibilitar a navegação por meio de ordenação hierárquica. Pode ser entendida, ainda, como instrumento de gestão e organização do conhecimento em ambientes organizacionais.

De acordo com Campos e Gomes (2008), as taxonomias caracterizam-se por: conter uma lista estruturada de conceitos/termos de um

domínio; incluir termos sem definição, somente com relações hierárquicas; possibilitar a organização e a recuperação de informação através de navegação; permitir agregação de dados, além de evidenciar um modelo conceitual do domínio; ser um instrumento de organização intelectual, atuando como um mapa conceitual dos tópicos explorados em um sistema de recuperação de informação, e ser um novo mecanismo de consulta em portais institucionais.

No contexto da Web Semântica, segundo as autoras, é possível que, com a utilização de taxonomias, sejam estabelecidos padrões de alto nível para a ordenação e a classificação de informação através do uso de mecanismos de herança, já que os mesmos possibilitam a captura de hierarquia das classes (generalização/especialização). Esse mecanismo se utiliza da herança entre as classes, ou seja, a subclasse herda as propriedades da superclasse, e assim é possível a representação de uma rede complexa de relacionamentos de forma simples.

Na organização e na recuperação de informação, as taxonomias são usadas para a criação de metadados ou termos para descrever um objeto. Na categorização,¹⁷ são empregadas para definir classes e subclasses, como esquemas que organizam conteúdos das páginas na Web e como lista de controle de dados para suporte de mineração de dados (CARLAN, 2010). Carlan (2010) relata ainda que Daconta, Obrst e Smith (2003) consideram a taxonomia semanticamente deficiente por não expressar significados complexos: contém apenas relações hierárquicas e partitivas, havendo necessidade de complementação pelo uso de mecanismo de busca para acesso a outras relações (CAMPOS; GOMES, 2008).

A taxonomia é utilizada como método de organização em *website* para facilitar interatividade, brouseando seções de *sites*, ou sustentar mecanismos de acesso. Em algumas situações, a relação hierárquica ampla é adotada para estruturar alguns sistemas de *menu* (CURRÁS, 2010).

São três os tipos de taxonomia: a canônica; a de domínio e a de processo ou tarefas. A taxonomia canônica é uma classificação que revela relações de família, gênero e espécie, como a taxonomia biológica, em que são representadas as categorias de seres vivos. Na taxonomia de domínio, um termo é

¹⁷ É realizada por meio da análise do domínio a partir de recortes que permitem identificação dos conceitos (categorias) que fazem parte do domínio (CAMPOS; GOMES, 2008, p. 4).

associado a tantas classes e subclasses quantas forem necessárias, dentro de um domínio do conhecimento (policotômica), como é o caso da Geologia. A taxonomia de processo também é policotômica e composta de conceitos que representam processos ou tarefas – identifica-se esse tipo de taxonomia na realização do processo de representação do conhecimento.

A taxonomia permite armazenar, recuperar e comunicar informações de maneira lógica, utilizando a navegação. E a vantagem em usá-la na recuperação da informação é o fato de permitir ao usuário, por meio da navegação, uma melhor seleção dos termos de busca. Como os termos são apresentados em classes, subclasses e subsubclasses, em quantos níveis de especificidade sejam necessários, é facilitada a identificação do termo pelo usuário. A desvantagem está em não poder utilizar operação booleana, ao contrário das listas de assuntos (CAMPOS; GOMES, 2008).

d) Esquema de categorização

A indexação em linguagem artificial não estruturada é conhecida como esquemas de categorização, em que são utilizadas técnicas para classificar documentos. É possível a definição automática de categorias, e, com base nelas, conjuntos de documentos podem ser classificados. Esse processo pode ser adotado tanto na fase do armazenamento da informação como na etapa de recuperação.

De acordo com Rizzi et al. (2000, apud GALHO; MORAES, 2012), esquemas de classificação possuem a vantagem de ser empregada para classificar diversos tipos de materiais, de mensagens a publicações, contribuindo para coleta, análise e distribuição de informação.

O processo de categorização pode ser dividido em duas etapas:

- preparação do texto.

Os textos são analisados e empregam-se técnicas para seleção de características do texto que possibilitam sua categorização. Para estabelecer essas características, são eliminadas palavras que não interferem na definição do texto (*stopwords* ou palavras negativas);

São identificadas palavras, ou conceitos, que melhor representam o texto para geração de um índice que represente a categoria.

– indexação automática

É feita uma comparação entre a lista de termos das categorias e a lista de termos de um novo documento. A categoria que possuir a lista mais similar à lista do novo documento é escolhida como classe.

A adoção dessa técnica reduz o tempo de processamento, porém corre-se o risco de eliminar termos relevantes para a representação dos textos. Esse processo pode ser desenvolvido com base em critérios estatísticos ou na semântica e/ou sintática.

Modelos de relacionamento

a) Tesouro

A palavra "tesouro" vem do latim *thesaurus*, que significa tesouro e foi empregada pela primeira vez por Peter Mark Roget, em 1832, ao publicar o *English Thesaurus of Words and Phrases*, que reunia palavras pela ordem alfabética, de acordo com as ideias que exprimiam e pelo seu significado (HODGE, 2000). Por outro lado, Lancaster (1986) afirmou que o tesouro tem suas origens no Unitermo, com influência da teoria da classificação facetada. A influência da teoria da classificação facetada americana para os tesouros se reflete na indexação alfabética de assunto adotada por Charles Cutter, e a base europeia apoia-se nos princípios analítico-sintético de Ranganathan.

Mortimer Taube, para desenvolver o Unitermo, em 1951, baseou-se no fato de que uma ideia poderia ser representada por uma única palavra. A partir do Unitermo, o Centro de Engenharia de Informação DuPont, em 1959, desenvolveu o primeiro tesouro. Na década de 60, os tesouros foram aperfeiçoados, tanto pela perspectiva americana como pela europeia.

No contexto americano, teve início o desenvolvimento de vários tesouros, principalmente na área de Engenharia, culminando com a elaboração de normas para a construção de tesouros pela American National Standards Institute - ANSI/NISO e pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO).

Na Europa, o Classification Research Group (CRG), em Londres, teve grande impacto no desenvolvimento de pesquisa em relação à classificação facetada de Ranganathan.

Para a construção de vocabulário, na perspectiva europeia, prevalece a apresentação sistemática (BSI 5723), enquanto, na americana, a ênfase é dada à base linguística (LANCASTER, 1986).

A linha mais recente para o desenvolvimento de tesauro se apoia na teoria dos conceitos de Dahlberg com a teoria da classificação, denominada de tesauro-com-base-em conceitos, ou tesauros terminológicos, em que prevalece o uso do conceito para o desenvolvimento de hierarquias (relações lógicas). Na teoria de Dahlberg, “[...] o conceito se estabelece quando três elementos estão presentes: o referente/objeto, suas propriedades e um termo que o sintetize, ou seja, que carregue com ele todas as propriedades do referente” (CAMPOS; GOMES, 2006, p. 354).

De acordo com Currás (2010, p. 99), o tesauro é “[...] uma linguagem especializada, normalizada, pós-coordenada, usada com fins documentários, onde os elementos linguísticos que a compõem – termos simples ou compostos – se relacionam entre si sintática e semanticamente”.

O tesauro definido em relação à estrutura, para Cunha e Cavalcanti (2008, p. 312), “É um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente cobrindo um domínio específico do conhecimento” e, com referência à sua função, é

[...] um instrumento da terminologia empregado para traduzir em linguagem artificial (linguagem documentária, linguagem de indexação) a linguagem natural usada nos documentos e pelos indexadores ou pelos usuários, assim como para voltar à linguagem natural a partir da linguagem artificial (CUNHA; CAVALCANTI, 2008, p. 362).

Quanto ao tipo, podem ser classificados em: a) monolíngues ou multilíngues; b) macrotesauros ou microtesauros; e c) multidisciplinares e de disciplina específica (CUNHA; CAVALCANTI, 2008, p. 362).

São sistemas totalmente estruturados, que têm por objetivo proporcionar um instrumento idôneo para o armazenamento e a recuperação da informação em áreas especializadas. Até o presente, podem ser consideradas as

ferramentas mais refinadas para a representação do conhecimento. Consideram e identificam as relações de sinonímia, de hierarquia e associativa entre conceitos de um modo consistente (BARITÉ ROQUETA, 2011). Esses relacionamentos são geralmente representados pela notação TG (termo genérico), TE (termo específico), UP (usado para) e USE (termos equivalentes) e TR ou TA (termo relacionado ou termo associativo). A finalidade dessas remissivas é encaminhar o usuário para os termos preferidos pelos sistemas (CINTRA et al., 2005).

Percebe-se que o tesauro é um sistema multiuso, podendo ser utilizado para estabelecer categorias para bibliotecários e pesquisadores, ajudando-os na análise e na representação dos conteúdos dos recursos, assim como na elaboração de índices de documentos e coleções.

Para Miranda (2005, p. 147),

[...] um tesauro distingue-se de um simples vocabulário por duas características. A primeira é que as palavras nele listadas não descrevem, mas antes significam: **cada termo nomeia um conceito**. Sendo assim, esses termos que designam conceitos não são simples palavras, tornam-se 'termos', ou ainda 'descritores'. A segunda característica é que todos os termos estão relacionados entre si; nenhum termo pode figurar no tesauro sem estar relacionado a algum outro, sendo essa relação determinada pela definição de seu conceito (grifo do autor).

Os tesauros devem ser constantemente atualizados, suprimindo termos em desuso, reagrupando descritores e adicionando novos termos, o que possibilita acompanhar os avanços do conhecimento e a modificação de significados de termos já existentes (CINTRA et al., 2005).

Eles usualmente empregam uma indexação descritiva e corresponde a um sistema de pesquisa. O tesauro pode ser utilizado ainda como um recurso para pesquisa com motores de busca nas pesquisas em textos livres (TUDHOPE; NIELSEN, 2006). Tende a ser definido para um domínio específico ou para uma família de produtos que pode ser extensa e são definidos por padrões internacionais. Os padrões utilizados são definidos pelas normas ISO 2788/86 e a ISO 5964, ambas revisadas pela ISO 25964-1:2011 e a ANSI/NISO Z39.19-2005 no que se refere a construção, formatação e manutenção de tesauros monolíngues e multilíngues (NISO, 2005), assim como a BS 8723 - British Standards (ROWLEY, 2007).

A norma ISO 25964: *Information and Documentation – Thesauri and interoperability with other vocabularies* é precursora ao buscar estabelecer padrões que possibilitem a interoperabilidade entre os tesauros e outros tipos de SOC (CLARKE, 2011).

A norma está dividida em duas partes: parte 1 – tesouro para recuperação da informação, publicada em agosto de 2011, abrange:

conteúdo e construção do tesouro mono e multilingue;
orientação na aplicação de análise de facetas para tesauros;
orientação para o gerenciamento e a manutenção de tesauros; e
requisitos funcionais de *software* para gerenciamento do tesouro.

De acordo com Clarke (2011), essas ações possibilitam a compatibilidade entre tesauros com base nos padrões defendidos pela World Wide Web Consortium (W3C) em relação aos SOC, ou seja, os Simple Knowledge Organization Schemes (SKOS).

O SKOS é uma área de desenvolvimento de especificações e padrões para apoiar o uso de SOC no âmbito da Web Semântica, utilizando o Resource Description Framework (RDF).

[...] O RDF, que é uma aplicação XML, é desenhado para facilitar o *software* perceber o suficiente sobre um *Web site*, de modo a que possa descobrir recursos, catalogar o conteúdo do *site*, escalonar esse conteúdo, perceber que possui o conteúdo e sob que termos e a que preço este pode ser usado, e fazer outras coisas que um *Web spider* ou um agente inteligente possa querer fazer (HAROLD, 1999, apud BAPTISTA; MACHADO, 2001, p. 3-4).

A parte 2, que ainda está sendo desenvolvida, contemplará a interoperabilidade com outros vocabulários que incluem esquemas de classificação, taxonomias, cabeçalhos de assunto, ontologias, terminologias, lista de autoridades de nomes e anéis de sinônimos. Essa norma não visa padronizar todos esses vocabulários, mas descrever ações que permitam a interoperabilidade entre tesouro, especialmente como implementar mapeamentos¹⁸ entre esses vocabulários.

Moura (2011) enfatiza que Lancaster, em 1993, já alertava para a questão do uso da tecnologia no auxílio à representação e à recuperação da informação. O objetivo era criar metodologias que possam aproveitar a linguagem

¹⁸ Processo de identificação de termos, conceitos e relações hierárquicas que são aproximadamente equivalentes (DOERR, 2001).

natural com o “rigor e a complexidade” dos vocabulários controlados para a consecução desse processo.

Para Clarke (2011), essas ações buscam integrar os SOC ao ambiente de rede, principalmente à Web Semântica.

b) Redes semânticas

As redes semânticas têm suas origens no associacionismo de Aristóteles e do reducionismo (RINGLAND, 1988, apud MIRANDA, 2005). Quillian, em 1968, procurou definir um modelo de como o significado das palavras era representado na memória humana, e, a partir daí, construiu uma rede associativa de nós (nodos) que representavam as relações entre conceitos.

De acordo com Miranda (2005), uma rede semântica é uma estrutura para a representação do conhecimento definida como um padrão de nós interconectados por arcos rotulados.

Embora exista uma grande variedade de terminologias e notações nos sistemas de representação do conhecimento por redes semânticas, as seguintes considerações são comuns na maioria delas:

- Nodos na rede representam conceitos de entidades, atributos, eventos e estados.
- Arcos na rede, geralmente chamados de relações conceituais, representam as relações mantidas entre os conceitos. Os rótulos sobre os arcos especificam os tipos de relações.
- Algumas relações conceituais representam casos de linguística, como agente, paciente, receptor, ou instrumento. Outros casos em outras relações representam conectivos espaciais, temporais, causais e lógicos. Outros ainda especificam o papel de uma entidade em relação a outro, como por exemplo, mãe, proprietário, residência etc.
- Os tipos de conceitos são organizados em uma hierarquia de acordo com os níveis de generalidade. Esta hierarquia é frequentemente chamada de *hierarquia de tipos* ou *hierarquia taxionômica*.
- Relações mantidas entre todos os conceitos de um determinado tipo são herdadas através da hierarquia por todos os seus subtipos (MIRANDA, 2005, p. 155-156).

O processamento da linguagem natural impulsionou importantes trabalhos na área de redes semânticas. Esse SOC apresenta conceitos e termos estruturados não como hierarquias, mas como uma rede. Os conceitos são dispostos como nós com vários relacionamentos que vão se expandindo. Os relacionamentos geralmente seguem o padrão termo geral (TG), termo próximo (NT) e termo relacionado (TR). Esses incluem relacionamentos todo-parte específicos, causa-efeito, pai-filho etc.

Uma das redes semânticas mais bem elaboradas é da Princeton's WordNet, que agora pode ser utilizada com uma variedade de mecanismos de busca. A WordNet é um banco de dados léxico eletrônico que contém informações sobre palavras e expressões compostas (*phrasal verbs*, colocações, frases idiomáticas etc.). As entradas são separadas de acordo com categorias sintáticas: substantivo, verbo, adjetivo e advérbio. Dentro de cada categoria, são armazenadas várias relações semânticas entre palavras e expressões compostas (CARLAN, 2010; MIRANDA, 2005).

c) Ontologia

Ontologia é um ramo da filosofia que lida com a natureza e a organização do ser. Esse termo foi introduzido por Aristóteles em *Metafísica*, que estudou o ser e suas propriedades (GRUBER, 1996).

No sentido filosófico, ontologia “[...] é o estudo do que existe e do que admitimos que existe, para conseguir uma descrição coerente da realidade”, e Currás (2010, p. 38) traz esse conceito para a Ciência da Computação de forma bastante simples: é o estudo da transformação da linguagem natural em uma linguagem codificada para obtenção de uma resposta coerente da realidade.

A inteligência artificial e a gestão do conhecimento adotaram a ontologia para se referir a conceitos e termos utilizados para descrever ou construir uma representação de alguma área do conhecimento.

Várias são as definições encontradas na literatura para a ontologia, e essas diferenças estão associadas aos diferentes usos das ontologias pelas diversas áreas de conhecimento. Diante desse fato, focaremos nas definições voltadas para a área da CI, como a adotada por Currás (2010, p. 41): “É a descrição sistemática das entidades e suas modalidades, e das regras que permitem

descrever um domínio específico de acordo com entidades e processo que permitem descrever ‘todas’ as coisas e os processos”. Para a autora, diante das várias definições encontradas na literatura, identifica-se uma relação entre as ontologias e os sistemas de classificação, por esse reproduzir a linguagem natural em códigos.

Já para Barité Roqueta (2011), ontologias são desenhos de estruturas funcionais, que contêm entidades ou elementos que se relacionam entre si, com um determinado objetivo em um ambiente digital.

As ontologias definem conceitos e relações de alguma área do conhecimento, de forma compartilhada e consensual, e promovem e facilitam a interoperabilidade entre sistemas de informação, em um processo ‘inteligente’ dos agentes (computadores) (BRASCHER; CARLAN, 2010, p. 160).

O termo ontologia, algumas vezes, é entendido como SOC, particularmente se este é representado utilizando Web semântica. No entanto, se adotado para modelagem da Inteligência Artificial e proposta de inferência, as ontologias tendem a ter mais precisão, definição e relação formal com os sistemas de conhecimento, como os serviços de terminologia¹⁹ (TUDHOPE; KOCH; HEERY, 2006).

No contexto da Web, a utilização de ontologias é crucial: permite que agentes de *software* compreendam a semântica embutida em definições e vocabulários especificados em relação a um domínio, sem ambiguidades, viabilizando o intercâmbio de informações através de consultas (MOURA, 2001, apud MIRANDA, 2005).

Miranda (2005) reproduz o quadro elaborado por Almeida e Bax (2003) em que os autores detalham tipos de ontologias (Quadro 3).

¹⁹ “É um conjunto de serviços de extração, apresentação e aplicação de vocabulários, controlados e não controlados, incluindo os conceitos membros, termos e relacionamentos que descrevem o significado dos termos e facilitam a interoperabilidade semântica para fins de pesquisa, navegação, raciocínio semântico, indexação e classificação de assunto” (TUDHOPE; KOCH; HEERY, 2006, p. 42, tradução nossa).

Quadro 3 - Tipos de ontologia

ABORDAGEM	CLASSIFICAÇÃO	DESCRIÇÃO
Quanto à função Mizoguchi, Vanwelkenhuysen e Ikeda (1995)	Ontologias de domínio	Reutilizáveis no domínio, fornecem vocabulário sobre conceitos, seus relacionamentos, sobre atividades e regras que os governam.
	Ontologias de tarefa	Fornecem um vocabulário sistemático de termos, especificando tarefas que podem ou não estar no mesmo domínio.
	Ontologias gerais	Incluem um vocabulário relacionado a coisas, eventos, tempo, espaço, casualidade, comportamento e funções. etc.
Quanto ao grau de formalismo Uschold e Gruninger (1996)	Ontologias altamente informais	Expressa livremente em linguagem natural.
	Ontologias semi-informais	Expressa em linguagem natural de forma restrita e estruturada.
	Ontologias semiformais	Expressa em uma linguagem artificial definida formalmente.
	Ontologias rigorosamente formal	Os termos são definidos com semântica formal, teoremas e provas.
Quanto à aplicação Jasper e Uschold (1999)	Ontologias de autoria neutra	Um aplicativo é escrito em uma única língua e depois convertido para uso em diversos sistemas, reutilizando as informações.
	Ontologias como especificação	Cria-se uma ontologia para um domínio, a qual é usada para documentação e manutenção no desenvolvimento de <i>softwares</i> .
	Ontologias de acesso comum à informação	Quando o vocabulário é inacessível, a ontologia toma a informação inteligível, proporcionando conhecimento compartilhado dos termos.
Quanto à estrutura Haav e Lubi (2001)	Ontologia de alto nível	Descrevem conceitos gerais relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objeto, evento, ação etc), os quais são independentes do problema ou domínio.
	Ontologias de domínio	Descrevem o vocabulário relacionado a um domínio, como, por exemplo, medicina ou automóveis.
	Ontologias de tarefa	Descrevem uma tarefa ou atividade, como, por exemplo, diagnósticos ou compras, mediante inserção de termos especializados na ontologia.
Quanto ao conteúdo Van-Heijst, Schreiber e Wielinga (2002)	Ontologias terminológicas	Especificam termos que serão usados para representar o conhecimento em um domínio (por exemplo, os léxicos).
	Ontologias de informação	Especificam a estrutura de registros de bancos de dados (por exemplo, os esquemas de bancos de dados).
	Ontologias de modelagem do conhecimento	Especificam conceitualizações do conhecimento, têm uma estrutura interna semanticamente rica e são refinadas para uso no domínio do conhecimento que descrevem.
	Ontologias de aplicação	Contêm as definições necessárias para modelar o conhecimento em uma aplicação.
	Ontologias de domínio	Expressam conceitualizações que são específicas para um determinado domínio do conhecimento.
	Ontologias genéricas	Similares às ontologias de domínio, mas os conceitos que as definem são considerados genéricos e comuns a vários campos.
	Ontologias de representação	Explicam as conceitualizações que estão por trás dos formalismos de representação do conhecimento.

Fonte: Almeida e Bax (2003 apud MIRANDA, 2005, p. 152)

As ontologias fornecem um vocabulário para representação do conhecimento, que possui um conceito que o sustenta, não permitindo ambiguidade. Em virtude da ontologia ser escrita em uma linguagem formal, não há distorções semânticas como acontece na linguagem natural. Dá para utilizar o mapeamento sem que seja alterado o conceito, que pode ser traduzido para outras línguas (GUIMARÃES, 2002).

Com base em Maedche (2002) e Gómez-Perez (1999), Guimarães (2002) explica que uma ontologia é formada por:

- a) um conjunto de conceitos e uma hierarquia entre esses conceitos, uma taxonomia;
- b) um conjunto de relacionamentos entre esses conceitos;
- c) um conjunto de funções;
- d) um conjunto de axiomas, regras que representam sempre a verdade; e
- e) um conjunto de instâncias, que é um conjunto prévio existente na ontologia.

Guarino (1998 apud RAMALHO, 2010) afirma que a taxonomia representa a espinha dorsal de uma ontologia.

Com base nos SOC analisados neste estudo, o Quadro 4 traz um resumo das principais características, vantagens e desvantagens na utilização desses SOC pelos RIs.

Quadro 4 - Síntese comparativa dos SOC

TIPO DE SOC	CARACTERÍSTICAS	VANTAGEM	DESVANTAGEM
Classificação e categorização			
Cabeçalhos de assunto	Possuem estrutura pré-coordenada Descrevem elementos e relações São baseados, em grande parte, em aspectos linguísticos não conceituais São usados para indexação/recuperação em ambientes virtuais Adicionam qualificadores aos termos e diminuem a ambiguidade	Apresentam maior precisão na recuperação Cobrem todos os campos do conhecimento	Requerem catalogadores e indexadores treinados Exigem, em ambientes digitais, sistemas com capacidade de manusear Apresentam inconsistência na estrutura Não possuem estrutura classificatória, o que impossibilita a sua utilização para o desenvolvimento de ontologia
Classificação de Bliss	Utiliza faceta Apresenta níveis integrativos Usa notação sucinta Tem estrutura pré-coordenada	É utilizada em ambiente Web	Não fornece relação sintática
Classificação de Dois Pontos	Possui esquema totalmente facetado Apresenta sintaxe rigorosa Usa símbolos alfanuméricos e de conexão Tem estrutura pré-coordenada	É utilizada em ambiente Web	Não fornece relação sintática
CDD	Usa números decimais para representar todas as ideias Adota a hierarquia Dá ênfase à prática Tem estrutura pré-coordenada e pré-facetada	É utilizado intencionalmente em larga escala e os usuários possuem alto grau de conhecimento sobre ele É bem provido de recursos É consistente Tem hierarquia numérica Está disponível em outros idiomas Está disponível em formato eletrônico	Não é atualizada com frequência Não tem circulação em áreas temáticas Tem custo alto Recebe forte influência americana
CDU	Usa hierarquia Possibilita cobertura abrangente de assunto Possui compromisso com atualização Está disponível em outros idiomas É amplamente utilizado em bibliotecas acadêmicas europeias Prioriza terminologia europeia/britânica Possui notação mista (alfanumérica) Apresenta sintaxe relevante Tem estrutura pré-coordenada e pré-facetada	É traduzida em 93 idiomas Tem projeção de disponibilização eletrônica	Possui notação complexa e pode ser vista como difícil pelos usuários Não é tão bem amparada em termos de ferramentas e recursos, se comparada com os esquemas americanos

Continua

Continuação

TIPO DE SOC	CARACTERÍSTICAS	VANTAGEM	DESVANTAGEM
LCC	Baseia-se na garantia literária Possui versão Web É amplamente utilizada internacionalmente É bem provida de recursos Apresenta estrutura pré-coordenada Dispõe de sistema de classificação enumerativa Apresenta notação mista (alfanumérica) Usa índice para cada classe	Pede grandes atualizações, já que é composto por séries de classificações individuais, realizadas por especialistas. É sistemática, elástica, expansiva e possui notação econômica e, em muitos casos, tem valor mnemotécnico Possui programa de compatibilização com a notação de Dewey É criada em uma base <i>ad hoc</i> para descrever documentos individuais em vez de domínios específicos do conhecimento Possui estrutura falha, não tem consistência Não apresenta especificidade sobre determinados temas É muito extensa e complexa para não especialistas Tem forte influência americana Pode ser mapeada para DDC, MeSH, AAT É disponibilizada em formato eletrônico	Está muito impregnada da cultura americana O esquema enumerativo adotado por esse sistema transforma-o em uma estrutura rígida que não permite hospitalidade. Faz atualizações em longo prazo
ESQUEMA DE CATEGORIZAÇÃO	Processa linguagem natural por técnica de indexação automática	Reduz o tempo de processamento	Possibilita eliminação de termos relevantes
TESAURO	É um sistema pré-coordenado e pós-coordenado de termos de linguagens documentárias Pode ser estruturado manual e mecanicamente Apresenta organização lógico-hierárquica Ordena os termos que o compõem, em hierarquias e relações semânticas e sintáticas Sistematiza o conhecimento, usando métodos lógicos, coerentes e científicos estabelecidos segundo normas prefixadas É utilizado mais pela área de CI Pode ser geral ou especializado; multidisciplinar e monodisciplinar; principal e auxiliar ou marginal É alfabético ou sistemático; macrotesauro ou microtesauro Pode ser hierárquico, facetado e gráfico, quando for tesauro sistemático Pode ter estrutura alfabética, sistemática e/ou gráfica Adiciona qualificadores aos termos e diminui ambiguidade	Apresenta especificidade, maleabilidade e capacidade de descrever as informações de forma completa Possibilita ao usuário identificar assuntos relacionados que podem ser de interesse na busca Apresenta termos acompanhados de definições expressas em linguagem natural Possui normas internacionais para a elaboração	Tem aplicação direta limitada no ambiente Web, devido a herogeneidade de formatos e estruturas e rápido crescimento de recursos e conteúdo. Não possui padrões adequados e modelos para representar diferentes níveis de abstração na Web (XML, RDF) e ainda restringe a sua utilização

Continua

Conclusão

TIPO DE SOC	CARACTERÍSTICAS	VANTAGEM	DESvantAGEM
Modelo de relacionamento			
TAXONOMIA	Tem lista estruturada de termos/conceitos de um domínio Apresenta termos sem definição, apenas relações hierárquicas (gênero-espécie e parte-todo) Possibilita a organização e a recuperação de informação por meio da navegação Evidencia o modelo conceitual de um domínio Permite agregação de dados É preparada por pessoal especializado com foco no usuário Tem gestão centralizada do sistema de classificação Apresenta elevado grau de precisão Sistematiza o conhecimento, usando métodos lógicos, coerentes e científicos estabelecidos segundo normas prefixadas É um sistema pré-coordenado e pós-coordenado de termos de linguagens documentárias É ordenado por unidades conceituais, semelhantes às palavras-chave, segundo uma hierarquia, e estabelece conexões de relações semântica e sintática É policotômica	É interativa e dinâmica É usada como mecanismo de pesquisa em conjunto com ferramentas de busca É adotada por bibliotecas digitais, portais organizacionais, Web semântica, ontologias, gestão da organização e do conhecimento Permite a visualização do conteúdo de forma integrada Proporciona uma busca rápida Expõe a relação entre os assuntos/documentos Apresenta a informação e a estrutura classificatória Permite uma interface amigável Possibilita a interoperabilidade entre sistemas quando fazem parte de uma ontologia Evita ambiguidade Identifica as relações de sinonímia, hierarquia e associativa entre conceitos de forma consistente	Possui formatos das tabelas voltados mais para uso em bibliotecas do que para processamento automático de recursos Não apresenta ausência de requisitos explícitos para sua construção, suas regras e seus métodos Necessita de uma estrutura sofisticada para uma busca eficaz É restritiva em suas possibilidades de exploração por conter apenas relações hierárquicas e partitivas Não permite a utilização de operadores booleanos
REDES SEMÂNTICAS	Apresenta conceitos e termos estruturados não como hierarquias, mas como uma rede. Têm conceitos dispostos como nós com vários relacionamentos que vão se expandindo		
ONTOLOGIA	Ordena os termos em relação às suas peculiaridades e propriedades com o objetivo de estabelecer categorias Possui as seguintes propriedades: • compartilhamento: entendimento sobre ideias – agentes de <i>software</i> devem ter um mesmo entendimento sobre um determinado conceito; • filtragem: define o que deveria ser extraído de um sistema – de forma a construir um determinado modelo desse sistema Apresenta propriedades dos conceitos Disponibiliza hierarquia de conceitos Apresenta conjunto de axiomas	Possibilita o reuso Preserva semanticamente conhecimentos e não permite ambiguidade Proporciona a interoperabilidade semântica Permite um conhecimento compartilhado de termos para ajudar no intercâmbio de informações entre pessoas e programas Apresenta comunicação e especificação de sistemas Confiabilidade Processa os textos em linguagem natural Possibilita usabilidade Representa e recupera a informação	É um trabalho complexo e com custo elevado

Fonte: Carlan (2010), Cintra et al. (2005), Curras (2010), Galho e Moraes (2012), Gomes e Guimarães (2012), Nicholson et al. (2001), Miranda (2005), San Segundo Manuel (1996), Souza (2004), Zeng (2008)

Como foi abordado, muitas pesquisas têm sido desenvolvidas não apenas relativas ao uso dos SOC, mas também relacionadas com os próprios sistemas. No próximo item, há uma descrição de trabalhos em que os SOC são explorados em relação a vários usos.

3.2 OS SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO E SEUS USOS

De acordo com Patel et al. (2005), há muito se fala que, com o advento da Internet, não seria mais necessário adotar SOC para indexação de assunto, devido à utilização de pesquisa em texto completo empregando motores de busca. Atualmente, pode-se verificar que isso não vem ocorrendo. Ao contrário, serviços de indexação e estruturação de informação coletiva têm sido desenvolvidos, como a folksonomia, a utilização de *tags* de imagens no Flickr.

Hodge (2000) apresentou um relatório em que faz considerações sobre a utilização dos SOC em ambientes de BDs. A princípio, os SOC foram utilizados por bibliotecários e por pesquisadores, porém, com o aumento dos dados eletrônicos, a explosão das publicações eletrônicas e, conseqüentemente, a dificuldade em localizar essas informações, o interesse por esses sistemas tem envolvido não apenas outros profissionais como também os usuários.

O autor descreve algumas experiências em que os SOC são utilizados para as mais diversas necessidades: por editoras, a fim de fornecer o acesso aos seus textos por meio de buscas, empregando esquemas de classificação e categorização; por organizações governamentais, com finalidades estatísticas, adotando esquemas de classificação e arquivos de autoridades.

Em relação às BDs, Hodge (2000), ao se reportar ao acesso alternativo de assunto, em que existe a possibilidade de busca por assunto por diferentes públicos que não compartilham de terminologia comum, indica que podem ser utilizados os seguintes mecanismos:

- a) indexação ou classificação dos recursos empregando múltiplos esquemas – classificam-se e indexam-se os recursos com múltiplos esquemas, porém há um custo mais elevado;
- b) manutenção de esquemas originais com outros que possam contribuir com a BD – termos coletados de outras organizações

podem ser visualizados por pontos de acesso alternativos, criando-se outras terminologias para acessar a coleção; e

- c) mapeamento entre o esquema principal e esquemas alternativos – é baseado no mapeamento entre terminologias utilizadas por diferentes SOC. Quando as BDs possuem registros bibliográficos ou registros de metadados com maior granularidade²⁰, podem ser desenvolvidos sistemas de comutação que traduzem conceitos adotados pela BD ou pelo recurso.

Os SOC podem adicionar significado aos recursos de informação, sejam estes textuais ou multimídias. Isso é possível porque os SOC podem fornecer novas dimensões, dispostas em camadas, a um ou mais objetos multimídias. Várias ferramentas e serviços podem ser desenvolvidos para que uma pesquisa em texto gere um resultado gráfico ou visual (uso de *gazetteers*).

Algumas pesquisas vêm enfocando a atualização ou o desenvolvimento de novas tabelas de classificação, assim como o emprego delas no meio eletrônico. As pesquisas, nesse aspecto, têm sido mais de interesse da TI (GOMES, 2009).

A atualização de tabelas, como a CDD e a CDU, tem um aspecto em comum, o de torná-las o mais facetadas possível. A CDU, em especial, por possuir códigos para definir a relação entre as diferentes facetas, está sendo vista como um instrumento de recuperação importante, já que os códigos possuem valor semântico reconhecível pelo computador. Existe a questão das cadeias falsas que também vem sendo discutida.

Gomes (2009) identifica que o foco dessas tabelas é o aspecto do vocabulário da classificação, que permite ser utilizado como fonte de vocabulário da linguagem natural para a indexação coordenada ou para a comparação de termos usados na classificação automática, podendo ser ampliada como fonte de vocabulário da linguagem natural para uma terminologia mais adequada ao serviço de documentação.

O desenvolvimento de novas tabelas, em vista de aspectos restritivos de outros sistemas de OC, como os tesauros – satisfatórios na indexação,

²⁰ Refere-se ao tamanho, à decomponibilidade e ao grau em que um recurso é planejado para ser usado como parte de um recurso maior (NISO, 2007).

mas com deficiências na recuperação –, é atualmente um grande desafio. Os sistemas de classificação, por exemplo, não possuem a flexibilidade necessária para representarem a multidisciplinaridade dos documentos. O interesse nas tabelas está no fato de tratar-se de um sistema pré-coordenado que, adotando um sistema de notação, poderia representar a multidisciplinaridade desses documentos, uma vantagem em relação aos tesauros. Entretanto, seria necessário adotar *softwares* sofisticados.

Como exemplos de projetos de tabelas de classificação, Rafferty (2001, apud GOMES, 2009) cita:

- a) Projeto FATKS – emprega regras sintéticas, faceta e ordem de citação da BBC2, adota algumas regras e modelo de notação decimal expressiva, indicadores de faceta, modelo de auxiliares da CDU e a base geral do Broad System of Ordering (SLAVIC; BROUGHTON, 2003); e
- b) Projeto ILC – mostra a preocupação com a classificação e com a inclusão de elementos, relações explícitas que refletem não apenas um sistema de OC, mas também um modelo de representação. É construído com base na teoria dos níveis integrativos (ISKÓ, 2004). De acordo com Gomes (2009, p. 68-69), nessa teoria, “[...] o mundo se organiza em níveis de complexidade crescente: cada nível só pode existir a partir do nível anterior, o que não significa que as propriedades sejam herdadas nos níveis subsequentes”.

Gomes (2009) ainda relata que tabelas de classificação, tesauros e ontologias têm sido construídas pelo método analítico-sintético, denominado também de facetado, e o mais adotado para construção dos SOC, incluindo também as taxonomias. O emprego desse método está baseado na vantagem de produzir classificação coerente, intuitiva e autoexplicativa, além de permitir o acesso múltiplo.

Fujita (2001) salienta os problemas que os sistemas de classificação apresentam em relação à OC, no que diz respeito à inflexibilidade hierárquica de assuntos e subassuntos contida nas grandes classes, e a dificuldade em trabalhar com a interdisciplinaridade dos assuntos, também provocada pela inflexibilidade das hierarquias de assuntos. Uma das explicações para esse resultado é respaldada pela adoção de mecanismos que utilizam conceitos que possibilitam o entendimento

por áreas de conhecimento, e dessa forma pode atender a demandas das diversas áreas do saber.

Gomes (2009) aborda a representação do conhecimento e as redes semânticas na visão de alguns autores, como Vickery (1997), que alertam sobre a importância da ontologia e das taxonomias para a área da computação. As ontologias, por incorporarem adjetivos, expressões sintáticas e variações morfológicas, atenderiam às necessidades de *software* no processamento das buscas. As taxonomias apoiariam a indexação automática e auxiliariam na navegação nos *sites* e no controle de processos nas empresas.

De acordo com Sure e Studer (2005), citados por Sayão e Marcondes (2008), as ontologias são representadas caracteristicamente por uma linguagem de representação processada por computador e consideradas uma das tecnologias-chave para o desenvolvimento da *Web Semântica*. Desse modo, por meio do estabelecimento de esquemas comuns na forma de ontologias, as tecnologias semânticas permitem a descrição de objetos em repositórios digitais com o objetivo principal de proporcionar a interoperabilidade semântica.

As ontologias são vistas como a nova geração de sistemas de informação²¹, que deverão ser capazes de resolver o problema da interoperabilidade semântica. Esses sistemas deverão ter habilidade para processar o entendimento que o usuário possui do mundo com as fontes de informação. Para tal, o projeto de ontologias exerce um papel de fundamental importância (FONSECA; ENGENHOFER; BORGES, 2000): a conceitualização é uma visão abstrata e simplificada de um domínio específico da realidade. Dessa forma, as ontologias são projetadas para possibilitar que o conhecimento seja compartilhado e reusado; elas explicam como um indivíduo, grupo, linguagem ou ciência entendem um determinado domínio. Ontologia é um componente importante no processamento de conhecimento por computador e, consequentemente, para a solução de interoperabilidade semântica entre repositórios digitais federados, principalmente no mapeamento entre esquemas de representação distintos. A DAMLOIL e a *Web Ontology Language* (OWL), publicada pela World Wide Web Consortium (W3C), são linguagens usadas para aplicações que necessitam compreender o conteúdo da

²¹ Sistema automatizado ou manual que abrange pessoas, máquinas e/ou métodos organizados para coletar, processar, transmitir e disseminar dados que representam informação para o usuário e/ou cliente (SISTEMA DE INFORMAÇÃO, WIKIPEDIA).

informação e frequentemente adotadas na implementação de arquiteturas federadas (DOERR, 2006; RODRIGUEZ MARTINEZ, 2006).

Para Poli (1996), referenciado por Gomes (2009, p. 77), “[...] as ontologias são o lado objetivo da realidade. Elas requerem um documento que defina formalmente as relações entre os termos”, enquanto os tesauros “[...] empregam somente termos (substantivos) e apenas indicam a existência de relações entre elas (no plano conceitual e da língua)”. Já Gilchrist escreve que “[...] as taxonomias e ontologias definem esses relacionamentos e podem incorporar adjetivos, expressões sintáticas, variações de sorte que possam ser usados no *software* para processamento das buscas” (apud GOMES, 2009, p. 77).

Como afirma Hjørland (2007, p. 5), “[...] os registros bibliográficos (metadados de registros completos) é o que define a OC na CI. A OC, no sentido mais abrangente, é relevante para a CI à medida que contribui para concepção, uso e avaliação de registros bibliográficos [...]”. Esses sistemas são utilizados para a organização e representação dos documentos e também para desenvolvimento de atividades, como descrição, indexação e classificação, executadas por profissionais da CI.

O estudo de Shiri (2003) apresentou tendências e desafios abordados em três conferências internacionais no ano de 2002: Join Conference on Digital Libraries; VI European Conference on Research and Advanced Technology for Digital Libraries and V International Conference on Asian Digital Libraries. Foram abordados os seguintes aspectos: arquitetura, sistemas, ferramentas e tecnologias; conteúdo digital e coleções; metadado; interoperabilidade; padrões e SOC; usuários e usabilidade; e questões em relação a direito, organização e aspectos econômicos e sociais.

Quanto aos metadados, a necessidade de esquemas bem estruturados para descrição dos objetos e conteúdo e os vários níveis de granularidade, assim como a utilização de algoritmos para fornecimento de metadados; a escolha de padrões de metadados e o mapeamento entre os diferentes formatos de metadados foram os aspectos mais pontuados.

A interoperabilidade destacou-se como um dos assuntos mais discutidos nesses eventos, assim como os SOC. A respeito dos SOC, as questões são: o uso de tesauro e sistemas de classificação para pesquisa cruzada e navegação (*browsing*) entre bibliotecas digitais; a criação de ontologias utilizando

tesauros já existentes; os sistemas de classificação e vocabulários controlados especializados para facilitar uma representação do conhecimento geral com diferentes tipos de materiais; o uso de taxonomia para fornecer acesso unificado e organizado entre variados repositórios, por meio de diferentes camadas da coleção digital. Houve discussão também sobre como as ferramentas podem interagir entre elas e a utilização de mapeamento e interoperabilidade entre os vários SOC.

Em vista dos temas debatidos, o autor apontou futuros desafios em pesquisa que envolvem as bibliotecas digitais (BDs):

- a) sistemas de federação de bibliotecas digitais, o uso de pesquisa avançada e características de recuperação, como relevância e tempo de recuperação; e
- b) utilização de SOC, como as ontologias, entre outros.

Assim como os temas abordados por Shiri (2003), um outro assunto se destaca nas pesquisas: as múltiplas terminologias, com o aumento exponencial das informações na *Web*, têm se tornado ponto inquestionável de preocupação entre os pesquisadores das mais diversas áreas em relação à recuperação da informação.

O acesso multilíngue é outro aspecto bastante relatado na literatura (NAGAMORI; SUGMOTO, 2006). Hodge (2000) afirma que uma variedade de fontes, que inclui dicionários ou tesauros multilíngues, pode propiciar esse tipo de acesso.

Pesquisas têm analisado o mapeamento como alternativa para a promoção da interoperabilidade entre as terminologias. Entre elas, o Projeto HILT, que estudou a questão da terminologia adotando uma abordagem de mapeamento para alcançar a compatibilidade entre sistemas utilizados pelo JISC (LCSH, UNESCO, MeSH, CDD).

O Aquarelle Terminology Service optou por implementar um sistema em que os termos são mantidos e os tesauros gerenciados localmente, com um servidor de termos central coordenando a recuperação. Essa técnica foi considerada bastante trabalhosa, na visão de Doerr e Fundulaki (1998, apud MCCULLOCH, 2004), pelo fato de que um servidor de termos deveria ser alimentado com expressões equivalentes ao significado dos termos em diferentes autoridades, por técnicos ou por meio de métodos linguísticos controlado pelo ser humano.

A União Europeia financiou o projeto Renardus, que implementou mapeamentos entre as terminologias utilizadas pelas centrais Resource Discovery

Network e a CDD para fornecer uma interface de navegação central para *gateways* de assunto. Esse projeto conseguiu demonstrar que o mapeamento pode ser uma alternativa para resolver os problemas de recuperação da informação causados pelas barreiras linguísticas. Assim, ficou demonstrado que a escalabilidade do método sugere ser uma solução universal aceitável. O projeto utiliza o protocolo Z39.50. No que se refere à interoperabilidade semântica em nível de dados categóricos, o Renardus adota o mapeamento para integrar a CDD com diferentes sistemas de classificação utilizados pelos participantes do portal (MCCULLOCH, 2004; TUDHOPE; KOCH; HEERY, 2006a). Uma das restrições encontradas nesse método, em referência à interoperabilidade, está relacionada com a identificação de áreas de interesse dentro da hierarquia da CDD (MCCULLOCH, 2004).

O custo com a utilização do mapeamento é um fator também considerado. Existe a necessidade de manutenção de um servidor de terminologias e implementação de atualizações do sistema e variações locais, devido à complexidade das buscas feita pelos usuários.

Duas outras técnicas são expostas pelo autor, a tecnologia Smartlogik, para desenvolver termos orientadores, que procuram terminologias múltiplas simultaneamente para retornar termos que satisfaçam aos usuários. Essa tecnologia utiliza vocabulários controlados, como taxonomia e ontologias, para o desenvolvimento de classificação automática de informações a partir do texto de resumos, removendo o elemento humano de identificação de equivalências em terminologias diferentes e melhorando a precisão da pesquisa e a integração de conteúdo.

A outra técnica é a de Clustering. Adotada pelo sistema de Cheshire, desenvolvido pela JISC/NSF e pesquisadores das Universidades de Liverpool e da Califórnia, em Berkely, ela retorna fontes semanticamente relacionadas para o usuário, criando vínculos entre os campos de metadados do material associado pelo assunto coberto (CHESHIRE, 1999, apud MCCULLOCH, 2004). A técnica não melhorou a compatibilidade entre as terminologias diretamente, mas serviu para criar conexões entre os registros do catálogo. O Sistema Cheshire II foi desenvolvido para tentar resolver questões com a pesquisa em catálogos *online* de busca em relação a insuficiência e sobrecarga de informações, e estabelecer ligações entre catálogos *online*, textos completos e material multimídia.

Por fim, McCulloch (2004, p. 300) sugere que, “[...] para a recuperação cruzada e multidisciplinar, as normas internacionais devem ser aplicadas quando várias terminologias estão em uso, ou se os recursos estão acessíveis através de *gateways* múltiplos”. E o autor acrescenta: “[...] onde quer que a atividade de catalogação seja realizada, o desenvolvimento e o aperfeiçoamento de métodos para harmonização dos termos de assunto de diferentes vocabulários controlados devem ser alcançados”.

A pesquisa a textos livres é uma das principais atividades na *Web*, e um número muito pequeno desses recursos utilizam metadados ou vocabulários controlados que possibilitem a sua recuperação, já que muitos empregam a linguagem natural, fator que prejudica a recuperação (PARK, 2009). O autor identifica que o SOC é uma alternativa que pode auxiliar na recuperação desse recurso. Ele seria utilizado por uma ferramenta de mineração. Entretanto, o autor salienta que, embora os SOC possam realizar buscas de texto livre em BD e integrar recursos existentes na *Web* e na BD, eles devem ser utilizados com cautela. Deve-se considerar o domínio em que os SOC são aplicados e a complexidade do sistema, pois, do contrário, o resultado pode ser irrelevante para o usuário.

A implementação de SOC em BD precisa considerar aspectos como:

- a) necessidade do usuário – deve ser analisada em termos de conteúdo e funcionalidade, de possibilidade e relacionamento com o ambiente interno e externo à biblioteca, observando se o acesso ao SOC vai ser disponibilizado ou não para o usuário;
- b) identificação e localização do SOC apropriado – o custo e o tempo de construção de um SOC devem ser analisados. Se for utilizado um SOC externo, deve-se observar como é feito o acesso, se está disponível ou não na *Web*, por exemplo, já que existem problemas com o acesso remoto. Caso seja instalado localmente, problemas com a atualização também devem ser levados em consideração;
- c) utilização de *hardware*, *software* – redes que suportem a integração e a manutenção;

- d) aquisição e direitos de propriedade intelectual – aspectos em relação ao uso devem ser avaliados para não incidir sobre os direitos de propriedade intelectual;
- e) manutenção e atualização – no estabelecimento de um SOC, torna-se necessário analisar como as informações dos recursos da BD serão relacionadas e identificadas. O mecanismo a ser utilizado vai depender de como essas informações são identificadas (metadados). Os metadados deverão ser analisados em relação à granularidade e ao nível de atribuição (livros, capítulos).

No autoarquivamento praticado pelos repositórios, em que os autores são os principais responsáveis pela inserção do metadado, há uma inversão de papéis quando a responsabilidade de um *expert* em executar essa tarefa é transferida para amadores (FERREIRA; BAPTISTA, 2005). Com essa alteração de paradigma, é criada uma situação em que os termos adotados por esses depositantes nem sempre terão consistência em relação aos que seriam adotados pelos profissionais, o que poderia dificultar a recuperação do recurso. Para Furnas et al., referenciados por Ferreira e Baptista (2005), a probabilidade de pessoas que não são especialistas em indexação utilizarem as mesmas palavras-chave que os especialistas é inferior a 20%. Embora também não exista uma completa concordância entre os indexadores especialistas. A partir desse contexto, os autores desenvolveram um *add-on* para o DSPACE que disponibiliza palavras-chave que podem ser utilizadas pelos usuários na fase de indexação. É apresentada uma taxonomia relacionada à comunidade (centros/departamentos) que disponibiliza termos que poderão ser adotados como descritores dos recursos. Dessa forma, a interoperabilidade entre os repositórios que utilizam o mesmo mecanismo é simplificada, pois a probabilidade de encontrar os mesmos vocabulários torna-se maior.

Park (2009) concluiu o relatório da sua pesquisa com algumas recomendações e alertas. Ele recomenda que os desenvolvedores de SOC busquem formas para que eles sejam utilizados no ambiente *Web* e alerta que é necessário o uso de um sistema de identificação único e persistente – recomendado pela NISO em 1999 –, além do estabelecimento de um conjunto padrão de metadados para descrição. Esse aspecto foi abordado também no Princípio 4 em

relação ao objeto digital formulado pela NISO, em 2007. Os metadados identificam recursos na *Web* como um SOC e fornecem informações importantes que permitem uma aplicação para uso remoto sem um conhecimento prévio do seu conteúdo ou da sua estrutura.

Na próxima seção, serão apresentados aspectos em relação aos metadados, como características, propriedades, utilização e benefícios, com o objetivo de buscar a associação entre os repositórios e a necessidade da interoperabilidade semântica. Também destacam-se as vantagens de se utilizar os SOC e o papel dos metadados para descrição de informações e documentos *on-line* na obtenção da interoperabilidade semântica.

4 METADADOS DESCRITIVOS

Nesta seção, são abordados os metadados descritivos, sua relação com os SOC e a sua importância para o acesso, a recuperação e o reuso da informação no ambiente Web.

Em 1969, Jack Myres criou o termo *metadata* (metadado) para definir os dados que descreviam registros de arquivos convencionais que, segundo Alvarenga (2003, p. 19), etimologicamente, representa “dado sobre dado”, dado que descreve a essência, os atributos e o contexto em que os recursos aparecem e também caracteriza suas relações, com o objetivo de fornecer acesso e uso potencial desse recurso.

De acordo com a NISO (2004, p. 1), os

Metadados são informações estruturadas que descrevem, explicam, localizam e, ainda, facilitam a recuperação, o uso ou o gerenciamento de recursos de informação. Metadados, constantemente, são denominados de dados sobre dados ou informação sobre informação (Tradução nossa).

Na visão de Campos (2007, p. 19), Gilliland-Swetland (2000, p. 1) empregou uma definição abrangente para metadados: é “[...] a soma total do que pode ser dito sobre algum objeto informacional em algum nível de agregação [...]”, o que traz um conceito de que o metadado é uma economia informacional.

Gilliland-Swetland (2000) acrescenta que, embora não estejam tão familiarizados com o termo quanto os profissionais da formação, os criadores e consumidores de conteúdo vêm se tornando hábeis criadores de metadado, a partir do momento que contribuem com as *tags* e as folksonomias. Estudantes de vários níveis estão sendo capacitados para utilizar programas que possam identificar a procedência de metadados, com o objetivo de verificar a credibilidade da informação. Ou seja, tanto os profissionais de informação como os criadores e os usuários digitais devem entender os papéis dos diferentes tipos de metadados: garantir a acessibilidade, prover autoridade e ser interoperável e escalável.

Dois aspectos são abordados pela autora: um chama a atenção dos profissionais da informação sobre a importância de identificar esquemas de metadados que auxiliem provedores e usuários, de forma a assegurar que os esquemas de metadados e os vocabulários controlados – tesauros e taxonomias,

incluindo a folksonomia – acompanhem as inovações e versões completas dos conjuntos de valores. O outro destaca que as terminologias utilizadas sejam apropriadas para a descrição dos conteúdos.

No entendimento de Relvão (2003, p. 5), os metadados aumentam o potencial informativo dos objetos por se tratar de dados que fornecem informação ou documentação de outros dados dentro de uma aplicação ou um ambiente, podendo conter informação descritiva em relação a contexto, qualidade, condição ou características dos dados.

Esses aspectos são estabelecidos por meio dos três principais tipos de metadados:

- a) estrutural – indica como os objetos são organizados e agrupados, possibilitando a navegação entre partes do objeto (livros, capítulos de livros);
- b) administrativo – fornece informações para o gerenciamento dos recursos, indicando quando e como foram criados, direitos e preservação;
- c) descritivo – descreve os recursos para descobertas e identificação, possibilitando a recuperação pelos usuários: título, autor e assunto (NISO, 2004, 2007).

Entre os benefícios provenientes da utilização de metadados, vale citar: a) documentar e organizar os recursos de forma estruturada; b) identificar informação sobre determinada área de interesse; c) facilitar e propiciar maior precisão na recuperação das informações; d) intercambiar informações entre aplicações e organizações; e) possibilitar a interoperabilidade entre as diversas fontes de dados e integrar recursos; f) permitir a agilidade e o acesso com qualidade na recuperação da informação; g) gerenciar a informação; h) preservar a informação; i) classificar conteúdos; j) promover acessibilidade, que possibilita a localização geográfica do recurso; k) garantir compatibilidade de uso, que indica um determinado fim; l) possibilitar acesso para aquisição de um conjunto de dados identificados; m) assegurar transferência, dados necessários para o processamento e o uso do conjunto de dados; n) organizar recursos eletrônicos; e o) auxiliar no armazenamento e preservação (GARCIA; MOURA; CAMPOS, 1999; NISO, 2004; PEREIRA; RIBEIRO JÚNIOR; NEVES, 2005; SOUZA; CATARINO; SANTOS, 1997).

De acordo com Campos (2007) e Gilliland-Swetland (2000):

- a) a fonte dos metadados pode ser interna (é gerado pelo criador do objeto informacional) e externa (gerado após ser criado o objeto, por outro que não é o criador);
- c) os metadados podem ser criados automaticamente pelos computadores ou manualmente por humanos;
- d) os metadados podem ter natureza leiga ou especializada;
- e) os metadados podem ter diferentes estados: estáticos, dinâmicos; longo prazo – para garantir seu acesso e uso; curto prazo – para transações;
- f) os metadados podem ter uma estrutura estabelecida (MARC) ou não ter nenhuma (simples anotações);
- g) a semântica pode ser controlada por meio de um vocabulário controlado ou não controlado; e
- h) os metadados podem se referir a uma coleção ou a um item individual.

Os metadados podem ser divididos pelas seguintes categorias: a) descritividade; b) modularidade; c) reflexividade; d) visibilidade; e) interoperabilidade; f) reusabilidade e extensibilidade (CHAN; ZENG, 2006). Essas categorias propiciam a flexibilidade e a automação. A flexibilidade possibilita o trabalho com inúmeras estruturas e padrões, de forma a produzir a interoperabilidade e a automação dos processos com o mínimo de intervenção humana. A flexibilidade também mantém um intercâmbio com padronização dos metadados e interoperabilidade. “Os metadados são intensamente explorados, em parte, devido à necessidade de padronizar mantendo a flexibilidade” (CAMPOS, 2007, p. 16).

A proliferação de informações no ambiente digital e a democratização de acesso às informações trazem um aspecto preocupante, que é o da padronização. Iniciativas isoladas dificultam a interoperabilidade, sendo necessário, no mínimo, o estabelecimento de elementos de metadados para que isso não aconteça (SHINTAKU; BRÄSCHER, 2007).

Nesse sentido, Campos (2007) desenvolveu uma revisão de literatura com o objetivo de identificar as funções de metadados nos diferentes contextos, com ênfase na CI, por meio da técnica de análise de conteúdo temática. Campos (2007) destacou que autores ressaltam diferentes utilizações dos metadados: descritividade, padronização, interoperabilidade, modularidade,

reflexibilidade, visibilidade, flexibilidade, automatização, administração e preservação. Os procedimentos adotados são usuais como padrões, modelos, linguagens e tecnologias, como a catalogação digital ML, RDF, Extensible Markup Language Schema (XMLS), ontologias, *data warehouses*, *Web* semântica e serviços *Web*. Entre os aspectos tratados, a autora identificou que os artigos sobre catalogação digital ou convenções de metadados abordam as categorias de descritividade e padronização. A interoperabilidade seria resultado, principalmente, da utilização do padrão.

De acordo com Gilliland-Swetland (2000) e Bax (2001), um objeto informacional apresenta quatro características principais: conteúdo, contexto, estrutura e apresentação. Conforme Campos (2007), a função descritiva dos metadados pode ser mais específica, a depender da identificação de cada característica. A categoria da descritividade do metadado enfatiza a descrição do conteúdo de um objeto, normalmente conferindo-lhe vários atributos, que o vinculam aos parâmetros da comunidade de catalogação.

A modularidade é considerada um componente do registro de metadado²², que pode ser entendida como uma peça de um quebra-cabeça. Pode ser colocada junto ou combinando peças de fontes de metadados provenientes de diferentes processos (humano ou máquina). Os metadados permitem, também, usar e reusar peça por peça quando novos registros precisarem ser gerados por humanos ou por máquinas. Esse processo de criação de metadados descritivos já vem sendo adotado por bibliotecários há muito tempo (ZENG; CHANG, 2006).

Duval et al. (2002) consideram a modularidade do metadado como um princípio fundamental para a organização de um ambiente caracterizado por diversas fontes de conteúdo, estilos de gestão de conteúdos e abordagens diferenciadas para a descrição de recursos. Isso permite que os formuladores de esquemas de metadados criem novos conjuntos de esquemas de metadados com base nas melhores práticas, o que é uma opção mais satisfatória do que criar novos elementos. No ambiente de metadados modulares, diferentes tipos de elementos de metadados (descritivo, administrativo, técnico, uso e preservação) de diferentes esquemas, aplicações de vocabulários e outros blocos de construção podem ser combinados de forma interoperável. A cada dia, as comunidades de informações

²² Conjunto de metadado de uma estrutura relevante, compreendendo todos os elementos relevantes, descrevendo um objeto.

veem a interoperabilidade como um dos princípios mais importantes para a implementação dos metadados (CHAN; ZENG, 2006).

Como definido no relatório da American Library Association Committee on Cataloging: Description and Access (CC:DA) Task Force on Metadata,

Um esquema de metadados fornece uma estrutura formal projetada para identificar a estrutura de conhecimento de uma determinada disciplina e para relacionar essa estrutura com a informação da disciplina por meio da criação de um sistema de informação que ajudará na identificação, na descoberta e no uso de informações (CHAN; ZENG, 2006, p. 2, tradução nossa).

O esquema de metadados é necessário para aumentar a interoperabilidade entre outros esquemas de metadados. O esquema de registro de metadados armazena esquemas de metadados e relações entre os termos de metadados para os usuários (NAGAMORI; SUGMOTO, 2006). São compostos pelos seguintes componentes: a) um conjunto de termos definidos para expressar propriedades de um recurso (elementos do Dublin Core - DC, *title*, *creator*, *subject*); b) um conjunto de termos que expressa os valores de um tipo de propriedades; c) um conjunto de regras que definem restrições estruturais e características neutras para alguns esquemas específicos de descrição para implementação (nível de mandato, repetibilidade); e d) um conjunto de regras vinculativas para uma linguagem de descrição específica (XML).

Ou seja, o esquema de metadados consiste em um conjunto de elementos designados para um objetivo específico, como a descrição de um tipo particular de recurso de informação (CHAN; ZENG, 2006; NISO, 2004).

Por exemplo, o elemento assunto é constante em quase todos os padrões de metadados. Nesses padrões, podem estar incluídos outros elementos que se relacionam com o assunto, em menor ou maior quantidade (ZENG, 2010). A definição ou significado dos elementos é conhecida como esquema semântico. Usualmente, os esquemas especificam os nomes dos elementos e a sua semântica. Podem, ainda, conter regras de conteúdo, como a utilização de termos de um determinado vocabulário controlado.

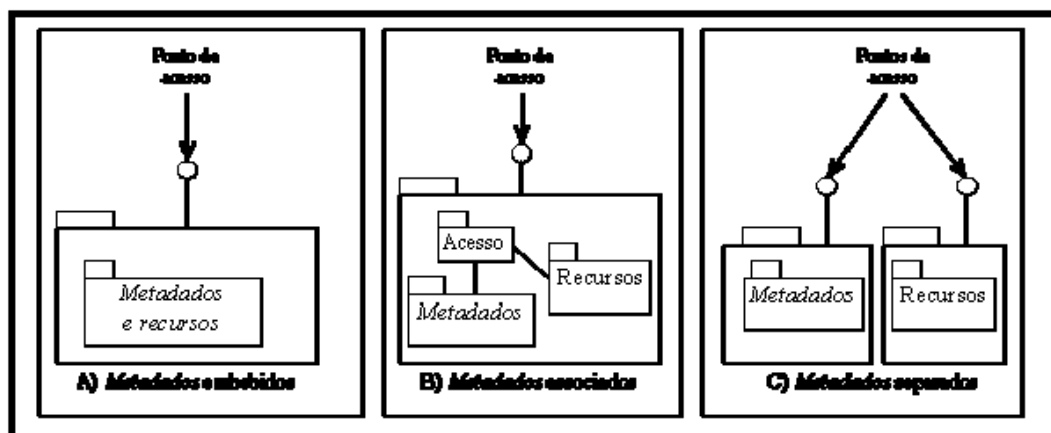
Em relação à forma de descrição dos recursos, Duval et al. (2002) elencam os seguintes tipos:

- a) metadado embebido ou intrínseco – pode ser pesquisado na Internet por um motor de busca, aumentando sua visibilidade. É uma forma de incentivar os criadores a implementarem metadados em seu recurso. De acordo com Sayão (2007), ele pode estar embutido em um objeto digital inscrito na sua codificação (documentos Hyper Text Markup Language - HTML) e ser aproveitado por outros sistemas que identifiquem esses elementos (BORBINHA; FREIRE, 2002);
- b) metadado associado ou extrínseco – é mantido em arquivos associados (extrínsecos) ao recurso que descreve. Pode ou não ser pesquisado por motores de busca. Tem como principal vantagem o controle sem alteração do conteúdo do recurso e possui a desvantagem de ser simplificado, requerendo cogerenciamento dos arquivos de recursos e dos arquivos de metadados.

O autor inclui, ainda, o metadado separado, mantido em uma base de dados isolado do recurso. Geralmente, é inacessível aos motores de busca. Borbinha e Freire (2002) acrescentam que esse é o modelo da biblioteca tradicional, bases de dados bibliográficas e livros dispostos nas estantes.

A Figura 4 ilustra como a estrutura descritiva dos metadados se relaciona com as estruturas de informação.

Figura 4 – Estruturas descritivas de metadados



Fonte: Borbinha e Freire (2002).

A definição de elementos de metadados contempla dois componentes:

- semântica – definição do significado dos elementos e dos seus refinamentos; e
- conteúdo – declarações ou instruções sobre o que e como valores podem ser atribuídos aos elementos (CHAN; ZENG, 2006).

É por meio desse último componente que este estudo busca identificar como a OC e os valores atribuídos ao elemento *Subject (Assunto)* podem contribuir para a otimização da interoperabilidade semântica entre os RIs.

Esse objetivo é alcançado pelo fato de os metadados darem um significado semântico aos dados, por meio de procedimentos técnicos, como catalogação, indexação e categorização dos conteúdos informacionais, permitindo que haja uma integração entre sistemas heterogêneos de informação (SOUZA; ALVES, 2009).

O esquema de registro de metadados consiste em elementos pré-definidos que representam atributos de um recurso. Também é fortemente reconhecido como uma ferramenta importante não apenas para compartilhar informação sobre vocabulários para metadados, mas também para enriquecer o reuso dos vocabulários. Desempenha um papel fundamental para a interoperabilidade semântica entre comunidades que falam diferentes idiomas ao

longo do tempo. O alcance da interoperabilidade é essencial para que os recursos de informação sejam compartilhados e descobertos (NAGAMORI; SUGIMOTO, 2006).

Para que o acesso e a recuperação dos conteúdos digitais sejam eficazes, é necessário caracterizar os documentos por meio da padronização da descrição das informações (CASTRO; SANTOS, 2007). Pode-se afirmar que, com esses procedimentos, os “[...] metadados são informação que resume, enriquece ou complementa os objetos ou serviços referenciados, produzindo assim um potencial incremento de informação” (BORBINHA; FREIRE, 2002, p. 1).

Na Internet, os metadados são projetados para tarefas que vão desde descrição de recursos, descobertas e filtragem até gestão de conteúdos. Essa diversidade de propostas reflete na multiplicidade de recursos de informação existentes na Internet (quantidade de páginas na *Web*, portal governamental, bibliotecas digitais e catálogos de compras) e na variedade de usuários, que vai de crianças a empresários e profissionais. Essas comunidades possuem seu próprio padrão de metadado. Os padrões de metadados “[...] são estruturas padronizadas para a representação do conteúdo informacional que será representado pelo conjunto de dados-atributos (metadados)” (ALVES, 2005, p.115). Como exemplos de padrões de metadados, pode-se citar:

- a) Dublin Core Metadata Element Set (DCMES) – conjunto de elementos de metadados Dublin Core, Dublin Core Metadata Terms – termos de metadados Dublin Core;
- b) Electronic Theses and Dissertations Metadata Standard (ETD-MS) – padrão de metadados de teses e dissertações eletrônicas;
- c) Learning Object Metadata (LOM) – metadados de objetos de aprendizagem;
- d) Metadata Object Description Schema (MODS) – esquema de descrição de objetos de metadados;
- e) Metadata Encoding and Transmission Standard (METS) – esquemas para objetos digitais de alta complexidade em bibliotecas (NAGAMORI; SUGIMOTO, 2006; ZENG, 2010).

De acordo com Shintaku, Ferreira e Robredo (2011), o DC, ETD-MS e o METS, por serem mais descritivos e codificados em XML, possibilitam uma melhor interoperabilidade.

A utilização de padrões de metadados permite, aos sistemas de informação e de gestão do conhecimento, a integração e o compartilhamento de informações e aplicações (SOUZA; ALVES, 2009).

Ressalta-se que a adoção de um padrão de metadados para um domínio específico atende às demandas da comunidade. Entretanto, pode comprometer a descoberta e a reutilização de metadados por outras comunidades, pois é necessário que não satisfaça apenas o domínio específico, mas também a interoperabilidade, já que esta é um importante fator para a comunicação entre comunidades (NAGAMORI; SUGMOTO, 2006). Comunicação que pode ser estabelecida por meio do perfil de aplicação.

Entre os padrões de metadados supracitados, será discutido apenas o padrão DC, por fazer parte do escopo desta pesquisa.

4.1 PADRÃO DUBLIN CORE

O padrão DC

[...] é um conjunto de quinze elementos de metadados, os quais podem ser definidos como o mais baixo denominador comum para a descrição de recurso de informação, equivalente aos dados de uma ficha catalográfica. Suas características principais são: o entendimento semântico universal dos elementos metadados, o escopo internacional e a extensibilidade visando permitir adaptações às necessidades adicionais de descrição (SOUZA; ALVES, 2009, p. 192).

Esse padrão, mantido pela OCLC e desenvolvido por meio de cooperação internacional, surgiu pela necessidade de equilibrar a quantidade de informações constantes nos registros bibliográficos, de forma a baratear e agilizar o processo de descrição (CAMPOS, 2007). A manutenção e o desenvolvimento do DC são de responsabilidade da DCMI, que busca criar mecanismos para facilitar a recuperação de recursos da Internet utilizando padrões de metadados (CATARINO, 2009).

O DCMI teve início em 1995, durante um *workshop* em Dublin, Ohio, com a participação de bibliotecários, pesquisadores da área de BDs, provedores de conteúdo e *experts* em linguagens de marcação que buscavam o aumento da descoberta de padrões para recursos de informações. Devido ao baixo custo e ao seu potencial de ampliar a visibilidade de recursos em coleções entre os setores e domínios de assuntos, o DC foi amplamente adotado pela comunidade científica. Em junho de 2000, já existiam 20 traduções para esse padrão, empregado pela European Committee for Standardization/Information Society Standardization System. É também o padrão oficial da WWW Consortium e do Z39.50 (NISO, 2001).

A implementação do DC buscou a interoperabilidade semântica, a flexibilidade e a modularidade na *Web*, o que levou a um consenso internacional da comunidade científica em relação à utilização desse padrão na obtenção da interoperabilidade (PEREIRA; RIBEIRO JÚNIOR; NEVES, 2005).

O DC compreende todos os termos mantidos pelo DCMI – o conjunto de elementos DCMI Metadata Terms (DCMI Terms) e outros termos que representam propriedades, subpropriedades, classes (DCMI Type Vocabulary) e esquemas de codificação Vocabulary Encoding Scheme (VES) e Syntax Encoding Scheme (SES) (CATARINO, 2009).

A descoberta de recursos por meio das comunidades é uma das questões mais importantes no ambiente digital. Uma forma simples de realizar a pesquisa de recursos entre comunidades é a adoção de um esquema de metadados, sendo o do DCMES um dos mais utilizados. Como existem outros tipos, as comunidades podem desenvolver um esquema para melhor atendê-las. Para que os recursos sejam recuperados entre essas comunidades, é necessário desenvolver mapeamentos (*crosswalk*)²³ que definam as relações entre um termo de metadados utilizado por diferentes esquemas (NAGAMORI; SUGMOTO, 2006; SHINTAKU; FERREIRA; ROBREDO, 2011).

Como vimos, o DC busca fornecer um conjunto básico de elementos de descrição que podem ser adotados por catalogadores ou não catalogadores para simples descrição de recursos de informações, tornando-as mais visíveis aos motores de busca e sistemas de recuperação (WEIBEL, 1997). Esses elementos fornecem um conjunto mínimo para descrição dessas informações para a Internet,

²³ "É um mapeamento semântico ou técnico, as vezes ambos, de uma estrutura de metadados para outro esquema de metadados" (PATEL et al., 2006, p. 21).

porém não substitui modelos como o AACR2 ou o MARC 21, que, embora sejam mais completos, possuem algumas restrições no ambiente *Web* (SOUZA; ALVES, 2009; WEIBEL, 1997).

O Dublin Core pode ser inserido em uma página HTML. Utiliza uma linguagem XML, adota a sintaxe RDF e é o formato padrão empregado para efetuar a interoperabilidade entre outros sistemas. A interação entre o DC e o RDF dá-se pelo fato do RDF apresentar as regras sintáticas nas quais o DC pode ser introduzido (BAPTISTA; MACHADO, 2001).

Os 15 elementos básicos do Dublin Core podem ser implementados pelos usuários ou pelo sistema, e cada um possui um valor (ex.: Esquemas de codificação VES e SES) (CATARINO, 2009). Têm uma ocorrência ilimitada, e a forma utilizada para distinguir os valores deles é a utilização de qualificadores. Pode ter um identificador (esquema) e/ou um modificador que informa, aos usuários e aos programas, como o valor daquele elemento deve ser interpretado. Como exemplo, há o elemento Assunto (*Subject*) e o qualificador palavras-chave (LCSH). A utilização dos sistemas de organização do conhecimento (SOC), que, de acordo com Brasher e Café (2008, p. 7), “[...] são sistemas conceituais que representam determinado domínio por meio da sistematização dos conceitos e das relações semânticas que se estabelecem entre eles, [...]”, aumentam as especificidades dos dados (ALVES; SOUZA, 2007; DUBLIN CORE METADATA INICIATIVE, 2011; HODGE, 2000; NISO, 2001, 2004). “O conjunto de descritores do Dublin Core é embutido no próprio documento descrito (HTML, XML), ou dependendo do recurso, a metainformação encontra-se separada por outro recurso catalogado” (SOUZA; VENDRUSCULO; MELO, 2000, p. 93), como parte do mecanismo de acesso (HODGE, 2000).

Para Pereira, Ribeiro Júnior e Neves (2005), a padronização de elementos mínimos para a descrição de recursos possibilita a obtenção de resultados de busca na Internet com baixa revocação e alta relevância.

Segundo Castro e Santos (2011), a consistência das informações existentes nos repositórios digitais, baseada na utilização de metadados como forma de representação da informação, favorece a melhoria do acesso e a recuperação das informações pelos usuários.

Para esses autores, é necessário ter formas diferenciadas de representação de um recurso informacional, e os metadados consistiriam em uma

delas, para que as informações em ambientes digitais sejam recuperadas de modo mais eficaz. O tratamento dado às informações deve ser mais acurado e atender aos requisitos da interoperabilidade entre sistemas.

Na promoção da interoperabilidade semântica, uma série de elementos do DC para descrição sugere a utilização do vocabulário controlado para cada valor do elemento *Subject*. Vocabulários controlados podem ser aplicados dentro de certo domínio local para o desenvolvimento da interoperabilidade (NISO, 2001, 2004), embora algumas pesquisas sugiram preocupações em relação a essa possibilidade, já que poderia acarretar dificuldades na recuperação da informação (SHINTAKU; BRASCHER, 2007; PARK; TOSAKA, 2010).

A Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), por meio das unidades de pesquisa para geração, difusão e aplicação de tecnologia da informação e comunicação, preocupa-se com o estabelecimento de padrões de descrição de conteúdo de informações (ALVES; SOUZA, 2007; SOUZA; ALVES, 2009; SOUZA; VENDRUSCULLO; MELO, 2000). No projeto Banco de Imagem, armazenamento e acesso de fotos e gravuras, adotou-se o DC, assim como em outros trabalhos. Foram incorporadas adaptações com o objetivo de promover adequação às necessidades da empresa.

Ao todo, foram utilizados 18 elementos do DC (título, autor ou criador, palavras-chave, categoria, descrição, publicador, colaborador, data, tipo, formato, acesso, identificador de recurso, fonte, idioma, relação, cobertura, direito autoral e contato). Em relação ao elemento *Creator* (autor), estabeleceram-se qualificadores - autor para textos escritos, fotógrafos ou ilustradores para recursos visuais. Sobre o elemento *Subject*, os recursos devem ser descritos por palavras-chave ou frase que descreva o assunto do conteúdo do recurso. A utilização de vocabulários controlados é encorajada (AGROVOC, THESAGRO). Para a atribuição de categoria de assunto, incentiva-se o uso de sistemas de classificação AGRÍCOLA, CDD e, para o uso das palavras-chave, recomenda-se que sejam extraídas do título ou do item descrição do recurso (SOUZA; VENDRUSCULLO; MELO, 2000). Para os RIs que utilizam o DSpace, o metadado *Subject*. “[...] possui certa independência entre o metadado e o documento [...]”. A função dele é organizar, classificar e hierarquizar os documentos no RI, para facilitar a recuperação do assunto por meio do agrupamento dos documentos que possuem o mesmo assunto (SHINTAKU; BRASCHER, 2007, p. 2).

A Agência de Informação adota a indexação e a categorização de assuntos baseada no uso de tesouros e vocabulários controlados, visando à geração de metadados com qualidade e à integração de fontes diversificadas e heterogêneas de informação. De acordo com a Agência, a aplicação do DC é favorecida por uma somatória de fatores: simplicidade, interoperabilidade e capacidade de abstrair informações relevantes de objetos, como imagens, sons e vídeos.

A Agência considera que o DC – que inclui aspectos de precisão, fidelidade, critérios de seleção, generalizações, consistência dos dados, definições utilizadas, metadados sobre fontes de dados e adoção de padrões de representação descritiva dos recursos de informação (catalogação) – possibilita o uso efetivo de regras de descrição que enriquecem as representações dos recursos e facilitam o compartilhamento dos dados. O resultado positivo é traduzido pela dinamização do processo de catalogação de recursos de informação e pela eliminação de esforços repetitivos, o retrabalho. O DC contribui também para o compartilhamento de dados e a interoperabilidade entre os sistemas utilizados pela Agência (ALVES; SOUZA, 2007).

A qualidade dos metadados utilizados em RIs é um assunto que vem despertando o interesse de pesquisadores, embora seja ainda pouco explorado na literatura. O estudo de Park (2009) traz uma pesquisa sobre o estado da arte em relação a esse assunto.

Segundo o autor, a qualidade dos metadados reflete em suas funções básicas, que são a descoberta, o uso, a proveniência, a atualização, a autenticação e a administração, ou seja, as funções de encontrar, identificar, selecionar e recuperar documentos.

Constatou-se que a qualidade dos metadados pode ser aferida pelos seguintes critérios:

- a) exatidão – reflete a exatidão de descrições, representação dos dados e conteúdo dos recursos, que abrange a precisão do conteúdo dos dados dos elementos (problemas com a semântica dos elementos, como *creator*, *contributor*) e direitos autorais;
- b) completude – reflete a capacidade de acesso a um objeto e a conexão com a coleção. Representa a descrição de um determinado objeto em relação a um domínio;

- c) consistência – está relacionada a semântica e sintaxe dos elementos adotados para descrição dos recursos (elemento *Date*, em que pode ser utilizado o formato americano ou nacional).

Algumas ações podem contribuir para qualidade dos metadados: a adoção de melhores práticas, registros de perfil de aplicação local e disponibilização de diretrizes (políticas) em relação aos recursos de uma coleção, como as dos RIs. Essas diretrizes auxiliam criadores, depositadores e administradores (bibliotecários) na criação de metadados de qualidade.

A partir desse estudo, Park (2009) concluiu que a necessidade de criação de um modelo de dados comum que seja interoperável entre bibliotecas exige uma avaliação das práticas atuais de criação de metadados e da documentação para aqueles criados localmente, além da utilização de perfis de aplicação.

É importante salientar algumas abordagens em relação à interoperabilidade realizadas pelo autor nesse estudo:

- a) a criação de um identificador semântico único para cada elemento;
- b) a flexibilidade na utilização de linguagem natural possibilita a representação de um conceito de várias formas (sinônimo, homônimo), e nem sempre a máquina soluciona;
- c) a importância do profissional da informação no processo de criação de metadados;
- d) a consistência semântica está relacionada a conceitos utilizados para descrever um mesmo valor ou um mesmo elemento do recurso (*subject/description*);
- e) a consistência estrutural que representa uma mesma estrutura ou formato utilizada para representar um mesmo atributo ou elemento de um recurso (ex.: datas).

O estudo realizado por Heery et al. (2006b) identificou que o nível de interoperabilidade entre os repositórios é muito baixo, após constatar, a partir dos RIs E-print, que não interagem com os repositórios de ensino-aprendizagem. A interoperabilidade semântica e a organização do conhecimento nos repositórios estão vinculadas a esforços individuais. Identifica-se, em nível de serviços, uma

falta de interesse em relação aos serviços oferecidos e a coleta de dados, principalmente para aqueles que utilizam o sistema de *harvesting*²⁴ para os metadados.

Alguns repositórios oferecem instrumentos para submissão ou inserção de metadados, adotam a padronização para uma classificação ou um sistema de cabeçalhos de assunto e disponibilizam sistemas de navegação por assunto ou tópicos, contribuindo para o aumento da qualidade de acesso.

A inclusão de metadados pode enriquecer os registros de metadados que utilizam vocabulários controlados. Serviços podem ser adotados para corrigir a desatualização de termos desses vocabulários e reconhecer valores de vocabulários recomendados, além de atribuir esquemas apropriados de vocabulários para a confirmação. Novos valores também podem ser gerados baseados no mapeamento entre esquemas de vocabulários (HILLMANN et al., 2004, apud KOCH, 2006).

Embora os repositórios compartilhem de aspectos comuns às bibliotecas digitais, Koch (2006) destaca que essas já se encontram em um estágio mais avançado de interoperabilidade semântica e organização do conhecimento em relação à eficiência do acesso à informação.

Diferentes níveis de cobertura e serviços dos repositórios requerem recomendações diferenciadas:

- a) coerência individual, frequentemente institucional, foca a padronização na descrição do seu próprio ambiente e pede uma semântica mais rica quanto possível;
- b) serviços de integração de repositórios heterogêneos deveriam buscar a preservação da riqueza das fontes participantes. Devem permitir a pesquisa e a busca utilizando sistema de filtragem e exibir os vocabulários utilizados;
- c) há uma necessidade fundamental e imediata de os repositórios indexarem ou integrarem o assunto sugerido pelo depositante a um vocabulário controlado. Devem buscar o mapeamento e outras formas de tratar essa heterogeneidade;

²⁴ Coleta automática de metadados que já é associada com recurso produzido automática ou manualmente.

- d) deve existir uma preocupação com a interoperabilidade semântica em um nível mais alto, empregando métodos de serviços de terminologia.

Corroborando a pesquisa de Heery et al. (2006a), Park e Tosaka (2010) apresentaram um estudo sobre o panorama atual, focalizando a criação de metadados em repositórios digitais, coleções e bibliotecas que incluíram recursos digitais e digitalizados. Utilizaram um questionário eletrônico (Vovice) para investigar os elementos de metadados descritivos adotados, se fizeram uso de vocabulários controlados para acesso aos assuntos e divulgação dos metadados e se aplicaram diretrizes para os metadados em ambientes externos. Foram formuladas as seguintes questões:

- a) Que esquema de metadado e padrão de conteúdo²⁵ foram utilizados no repositório e nas coleções?
- b) Quais esquemas de vocabulários controlados são utilizados para facilitar o acesso ao assunto?
- c) Quais os critérios adotados para seleção dos metadados e dos esquemas de metadados?
- d) Até que ponto os metadados foram expostos e quais as práticas de compartilhamento de metadados?

Os autores esclarecem que pesquisas recentes mostraram que as práticas atuais de criação de metadados são problemáticas, há o uso de vários esquemas de metadados, padrões de conteúdo e vocabulários controlados, dificultando o aumento do nível de interoperabilidade entre as coleções digitais e os repositórios.

Os problemas são aumentados nos ambientes onde muitas instituições não registram o delineamento do processo de criação de metadados. Um outro aspecto é que só recentemente pesquisadores começaram a estudar essas questões. A literatura sobre o assunto ainda se encontra em um estágio incipiente. Pesquisas foram realizadas tendo como foco populações diferentes e com tamanhos de amostras distintas (algumas muito pequenas). Em alguns casos, a literatura traz resultados contraditórios, dificultando o entendimento sobre o estado atual da criação de metadados.

²⁵ Padrão de autoridades ou conjunto de regras que determinam vocabulário, sintaxe ou formato em que o dado ou elemento de metadado está embutido (Tesauros, vocabulário controlado, AACR).

Esse estudo identificou que a criação de metadados ainda não propicia a interoperabilidade. Enquanto elementos de metadados são elaborados para atender às necessidades locais, eles podem dificultar a interoperabilidade dos metadados nos RIs e nas coleções, quando os mecanismos não compartilham as extensões e variantes definidas localmente. Diretrizes para a criação de metadados e de perfis de aplicação são essenciais para o estabelecimento e a garantia de qualidade, entretanto a maioria dos conteúdos das diretrizes e as melhores práticas não são disponibilizadas publicamente. A falta de mecanismos para facilitar o acesso público às aplicações locais e às diretrizes pode prejudicar a consistência e a interoperabilidade entre os RIs. O desenvolvimento de registros pesquisáveis disponibilizados publicamente tem potencial para o aumento da interoperabilidade.

O fator limitante desse estudo foi a população participante, o que impede a generalização dos achados do estudo. Por outro lado, os resultados indicam uma necessidade urgente de elaboração de um modelo de dados comum, compartilhável e interoperável entre os RIs e as coleções digitais. O desenvolvimento de um modelo de dados em comum demanda futuras pesquisas sobre mecanismos de mediação subjacentes para a implementação de elementos de metadados semânticos, padrões de conteúdo e vocabulários controlados no mundo onde os metadados podem ser distribuídos e compartilhados de forma ampla, além do ambiente local e dos usuários da comunidade. Outras alternativas são: aplicação de metadados semiautomáticos, semântica de metadados DC, elementos de metadados customizados e desenvolvimento de metadados pelos profissionais de catalogação.

Os autores ressaltam que é necessário desenvolver estudos que incorporem outros métodos, como pesquisas utilizando telefone e entrevistas com grupo focal, para melhor entendimento sobre o estado atual da criação de metadados. A variação institucional também deverá ser considerada nos próximos trabalhos.

Zeng (2010, p. 181) esclarece que a aplicação de metadados e os novos esquemas de metadados podem aumentar e assegurar os elementos específicos por domínio (“dc:subject.MeSH ou “NLMDC.Subject.NLMClass”, que contêm conceitos e classificações altamente específicos, que permitem acesso consistente e sistemático aos recursos de áreas específicas). Porém, a autora

chama a atenção para dois problemas em relação a esse modelo: “1) os elementos de metadados relacionados com assunto são muito restritos nos conjuntos de elementos de metadados; 2) para um ‘*container*’²⁶ – um item –, uma ocorrência de metadados está sempre na posição de um substituto” (tradução nossa). Esses problemas poderiam ser evitados utilizando a decomposição de estrutura do conteúdo, que pode trazer melhor acesso ao assunto, pois os tópicos tratados no documento emergiriam quando o registro de metadados descritivos refletisse as partes dos documentos.

Em relação à decomposição de estrutura dos componentes, o desmembramento desses componentes pode ser tratado como recursos individuais com descrições de metadados individuais, que seriam ligados por um elemento de “relação”. Dessa forma, os metadados seriam adaptados para tratar as tarefas de descrição dos componentes que apresentam relações estruturais entre dois ou mais componentes. Essa decomposição do conteúdo seria realizada por meio de linguagens de marcação específicas por domínio, como no exemplo a seguir, a partir do esquema MatML (Linguagem de marcação extensível para dados de propriedades de materiais):

As informações contidas pelo elemento “Material” são compartimentadas em quatro elementos principais:

1. O elemento ‘BulkDetails’ contém uma descrição do material completo;
2. O elemento ‘ComponentDetails’ contém uma descrição de cada componente do material completo (útil para sistemas de materiais complexos, como compostos ou soldas);
3. O elemento ‘metadata’ contém descrições de dados encontrados no documento;
4. O ‘Glossary’.

Utilizando o elemento ‘BulkDetails’, podem ser encontrados subelementos capazes de ter seus próprios “subsubelementos”, no exemplo ‘Characterization’:

²⁶ Um item que armazena e gera dinamicamente metadados relacionados com um único recurso em múltiplos formatos, cada um pode ser recolhido como registro no protocolo OAI-PMH (DRIVER, 2009).

BulkDetails
 * Nome
 * Classe
 * Subclasse
 * Especificação
 * Caracterização
 Fórmula
 Composição Química
 Composição da Fase
 Detalhes Dimensionais
 Notas
 * Fonte
 * Forma
 * Detalhes de Processamento
 * Dados de Propriedade
 * Notas (ZENG, 2010, p. 181)

A autora supracitada acredita que essa é a nova alternativa para a descoberta de conteúdo e que novas pesquisas devem ser desenvolvidas nesse sentido.

OS METADADOS E A POLÍTICA INSTITUCIONAL DE INFORMAÇÃO

Drake (2004) afirma que, entre os aspectos que devem ser considerados no desenvolvimento dos repositórios, figuram: cultura institucional, escopo do repositório, conteúdo, níveis de acesso, aspectos legais, sustentabilidade, recursos financeiros e padrões de interoperabilidade. Esses pontos devem ser contemplados pela política institucional de informação de cada instituição.

A adoção de uma política institucional de informação é outro fator apontado como necessário para implantação, gerenciamento e sucesso dos RIs. Como política de informação, entende-se

[...] um conjunto de princípios, leis, diretrizes, regras, regulamentos e procedimentos inter-relacionados que orientam a supervisão e **gestão do Ciclo Vital da Informação**: a produção, coleção, organização, distribuição/disseminação, recuperação e eliminação da informação. Política de informação compreende o acesso à, e uso da informação (HERNON; RELYEA, 1991, p. 176 apud JARDIM; SILVA; NHARRELUGA, 2009, p. 6, grifo nosso).

Essa assertiva corrobora a pesquisa desenvolvida por Patel et al. (2005), em que os autores relacionam a interoperabilidade semântica com a gestão do processo de Ciclo de Vida da Informação.

Andrade, Silva e Cervantes (2011), ao citarem Vargas (2008), demonstram que a política institucional de informação deve contemplar também funções da biblioteca, como o estabelecimento de metadados, que podem ser administrativos, estruturais e descritivos.

Para Park e Tosaka (2010), a disponibilização de diretrizes e de melhores práticas em relação aos metadados utilizados pelo RI também é um fator preponderante para o alcance da interoperabilidade.

Conforme já abordado, a interoperabilidade semântica é o foco deste trabalho e, para que seja alcançada, é necessário entender que o processo tem início em alguns pressupostos que possibilitam qualidade em nível de objeto, metadados e processos, incluindo os SOC. Esses aspectos foram percebidos pela NISO (2007), que, ao buscar o estabelecimento de boas práticas para o desenvolvimento de coleções digitais, propôs:

- a) fornecer uma visão de alguns dos componentes e atividades mais envolvidos na criação de boas coleções digitais;
- b) identificar recursos que auxiliem no desenvolvimento de boas práticas locais para criação e gerenciamento de coleções digitais; e
- c) incentivar a participação da comunidade no desenvolvimento de melhores práticas para a construção de coleções digitais, que consistem, “[...] de objetos digitais que são selecionados e organizados para facilitar sua descoberta, acesso e uso” (NISO, 2007, p. 4).

A partir dessas propostas, que visam à organização de coleções digitais – coleções, objetos, metadados e iniciativas –, a NISO (2007) definiu princípios que favorecem a criação de uma coleção digital de boa qualidade, que podem ser conferidos no Quadro 5.

Quadro 5 – Princípios que auxiliam na formação de uma coleção digital de boa qualidade.

PRINCÍPIOS	CONTEÚDO
Princípio 1	Uma coleção digital é criada baseada numa política de coleções
Princípio 2	As coleções devem ser descritas para que os usuários identifiquem: escopo, formato, restrições e acessos, integridade e interpretação
Princípio 3	A coleção deve ser administrada e os recursos devem ser gerenciados por toda a vida útil
Princípio 4	A coleção deve estar amplamente disponível, deve proporcionar usabilidade e acessibilidade
Princípio 5	Os direitos autorais devem ser respeitados
Princípio 6	Uma coleção digital deve criar mecanismos para a elaboração de estatísticas
Princípio 7	Uma coleção digital deve ser interoperável
Princípio 8	Uma coleção digital deve integrar os usuários aos fluxos de trabalho
Princípio 9	Uma coleção digital deve ser sustentável

Fonte: NISO (2007)

Serão apresentados apenas os Princípios 2, 3 e 7, em que são abordados aspectos que possuem uma relação mais estreita com esta pesquisa.

O Princípio 2, que versa sobre a descrição de coleção, ou seja, os metadados, possui como objetivo o auxílio na descoberta e no entendimento dos usuários sobre o conteúdo a que eles estão tendo acesso. Esses metadados devem descrever: escopo, responsabilidades/contatos, condições de uso, restrições de acesso e informações sobre contatos.

O Princípio 3, a administração da coleção, compreende o armazenamento, o gerenciamento e a preservação dos dados. Propicia o uso e reúso a qualquer tempo. Inclui criação, acertos e enriquecimento de metadados, correção e enriquecimento dos próprios dados, adição de anotações, *links* para outros materiais e outras informações.

A interoperabilidade é abordada no Princípio 7. Os desenvolvedores de coleções devem oferecer serviços que dão suporte à interoperabilidade, principalmente no compartilhamento de metadados com ferramentas de buscas externas. Esforços devem ser perseguidos para a obtenção da qualidade de conteúdos e dos metadados, por meio de uma adequada descrição da coleção.

Com referência à qualidade do objeto, a NISO (2007) também disponibiliza princípios em relação ao que foi denominado de “objeto digital adequado” em coleções digitais, apresentados no Quadro 6.

Quadro 6 – Objeto digital adequado

PRINCÍPIOS	CONTEÚDO
Princípio 1	Prioridades da coleção, interoperabilidade e reusabilidade. A produção do objeto deve buscar a manutenção de prioridades da coleção no qual está inserido, ao mesmo tempo em que deve manter as qualidades necessárias a interoperabilidade e reusabilidade
Princípio 2	Persistência. Deve proporcionar condições para que o objeto permaneça acessível indefinidamente, independentemente de mudanças tecnológicas
Princípio 3	Padrões e melhores práticas no processo de digitalização. O objeto digital deve ser digitalizado de acordo com as melhores práticas, padrões reconhecidos e formatos que tenham perspectiva de utilização por longo prazo. Caso isso não seja possível, as razões devem ser documentadas.
Princípio 4	Identificadores únicos e persistentes. Um objeto deve ser identificado por um identificador persistente único em conformidade com esquemas bem documentados.
Princípio 5	Autenticidade, integridade, proveniência e contexto. O usuário deve ter a possibilidade de determinar a proveniência e o contexto, a autenticidade e a integridade do objeto.
Princípio 6	Metadados descritivos, administrativos e estruturais. Um objeto digital deverá ter metadados descritivos e administrativos e os objetos digitais complexos, compostos por vários arquivos, deverão estar associados aos metadados estruturais.

Fonte: NISO (2007)

Em busca das melhores práticas para os metadados, o Quadro 7 apresenta os princípios para a obtenção da qualidade dos metadados concebidos pela NISO (2007).

Quadro 7 – Qualidade dos metadados

PRINCÍPIOS	CONTEÚDO
Princípio 1	A qualidade do metadado está relacionada com a conformidade dos padrões da comunidade com os materiais, usuários e usos da coleção
Princípio 2	Deve propiciar a interoperabilidade
Princípio 3	Deve utilizar controle de autoridade e padrões de conteúdo para descrição dos objetos e disponibilizar objetivos relacionados
Princípio 4	Deve incluir uma declaração de condições e termos para uso dos objetos digitais
Princípio 5	Deve suportar a gestão e a preservação dos objetos
Princípio 6	Os objetos devem possuir qualidade, incluindo autoridade, autenticidade, persistência e identificação única

Fonte: NISO (2007)

Entre os princípios preconizados para os metadados pela NISO (2007), foram selecionados os Princípios 1, 2 e 3 por estarem mais próximos dos temas analisados neste estudo.

No que se refere ao Princípio 1, esquema de metadados, vocabulários controlados e tesauro desenvolvidos para uma comunidade e tipos de materiais específicos devem ser cuidadosamente avaliados para que o mais apropriado seja adotado. A seleção do esquema de metadados deve ser baseada no tipo de coleção, no nível de recursos que a instituição pretende adotar, no grau de *expertise* dos criadores, na expectativa de uso e do usuário e no objetivo de alcançar a interoperabilidade e o compartilhamento da coleção digital.

As instituições devem analisar a granularidade da descrição em nível de conjunto, série ou grupo, item ou nos vários níveis, visando à profundidade desejada e ao alcance do acesso aos materiais antes de adotar um determinado esquema de metadados. É recomendado também que os esquemas adotados por instituições similares sejam considerados, pois aumentam a interoperabilidade entre coleções.

Conforme o Princípio 2, os metadados, para terem qualidade, devem ser coerentes, significativos e úteis no contexto global. Devem incluir todas as informações pertinentes sobre os objetos, como os pressupostos sobre o contexto no qual ele é acessado localmente não ter mais validade no amplo ambiente de rede. Os esquemas de metadados devem ser documentados em um registro²⁷ que forneça informações padronizadas para definição, identificação e utilização de cada elemento de dado. Um registro define características dos metadados e os requisitos de formação, assegurando que um esquema de metadados e elementos de dados utilizados por uma instituição possam ser aplicados de forma consistente, por ela ou pela comunidade, reutilizados por outras comunidades e interpretados pela máquina ou por humanos, ou seja, deve contemplar um perfil de aplicação.

Quanto ao Princípio 3, os atributos dos objetos devem ser expressos utilizando termos controlados sempre que possível. Incluem-se aí nomes de pessoas, nomes de empresas, nomes de lugares, títulos e assunto.

²⁷ São metadados refletidos num formato único. Um registro é recuperado num XML como resposta a um pedido OAI-PMH de metadados de um item (DRIVER, 2009).

Os esquemas de classificação, uma forma de controle de vocabulário, são vistos como modo de proporcionar o acesso aos assuntos, assim como tesouros, taxonomias²⁸, lista de autoridades e tantas outras soluções.

A escolha de um vocabulário dependerá de alguns fatores, como esquemas de metadados adotados, natureza da coleção, recursos e expectativas. Alguns fatores podem ser considerados:

- a) Os usuários são crianças, adultos, especialistas ou generalistas?
- b) Qual o idioma utilizado?
- c) Quais recursos poderiam ser utilizados? Qual o vocabulário adotado?
- d) Ferramentas para apoiar o uso do vocabulário: existe um tesouro *on-line*? Pode ser incorporado no sistema de pesquisa da coleção?
- e) Existe mapeamento de referências e termos relacionados?
- f) Manutenção: novos termos são criados, velhos termos se tornam arcaicos ou obsoletos. Quem mantém o vocabulário e como é atualizado?

Uma boa estratégia seria proporcionar ao usuário final a implementação local de pesquisa de autoridade e vocabulários, a exposição da regras adotadas e a disponibilização da documentação contendo informações que possam auxiliar os criadores a identificar os termos mais consistentes.

A NISO (2007) coloca que os metadados estão entre os aspectos mais desafiadores para a identificação de recursos disponíveis na *Web*. A existência de metadados descritivos aumenta a probabilidade de os conteúdos digitais serem identificados e utilizados, maximiza o acesso e atrai novos usuários.

É importante que as instituições reconheçam que os esquemas de metadados e os protocolos de informação não são suficientes; é necessário que vocabulários controlados e tesouros sejam implementados, para assegurar a semântica dos termos e que os padrões de conteúdos de dados sejam adotados (catalogação), como a Anglo-American Cataloguing Rules (AACR)/Resource Description and Access (RDA), Describing archives: a Content Standard (DACS), a

²⁸ Pode referir-se tanto a uma classificação hierárquica das coisas, como aos princípios que baseiam a classificação (HODJE, 2000).

Cataloging Cultural Objects (CCO), com o objetivo de garantir a padronização de formatos para a autoria.

Um dos pontos abordados pela NISO (2007) é a elevação do custo ao serem adotados esses procedimentos, entretanto esse custo é justificado pelo aumento da eficácia e eficiência para o usuário final.

As decisões sobre os padrões de metadados e os níveis de descrição devem ser aplicadas de acordo com o contexto das instituições, das disponibilidades dos recursos técnicos e humanos, das necessidades dos usuários e das abordagens, no geral, ou dentro de domínio de conhecimento específico (NISO, 2007).

No entender de Sayão (2007, p. 21),

A efetiva aplicação desses princípios é uma etapa importante na viabilização de repositórios digitais abertos e interoperáveis, mas para tal, as orientações estabelecidas devem ser traduzidas em melhores práticas e o uso de padrões abertos e de ampla aceitação.

Em síntese, os princípios recomendados pela NISO (2007) reforçam a necessidade de estudos na perspectiva da OC, em busca da interoperabilidade semântica. Em consonância com o pensamento de Patel et al. (2005), Broughton et al. (2005) e Zeng (2010), entre outros, que pontuam o papel fundamental da OC no acesso e na recuperação da informação no ambiente digital, na próxima seção, serão buscados, nos princípios da OC e nas abordagens em relação aos SOC, elementos para a otimização dessa interoperabilidade.

Como se pode observar, os metadados têm sido reconhecidos como solução para problemas de recuperação de informação na Internet, principalmente por oferecerem um padrão de descrição de informação. A utilização do padrão Dublin Core, implementado pelo protocolo OAI-PMH, é um dos principais fatores para implementação da interoperabilidade entre os participantes da iniciativa OAI. Soma-se a esse contexto a padronização da terminologia, utilizando os SOC para a organização de objetos digitais, como um dos aspectos mais importantes para a integração de sistemas, ou seja, a interoperabilidade semântica, abordada na seção a seguir.

5 A QUESTÃO DA INTEROPERABILIDADE

A interoperabilidade é apresentada por meio de definições que buscam demonstrar a sua importância para o acesso e a recuperação de recursos no contexto dos RIs. A interoperabilidade semântica como meio para o compartilhamento do conhecimento em um ambiente digital com proporções globais é retratada nessa seção. Relatos de pesquisas utilizando os sistemas de organização do conhecimento como forma de enriquecê-la também são descritos.

Desde meados do século XX, para acompanhar o fenômeno chamado “explosão informacional”, as bibliotecas precisaram estabelecer serviços cooperativos que possibilitassem trocas de informações. Para tanto, foi necessário criar mecanismos que sustentassem esse compartilhamento de informações e cooperação entre bibliotecas de forma globalizada. Essas ações já traziam o conceito de interoperabilidade, em que os sistemas conversavam não apenas entre eles, mas com outros sistemas de informação. Essa comunicação, entretanto, pressupõe que exista interoperabilidade técnica, já que esse tipo de interoperabilidade permite que os sistemas e os *hardwares* se reconheçam. Os fatores que possibilitam esse tipo de interoperabilidade são: utilização de um protocolo comum, OAI-PMH; adoção de um conjunto mínimo de metadados proporcionado pelo Dublin Core; utilização de sintaxe comum XML (TRISKA; CAFÉ; 2001; SHINTAKU; FERREIRA; ROBREDO, 2011).

Identifica-se, no ambiente acadêmico, a existência não apenas dos tradicionais catálogos de bibliotecas, mas serviços que disponibilizam bases de dados, bibliotecas virtuais, RIs, serviços de referências virtuais e novos recursos acadêmicos, como objetos de aprendizagem. Embora estejam disponíveis, existe uma dificuldade em integrá-los em um único espaço de informação. Essa integração pressupõe que seja possível o acesso aos diversos tipos de recursos por meios de vários pontos de acesso, como autor, título, assunto, tipo de trabalho e formato. Para que esse tipo de acesso seja possível, faz-se necessário integrar essas informações de forma interoperável.

O conceito dado por Ralyté et al. (2008, p. 754) é que a interoperabilidade pode ser vista como “[...] a habilidade para o sistema ou um produto para trabalhar com outros sistemas ou produtos sem esforço especial por parte do cliente”.

Pesquisas atuais sobre a interoperabilidade em arquitetura de BDs estão relacionadas com os desafios da criação de uma ferramenta para o acesso e a integração da informação, nos seguintes domínios: comunidade particular (bibliotecas, comunidades científicas); classificação específica de informação (relatórios técnicos, *software*); e informações específicas na área tecnológica (imagens digitais, bases de dados relacionais) (PAYETTE et al., 1999).

A necessidade da interoperabilidade, abordada pela maioria das pesquisas sobre bibliotecas digitais, está relacionada com o fato de várias bibliotecas que possuem diferentes arquiteturas, formatos de metadados e tecnologias poderem se comunicar efetivamente por meio de aplicações comuns (SHIRI, 2003).

No contexto do trabalho desenvolvido pela Cornell University/Corporation for National Research Initiatives (CNRI), a interoperabilidade é definida como a habilidade de componentes de uma BD ou serviços serem funcionais e logicamente intercambiáveis, em virtude de eles terem sido implementados em acordo com um padrão bem definido e interfaces publicamente conhecidas (PAYETTE et al., 1999). Com esse modelo, diferentes serviços e componentes conseguem se comunicar por meio de interfaces abertas, e usuários podem interagir entre si de forma equivalente.

Sayão (2007) salienta que as “[...] bibliotecas não são ilhas”, portanto a questão da interoperabilidade técnica é imprescindível, pois a partir dela é que bibliotecas, bibliotecas digitais e repositórios podem cumprir a sua função de compartilhar e possibilitar o acesso e a recuperação da informação com eficiência.

É necessário atenção a um aspecto em particular pertinente às ações de acesso livre, principalmente no que se refere aos repositórios digitais, e, no caso específico desse trabalho, dos RIs: a interoperabilidade técnica. Como já apresentado na introdução deste trabalho o número de repositórios vem crescendo, e com isso a preocupação com a interoperabilidade. Nesse sentido, Sayão (2007, p. 4) salienta que devem ser adotadas ações urgentes que forneçam estruturas e regulamentação para que esses repositórios funcionem “plenamente abertos e interoperáveis”, ou seja, para que os sistemas deixem de ser ilhas, é fundamental a atenção por parte dos profissionais da informação e do governo em relação à interoperabilidade organizacional e semântica em seus diferentes níveis e, sobretudo, a técnica.

Quando os repositórios e objetos digitais são criados dessa forma, o efeito global pode ser uma federação de repositórios que agregam conteúdo com muitos atributos diferentes, mas que podem ser tratados da mesma forma devido às suas definições de interface compartilhada (SAYÃO; MARCONDES, 2008).

Alipour-Hafeziet al. (2010) lembram que a OCLC é um exemplo de sistema de interoperabilidade. A OCLC criou um sistema em que bibliotecas de todo o mundo dividem seus dados bibliográficos no contexto do WorldCat. O produto WorldCat foi estabelecido há mais de 20 anos e ainda é usado.

Arms et al. (2002) identificam três estratégias de interoperabilidade: federação (sistemas cooperativos nos quais componentes individuais são projetados ou operados de forma autônoma), *harvesting* (coleta automática de metadados) e *gathering* (agregação automática de informação). Cada nível tem diferentes procedimentos de operação, padrões e protocolos.

Quanto ao Modelo Federado, a federação pode ser considerada a abordagem convencional para a interoperabilidade. Na federação, um grupo de organizações concorda que seus serviços estão de acordo com algumas especificações e integração de atividades, serviços, práticas e procedimentos, conteúdo e tecnologia (ARMS, 2002; ARMS et al., 2002; FERREIRA; SOUTO, 2006).

O termo federação, apesar de expressar uma estratégia específica de interoperabilidade, tem sido muito frequentemente usado para indicar, genericamente, a integração e a interoperabilidade entre repositórios digitais em diferentes níveis (SAYÃO; MARCONDES, 2008).

O protocolo Z39.50, que é um modelo federado, foi desenvolvido nos anos 80 e ficou conhecido por Linked Systems Project (LSP). Esse projeto envolveu a Biblioteca do Congresso Americano, OCLC, Research Libraries Information Network (RLIN) e a Western Library Network com o objetivo de transferir registros bibliográficos entre os participantes do projeto. Posteriormente, o protocolo desenvolvido pela LSP foi adotado pela NISO, acreditado pela ANSI como padrão para bibliotecas, editores e provedores de serviços de informação, o que pode ser considerado um padrão americano (NEEDLEMAN, 2000).

Alguns dos protocolos utilizados para o compartilhamento de informações são:

- a) o ANSI/NISO 39.50 (ISO 23950). É um protocolo de comunicação entre computadores que pode ser implementado sobre qualquer plataforma. Ele tem como propósito a pesquisa e a recuperação de informações (SAYÃO; MARCONDES, 2008). “A implementação do protocolo permite que, por meio de uma única interface, seja possível o acesso uniforme a uma diversidade de fontes de informações heterogêneas de modo síncrono e transparente para o usuário-final” (NISO, 2002, apud SAYÃO; MARCONDES, 2008, p. 138-139);
- b) o Search Retrieval Web Service (SRW) – um protocolo que se propõe a ser o sucessor do Z39.50 (SAYÃO; MARCONDES, 2008).

O Modelo *Harvesting*, que emprega o conceito de *metadata harvesting*, coleta automática de metadados, tem origem nas dificuldades encontradas na criação de grandes federações. Esse conceito estabelece que os participantes concordem em somar esforços que permitam compartilhar serviços básicos, sem a obrigação de adotar um conjunto completo de acordos (ARMS et al., 2002). O *metadata harvesting* foi estabelecido pelo protocolo OAI-PMH. Os serviços baseados em *harvesting* são assíncronos e muito mais simples de operar (SAYÃO; MARCONDES, 2008). Nesse modelo, cada biblioteca constrói o metadado sobre suas coleções disponíveis em um formato de troca simplificado.

Ainda que um determinado grupo de organizações não estabeleça nenhum grau formal de cooperação, um nível básico de interoperabilidade é ainda possível por meio de agregação automática de informações disponíveis publicamente, utilizando-se metabuscadores, robôs, máquinas de busca e ainda protocolos que suportem *Web services* e outros padrões da indústria de TI. A agregação requer, essencialmente, pouco ou nenhum esforço por parte dos participantes, entretanto oferece um grau baixo de interoperabilidade (ARMS et al., 2002, apud SAYÃO, MARCONDES, 2008, p. 138-139).

No Modelo *Gathering*, mesmo que a cooperação formal entre as organizações não seja possível, a base do modelo de interoperabilidade é ainda viável, por meio da coleta de informações abertamente acessíveis usando os motores de busca na *Web* (ARMS et al., 2002).

A padronização, que deve ser um compromisso cooperativo e repetitivo a partir do estabelecimento de convenção, é necessária para o entendimento de estruturas, convenções, sintaxe e significado dos metadados. O resultado dessa padronização estará refletido em uma comunicação efetiva, no melhor compartilhamento e reutilização das informações e nas reduções de custos e de retrabalho (CAMPOS, 2007).

Por utilizar a padronização como recurso, todos os aspectos da interoperabilidade devem ser formalmente definidos e cada organização tem o rígido compromisso de seguir exatamente o conjunto de padrões e procedimentos convencionados. Na prática, isso pode determinar o uso da mesma plataforma computacional – *hardware*, aplicativos e sistema operacional – e das mesmas condicionantes administrativas, reduzindo drasticamente a autonomia dos componentes individuais (ARMS, 2002, apud SAYÃO; MARCONDES, 2008). De acordo com Sayão e Marcondes (2008), as redes cooperativas de bibliotecas que adotam formatos como o da Lilacs e o MARC utilizam o modelo *gathering*. É um modelo que se restringe a esse tipo de rede.

Shintaku, Ferreira e Robredo (2011) realizaram um estudo que buscou identificar a questão da interoperabilidade entre os repositórios nacionais no que se refere aos esquemas de metadados utilizados. Os autores analisaram aspectos relativos à implementação dos esquemas de metadados e como esses repositórios estão tratando a catalogação e a interoperabilidade.

O Repositório ARCA, mantido pela Fiocruz, utiliza esquemas de metadados Open Archives Initiative-Dublin Core (OAI-DC), RDF e METS. Os autores salientam que esse foi o primeiro RI nacional a utilizar o sistema Open Archives Initiative-Object Reuse and Exchange (OAI-ORE), que permite, além do metadado, a disseminação do objeto digital.

Os esquemas de metadados são utilizados de acordo com o tipo de documento, o que facilita a interoperabilidade com iniciativas internacionais.

O Repositório Institucional da Universidade de Brasília (RiUnB) recorre aos esquemas OAI-DC, RDF e METS, que atende os principais requisitos de interoperabilidade.

O LUME, da Universidade Federal do Rio Grande do Sul, adota, além dos esquemas de metadados utilizados pela UnB, o Qualified Dublin Core (QDC) e o MTD-BR, que possibilitam interoperar com a BDTD.

O Banco Internacional de Objetos Educacionais (BIOE) disponibiliza o LOM, já que tem como objetivo descrever os objetos educacionais, e o Dublin Core.

Com esses exemplos, Shintaku, Ferreira e Robredo (2012) demonstram que, com o emprego de mecanismos como o *crosswalk*, é possível coletar (*harvesting*) as informações disponibilizadas pelos diferentes esquemas de metadados, e que a adoção de um determinado esquema é uma decisão do repositório que deve constar da política de informação institucional. Questão também abordada anteriormente pelos autores Hernon e Relyea (1991, apud JARDIM; SILVA; NHARRELUGA, 2009), Patel et al. (2005), Vargas (2008 apud ANDRADE; SILVA; CERVANTES, 2011) e Para Park e Tosaka (2010).

5.1 INTEROPERABILIDADE SEMÂNTICA

Há 17 anos, a comunidade científica já expressava que os desafios para as BDs eram: sua conceitualização e a sua relação com bibliotecas tradicionais e editoras; descrição de objetos; gerenciamento e organização de coleções; interfaces para usuários; interação homem-computador e interoperabilidade.

As discussões, durante o IITA Digital Libraries Workshop, foram em torno da infraestrutura focada em ferramentas comuns que permitissem o uso de tecnologias e padrões que fornecessem base para as questões de interoperabilidade, facilitando o crescimento das BDs na Internet. Porém, essa discussão só seria viável após o crescimento de BDs, para que pudessem se apoiar em fatos. A discussão sobre a interoperabilidade semântica teve uma atenção especial por parte dos debatedores, a partir do momento em que esta versa sobre a capacidade do usuário em acessar, de forma consistente e coerente, os documentos distribuídos em sistemas heterogêneos (LYNCH; GARCIA-MOLINA, 1995).

Como previsto em 1995, com o crescimento das bibliotecas digitais, a interoperabilidade passou a ter uma posição de destaque. Esse interesse vem crescendo em vista da indústria de conteúdos, como as editoras. “As pesquisas mais recentes sobre interoperabilidade no âmbito da arquitetura de bibliotecas digitais estão concentradas no desafio de criar uma infra-estrutura para acesso e integração de informação transversalmente aos domínios específicos” (SAYÃO; MARCONDES, 2008, p. 136).

Nos últimos anos, o número crescente de sistemas linguísticos, estruturais e até mesmo tipologicamente diferentes, que exigem uma integração, tem levado ao desenvolvimento de um modelo para um sistema internacional de organização do conhecimento abrangente, envolvendo um "sistema central" com o potencial de interconexão de vários sistemas de forma global (BOTERAM, 2010).

A interoperabilidade semântica define o significado de que o dado pode ser compreendido inequivocadamente por humanos e programas de computador, e que a informação pode ser processada significativamente da mesma forma. A integração semântica é o meio para se atingir a interoperabilidade semântica e pode ser considerada como subconjunto da integração de informações, que inclui acesso aos dados, agregação, correlação e transformação (SELVAGE et al., 2006).

Alemu, Stevens e Ross (2012) identificaram que alguns fatores que prejudicam a interoperabilidade estão relacionados à questão das bibliotecas terem dificuldade em adotar a padronização. A dificuldade das bibliotecas não seguirem um único padrão está relacionada ao fato de que são instituições culturais, e cultura é um fenômeno fluido. Esta fluidez torna difícil, se não impossível, fornecer definições e explicações objetivas para os objetos. Os objetos armazenados nas bibliotecas transmitem diferentes significados para os diferentes grupos de usuários e profissionais, que podem ser interpretados de diferentes formas. Os autores acreditam que seja improvável para os seres humanos concordarem com uma classificação singular, autoritária e hierarquizada dos objetos. Daí a dificuldade que esses autores têm em ver informação, sistemas de organização e padrões de metadados capazes de refletir as diversas interpretações da realidade. De acordo com Veltman (2004, apud ALEMU; STEVENS; ROSS, 2012, p. 41), as normas atuais tendem a aderir à interoperabilidade semântica dos metadados baseadas em um “[...] ponto de vista ontológico e objetivamente verdadeiro, que falha ao capturar e representar perspectivas e interpretações locais e/ou regional”.

Além desses aspectos, três fatores agravam essa situação: a) as normas se preocupam com a representação física do recurso; b) a falta de especialidade do bibliotecário prejudica a descrição adequada no que diz respeito ao aspecto semântico; e c) a dificuldade de acompanhar o crescimento das coleções digitais (ALEMU; STEVENS; ROSS, 2012).

Seguindo o pensamento de Alemu, Stevens e Ross (2012), identifica-se, em Degoulet et al. (2011, apud MOURA, 2011, p. 168), o que está intrínseco à complexa relação padronização dos vocabulários x interoperabilidade semântica: “[...] a interoperabilidade semântica requer a articulação de três elementos: a terminologia, a explicitação das relações semânticas dos elementos da terminologia e a representação dos conceitos nas mensagens”. De acordo com Moura (2011, p. 168), a terminologia diz respeito ao

[...] vocabulário controlado relacionado a um domínio de conhecimento ou de atividade. As relações semânticas articulam a rede de conceitos do domínio e expressam a natureza de cada conceito que a integra.

Isso fica claro até em relação a alguns termos utilizados na área da CI, como o termo serviços de informações que pode ser entendido e adotado pelos profissionais da área da CI e de TI de forma diferenciada.

Para Patel et al. (2005), o principal objetivo da interoperabilidade semântica é justamente apoiar processamentos complexos e avançados de consulta aos recursos em um contexto de informações heterogêneas. O estudo acerca do estado da arte sobre o tema produzido pela DELOS WP5 (Delos Network of Excellence on Digital Libraries) analisou diversos aspectos em relação à interoperabilidade semântica no ambiente das bibliotecas digitais: possibilidade de precisão da pesquisa, a compreensão sobre coleções e bases de conhecimento e a integração de recursos heterogêneos. Também enfatiza o seu papel na gestão do processo de ciclo de vida da informação.

O tema interoperabilidade semântica está presente em diferentes contextos, comunidades e disciplinas e há muito tempo faz parte do processo de indexação e do acesso à informação nas buscas que exigem algum grau de precisão e, mais recente, na integração de fontes de pesquisas distribuídas e heterogêneas no ambiente digital.

De acordo com Boteram (2010), o grau e a qualidade dos sistemas interoperáveis semanticamente e os modelos que fornecem fundamentos teórico e estrutural para a sua construção e implementação sempre têm que ser avaliados no que diz respeito à extensão de sua adequação para apoiar estratégias de recuperação. No entanto, a associação de um modelo específico de recuperação

com os respectivos modelos de interoperabilidade semântica não é para ser entendida como uma atribuição personalizada de forma rígida, mas sim como uma atribuição livre, que funciona de forma independente.

A ausência de uma linguagem de indexação controlada e estruturada torna muito mais difícil, ou mesmo impossível, para o usuário prever os elementos de um vocabulário específico usado na indexação – uma condição básica na recuperação da informação. Mesmo quando se usa um índice de linguagem altamente desenvolvido e complexo, as limitações de uma estratégia de busca utilizando um conjunto de palavras simples são evidentes. Restringir as estratégias de busca a jogos de palavras não é muito eficiente devido à alta especificidade dos conceitos individuais utilizados. A inexatidão, ou uma expansão deliberadamente inespecífica da consulta inicial, leva a um truncamento simples da busca. Pode-se identificar, no tesouro, um instrumento rígido para a recuperação da informação, que impossibilita por exemplo, o uso de busca booleana.

A interoperabilidade semântica necessária para suportar essa estratégia pode ser fornecida por modelos básicos dessa interoperabilidade, que estabelecem relações de equivalência entre os diferentes conceitos semanticamente idênticos em uma relação um-para-outro, sem especificar quaisquer tipos de relações.

Assim, o conteúdo informacional e a estrutura semântica de todo o campo têm que ser acessíveis a uma abordagem exploratória apoiada por uma navegação controlada e orientada.

Modelos específicos e altamente eficientes para uma interoperabilidade semântica podem melhorar o acesso a vários sistemas heterogêneos, estabelecendo uma ligação entre os conceitos documentados e o contexto temático.

Dessa forma, projetar, ajustar e alinhar linguagens documentárias é particularmente importante quando se combinam e integram vários tipos heterogêneos de tais sistemas em um esquema interligado.

Um amplo sistema internacional de organização do conhecimento, que consiste em uma rede heterogênea de sistemas interligados, exige uma estrutura altamente funcional e relacional, a fim de proporcionar uma infraestrutura eficiente para a recuperação bem-sucedida da informação.

Como os princípios e estratégias dos sistemas avançados de recuperação de informações dependem, em grande parte, de informações

semânticas, novos conceitos e estratégias para atingir a interoperabilidade semântica têm que ser desenvolvidos. Como as fronteiras tradicionais entre as instituições e as disciplinas começam a se confundir, os pesquisadores exigem cada vez mais acesso à informação a partir de uma ampla variedade de fontes, tanto dentro como fora de sua própria área de assunto.

Para entender a interoperabilidade semântica, é necessário conhecer as várias tecnologias e metodologias por trás dela, que são coletivamente chamadas de espectro semântico, o qual descreve uma série de tecnologias e metodologias para criar definições para dados cada vez mais precisos. Logo, deve existir um equilíbrio entre precisão e imprecisão – maior precisão não é sempre o melhor, e algumas questões precisam ser consideradas, como tempo, custo e esforço (SELVAGE et al., 2006).

De acordo com Baptista (2010), a questão da interoperabilidade semântica para os repositórios e bibliotecas digitais não pode ser restrita apenas à utilização do protocolo OAI-PMH; é necessário estabelecer e seguir regras apropriadas. A DCMI e o W3C têm desenvolvido algumas iniciativas para que se atinja esse objetivo. O projeto DRIVER 2.0, baseado nessas iniciativas, criou um conjunto de diretrizes para os fornecedores de conteúdo, que estão estreitamente relacionadas com a interoperabilidade semântica.

Projeto Driver 2.0

O DRIVER é um projeto financiado pela União Europeia, que contou com participação da Hyper Articles en Ligne - HAL (França), Digital Academic Repository - DARE (países baixos), German Initiative for Networked Information - DINI (Alemanha), Securing a Hybrid Environment for Research Preservation and Access - SHERPA (Reino Unido), BASE e o grupo OAI Best-Practice, que

[...] visa à constituição de uma estrutura organizacional e tecnológica para implementar uma camada de dados pan-europeia que permita o uso avançado de recursos de conteúdos na área da investigação no ensino superior. O DRIVER desenvolve uma infra-estrutura de serviços e uma infra-estrutura de dados [...] concebidas para instrumentar os recursos e serviços existentes na rede de repositórios (DRIVER, 2009, p. 9).

De acordo com as Diretrizes DRIVER 2.0 (2009, p. 11), o quadro atual dos repositórios não permite ainda que a recuperação de um documento seja rápida e direta, utilizando apenas alguns *clicks*. Para se obter uma recuperação simultânea do registro e do texto integral, é preciso o desenvolvimento de serviços avançados e integrados, como a “[...] pesquisa por assuntos combinada com a navegação através de classificações, análise de citações [...]”.

As diretrizes abordam cinco questões: coleções, metadados, implementação do protocolo OAI-PMH, práticas recomendadas para vocabulários e semânticas. Na questão dos metadados, foram definidas características obrigatórias e recomendações para resolver problemas relacionados com a semântica.

Neste estudo, são abordadas as recomendações do DRIVER em relação aos elementos de metadados do Dublin Core, *Subject*.

Essas diretrizes recomendam a utilização de vocabulários e semânticas, com o objetivo de reduzir a ambiguidade dos conceitos e termos empregados na comunicação científica, e a adoção do idioma inglês para informação descritiva, aspecto que gera algumas discussões por parte de países que têm outro idioma (BAPTISTA, 2010).

O elemento *Subject* retrata o assunto do recurso, geralmente expresso por palavras-chave, frases ou códigos de classificação. Pode-se utilizar um ou outro, mas, se possuir os dois, eles devem ser registrados separadamente. Recomenda-se evitar o uso de palavras genéricas para descrever o recurso.

Quando os termos são retirados de vocabulários controlados (LCSH) e de esquemas de classificação (CDD), deve-se utilizar um URI para identificação. Caso não exista nenhum esquema de classificação específica, o DRIVER recomenda a utilização da CDD.

De acordo com Baptista (2010), os esquemas de classificação adotados no contexto OAI são: Classificação da Biblioteca do Congresso Americano (LCC), CDD e CDU. Os sistemas de cabeçalhos de assuntos são: LCSH e Schlagwortnormdatei (SWD). Os metadados OAI contêm códigos de áreas a partir de esquemas como Mathematics Subject Classification (MSC) e MeSH.

Uma das observações do DRIVER (2009, p. 120) é que esses esquemas sejam expostos pelo RI. Com a utilização desses elementos de classificação, é possível implementar serviços de navegação por classificação, substituir “[...] códigos de classificação por termos em inglês, tradução de termos

para outros idiomas e fazer uma junção dos códigos de classificação utilizando regras de mapeamento”.

Para Baptista (2010), as Diretrizes DRIVER são extremamente importantes para a consecução da interoperabilidade, pois advogam a necessidade de atribuir significados comuns às propriedades e a definição de regras e esquemas de codificação (elementos/valores do elemento *Subject*). A identificação por meio de URI e a adequação de esquemas de codificação (esquemas de sintaxe e de vocabulários) são os requisitos necessários para o alcance da interoperabilidade semântica entre os repositórios digitais. De acordo com Baptista (2010, p. 82), “[...] a obtenção de melhores resultados de interoperabilidade só será possível por meio da utilização de esquemas de codificação comuns”. A utilização de vocabulários controlados permite que haja uma completa integração entre os repositórios. A autora sugere alguns trabalhos futuros, entre eles, “[...] a criação de mapeamentos directos para descrições em RDF, a criação de perfis de aplicação locais, o mapeamento entre vocabulários controlados locais e globais”.

Koch (2006) recomenda algumas ações para a otimização da interoperabilidade, que devem ser adotadas de acordo com o escopo e o objetivo dos serviços oferecidos pelos diferentes tipos de repositórios. Em relação ao escopo, um repositório deve focar no armazenamento e na simplicidade do acesso; integração de dados e fontes heterogêneos por meio de um serviço comum; as descobertas e o acesso aos serviços, melhorando a qualidade dos metadados; serviços avançados como as bibliotecas digitais devem incentivar a interoperabilidade semântica e a organização do conhecimento a fim de fornecer um acesso à informação mais preciso. Em relação aos níveis de cobertura de um repositório ou serviço, pode-se adotar a padronização das descrições dos seus recursos e concentrar na melhoria semântica. Serviços de compartilhamento entre repositórios devem preservar as peculiaridades de cada coleção. Devem ser disponibilizadas opções de busca, navegação e filtragem que possibilitem o acesso à multiplicidade de vocabulários, utilizando-se o mapeamento para tratar a heterogeneidade entre os diferentes vocabulários existentes.

O mapeamento pode ser de elementos, semântico ou sintaxe, provenientes de esquemas de metadados. Ele permite que um metadado usado por uma comunidade seja adotado por um padrão de metadados diferente, mas depende da similaridade entre os esquemas, da granularidade dos elementos e da

compatibilidade de regras de conteúdo (NISO, 2004). Como exemplo, há o mapeamento utilizando o MARC 21 e o DC, em que o elemento título no DC – *title* – corresponde ao campo 245 00\$ no MARC 21.

A existência de recomendações nacionais que favoreçam a interoperabilidade semântica possibilita pesquisas e navegação em serviços que agregam assuntos e beneficia a disponibilização de documentos para pesquisas em domínios e *download* em escala internacional. No entanto, os serviços em disciplinas específicas e esquemas de vocabulários nacionais disponibilizados atualmente não são suficientes para a recuperação eficaz dos recursos. A normalização internacional de esquemas de vocabulários e a padronização internacional para uma ampla cobertura de assuntos são recomendadas quando se quer ir além da pesquisa por títulos, autores, instituições (HEERY et al., 2006b). Iniciativas em termos de políticas que visam a interoperabilidade semântica adotadas por organismos internacionais podem ser acompanhadas no Quadro 8.

Quadro 8 - Exemplos de boas práticas para obtenção da interoperabilidade semântica

PAIS	ÓRGÃO	RECOMENDAÇÕES
Alemanha	German Initiative for Networked Information (DINI)	São pré-requisitos mínimos para os repositórios obterem o Certificado de Qualidade oferecido pelo DINI: disponibilidade de uma política bem definida para a indexação de assuntos, indexação de documentos com a utilização de palavras-chave e utilização da CDD como esquema de classificação para os documentos. Saliena que apenas a utilização de vocabulário controlado não será suficiente para atender os requisitos para uma recuperação com qualidade e precisão. Será necessária a adoção de sistemas internacionais de indexação.
Suécia	Coordination of e-Publishing within Swedish Higher Education (SVEP)	Recomenda um perfil de metadados para a interoperabilidade entre repositórios institucionais suecos em nível OAI e para a geração de relatórios de metadados para a biblioteca nacional, codificado em XML. Um nível mínimo é exigido em relação ao DC unqualified para o elemento assunto e para os serviços SVEP, um sistema de classificação. A classificação de assunto para as publicações científicas é baseada no sistema da Central Statistical Office and the Science Council. Inclui mapeamento para outros SOC. Todos os repositórios que seguem a SVEP devem utilizar uma classificação comum de assunto.
Austrália	Australian Research Repositories Online to the World (ARROW)	Os registros de metadados são coletados diariamente e disponibilizados para busca, navegação, citação ou estatística. Recomenda a American Subject Rules Classification (ASRC) para a classificação de assuntos, podendo utilizar sistemas de palavras-chave (LCSH) ou livres. Algumas bibliotecas adotam a LCC, além da ASRC, com o objetivo de alcançar a interoperabilidade internacional.
Grã-Bretanha	JISC's Digital Repositories Review (HEERY; ANDERSON, 2005)	Inclui a utilização de ferramentas inteligentes para a extração automática de dados e a classificação automática. Utiliza ferramentas como <i>clustering</i> e serviços de sinônimos e ferramentas para facilitar a exploração interdisciplinar.

Fonte: Koch (2005).

5.2 OS SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO E A INTEROPERABILIDADE SEMÂNTICA

Como constatado na literatura apresentada até aqui, muitos vocabulários controlados são construídos ou adaptados às coleções, e cada vez mais esses vocabulários vêm sendo empregados na Web. A padronização, ou esquemas universais, nem sempre é bem aceita pela comunidade em geral. O ideal,

para Tennis (2004), seria que esses vocabulários pudessem trabalhar juntos, de forma interoperável, e possibilitar o acesso por assunto aos recursos de várias coleções. Porém, essa situação ainda não existe, e uma forma de facilitar esse tipo de interoperabilidade é construir mecanismos que permitam diferentes vocabulários se comunicarem, utilizando significado, relações e níveis de extensão²⁹ e intenção³⁰ por distintos grupos de usuários por meio de diferentes vocabulários controlados em múltiplas coleções no ambiente Web (BAPTISTA; MACHADO, 2001).

A interoperabilidade semântica não pode ser implementada tão facilmente. Não basta saber que é necessária ou possuir conhecimento sobre o assunto ou adotar processamento de dados ou criação de serviços. É imprescindível possuir uma infraestrutura, ou seja, arquitetura, protocolos, soluções sintáticas, esquemas de codificação e sistemas de identificação, padrões de organização, sustentação jurídica e serviços de apoio – como acesso a bases de conhecimento e semânticas e os SOC, em que se incluem as ontologias fundacional ou de fundamentação e de núcleo.

Segundo Guizzardi (2005, apud CAMPOS; CAMPOS; MEDEIROS, 2011),

[...] a ontologia de fundamentação tem por objetivo identificar categorias gerais de certos aspectos da realidade que não são específicos a um campo científico, descrevendo conhecimento independentemente de linguagem, de um estado particular das coisas ou ainda do estado de agentes,

e a ontologia de núcleo indexa bibliotecas de ontologias de domínio e serve de base para a construção de linguagens de representação e sistemas de inferência para um domínio de conhecimento.

De acordo com Tudhope, Kock e Heery (2006), a utilização dos SOC melhora a interoperabilidade semântica. Entretanto, às vezes é necessária a criação ou a extração dos vocabulários e mapeamento, a fim de tornar os sistemas interoperáveis semanticamente, sobretudo quando os SOC possuem estrutura, linguagem de domínio ou granularidades distintas. Não existe um consenso sobre

²⁹ Extensão é o conjunto de gêneros ou espécies possuidoras das mesmas qualidades (PIEADÉ, 1977, p. 28).

³⁰ Intenção é o número de qualidades que distinguem as espécies, a soma de todas as propriedades comuns aos membros das classes (PIEADÉ, 1977, p. 28).

qual esquema, sistema de vocabulário ou abordagem deve ser adotado (NICHOLSON et al., 2001; PATEL et al., 2005; TUDHOPE; KOCH; HEERY, 2006). Essa falta de concordância favorece a utilização de um mecanismo de integração que possa melhorar a interoperabilidade, possibilitando uma independência em relação à descrição de assuntos em vários domínios. Exemplos bem-sucedidos de utilização de abordagens biblioteconômicas são verificados no contexto de museus e arquivos, no entanto verifica-se que muitos trabalhos sobre o assunto se concentram em avaliações de alguns sistemas utilizados.

A interoperabilidade semântica no ambiente de bibliotecas digitais está relacionada ao compartilhamento consistente do uso de terminologias, de modo que a interoperabilidade no seu nível mais alto ou domínio possa ser obtida por meio de negociação e aceitação dos conceitos compartilhados, os termos e seus significados. Nesse sentido, os serviços de terminologia desempenham papel importante para a facilitação do uso consistente e coerente de termos comuns e sua semântica (PATEL et al., 2005).

De acordo com Koch (2006), um número reduzido de repositórios e de agregadores de serviços adota medidas para harmonizar a heterogeneidade das informações ou oferecem serviços que facilitem a pesquisa e a navegação por assunto ou tópicos para recuperação dos recursos.

O emprego da padronização para alcançar a interoperabilidade semântica em uma biblioteca digital inclui forma e significado dos metadados, esquemas de conteúdo, compartilhamento de conceitos definidos pelos SOC, uso de nomes e construção de identificadores para conceitos e itens.

Koch (2006) cita algumas vantagens ao adotar a padronização como forma de alcançar a interoperabilidade:

- a) as informações podem ser imediatamente transferidas, integradas e incorporadas sem transformação;
- b) a informação pode ser mantida sob uma única forma;
- c) fontes incorporadas podem ser aplicadas para serem funcionalmente completadas por um serviço integrado.

Como desvantagens:

- a) as fontes de informação precisam de adaptação ao padrão;
- b) o empreendimento para um padrão terminológico pode ser muito alto;
- c) as normas têm que prever o uso futuro, e, com as constantes mudanças, há necessidade de frequentes avaliações;
- d) é necessária uma decisão política em relação ao padrão, pois esse, às vezes, não é ideal para outras aplicações; e
- e) a adaptação de informação para um padrão exige uma interpretação que pode ser manual ou automática, levando à perda de informação.

Por outro lado, quando não se emprega a padronização, pode-se adotar a interpretação mecânica, que compreende:

- a) mapeamento;
- b) correlação de conceitos definidos pelos SOC; e
- c) tradução de nomes e reformatação de identificadores (Digital Object Identifier – DOI, International Standard Book Number – ISBN, International Standard Serial Number – ISSN, URI).

Assim como a padronização, a interpretação mecânica, que engloba mapeamento, correlação, tradução e reformatação de identificadores, possui suas vantagens e desvantagens, a saber:

- a) fontes de informação em dados originais não precisam de adaptação;
- b) as fontes podem servir como função local adicional;
- c) apenas partes de aplicativos necessitam de interpretação;
- d) a interpretação pode ser otimizada por múltiplas funções; e
- e) interpretadores podem ser facilmente adaptados às alterações.

São citadas as seguintes desvantagens:

- a) é necessário tempo de comunicação para o processamento da interpretação;
- b) o custo da produção manual da base de conhecimento, como tabelas de correlação terminológica, pode ser elevado; existe a possibilidade do uso de processos automatizados;
- c) pode haver perda de informação durante a interpretação ou comprometer a precisão do sistema; e
- d) existe a possibilidade da interpretação mecânica não ser extensiva a todos os itens.

Há ainda a opção de trabalhar com a combinação dos dois processos, desde que ocorram um equilíbrio e um diagnóstico sobre o objetivo e a estrutura de cada sistema.

A adoção de padrões comuns para representação e acesso a vocabulários tem a vantagem de permitir a interoperabilidade em ambientes de rede e uma divisão de esforços. Vocabulários e recursos de informação e pesquisa/indexação/ferramentas de mapeamento podem ser desenvolvidos em instituições separadas e hospedados em diferentes locais (TUDHOPE; KOCH; HEERY, 2006).

De acordo com Tudhope, Koch e Heery (2006), a combinação de sistema de classificação com tesouro para indexação promove recursos excelentes. Outra possibilidade é o mapeamento de dois vocabulários juntos. Isso pode facilitar as pesquisas avançadas, com a expansão de busca ou a integração de pesquisa e navegação. O serviço de terminologia aumenta a qualidade da pesquisa ao proporcionar pontos adicionais para a navegação, e a taxonomia possui a função didática de auxiliar o usuário a identificar áreas de interesse, ajudando-o a descobrir a informação (CAMPOS; GOMES, 2008).

Serviços de terminologias podem corrigir termos desatualizados de vocabulários controlados e identificar valores de vocabulários recomendados, bem como atribuir esquema de codificação apropriado para a recuperação de informações (HILLMAN et al., 2004, apud KOCH, 2006); divulgar conceitos, termos e relações semânticas; promover o uso consistente de vocabulários; explicitar a

semântica; fornecer exemplos de utilização e de melhores práticas, e tornar a informação acessível em relação a proveniência (informação sobre a fonte de dados, como atualização), circulação, autoridade, dedução e processos de raciocínio (PATEL et al., 2005, p. 35).

Zeng e Chan (2004) e Chan e Zeng (2002) citaram alguns métodos de acesso aos recursos digitais que melhoram a interoperabilidade:

- a) derivação/modelagem – utiliza-se um vocabulário existente mais abrangente para gerar um vocabulário especializado ou mais simples. Ex.: Lista de cabeçalho canadense em inglês (CSH), utilizada pelas bibliotecas canadenses, foi desenvolvida como uma lista complementar para LCSH.
- b) tradução/adaptação – é realizada a tradução de um vocabulário controlado para outro idioma. Ex.: O repertório de Vedettes-Matière (RVM), desenvolvido pela Université de Laval (Canadá), é uma adaptação francesa de LCSH e CSH.
- c) mapeamento (intelectual) – consiste no estabelecimento de equivalência entre termos de diferentes vocabulários controlados ou entre termos verbais e classificações. Esse método, geralmente, demanda grande quantidade de esforço intelectual. Ex.: Renardus, WebDewey.
- d) ligação – vocabulários especializados são tratados como uma superestrutura de ligação (Ex.: European GEMET thesaurus).
- e) *switching* – um sistema de comutação ou esquema para intermediar a movimentação de termos equivalentes entre diferentes vocabulários. Ex.: Megathesaurus da HW Wilson Company (CHAN; ZENG, 2002).

Para melhor entendimento, compreende-se aqui, como mapeamento, a atividade mental de comparar e analisar duas ou mais linguagens documentárias. E o *crosswalk* é a expressão visual e textual do processo de mapeamento (MOYANO; GRIMALDO, 2011).

Diferentes expressões linguísticas para o mesmo conceito, distintos graus de especificidades e polissemia são dificuldades para o mapeamento de

vocabulários multilinguais ou multidisciplinares. São essas incompatibilidades que dificultam o mapeamento (CHAN; ZENG, 2002), entretanto algumas ações como a construção de tesouros com base na norma ISO 2788 e outras normas garantem que a estrutura e a gramática de um determinado vocabulário seja consistente ou compatível. Os autores veem a interoperabilidade dos SOC como inevitável no ambiente Web e que os resultados de projetos que visam à interoperabilidade entre os diferentes vocabulários são encorajadores.

Vários tipos de serviços de terminologia têm sido implementados no contexto das bibliotecas digitais como forma de auxiliar a reconciliação semântica, que inclui o uso de vocabulários, classificações, taxonomias e tesouros – garantindo a interpretação semântica correta dos termos. Esses serviços potencializam a interoperabilidade semântica de registros, repositórios de metadados, esquemas de metadados, registros de *crosswalk*, ou mapeamentos entre vocabulários, e ontologias.

Para que seja possível o aumento da qualidade das informações heterogêneas em serviços disponíveis na Web, humanos e agentes (máquinas) precisam compreender o significado do que é recuperado. O resultado das buscas poderia ser beneficiado pelas funcionalidades proporcionadas pelos serviços de informação, que utilizam, como suporte semântico, vocabulários, classificações, taxonomias e tesouros.

Zeng e Chan (2004) mostram que a busca por uma abordagem ideal de uma consulta “*one-stop*”, ou seja, um processo de busca contínua, em que o usuário poderá recuperar informações utilizando diversos recursos heterogêneos sem a necessidade de pesquisar em bases de dados ou coleções separadamente, apresenta resultados animadores. Os autores fizeram uma breve retrospectiva de como os SOC se mostraram eficazes para a recuperação da informação. Na era digital, com o aumento da produção de informação, se faz necessária a atenção para a compatibilidade entre esses sistemas. Embora existam alguns que possuem normas para sua elaboração, como no caso dos tesouros, não deixa de haver diferenças e variações de terminologias e sintaxe inerentes a domínios específicos.

Para os profissionais da informação, o panorama atual não reflete somente o crescimento exponencial de informação, que não permite acompanhar o seu processamento, seja manual ou automático. Também se depara com a proliferação de padrões de metadados e SOC utilizados nos diversos serviços de

informação disponíveis na Internet. Diante desse panorama, foram criados os sistemas de organização do conhecimento em rede (NKOS³¹). Como dispositivos de informação interativos, os NKOS são destinados a apoiar a descrição e a recuperação de recursos de informação heterogêneos na Web.

Um das primeiras questões sobre a interoperabilidade semântica diz respeito ao mapeamento dos SOC em diferentes idiomas. O projeto Multilingual Access to Subjects (MACS) relaciona vocabulários controlados em três idiomas (inglês, francês e alemão) nos catálogos de quatro bibliotecas nacionais, e Mérimée reúne vocabulários em uma base de dados sobre patrimônio francês em diversos idiomas. Já outros projetos adotam o mecanismo de comutação entre os vocabulários envolvidos:

- a) UMLS Metatesouro – projeto desenvolvido pela National Library of Medicine, EUA. O Metatesouro possui mais de 100 vocabulários da área biomédica mantidos por diversas organizações em vários idiomas.
- b) UC Berkeley DARPA (EUA) – cria um índice para os vocabulários de entrada, mapeando-os para vocabulários de consulta, e oferece suporte a consultas em vários idiomas.

No que diz respeito ao estabelecimento da interoperabilidade entre diferentes SOC, existem aspectos distintos que devem ser considerados, como características e estruturas.

Zeng e Chan (2004) elencaram exemplos de projetos desenvolvidos com SOC que possuem diferentes estruturas:

- a) Megathesaurus – da HW Wilson Company – é resultado de fusão entre distintos vocabulários controlados; e
- b) WebDewey (OCLC) – mapeamento entre LCSH e CDD.

Os SOC que possuem tipos de estruturas semelhantes:

- a) Renardus – mapeamento de vários esquemas de classificações locais com a CDD;
- b) MACS – cabeçalhos de assunto em três listas monolíngues e LCSH; e

³¹ Modelo funcional e de dados para permitir que os sistemas de conhecimento da organização e os serviços (KOS), como sistemas de classificação, tesauros, *gazetteers* e ontologias, funcionem como rede de serviços interativos de informação para apoiar a descrição e a recuperação de recursos de informação no ambiente digital.

- c) LCSH e MeSH – (Northwestern University – EUA) – mapeamento entre LCSH e MeSH.
- d) CAMed (Columbia University e Kent State University) – *links* para quatro tesouros na área de medicina complementar e alternativa. As estruturas dos tesouros são normalizadas antes de serem armazenadas no repositório.

Gomes e Guimarães (2012) também elencam vários trabalhos envolvendo utilização de tabelas de classificação, cabeçalhos de assunto, vocabulários controlados. Um dos exemplos citados é a utilização de mapas conceituais adotados para a classificação automática, e a geração de *topic maps* por meio de dispositivos gerados automaticamente desenvolvido pela OCLC.

O desenvolvimento de ontologias a partir de um mapa visual do conhecimento com a CDD, de acordo com Chowdhury e Chowdhury (2004) e Chowdhury (2002), citados por Gomes e Guimarães (2012), facilitaria a formulação das buscas promovendo melhores resultados na recuperação. Já a utilização da CDD é recomendada para elaboração de taxonomia para facilitar a navegação na rede a partir da categorização do conhecimento.

A OCLC é responsável por apoiar grande número de projetos que enfocam o desenvolvimento de produtos e serviços no ambiente digital, entre eles a utilização das tabelas de classificação CDD e LCC e listas de cabeçalhos de assunto, LCSH para classificação automática e descrição de recursos (GOMES; GUIMARAES, 2012).

A ontologia não é aceita como um vocabulário controlado, mas pode ser utilizada para desenvolvê-lo. Já um tesouro pode ser entendido como uma forma simples de ontologia (PATEL et al., 2005).

Os autores citados por Heery et al. (2006a, 2000b) revelam que existem muitas propostas para melhorar a interoperabilidade entre vocabulários de domínios específicos, tesouros e ontologias no contexto da recuperação da informação. Nos serviços de informação e no campo da representação do conhecimento, o conceito de ontologia mostra um modelo que representa um domínio específico de interesse. Esse tipo de solução é baseado nos modelos que ajudam a definir uma base comum entre diferentes comunidades de informação.

Nesse sentido, o Sistema Observer, desenvolvido por Mena (1998), permitiu uma arquitetura de busca em um sistema de informação global

interoperável por conta da ontologia. Além de cada ontologia definir os termos utilizados no conceito de um recurso específico do repositório, por meio da compilação feita pela ontologia de termos que são posteriormente mapeados para uma estrutura de dados específica (nome da entidade e atributos – elementos de metadados). A “interontologia”³² contém a relação que conecta os termos entre diferentes ontologias e permite a tradução pelo usuário para a ontologia de cada repositório.

Em outra pesquisa desenvolvida por Hunter (2001), foi implementada uma ontologia baseada nos significados de um tesouro (MetaNet) aplicado ao modelo ABC. O principal objetivo foi permitir que o computador, por meio de conhecimento semântico, entendesse a relação de equivalência e hierarquia entre os termos de metadados de diferentes domínios.

O MetaNet foi desenvolvido para pesquisas, na WordNet, de termos utilizados nos diferentes domínios. Acredita-se que os tesouros sejam implementados com ferramentas como RDF e RDFS (RDF Schema) emprestados da Web semântica, ao qual está estreitamente relacionado (TOLOSANA-CALASANZ et al., 2006).

O projeto HILT, desenvolvido por Nicholson et al. (2001), buscou identificar a utilização dos SOC nas bibliotecas, museus e arquivos do Reino Unido. A CDD, LCSH, o dicionário de sinônimos da UNESCO e o Hasset (tesouro eletrônico da área de Ciências Sociais e Humanas) foram identificados como os sistemas mais utilizados, assim como os sistemas digitais Art & Architecture Thesaurus (AAT), Agricultural Information Management Standards (AGROVOC), CAB Abstract Thesaurus (CABI), General Multilingua Environmental Thesaurus (GEMET), o MDA Archaeological Objects (MDA) Thesaurus e o MeSH, embora tenha sido detectado um número maior de utilização de sistemas desenvolvidos internamente.

Vários projetos de mapeamento multilíngue e multiestrutura já foram iniciados, utilizando uma variedade de métodos. Vislumbra-se a existência de produtos e serviços multilíngues envolvendo multiestruturas, como tesouros, classificações, cabeçalhos de assunto e lista de termos atribuídos aos registros de banco de dados (CHAN; ZENG, 2002, p. 6).

³² Uma interontologia pode ser entendida como uma intersecção (sobreposição) a relação entre as entidades afins.

Dessa forma, a necessidade da padronização dos SOC no ambiente digital parece ser inevitável para que coleções, serviços e SOC possam trabalhar conjuntamente em arquiteturas de bibliotecas digitais (HILL et al., 1999; TENNIS, 2004). Diante da inexistência de uma padronização dos SOC, pode-se recorrer: a) à busca pela interoperabilidade entre eles; b) à utilização de métodos, como o *switching*, para a obtenção da interoperabilidade entre os SOC, como uma alternativa, e também à utilização de registros de autoridade de assunto de vários sistemas *on-line*; e c) ao emprego do mapeamento, que conta com o auxílio de inovações tecnológicas para o gerenciamento de grandes arquivos de assuntos e gerenciamento de *links* (HILL et al., 1999).

Outro projeto é Scorpion, da OCLC, que possui ferramentas para reconhecimento automático de assuntos, utilizando a CDD.

Há vários tipos de vocabulários servindo diferentes propostas, com diferentes graus de controle, riqueza de relação semântica, formalidade e controle editorial. Existem muitas opções de serviços de terminologia, interativos e automáticos (TUDHOPE; KOCH; HEERY, 2006).

Com base nessa exposição, é possível identificar que a Organização do Conhecimento, por meio dos SOC, propicia o compartilhamento consistente de terminologias necessário à interoperabilidade semântica no contexto dos repositórios. A partir da adoção e da identificação de vocabulários controlados por esses sistemas é possível a utilização de serviços de terminologia, que, por sua vez, potencializam a interoperabilidade semântica.

6 METODOLOGIA

O problema abordado neste trabalho vai ao encontro de outras iniciativas nacionais e internacionais. Mas, no Brasil, elas ainda são incipientes, principalmente quando comparadas com iniciativas internacionais. A forma como a organização do conhecimento pode contribuir para a interoperabilidade semântica nos RIs está sendo abordada por vários organismos nacionais e internacionais e por diversos autores, como pode ser verificado no referencial teórico, motivo que nos levou a pesquisar não apenas a literatura como também avaliar o que está sendo realizado pelos RIs no Brasil.

O desenvolvimento deste estudo foi baseado na literatura. A pesquisa realizada é respaldada em uma sistemática fundamentada na pesquisa bibliográfica. O levantamento bibliográfico preliminar buscou identificar os temas organização do conhecimento e interoperabilidade semântica. Para isso, foram utilizadas as seguintes palavras-chave: *metadado/metadata*; *interoperabilidade/interoperability*; *interoperabilidade semântica/semantic interoperability*. Esses termos foram pesquisados isoladamente e relacionados com o foco do estudo, repositório institucional/*institutional repository*. Não houve restrição em relação ao período de cobertura da pesquisa ou idioma.

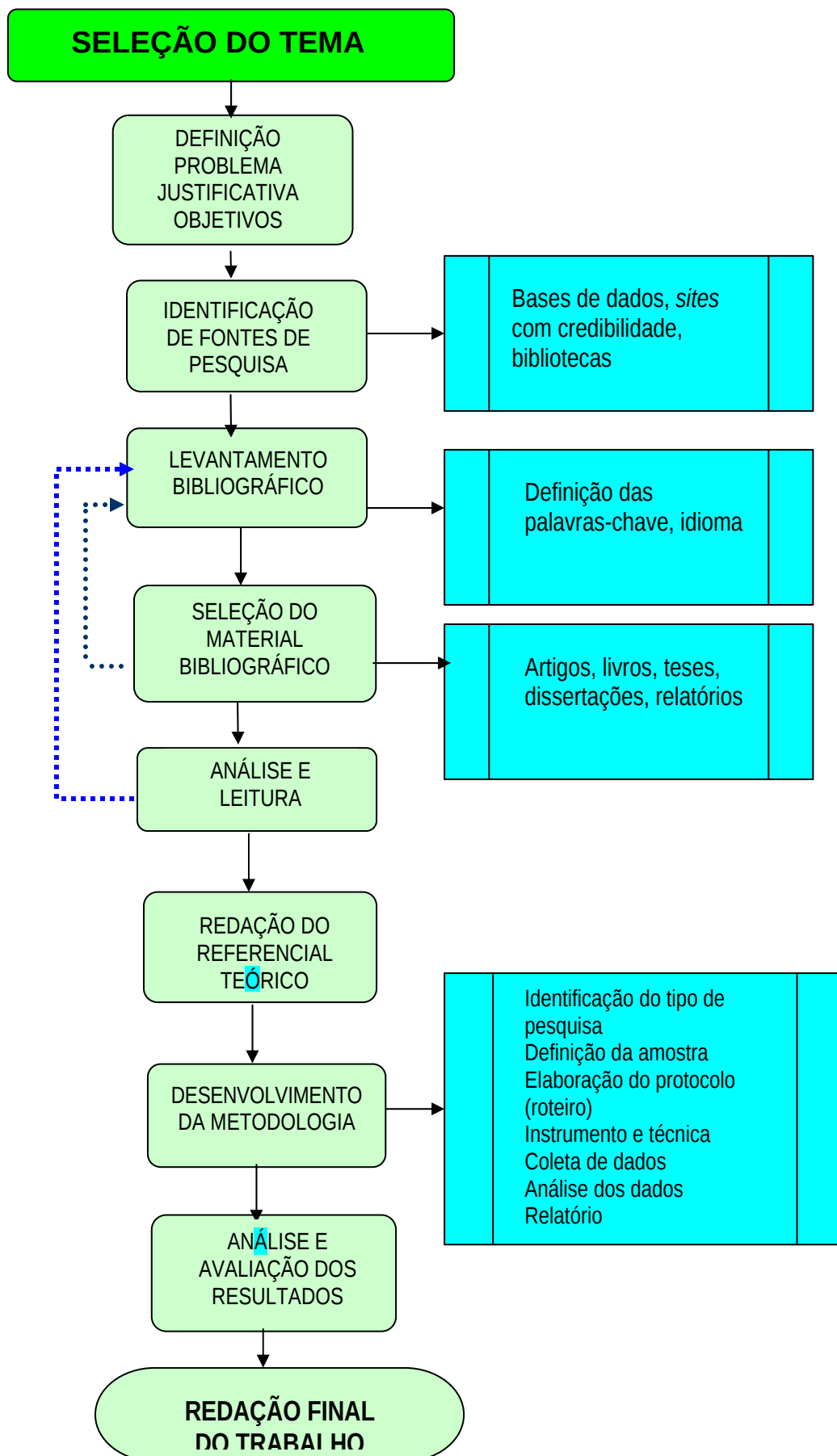
As bases de dados e *sites* selecionados para a pesquisa foram:

- periódicos CAPES – área de conhecimento, ciências sociais aplicadas, subárea, ciência da informação, bases de dados: Academic Search Premier (ASP); Annual Reviews; Emerald Fulltext (Emerald); Information Science & Technology Abstracts (ISTA); Library Literature and Information Science Full Text (Wilson); Library, Information Science & Technology Abstracts with Full Text (EBSCO); Library and Information Science Abstracts - LISA (CSA); SciELO.ORG; ScienceDirect (Elsevier); Web of Science;
- b) Mendeley, gerenciador de referências – podem ser realizadas estudos disponibilizados por grupos de pesquisas em *sites* de bibliotecas e bases de dados. Também possibilita a interação com grupos de pesquisas (www.mendeley.com);

- c) W3C, o Consórcio World Wide Web – é uma comunidade internacional que desenvolve padrões com o objetivo de garantir o crescimento da *Web* (www.w3.org);
- d) DSpace – é um *software* livre elaborado para aplicação em repositórios institucionais (www.dspace.com);
- e) EPrints – *software* desenvolvido para aplicação em repositórios institucionais (www.eprints.org);
- f) OpenDoar e ROAR – diretórios de registro de repositórios (www.opendoar.org e www.roar.eprints.org);
- g) Dublin Core Metadata Initiative (DCMI) – é uma organização aberta preocupada com o desenvolvimento da interoperabilidade dos padrões de metadados;
- h) DRIVER, Digital Repository Infrastructure Vision for European Research (www.driver-repository.eu); e
- i) JISC, Joint Information Science Committee – instituição britânica especializada em informação e tecnologia digital para educação e pesquisa (www.jisc.ac.uk).

Com o objetivo de identificar os elementos e instrumentos da OC que contribuem para a interoperabilidade semântica entre RIs, foram analisados os metadados descritivos utilizados para descrição do recurso e os elementos da representação temática que propiciam padronização e normalização das informações disponibilizadas pelos repositórios. Com o intuito de atingir esse objetivo, é apresentado, a seguir, o fluxo utilizado para o desenvolvimento da pesquisa (Fluxograma 1).

Fluxograma 1 – Desenvolvimento da pesquisa



6.1 PROCEDIMENTO METODOLÓGICO

De acordo com Gil (2008, p. 28), pode-se afirmar que, quanto aos objetivos, este estudo é classificado como uma pesquisa descritiva, pois visa à “[...] descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis”. A escolha pela análise qualiquantitativa deve-se à sua melhor adequação à natureza da questão deste estudo, pois permite uma relação entre o pesquisador, seu objeto de investigação e o contexto social. A pesquisa qualitativa possibilita a análise do conhecimento e das práticas adotadas pelos participantes, “[...] considera que pontos de vista e práticas no campo são diferentes devido às diversas perspectivas subjetivas e ambientes sociais a eles relacionados” (FLICK, 2002, apud BRAGA, 2007, p. 28). Esse tipo de pesquisa utiliza

[...] valores, crenças, representações, hábitos, atitudes e opiniões e adequa-se a aprofundar a complexidade de fenômenos, fatos e processos particulares e específicos de grupos mais ou menos delimitados em extensão e capazes de serem abrangidos intensamente (MINAYO; SANCHES, 1993, apud BRAGA, 2007, p. 28).

Por sua vez, o método quantitativo é apropriado para mensurar a extensão do objeto de estudo, sob a ótica do público pesquisado. Esse tipo de abordagem também é recomendado para identificar o grau de conhecimento: as impressões, os hábitos, os comportamentos em relação a produto, comunicação, serviço ou instituição (ALVES-MAZZOTTI; GEWANDSZNAJDER, 1999).

O método escolhido foi o estudo de casos múltiplos, que possibilita uma análise aprofundada de uma realidade delimitada (TRIVIÑOS, 2007), comparando semelhanças e destacando diferenças. “No caso das semelhanças, examinam-se tendências que presumivelmente estão afetando a todos, quando se tratar do destaque das diferenças, estas estão definidas *a priori* e exploradas com o fim específico de se construir uma teoria” (FERNANDES; GOMES, 2002, p. 15).

Para Yin (2005), o estudo de caso tem algumas vantagens, como permitir a explicação de ligações causais em situações complexas demais para utilizar levantamento de dados e estratégias experimentais. Possibilita, ainda, a

descrição de contexto real que sofreu intervenção e a avaliação de uma intervenção em curso.

A essência de um Estudo de Caso ou a tendência central de todos os tipos de Estudo de Caso é que eles tentam esclarecer ‘uma decisão ou um conjunto de decisões: por que elas foram tomadas? como elas foram implementadas? e, quais os resultados alcançados (YIN, 2005, p. 14).

Laville e Dionne (1999) também acrescentam que uma das maiores vantagens do estudo de caso é a possibilidade de aprofundamento pela oportunidade de adaptação dos instrumentos e modificação da abordagem para explorar elementos imprevistos.

Para garantir a eficácia e a credibilidade do estudo de casos múltiplos, é necessário apresentar as seguintes características:

- a) validade do construto – revisão de literatura, utilização de várias fontes (análise de *site*, questionário);
- b) validade interna ou credibilidade – utilização de triangulação, revisão da literatura, pesquisa documental e questionário, um processo que permite evitar ameaças à validade interna inerente à forma como os dados de uma pesquisa são recolhidos;
- c) validade externa – descrições sobre o caso de estudo, procedimentos para codificação e análise;
- d) confiabilidade – registro dos dados coletados; e
- e) relatório final da pesquisa (CALAZANS, 2007).

A pesquisa documental também foi utilizada por tratar-se de uma fonte rica e estável de dados, propiciando uma visão mais clara do problema (FERNANDES; GOMES, 2002).

6.1.1 Definição da Amostra

De acordo com o OpenDOAR, foram registrados 114 repositórios digitais no Brasil até abril de 2012, sendo 53 institucionais/departamentais (OPENDOAR, 2012). Trinta e oito foram implantados com o incentivo do

IBICT/FINEP por meio de convite para participar do projeto piloto (4 instituições) e posteriormente por editais (02/2009 e 03/2010).

O projeto IBICT/FINEP, por meio desses editais, buscou apoiar iniciativas de implantação de RIs e publicações periódicas científicas eletrônicas nas instituições públicas (federais, estaduais e municipais) de ensino e pesquisa e sua integração ao Portal Oasis.Br, com o objetivo de possibilitar o registro e a disseminação da produção científica dessas respectivas instituições, proporcionando-lhes uma maior visibilidade.

Seguindo o pensamento de Espírito Santo (1992), em que a amostra é definida como “[...] um grupo de sujeitos selecionados de um grupo maior e incluindo menos que todos os sujeitos naquele grupo maior”, foram determinados, como amostra deste estudo, os RIs contemplados pelos editais do IBICT/FINEP. Dessa forma, acredita-se que é possível identificar a atual situação desses repositórios em relação ao contexto deste estudo e confrontar com algumas iniciativas internacionais a fim de se obter um parâmetro para a definição de diretrizes que possibilitem a otimização da interoperabilidade semântica, não só em relação a esse grupo de repositórios, mas em uma perspectiva mais abrangente.

O critério de inclusão teve como base selecionar RIs contemplados pelo projeto piloto do IBICT/FINEP e pelos Editais IBICT/FINEP/PCAL/XBDB n.º 02/2009 e 03/2009, que resultaram em 4, 27 e 7 instituições de ensino/pesquisa, respectivamente, totalizando 38 RIs. Tal critério foi adotado por se entender que todos tiveram os mesmos incentivos em relação a treinamentos, equipamentos e *softwares* para a sua implantação. Buscou-se identificar o que eles estão utilizando em termos de metadados e sistemas de organização do conhecimento, com o propósito de potencializar a interoperabilidade semântica.

Foram excluídos da amostra os RIs não contemplados por esses editais.

Como os RIs estão vinculados às instituições, e nem todos têm uma identificação, foram elencadas aquelas contempladas pelos editais (Quadro 9).

Quadro 9 – Repositórios institucionais contemplados pelos editais IBICT/FINEP

INSTITUIÇÕES QUE PARTICIPARAM DO PROJETO PILOTO	
1	Universidade de Brasília (UNB)
2	Universidade Federal da Bahia
3	Universidade Federal de Pernambuco
4	Universidade Federal do Rio Grande do Sul
INSTITUIÇÕES CONTEMPLADAS PELO EDITAL N.º 02/2009	
5	Centro Renato Archer (CTI)
6	Fundação Centro Tecnológico de Minas Gerais
7	Fundação Santo André
8	Fundação Universidade Federal do Rio Grande
9	Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO)
10	Instituto Nacional de Tecnologia (INT)
11	Museu Goeldi
12	Universidade de São Paulo
13	Universidade Estadual de Ponta Grossa
14	Universidade Federal de Alagoas
15	Universidade Federal de Grande Dourados
16	Universidade Federal de Ouro Preto
17	Universidade Federal de Pelotas
18	Universidade Federal de São Caetano do Sul
19	Universidade Federal de São Carlos
20	Universidade Federal de Sergipe
21	Universidade Federal de Uberlândia
22	Universidade Federal de Viçosa
23	Universidade Federal do Acre
24	Universidade Federal do Goiás
25	Universidade Federal do Maranhão
26	Universidade Federal do Mato Grosso
27	Universidade Federal do Rio Grande do Norte
28	Universidade Federal do Tocantins
29	Universidade Federal dos Vales de Jequitinhonha e Mucuri
30	Universidade Federal Fluminense
31	Universidade Tecnológica Federal do Paraná
INSTITUIÇÕES CONTEMPLADAS PELO EDITAL N.º 03/2010	
32	Fundação João Pinheiro
33	Universidade Federal de Mato Grosso do Sul
34	Universidade Federal de São Paulo
35	Universidade Federal do Ceará
36	Universidade Federal do Espírito Santo
37	Universidade Federal do Pará
38	Universidade Federal Juiz de Fora

Fonte: IBICT (2012).

6.1.2 Instrumento de Coleta de Dados

De acordo com Barros e Lehfeld (2006, p. 89, apud TORINO, 2010, p. 82), “[...] a coleta de dados significa a fase da pesquisa em que se indaga e se obtêm dados da realidade pela aplicação de técnicas utilizando para isso de instrumentos de coleta de dados”.

6.1.2.1 Elaboração do protocolo

Com a sustentação teórica e a definição da metodologia, foi possível traçar as etapas e os procedimentos a serem realizados para a consecução desta pesquisa, que é a análise dos RIs em busca de elementos que potencializem a interoperabilidade semântica.

A utilização de um protocolo de coleta de dados buscou orientar e regular a estratégia de pesquisa, com o objetivo de garantir que os resultados possam ser replicados (CALAZANS, 2007).

O protocolo de coleta de dados é composto por:

- a) roteiro para a análise dos RIs contemplando os aspectos observados e identificados retratados em um relatório final, destacando as seguintes informações: forma de arquivamento, responsáveis pelo arquivamento, tipo de material aceito para publicação, identificação dos metadados descritivos e existência de política de informação, principalmente no que se refere ao fluxo de submissão para auxiliar não somente os procedimentos, mas também instrumentos que facilitam essas submissões (APÊNDICE A);
- b) planilha para identificação de instrumentos e elementos da representação temática, referenciados pela literatura como facilitadores da interoperabilidade (APÊNDICE B). De posse dessa planilha, os RIs selecionados foram analisados para a identificação dos metadados descritivos e elementos da representação temática adotados. Essas informações foram obtidas por meio da visualização da descrição completa dos metadados de cada tipo de documento disponibilizado pelo RI.

Buscaram-se também informações disponíveis nos *sites* para identificação dos procedimentos ou políticas adotadas pelo RI na perspectiva da organização do conhecimento

Definição de parâmetros para a análise dos repositórios

Ao partir da premissa de que a adoção de padrões de tratamento da informação no ambiente digital possibilita a interoperabilidade semântica com outros sistemas e mecanismos de busca, foi analisado o elemento dos metadados descritivos assunto (*dc.subject*), um dos 15 elementos do padrão DC (Quadro 10), e seus valores, por entender que eles viabilizam procedimentos que maximizam a integração de fontes diversificadas e heterogêneas de informação entre os sistemas.

As categorias de assunto podem se apresentar segundo a classificação ou a indexação dos documentos. Os procedimentos de descrição temática permitem que os documentos sejam dispostos de acordo com o interesse dos usuários do sistema de informação ou com as tarefas que estes sistemas oferecem. Essa descrição pode servir de base para associar uma descrição ao conhecimento prévio do usuário que busca informação na interface do sistema [...] (MONTEIRO, 2008, p. 83).

Em se tratando das estruturas de organização e das diferentes 'ligações' que podem se estabelecer entre conteúdos [...], é pertinente a correlação entre estes elementos e os padrões, instrumentos ou esquemas utilizados na descrição física ou temática [...] (MONTEIRO, 2008, p. 83).

Quadro 10 – Elementos metadados Dublin Core

ELEMENTOS	DESCRIÇÃO
Title (Título)	Nome atribuído ao recurso.
Creator (Autor)	A principal entidade responsável pela criação dos conteúdos do recurso. Normalmente, o nome de um “Creator” deverá ser utilizado para indicar a entidade
Subject (Assunto)	O assunto do recurso. Normalmente, o elemento “subject” será expresso como palavras-chave, frases-chave ou códigos de classificação que descrevam o conteúdo intelectual do recurso
Description (Descrição)	Informação sobre o conteúdo do recurso. A descrição pode incluir resumo, sumários, referências a representações gráficas do conteúdo e texto livre com informação do conteúdo, entre outros dados
Publisher (Editora)	Uma entidade responsável pela disponibilização dos recursos. Uma pessoa, uma organização ou um serviço podem constituir exemplos de editor. Normalmente, o nome de um editor deverá ser utilizado para mencionar a entidade
Contributor (Colaborador)	Uma entidade responsável por realizar contributos para o conteúdo do recurso. Uma pessoa, uma organização ou um serviço podem constituir exemplos de um colaborador. Normalmente, o nome de um colaborador deverá ser utilizado para referenciar a entidade
Date (Data de publicação)	A data referente a um evento no ciclo de vida do recurso. Normalmente, a data está associada à criação ou à disponibilização do recurso. A prática recomendada para codificar o valor data é definida na norma ISO 8601 [W3CDTF] e segue o formato YYYY-MM-DD
Type (Tipo)	O tipo de resultado científico do qual o recurso é uma manifestação. No elemento DC “type”, descreve-se o tipo de divulgação ou o tipo de conteúdo intelectual do recurso. Emprega-se para explicar ao utilizador que tipo de recurso se visualiza, se é um livro ou um artigo, se foi escrito para uso interno ou externo etc.
Format (Formato)	Manifestação física ou digital do recurso. Normalmente, o elemento “Format” pode incluir o tipo de meio ou as dimensões do recurso. O elemento “Format” pode ser utilizado para determinar o <i>software</i> , o <i>hardware</i> ou outro equipamento necessário para mostrar ou operar o recurso. Exemplos de dimensões são o tamanho e a duração. A prática recomendada é seleccionar um valor de um vocabulário controlado (por exemplo, a lista de tipos de meios da Internet [MIME] que define os formatos para os equipamentos informáticos)
Identifier (Identificador)	Referência inequívoca do recurso num contexto determinado
Source (Fonte)	Referência a um recurso do qual deriva o recurso atual
Language (Idioma)	Idioma do conteúdo intelectual do recurso
Relation (Relação)	Referência a um recurso relacionado
Coverage (Cobertura)	Extensão ou âmbito do conteúdo do recurso. Normalmente, a cobertura inclui a localização espacial (nome do lugar ou coordenadas geográficas), um período temporal (etiqueta de período, data ou intervalo de datas) ou a jurisdição (por exemplo, o nome de uma entidade administrativa)
Rights (Direito)	Informação sobre os direitos no recurso e sobre o recurso
Audience (Público)	Um tipo de entidade para a qual o recurso é dirigido ou útil

Fonte: Adaptado de Driver (2009)

No Quadro 11, são apresentados esquemas de codificação possíveis de serem utilizados para a representação temática.

Quadro 11 – Esquemas de codificação para descrição temática

Term Name: CMIType	
URI:	http://purl.org/dc/terms/DCMIType
Label:	DCMI Type Vocabulary
Definition:	The set of classes specified by the DCMI Type Vocabulary, used to categorize the nature or genre of the resource.
See:	http://dublincore.org/documents/dcmi-type-vocabulary/
See:	http://purl.org/dc/dcmitype/
Type of Term:	http://purl.org/dc/dcam/VocabularyEncodingScheme
Version:	http://dublincore.org/usage/terms/history/#DCMIType-003
Term Name: DDC	
URI:	http://purl.org/dc/terms/DDC
Label:	DDC
Definition:	The set of conceptual resources specified by the Dewey Decimal Classification.
See:	http://www.oclc.org/dewey/
Type of Term:	http://purl.org/dc/dcam/VocabularyEncodingScheme
Version:	http://dublincore.org/usage/terms/history/#DDC-003
Term Name: IMT	
URI:	http://purl.org/dc/terms/IMT
Label:	IMT
Definition:	The set of media types specified by the Internet Assigned Numbers Authority.
See:	http://www.iana.org/assignments/media-types/
Type of Term:	http://purl.org/dc/dcam/VocabularyEncodingScheme
Version:	http://dublincore.org/usage/terms/history/#IMT-004
Term Name: LCC	
URI:	http://purl.org/dc/terms/LCC
Label:	LCC
Definition:	The set of conceptual resources specified by the Library of Congress Classification.
See:	http://lcweb.loc.gov/catdir/cpsolcco/lcco.html
Type of Term:	http://purl.org/dc/dcam/VocabularyEncodingScheme
Version:	http://dublincore.org/usage/terms/history/#LCC-003
Term Name: LCSH	
URI:	http://purl.org/dc/terms/LCSH
Label:	LCSH
Definition:	The set of labeled concepts specified by the Library of Congress Subject Headings.
Type of Term:	http://purl.org/dc/dcam/VocabularyEncodingScheme
Version:	http://dublincore.org/usage/terms/history/#LCSH-003
Term Name: MeSH	
URI:	http://purl.org/dc/terms/MeSH
Label:	MeSH
Definition:	The set of labeled concepts specified by the Medical Subject Headings.
See:	http://www.nlm.nih.gov/MeSH/MeSHhome.html
Type of Term:	http://purl.org/dc/dcam/VocabularyEncodingScheme
Version:	http://dublincore.org/usage/terms/history/#MeSH-003
Term Name: NLM	
URI:	http://purl.org/dc/terms/NLM
Label:	NLM
Definition:	The set of conceptual resources specified by the National Library of Medicine Classification.
See:	http://wwwcf.nlm.nih.gov/class/
Type of Term:	http://purl.org/dc/dcam/VocabularyEncodingScheme
Version:	http://dublincore.org/usage/terms/history/#NLM-002
Term Name: TGN	
URI:	http://purl.org/dc/terms/TGN
Label:	TGN
Definition:	The set of places specified by the Getty Thesaurus of Geographic Names.
See:	http://www.getty.edu/research/tools/vocabulary/tgn/index.html
Type of Term:	http://purl.org/dc/dcam/VocabularyEncodingScheme
Version:	http://dublincore.org/usage/terms/history/#TGN-003
Term Name: UDC	
URI:	http://purl.org/dc/terms/UDC
Label:	UDC
Definition:	The set of conceptual resources specified by the Universal Decimal Classification.
See:	http://www.udcc.org/
Type of Term:	http://purl.org/dc/dcam/VocabularyEncodingScheme
Version:	http://dublincore.org/usage/terms/history/#UDC-003

Fonte: DUBLIN CORE METADATA INICIATIVE (2011)

Como os metadados descritivos podem conter informações sobre um recurso, e essas podem estar baseadas em mais de um procedimento ou um elemento, a análise preestabelece a inclusão de elementos/qualificadores que não estejam definidos na planilha de análise.

Ainda nessa perspectiva, serão analisados os procedimentos técnicos de indexação, catalogação e categorização dos conteúdos informacionais.

6.1.2.2 Questionário

Optou-se pela utilização do questionário eletrônico como um instrumento da pesquisa com a finalidade de complementar a análise dos *sites*, por se entender, diante de análise preliminar, que as informações disponibilizadas podem não ser suficientes para uma identificação segura do que é executado pelo RI no que diz respeito à utilização de SOC (APÊNDICE C).

O questionário, nesta pesquisa, foi esboçado com base em três questionários que também tiveram por objetivo identificar elementos em RI que viabilizem a interoperabilidade semântica (APÊNDICE C). Tal procedimento se apoia no fato de serem questionários validados, aplicados por organismos reconhecidos, como o Centre for Digital Library Research da University of Strathclyde (BIRREL; DUNSIRE; MENZIES, 2009); SurfFoundation (EIJINDHOVEN; GRAAF, 2007); e o projeto CLARA RG T1684 - CLARA, Cooperação Latino-Americana de Redes Avançadas e Banco Interamericano de Desenvolvimento – BID (ALTAMIRANDO LÓPEZ, 2011).

O questionário eletrônico Encuestafacil³³, composto de questões fechadas e abertas, foi enviado para os 38 *e-mails* dos administradores dos RIs (Quadro 9, p. 147). As questões fechadas, com opção de múltipla escolha, buscaram identificar os procedimentos adotados pelos RIs acerca dos metadados, tipo de vocabulário controlado e instrumentos utilizados para organização do conhecimento. A pergunta aberta permitiu que o administrador do repositório contribuísse com sua opinião em relação aos sistemas de organização do conhecimento. A análise proposta e a sistematização dos resultados buscam fornecer subsídios para a elaboração de diretrizes, tendo em vista a otimização da interoperabilidade

³³ <http://www.encuestafacil.com>

semântica. Por conseguinte, os aspectos tratados por essas diretrizes facilitarão a recuperação da informação pelos usuários, ao dirimir barreiras, como a inconsistência semântica e sintática e a presença de diferentes idiomas.

O questionário eletrônico enviado aos 38 administradores dos RIs, no período de 20 de maio a 27 de junho de 2012, incluiu o Termo de Consentimento Livre e Esclarecido (TCLE) (APÊNDICE D). O preenchimento do questionário eletrônico e a análise dos *sites* foram vinculados à aceitação do TCLE.

A coleta de dados foi realizada após a aprovação do projeto pelo Comitê de Ética da Universidade Estadual de Londrina (Processo n.º 22511, ANEXO A).

O questionário é composto por 18 questões fechadas, incluindo de múltipla escolha, e 1 aberta. O questionário foi dividido em quatro seções, com o objetivo de facilitar a identificação por parte dos respondentes, quanto às questões e ao aspecto maior da questão (BIRRELL; DUNSIRE; MENZIES, 2009). Estruturou-se da seguinte forma: Seção 1, Identificação do RI e aspectos administrativos; Seção 2, Conteúdo – material depositado e metadados; Seção 3, Tratamento temático da informação; e Seção 4, Interoperabilidade. As questões apresentadas nessas seções buscaram atender os objetivos específicos: de identificar metadados descritivos para a organização do conhecimento nos repositórios institucionais; detectar elementos e instrumentos da representação temática que favoreçam a interoperabilidade semântica; e fornecer subsídios para estabelecer uma conexão entre os sistemas de organização do conhecimento e a interoperabilidade semântica.

7 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Como relatado no início deste estudo, os RIs otimizam o avanço científico e tecnológico em atividades de pesquisa e ensino e reúnem os conhecimentos produzidos pelas instituições que estão dispersos pelo mundo em um único local, facilitando, assim, o acesso à produção intelectual das instituições. É como sintetiza Crow (2002), ao ressaltar as mudanças provocadas no sistema de comunicação científica:

- a) expande exponencialmente o acesso a pesquisa;
- b) reafirma o controle sobre o saber pela academia;
- c) serve como indicadores tangíveis da qualidade de uma universidade;
- d) demonstra as relevâncias científica, social e econômica das atividades de pesquisa da instituição;
- e) aumenta a visibilidade, o *status* e o valor público da instituição.

Gomes e Guimarães (2012, p. 2) afirmam que a busca de conhecimento pela sociedade atual representa muito mais do que a produção e o consumo do conhecimento. Vai além: necessita da “[...] sua organização e tratamento a fim de torná-lo representável e recuperável igualmente em escala global”. Por conta desse panorama, a interoperabilidade semântica assume um papel de extrema relevância.

Nesse sentido, este estudo se propôs a “Analisar o elemento de metadado *Subject* do Dublin Core (*dc.subject*) e os qualificadores utilizados para a organização do conhecimento com vistas a fornecer subsídios para a potencialização da interoperabilidade semântica dos repositórios institucionais”.

A lista dos administradores foi construída com base em informações fornecidas pelo IBICT (38 RIs). Após atualização dos dados dos administradores, em vista de alguns terem sido substituídos, o questionário foi enviado para os seus *emails*, acompanhado do (junto ao) TCLE. Os *sites* dos repositórios só foram analisados após a autorização da realização da pesquisa pelos 26 RIs. Dois dos RIs não foram identificados, seis não responderam e quatro justificaram a não

participação na pesquisa em razão de os repositórios se encontrarem em fase de implantação ou mudança de coordenação.

Os resultados obtidos com a pesquisa foram analisados em relação à literatura consultada e assim explicitados:

IDENTIFICAÇÃO DO RI

Dos 26 repositórios analisados, todos utilizam o *software* DSpace. Identificou-se que sete ainda não estão registrados nos diretórios OpenDoar e Roar.

ESQUEMA DE METADADOS

Dos 26 (100%) respondentes, 12 (38%) utilizam o esquema de metadados Dublin Core Qualified; 5 (16%) Dublin Core simplificado; 1 (3%) adota o *context_object*; 1 (3%), o MARC; 3 (9%), o METS; 6 (20%) não souberam responder. Três (10%) repositórios responderam que adotam o Dublin Core (entende-se, mediante análise do *site*, que se trata do Dublin Core simplificado) (Tabela 1).

Tabela 1 – Esquema de metadados utilizado

Esquema de metadados	N.º participantes
<i>context_object</i>	1 (3%)
DC qualified	12 (38%)
DC simple	5 (16%)
MARC	1 (3%)
METS	3 (9%)
Interno	1 (3%)
não sabe	6 (19%)
Outra	3 (9%)

Nota: Questão de múltipla escolha.

Pode-se inferir que, com referência à utilização do esquema de metadados, o Dublin Core é o esquema mais utilizado. Alguns dos repositórios agregam também outros padrões, como o *context-object* (3%), o MARC (3) e o METS (6).

Também é possível fazer uma correlação entre os padrões de metadados utilizados e o tipo de material depositado pelos RIs, conforme Tabela 4

(p. 158), já que alguns padrões estão voltados a objetivos específicos, como objetos educacionais e materiais audiovisuais.

O Dublin Core possui algumas características que o tornaram um dos mais utilizados pela comunidade internacional (NAGAMORI; SUGMOTO, 2006): é descritivo e codificado em XML, o que permite uma melhor interoperabilidade (SHINTAKU; FERREIRA; ROBREDO, 2011); interage com o RDF (BAPTISTA; MACHADO, 2001); atende ao modelo *Linked Data*; aumenta a visibilidade de recursos em coleções entre os setores e domínios de assuntos (NISO, 2001), por compreender os elementos DCMI terms e outros termos que representam propriedades, subpropriedades, classes (DCMI Type Vocabulary) e os esquemas de codificação Vocabulary Encoding Scheme (VES) e Syntax Encoding Scheme (SES) (CATARINO, 2009).

Com o conjunto de elementos de descrição do DC, é possível que catalogadores ou não catalogadores utilize-o para simples descrição de recursos de informações, tornando-as mais visíveis aos motores de busca e sistemas de recuperação (WEIBEL, 1997). Podem ser reproduzidos de forma ilimitada utilizando qualificadores que informam o usuário sobre como cada item pode ser interpretado (ALVES; SOUZA, 2007; DUBLIN CORE METADATA INICIATIVE, 2011).

ADMINISTRAÇÃO DO RI

A administração dos RIs é, na sua maioria, realizada pela biblioteca (17, 61%), enquanto 5 (18%) dos RIs utilizam a administração compartilhada – biblioteca e outro órgão –, 6 (21%) indicaram que o RI é administrado por outro órgão da universidade (Tabela 2).

Tabela 2 – Quem é o responsável pela administração do RI?

Responsável	N.º de participantes
Biblioteca	17 (61%)
Outro órgão em conjunto com a biblioteca	5 (18%)
Outros	6 (21%)

Nota: Questão de múltipla escolha.

Corroborando o trabalho de Carvalho (2009), esta pesquisa identificou que a biblioteca é o órgão responsável pelos RIs em 17 dos RIs analisados, ou seja, a maioria dos repositórios, e cinco compartilham a administração com a biblioteca e seis citaram outro órgão da instituição.

Do total de respondentes (26), identificou-se que o depósito dos recursos realizado pelos autores com controle do profissional da informação corresponde a 4 (12%); em 1 (3%) o depósito é realizado pelo autor; 12 (35%) recebem dos autores e o depósito é feito por profissionais da informação; em 12 (35%) a coleta é feita pela administração e depositado por profissionais da informação. Cinco (15%) respondentes indicaram outras possibilidades, como: os recursos são coletados e depositados por bolsistas de iniciação científica; a coleta e o controle de qualidade são realizados pelo administrador da comunidade; e o depósito é feito por pessoa indicada pelo administrador ou por outros órgãos responsáveis pelo acervo (Tabela 3).

Tabela 3 – Quem realiza o depósito?

Tipo de depósito	N.º de participantes
Autoarquivamento pelos autores	1 (3%)
Autoarquivamento pelos autores e controle de qualidade por profissionais da informação	4 (12%)
Entregue pelos autores e depositado por profissionais da informação	12 (35%)
Coletados pela administração e depositado por profissionais da informação	12 (35%)
Outro	5 (15%)

Nota: Questão de múltipla escolha.

Um outro aspecto abordado pela questão verificou que a maior parte da coleta e da submissão dos itens é realizada pelo pessoal administrativo, corroborando a pesquisa de Melero et al. (2009). O índice de autoarquivamento ficou muito abaixo dos encontrados por Darby et al. (2008) em pesquisa realizada com RIs no Reino Unido. A ausência do autoarquivamento diverge dos princípios norteadores do movimento de acesso aberto aos RIs (via verde) (HARNARD, 2007; KURAMOTO, 2011; MARCONDES; SAYÃO, 2009).

Em relação ao autoarquivamento, Shearer (2003) relata que ele tem sido mais bem-sucedido em algumas áreas do que em outras, resultado de uma

diferença significativa no tipo de comunicação científica adotado por alguns campos de conhecimento. Esse aspecto também foi salientado por Costa (2008) e por outros autores que veem como imprescindível a participação de pesquisadores no processo de autoarquivamento, embora ainda seja muito baixo o número de pesquisadores que adotam essa prática, tornando-se um dos aspectos mais desafiadores para a implantação e o sucesso dos RIs (LYNCH; LIPPINCOTT, 2005). Dados internacionais mostram que apenas 15% a 20% dos depósitos são praticados dessa forma (ENABLING OPEN SCHOLARSHIP, [2010b]; HARNARD, 2007). A realidade brasileira é observada por Kuramoto (2011), que alerta sobre o baixo índice de depósito.

CONTEÚDO

Livros e capítulos de livros são aceitos para depósito por 18 (13%) repositórios pesquisados; 5 (4%) depositam revisões de livros; 20 (15%) depositam apresentação em congresso; 7 (5%) submetem pôsteres apresentados em congresso; 23 (17%), artigos de periódicos; 7 (5%), artigos em submissão (*pré-print*); 20 (15%), dissertações e teses; 7 (5%) incluem monografias especialização/aperfeiçoamento; 4 (3%) TCCs; 2 (1%), objetos de aprendizagem; 2 (1%) depositam patentes; 5 (4%), relatórios; 4 (3%), dados de pesquisa; 1 (1%) inclui documento de discussão de trabalhos; 2 (1%) depositam documentos históricos; 7 (5%), materiais audiovisuais; 2 (1%), *softwares*; 1 (1%) está testando a inclusão de monografias (Tabela 4).

Tabela 4 – Que tipo de material pode ser inserido no RI?

Tipo de Material	N.º de participantes
Livros/capítulos de livros	18 (13%)
Revisões de livros	5 (4%)
<i>Papers</i> apresentados em congressos	20 (15%)
Pôsteres apresentados em congressos	7 (5%)
Artigos de periódicos	23 (17%)
Artigos de periódicos submetidos (<i>pre-print</i>)	7 (5%)
Dissertações ou teses	20 (15%)
Monografias (Especialização/Aperfeiçoamento)	7 (5%)
Trabalhos de graduação	4 (3%)
Objetos de aprendizagem	2 (1%)
Patentes	2 (1%)
Relatórios	5 (4%)
Dados de pesquisa	4 (3%)
Documentos de discussão e trabalho	1 (1%)
Documentos históricos	2 (1%)
Materiais audiovisuais	7 (5%)
<i>Softwares</i>	2 (1%)
Outros	1 (1%)

Nota: Questão de múltipla escolha.

Prevalece, entre os RIs pesquisados, o depósito dos artigos de periódicos, *papers* apresentados em congressos, livros e capítulos de livros, dissertações e teses, seguindo a tendência mundial (MELEREO et al., 2009). O índice de depósitos de objetos de aprendizagem foi baixo, merecendo análise posterior para identificar se esse fato está associado à existência de repositórios específicos para objetos de aprendizagem.

METADADOS

Dezessete (65%) dos 26 RIs indicaram que a administração controla ou gerencia os metadados; 4 (15%) fazem esse controle de forma parcial, 1 (4%) não participa; e 4 (15%) não souberam responder (Tabela 5).

Tabela 5 – Os metadados são mediados ou controlados pela administração do RI?

Nível de mediação	N.º de participantes
Sim	17 (65%)
Parcialmente	4 (15%)
Não	1 (4%)
Não sabe	4 (15%)

O padrão de conteúdo de metadados utilizado por 14 (54%) dos repositórios é o AACR2/RDA; 3 (12%) utilizam padrão interno; 1 (4%) adota o Dublin Core Application Profile (DCAP), 7 (27%) não sabem (Tabela 6).

Tabela 6 – Que padrão de conteúdo de metadados é utilizado?

Padrão de conteúdo	N.º de participantes
AACR/RDA	14 (54%)
DCAP	1 (4%)
Interno	3 (12%)
NS/NC	1 (4%)
Outro	7 (27%)

Já em relação ao padrão de conteúdo de metadados, 14 (54%) repositórios utilizam o AACR2/RDA, embora a análise de registros pelo *sites* de alguns RIs não apresentam a utilização do AACR2/RDA nos metadados de título e autor, e 3 (12%) adotam padrão interno. O alto índice de utilização do AACR2/RDA corrobora o resultado da pesquisa de Park e Tosaka (2010). O uso desses padrões está em consonância com as orientações da NISO (2007), entretanto a utilização de um padrão interno contraria as considerações feitas por Nagamori e Sugimoto (2006), Park e Tosaka (2010) no que diz respeito à interoperabilidade. Os autores salientam que a adoção de um padrão de conteúdo metadados para um domínio específico atende às demandas da comunidade, mas pode comprometer a descoberta e o reuso de metadados por outras comunidades.

Para Souza e Alves (2009), esses padrões possibilitam a integração e o compartilhamento de informações e aplicações pelos sistemas de informação. Heery et al. (2006a, 2006b), NISO (2007), Park e Tosaka (2010) despertam para a importância de desenvolver a explicitação do perfil de aplicação para que a comunicação possa ser estabelecida entre as diferentes comunidades.

INDEXAÇÃO DOS DOCUMENTOS

O controle de vocabulário ainda tem um índice baixo. Dos 25 respondentes, 2 (7%) informaram que não é feita a indexação de assunto; 8 (28%) disseram que a atribuição de palavras-chave é livre e utiliza o idioma português; 7 (24%) permitem livre atribuição das palavras-chave e adotam tanto o idioma do documento como o inglês; 6 (21%) afirmaram que são atribuídas palavras-chave ou sistema de classificação e idioma do país; em 4 (14%) as palavras-chave ou sistema de classificação são atribuídas ao idioma do país e em inglês. Foi informado ainda que as palavras-chave são atribuídas pelo autor em outros idiomas e não apenas no idioma do documento; e são retiradas do documento, empregando-se o inglês, o português e o idioma do documento (Tabela 7).

Tabela 7 – Qual das afirmações melhor descreve a indexação de assuntos

Indexação de assunto	N.º de participantes
Não é feita a indexação de assunto	2 (7%)
A atribuição de palavras-chave é livre, utilizando o idioma do país	8 (28%)
A atribuição de palavras-chave é livre, utilizando o idioma do país e em inglês	7 (24%)
Palavras-chave ou classificação são atribuídas de acordo com sistemas padronizados no idioma do país	6 (21%)
Palavras-chave ou classificação são atribuídas de acordo com sistemas padronizados no idioma do país e em inglês	4 (14%)
Outra	2 (7%)

Nota: Questão de múltipla escolha.

Darby et al. (2008) identificaram que 83% dos administradores de RIs entrevistados no seu estudo realizam controle de qualidade pós-depósito. Birrell, Dunsire e Menzies (2009), ao analisar RIs no Reino Unido, constataram que o controle de autoridade ainda é baixo e um número considerável adota a classificação.

Diante do crescimento no número de RIs e, consequentemente, dos depósitos, principalmente praticados pelo autoarquivamento, quando não existe controle da padronização de terminologias ou vocabulários e identificação dos valores

dos metadados, há um prejuízo na obtenção da interoperabilidade semântica entre os vários sistemas de informação. A tradução dos termos é um fator relevante para a recuperação em relação ao contexto global. A simples reutilização das palavras-chave adotadas pelas publicações também pode ser um fator prejudicial na recuperação, já que nem todas as publicações utilizam o vocabulário controlado para indexação dos documentos.

A falta de controle de vocabulário por parte dos RIs brasileiros foi abordada pela Professora Malgorzata Lisowska, durante a Segunda Conferência de Acesso Aberto em 2011, no Rio de Janeiro, ao apresentar o relato “Estrategia regional y marco de interoperabilidad y gestión para una Red Federada Latinoamericana de Repositorios Institucionales de documentación científica”. Lisowska exteriorizou sua preocupação em relação a esses aspectos para ações futuras que contemplem os repositórios nacionais.

Existe, por parte de vários organismos internacionais (NISO, 2001, 2004; DRIVER, 2009) e autores (BAPTISTA, 2010; HEERY et al., 2006a, 2006b; PATEL et al., 2005; PARK; TOSAKA, 2010; PEREIRA; RIBEIRO; NEVES, 2005), orientação para que sejam utilizados vocabulários controlados e o SOC adotado seja exposto, assim como o uso do idioma inglês (DRIVER, 2009). O desenvolvimento de vocabulário controlado para um domínio específico gera preocupação para alguns autores (SHINTAKU; BRASCHER, 2007; PARK; TOSAKA, 2010). Para Castro e Santos (2011), o tratamento das informações deve visar à interoperabilidade entre os sistemas de informação.

Zeng e Chan (2004) veem nos SOC a possibilidade da consulta “*one stop*”, ou seja, o acesso a informações por uma única opção de busca, consultando diferentes sistemas e informações heterogêneas. Nesse sentido, Bhat (2010) e Botheram (2010) afirmam que a interoperabilidade semântica está estritamente relacionada aos SOC.

Zeng (2010) adiciona, ainda, a possibilidade da decomposição de conteúdo por meio de linguagem de marcação, o que enriqueceria a descoberta de recursos. Isso representa a identificação do conteúdo de partes de um trabalho, como os capítulos de livros.

Em relação à responsabilidade de atribuir as palavras-chave, identificou-se que, dos RIs que responderam a essa questão, 19 (46%) indicaram os autores; 18 (44%) citaram bibliotecários; 1 (2%) indicou pessoal administrativo

e 3 (7%) explicaram: autores indicam os termos e os bibliotecários revisam com o objetivo de identificar termos autorizados; o recurso é inserido por responsável pelo depósito com revisão do bibliotecário; ou utilizam os termos do próprio documento (Tabela 8). Fica demonstrado um equilíbrio na responsabilidade da atribuição das palavras-chave, mas, quando comparado com os índices da Tabela 7, pode-se deduzir que parte dos termos atribuídos pelos autores não é checada pelos bibliotecários.

Tabela 8 – A atribuição de palavras-chave

Atribuição de palavras-chave	N.º participantes
Pelos autores.	19 (46%)
Por bibliotecários.	18 (44%)
Por pessoal administrativo.	1 (2%)
Não é obrigatório.	0
Outra	3 (7%)

Nota: Questão de múltipla escolha.

Vinte e seis RIs responderam à questão sobre controle de autoridade de assunto, em que a CDU (11%) aparece como o sistema de classificação mais utilizado, em seguida vem a CDD (8%) e a LCC (3%), adotada apenas por um repositório. O vocabulário controlado mais utilizado na área das Ciências da Saúde é o DECS (11%), demonstrando que não existe uma tradução dos termos para o inglês, o que corresponde aos índices sobre o controle de vocabulário (Tabela 7). Os vocabulários controlados utilizados pela maioria dos RIs são o da Biblioteca Nacional (24%) e da Bibliodata (8%), enquanto o LCSH (5%), que refletiria a tradução dos termos em idioma inglês, é empregado por apenas dois repositórios (Tabela 9).

Tabela 9 – Que controle de autoridade de assunto é utilizado?

Vocabulário controlado	N.º participantes
DDC	3 (8%)
CDU	4 (11%)
LCC	1 (3%)
LCSH	2 (5%)
Bibliodata	3 (8%)
Biblioteca Nacional (Brasil)	9 (24%)
MeSH	2 (5%)
DECS	4 (11%)
Outros	9 (24%)

Nota: Questão de múltipla escolha.

A utilização desses vocabulários controlados está em consonância com a maioria dos trabalhos citados neste estudo, porém, ao analisar os *sites*, apenas um repositório adota a identificação e a exposição do vocabulário utilizado, como orientam as diretrizes DRIVER e autores como Crow (2002), Zeng e Chan (2004), Koch (2006). Ao expor os vocabulários adotados, os RIs facilitariam a utilização de mecanismos e serviços de terminologia que melhoram a interoperabilidade semântica (Figura 5). Outra vantagem é utilizar esquemas de classificação, como tesauros e taxonomias, que proporcionam a melhoria na qualidade de acesso aos assuntos (NISO, 2007).

No exemplo da Figura 5, é possível visualizar como o repositório identifica os valores do metadado *Subject*.

Figura 5 – Metadados DC e qualificadores utilizados pelo Dspace@MIT

DSpace@MIT MITLibraries

Search DSpace@MIT [Home](#) [MIT Press](#) [MIT Press Out of Print Books](#)

☒ Search DSpace@MIT
☐ This Collection

[Advanced Search](#)

Browse

All of DSpace@MIT
[Communities & Collections](#)
[By Issue Date](#)
[Authors](#)
[Titles](#)
[Subjects](#)

This Collection
[By Issue Date](#)
[Authors](#)
[Titles](#)
[Subjects](#)

My Account
[Login](#)
[Register](#)

Links
[About DSpace@MIT](#)

Parallel MIMD computation : the HEP supercomputer and its applications

[Show simple item record](#)

dc.contributor	Kowalik, Janusz S.	en
dc.date.accessioned	2002-10-03T20:52:49Z	
dc.date.available	2002-10-03T20:52:49Z	
dc.date.copyright	1985	en
dc.date.issued	1985	en
dc.identifier.isbn	0262111012	en
dc.identifier.uri	http://hdl.handle.net/1721.1/1745	
dc.description	Includes bibliographies and index.	en
dc.description.statementofresponsibility	edited by J.S. Kowalik.	en
dc.format.extent	411 p.	en
dc.format.extent	41003543 bytes	
dc.format.mimetype	application/pdf	
dc.language.iso	en_US	
dc.language.iso	eng	en
dc.publisher	MIT Press	en
dc.relation.ispartofseries	The MIT Press series in scientific computation	en
dc.subject.ddc	001.64	en
dc.subject.lcc	QA76.8.D436	en
dc.subject.lcsh	Denelcor HEP (Computer)	en
dc.subject.lcsh	Parallel processing (Electronic computers)	en

A utilização de esquemas de classificação possibilita a pesquisa pela tabela de classificação (Figura 6).

Figura 6 – Pesquisa por sistema de classificação

DSpace@MIT MITLibraries

Search DSpace@MIT [Home](#)

[Advanced Search](#)

Browse

All of DSpace@MIT
[Communities & Collections](#)
[By Issue Date](#)
[Authors](#)
[Titles](#)
[Subjects](#)

My Account
[Login](#)
[Register](#)

Links

Browsing by Subject

[0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

Or enter first few letters:

Order: [ascending](#) Results: [20](#)

Now showing items 1-20 of 31667 [Next Page](#)

Subject
001.53/4
001.64
004/.36
011202
016.332
016.664/02

Vinte e cinco RIs responderam à questão sobre a existência de instrumentos ou procedimentos que auxiliam a descrição dos recursos, demonstrando uma heterogeneidade no tipo de SOC utilizado pelos repositórios. Entretanto, não foi mencionada a utilização de tabelas de classificação (CDU, CDD, LCC), folksonomia, ontologia ou indexação automática, e nove repositórios não adotaram nenhum SOC (Tabela 10). Eis a questão, eles usam CDU/CDD para uso interno, mas não disponibilizam esses instrumentos para auxiliar na descrição, ou seja, são utilizados apenas pelos bibliotecários.

Tabela 10 – Há algum instrumento ou procedimento para a descrição dos recursos?

Descrição de recursos	N.º participantes
Disponibilização de vocabulário controlado	7 (25%)
Tesauro	2 (7%)
Esquema de categorização	1 (4%)
Taxonomia	1 (4%)
Lista de descritores de assunto	6 (21%)
Nenhum	9 (32%)
Outra	2 (7%)

Nota: Questão de múltipla escolha.

Quando disponibilizados para o usuário (Figuras 7-8), tanto no processo de busca como de depósito, esses instrumentos viabilizam a harmonização terminológica necessária para a interoperabilidade semântica (HEERY, 2006a, 2006b; KOCH, 2006; MCCULLOUGH, 2004), como a possibilidade de desenvolvimento de ontologias a partir desses instrumentos (GOMES; GUIMARÃES, 2012; MOURA, apud MIRANDA, 2005). Não são identificadas facilidades, como a disponibilização de instrumentos, que auxiliam a identificação de termos utilizados pelos sistemas, conforme pode ser conferido no menu de alguns repositórios.

Figura 7 – A utilização da LCSH como instrumento de recuperação da informação

The screenshot displays the University of Southampton ePrints Soton website. On the left is a navigation menu with links: University Home, ePrints Soton (active), Policies, Latest Additions, Browse by Year, Browse by Subject, Browse by School, and Login. Below the menu are RSS 1.0, RSS 2.0, and Atom feeds. The main content area is titled 'Browse by Subject' and includes a search bar at the top right with the text 'Search ePrints Soton' and a magnifying glass icon, and a link to 'Advanced Search'. Below the search bar, it says 'Please select a value to browse from the list below.' The list of subjects is as follows:

- [Library of Congress Subject Areas](#) (67810)
 - [A General Works](#) (45)
 - [AC Collections, Series, Collected works](#) (7)
 - [AI Indexes \(General\)](#) (5)
 - [AM Museums \(General\), Collectors and collecting \(General\)](#) (24)
 - [AS Academies and learned societies \(General\)](#) (2)
 - [AZ History of Scholarship The Humanities](#) (8)
 - [B Philosophy, Psychology, Religion](#) (3818)
 - [B Philosophy \(General\)](#) (353)
 - [BC Logic](#) (4)
 - [BD Speculative Philosophy](#) (7)
 - [BF Psychology](#) (3052)
 - [BH Aesthetics](#) (52)
 - [BJ Ethics](#) (140)
 - [BL Religion](#) (66)
 - [BM Judaism](#) (117)
 - [BP Islam, Bahaism, Theosophy, etc](#) (41)
 - [BQ Buddhism](#) (1)
 - [BR Christianity](#) (48)
 - [BS The Bible](#) (13)
 - [BT Doctrinal Theology](#) (2)
 - [BV Practical Theology](#) (2)
 - [BV1460 Religious Education](#) (1)
 - [BX Christian Denominations](#) (13)
 - [C Auxiliary Sciences of History](#) (823)
 - [C Auxiliary sciences of history \(General\)](#) (66)
 - [CB History of civilization](#) (27)
 - [CC Archaeology](#) (675)

Nesse formulário, pode-se visualizar a possibilidade de pesquisa utilizando termos livres e vocabulário controlado. A utilização de termos livres atende ao que Koch (2006) chama de preservação da riqueza das fontes participantes, ou seja, a linguagem natural empregada pelo autor pode ser conservada, mas a opção por essa linguagem ser traduzida para uma linguagem controlada deve ser incentivada. As diretrizes Driver 2.0 apresentam recomendações de como isso pode ser feito (DRIVER 2.0).

Figura 8 – Formulário para pesquisa avançada

Search ePrints Soton

Advanced Search

University Home

ePrints Soton

Policies

Latest Additions

Browse by Year

Browse by Subject

Browse by School

Login

RSS 1.0

RSS 2.0

Atom

Advanced Search

Just leave the fields you don't want to search blank. [Click here for a simple search.](#)

Full Text: all of

Item ID:

Title: all of

Name: all of

Description/Abstract: all of

Date:

Uncontrolled Keywords: all of

Subjects:

- A General Works
- ...AC Collections. Series. Collected works
- ...AI Indexes (General)
- ...AM Museums (General). Collectors and collecting (General)
- ...AS Academies and learned societies (General)
- ...AZ History of Scholarship The Humanities
- B Philosophy. Psychology. Religion
- ...B Philosophy (General)
- ...BC Logic
- ...BD Speculative Philosophy
- ...BF Psychology
- ...BH Aesthetics

Any of these

Identificou-se que a prática de compartilhamento de metadados com a Online Public Access Catalog (OPAC) pelos repositórios ainda é reduzida, apenas 3 (12%) dos RIs adotam esse procedimento com 100% dos seus dados; e 5 (19%) compartilham apenas 50% (Tabela 11), lembrando que esse compartilhamento aumenta a visibilidade dos RIs.

Tabela 11 – Os metadados são compartilhados com a biblioteca OPAC?

Compartilhamento	N.º participantes
Sim (100%)	3 (12%)
Mais de 50%	5 (19%)
Menos de 10%	2 (8%)
Não	16 (62%)

Para Modesto (2010, p. 2), “[...] os OPACs ganham mais importância pelo valor agregado que adicionam ao produto que transmitem”. Birrell, Dunsire e Menzies (2009) afirmam que, devido ao fato de os RIs utilizarem um padrão aberto, têm maior possibilidade de integrar a OPAC.

INTEROPERABILIDADE

Dos 25 respondentes (um não respondeu), 17 (68%) RIs utilizam o protocolo OAI-PMH para exportação dos dados; 2 (8%) ainda se encontram em processo; 1 (4%) não utiliza; e 5 (20%) não souberam responder (Tabela 12).

Tabela 12 – O RI exporta metadado em OAI-PMH?

Nível de exportação de OAI-PMH	N.º participantes
Sim	17 (68%)
Pendente	2 (8%)
Não	1 (4%)
Não sabe	5 (20%)

O protocolo OAI-PMH, embora não assegure interoperabilidade semântica, interfere na interoperabilidade técnica quando não disponibilizado, comprometendo a comunicação entre os sistemas de informação (BAPTISTA, 2010).

Ainda dos 25 RIs, apenas 3 (11%) responderam que o repositório interopera com o sistema de gerenciamento de biblioteca/arquivo – questão que deve ser avaliada se é um problema de política do repositório ou se o sistema não possibilita a interoperabilidade –, como observado por Carvalho (2009) ao relatar que alguns *softwares* não permitem a interoperabilidade. Dois repositórios (7%) interoperam com a Administração; 2 (7%), com os recursos humanos; 1 (4%) interopera com o Departamento de Pesquisa Administrativa e 1 (4%), com sistema virtual de pesquisa (Tabela 13). A possibilidade de o RI interoperar com outros sistemas internos e externos está alinhada à proposta da concepção desse tipo de repositório, que é reunir, organizar, preservar, divulgar e garantir o acesso confiável e permanente da produção intelectual gerada na universidade em um único local (LEITE, 2009).

Tabela 13 – O seu repositório interopera com algum outro sistema interno?

Sistema	N.º participantes
Sistema de gerenciamento de biblioteca/arquivo	3 (11%)
Administração	2 (7%)
Gerenciamento de recursos eletrônicos	0
Empresa	0
Financeiro	0
Recursos humanos	2 (7%)
Departamento de pesquisa administrativa	1 (4%)
Ambiente de aprendizagem virtual	0
Ambiente virtual de pesquisa	1 (4%)
Outra	18 (67%)

Nota: Questão de múltipla escolha.

FUNÇÕES E SERVIÇOS

Dos 25 RI que responderam à questão se o conteúdo do repositório é pesquisável por alguma opção de busca, 1 (2%) indicou a BASE; 16 (36%) declararam motores de busca; 8 (18%) são pesquisáveis pelo OAIster e 15 (34%) são consultados pelo diretório Open DOAR. Verificou-se que os acervos de bibliotecas de algumas instituições são pesquisáveis pelo OAIster, entretanto não foi possível recuperar o conteúdo do RI. Alguns repositórios (7) ainda não foram cadastrados no Diretório Open DOAR, impossibilitando a pesquisa ao seu acervo (Tabela 14).

A utilização do DC promove a recuperação por motores de busca (WEIBEL, 1997), o que aumenta a visibilidade. Às vezes, é necessário que o administrador realize o cadastramento nas bases (ex.: OAIster).

Tabela 14 – Opções de pesquisa no repositório

Opções de busca	N.º de participantes
BASE	1 (2%)
Citeseer	0
motores de busca	16 (36%)
MEIND	0
MetaGer	0
AI search	0
OAlster	8 (18%)
OASE	0
Open DOAR	15 (34%)
OPUS	0
PLEIADI	0
Scirus	0
Outra	4 (9%)

Nota: Questão de múltipla escolha.

A última questão do instrumento de coleta foi aberta, com o objetivo de identificar o conhecimento dos pesquisados em relação aos metadados e aos instrumentos de organização do conhecimento que poderiam ser utilizados para a viabilização da interoperabilidade semântica.

Nove RIs sugeriram o uso do padrão de metadados Dublin Core, semântica; linguagem HTML, XML, RDF; protocolo Z39.50; ISSO; os esquemas de classificação CDD e CDU; tesauro; cabeçalho de assunto; taxonomia; instrumentos como lista de descritores, catálogos de assuntos; DSpace; resumo, vocabulário controlado; autor; título; data de publicação e dados da fonte de onde se originou o documento.

As respostas fornecidas pelos administradores indicam que a utilização dos metadados descritivos e os SOC auxiliam na melhoria da interoperabilidade semântica, e que algumas questões identificadas pela pesquisa que não estão contribuindo ainda com esse tipo de interoperabilidade é resultado de operacionalização técnica e/ou suporte profissional.

Com essas sugestões, finaliza-se esta pesquisa, para dar início à fase de reflexão do que foi consultado, analisado e ponderado e, assim, traçar considerações que possibilitem um aprofundamento maior sobre o assunto por parte dos gestores de RIs.

7.1 LIMITAÇÕES DA PESQUISA

O questionário constituiu-se em fator essencial na coleta de informações por permitir a identificação de dados que não estavam explícitos nos *sites* dos repositórios. Algumas questões apresentaram inconsistências nas respostas, quando confrontadas com o *site* do RI, o que sugere acrescentar, às próximas pesquisas, o método de entrevistas, não apenas com os administradores, mas com os profissionais engajados com as atividades do repositório – administradores, bibliotecários e autores (depositantes). Não havendo essa possibilidade, sugere-se que, por se tratar de tema recente na prática bibliotecária, sejam adicionados conceitos ou esclarecimentos sobre os tópicos abordados.

A ausência de conteúdo sobre política de informação e orientações para submissão no repositório torna-se um fator restritivo à identificação de procedimentos adotados nos repositórios.

Um outro aspecto limitante foi o número de questionários não respondidos. Identificou-se que alguns repositórios ainda se encontram em fase de implantação, e, em outros, houve mudança na coordenação ou na equipe técnica. Assim, por desconhecerem os procedimentos adotados, esses sujeitos se sentiram impossibilitados de responder às questões. A disponibilização de emails nos sites também comprometeu a identificação dos responsáveis e do acesso a eles. Esses fatores servem de alerta para a questão da descontinuidade do projeto.

A pesquisa não pode assegurar se, de fato, os vocabulários controlados são identificados (valores de metadados), o que proporcionaria um avanço em termos de utilização de serviços de terminologia entre os repositórios.

8 CONSIDERAÇÕES FINAIS

Identifica-se, já nas primeiras leituras deste estudo, que os temas aqui tratados não são atuais, há muito fazem parte do contexto das bibliotecas e dos profissionais da informação. Porém, algumas mudanças em relação a ambiente informacional, novas tecnologias e paradigmas relacionados à comunicação científica vêm desencadeando novas pesquisas e reflexões em relação a eles.

A interoperabilidade semântica, antes discutida em relação a uma perspectiva mais restrita, hoje faz parte de um novo modelo de produção, organização e recuperação da informação, em que os profissionais e os programas de computadores não estão dando conta de processarem tais conteúdos. Num contexto em que o crescimento da informação digital é medido em *exabytes* e os dados são gerados por indivíduos, novas estratégias para que os sistemas e as pessoas “conversem e se entendam” estão sendo desenvolvidas e/ou consolidadas. A interoperabilidade semântica no que concerne à recuperação de conteúdo é um dos aspectos mais discutidos por profissionais de áreas como Ciência da Informação, Informática e Linguística.

Novas formas de armazenamento, organização, tratamento e recuperação da informação surgiram e foram aperfeiçoadas a partir da Web Semântica de Tim Barnes Lee. As recentes possibilidades de utilização de antigos Sistemas de Organização do Conhecimento e o desenvolvimento de novos apontam para as inúmeras perspectivas em relação à organização do conhecimento em um ambiente onde prevalecem a abundância e a heterogeneidade de informação.

A chance de disponibilizar qualquer tipo de informação em meios antes inimagináveis traz novas preocupações, como as questões de direito autoral, preservação e armazenamento. Nesse sentido, os repositórios institucionais trazem em seu bojo uma nova roupagem para uma filosofia antiga, remontando aos discursos de Platão: divulgar o conhecimento livremente de forma a ser compartilhado por todos os interessados, como preconiza o movimento de acesso livre.

Na realidade, a epígrafe do início deste trabalho ilustra, em poucas palavras, o conteúdo e o sentido deste estudo.

Já nas primeiras tentativas de busca na Internet, fica claro, para qualquer usuário, que a dificuldade encontrada para obtenção de uma determinada

informação não é o acesso, mas obtê-la de modo que o satisfaça. Fato que está relacionado à semântica dos termos utilizados por quem disponibiliza a informação e por quem a procura.

A enorme variedade de sistemas resultou em implementação de diferentes terminologias por todos os diversos segmentos da comunidade *online*, causando diferenças acentuadas entre setores e áreas temáticas e transformando qualquer busca transregional ou multidisciplinar em uma tarefa árdua para o usuário final.

A interoperabilidade semântica vem sendo discutida por vários estudiosos e aqui buscou-se identificar como a organização do conhecimento pode contribuir para esse tipo de interoperabilidade.

A interoperabilidade semântica entre repositórios pode se referir a dois aspectos: a semântica dos elementos e a semântica dos valores. Neste estudo, o foco foi a semântica dos valores, relacionada ao uso de elementos da organização do conhecimento. Assim, a interoperabilidade semântica abordada neste estudo está baseada no significado comum dos termos contidos nos metadados e nos documentos.

A interoperabilidade semântica entre repositórios institucionais se traduz em benefícios, como acesso à informação pelo usuário, independente de idioma, *software* e *hardware*, de forma mais rápida e eficaz, com o aumento da visibilidade institucional e do pesquisador e com a possibilidade de reutilização da informação, propiciando o que os pesquisadores denominam de uma economia informacional.

Nessa perspectiva, a OC, por meio dos SOC, permite que os sistemas trabalhem não apenas com o controle terminológico, mas com a possibilidade de, por meio dos SOC, utilizar a linguagem natural para extração de termos controlados, proporcionando, assim, a interoperabilidade semântica em nível de dados categorizados.

A heterogeneidade das informações existentes nos repositórios pode ser administrada pelos gestores, a partir do estabelecimento de boas práticas em relação aos metadados; de criação e de disponibilização de perfis de aplicação como o Scholarly Work Application Profile (SWAP).

No contexto nacional, assim como ocorre em outros países, a interoperabilidade semântica ainda não atingiu o nível aceito como ideal. A partir dos

objetivos propostos por este estudo, é possível destacar os resultados provenientes da análise da literatura e dos resultados obtidos com a pesquisa desenvolvida com os 26 RIs que participaram deste estudo:

- a) os metadados descritivos utilizados pelos RIs permitem a identificação do título (elemento *Title*), autor (elemento *Creator*), do assunto (elemento *Subject*) e do resumo (elemento *Description*). O número de RIs que adotam o controle de vocabulário e a tradução dos termos para o inglês, considerado relevante por ser um idioma mundialmente aceito, ainda é baixo;
- b) elementos da representação temática, como sistemas de classificação e vocabulários controlados, são utilizados, porém não explicitados como valores/qualificadores do metadado *Subject*, o que favoreceria a descoberta dos recursos, assim como o uso de serviços de terminologias, otimizando esse tipo de interoperabilidade;
- c) instrumentos da representação temática, como vocabulário controlado, tesauro esquema de categorização, taxonomia e lista de descritores de assunto, foram indicados pelos administradores, porém poucos repositórios disponibilizam esses instrumentos para o usuário, o que poderia facilitar a utilização de termos adotados no sistema pelos depositantes e por aqueles que buscam os recursos; e
- d) a utilização dos SOC possibilita o desenvolvimento de serviços de terminologia, que, por sua vez, potencializam a interoperabilidade semântica.

Questões práticas e de qualidade em relação ao emprego de palavras-chave, assunto e classificações são preocupantes, não apenas pela prática do autoarquivamento, quando esse não é seguido de práticas que viabilizam o controle terminológico, mas também pelos procedimentos adotados pelos profissionais da informação. A disponibilização de instrumentos e mecanismos baseados na necessidade de recuperação da informação deveria ter em vista a harmonização de terminologias e a utilização de serviços de terminologias que podem agregar valor aos conteúdos dos repositórios por meio de uso dos SOC.

Vale salientar que a adoção dos SOC resulta em algumas possibilidades, como a utilização de *add-on* – recurso extra do DSpace – pelos repositórios, facilitando o depósito e a busca, além do desenvolvimento de ontologias.

Não existe um SOC ideal, mas alguns SOC proporcionam mais benefícios no ambiente Web do que outros, por isso devem ser considerados quando adotados ou reavaliados. No caso de construção de um SOC, é importante ter acesso às normas internacionais que versam sobre a sua elaboração, como a ISO 25964 para tesouros, que abarca a questão da interoperabilidade entre esses sistemas.

A interoperabilidade dos SOC é a situação ideal para o acesso por assunto aos recursos, entretanto essa possibilidade ainda se encontra distante. Uma forma de tornar a recuperação da informação eficiente consiste na utilização de serviços e mecanismos que permitam diferentes vocabulários se comunicarem com termos controlados. Baseado na literatura consultada, pode-se inferir que o emprego da CDD e de taxonomias contribui para o processamento automático da linguagem natural e o desenvolvimento de ontologias.

Estudos revelaram que muitas práticas estão sendo adotadas em nível internacional e devemos avaliar essas orientações, procurando empregá-las mediante as condições de cada repositório, mas pensando sempre no conjunto, pois, nesse contexto, fica bastante claro que “uma andorinha só não faz verão”.

8.1 FUTUROS TRABALHOS

Sugere-se o aprofundamento de questões relativas à representação temática adotada pelos RIs existentes no País, como forma de alcançar a interoperabilidade semântica, incluindo a possibilidade de utilização de ontologias.

A definição de política de interoperabilidade entre os repositórios deveria ser buscada para que futuros projetos, como a rede de repositórios nacional e de participação em iniciativas internacionais, não se deparem com aspectos relacionados à interoperabilidade técnica (qualidade dos metadados e utilização de protocolos) e à interoperabilidade semântica, no que concerne à ausência de controle de vocabulário ou à utilização de vocabulários com abrangência internacional.

Serviços e instrumentos para extração e seleção de termos para a indexação dos documentos pelos depositantes poderiam ser desenvolvidos com o objetivo de minimizar problemas com a ausência de padronização da terminologia adotada pelos autores.

Também se faz necessário o desenvolvimento de estudo comparativo entre RIs que utilizam e que não utilizam SOC em relação ao processo de recuperação da informação.

REFERÊNCIAS

ALEMU, G.; STEVENS, B.; ROSS, P. Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: a social constructivist approach. **New Lib. World**, v. 111, n. 1, p. 38-54, 2012.

ALIPOUR-HAFEZI, M. et al. Interoperability models in digital libraries: an overview. **The Electronic Library**, Bingley, v. 28, n. 3, p. 438-452, 2010.

ALTAMIRANDO LÓPEZ, T. **Segunda reunião para a formação de uma rede de repositórios**. 4 mar. 2011. Disponível em: <http://www.redclara.net/index.php?option=com_content&view=article&id=589:segunda-reuniao-para-la-conformacao-de-uma-red-de-repositorios-en-al&catid=5:importantes&Itemid=353&lang=pt>. Acesso em: 2 jun. 2011.

ALVARENGA, L. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais a teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. **DataGramaZero**, v. 2, n. 6, p. 2001. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000001252&dd1=b2c56>>. Acesso em: 24 jul. 2011.

ALVARENGA, L. Representação do conhecimento na perspectiva da Ciência da Informação em tempo e espaço digitais. **Econtrols Bibli**, Florianópolis, n. 15, p. 1-23, jan./jun. 2003. Disponível em: <http://www.encontrosbibli.ufsc.br/Edição_15/alvarenga_representação.pdf>. Acesso em: 12 jan. 2004.

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

ALVES, M. D. R.; SOUZA, M. I. F. Estudo de correspondência de elementos metadados: Dublin Core e Marc 21. **Rev. Digital Bibliotec. Cienc. Inf.**, Campinas, v. 4, n. 2, p. 20-38, 2007.

ALVES-MAZZOTTI, A. J.; GEWANDSZNAJDER, F. **O método nas ciências naturais e sociais: pesquisa quantitativa e qualitativa**. 2. ed. São Paulo: Pioneira, 1999.

ANDRADE, M. C.; SILVA, T. E.; CERVANTES, B. M. N. Política de informação para repositórios institucionais: um estudo comparativo. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 24., 2011. **Anais...** Maceió. Disponível em: . Acesso em: 20 ago. 2011.

ARMS, W. Y. **Thoughts about interoperability in the NSDL**: draft for discussion. Aug. 2002. Disponível em: <<http://www.cs.cornell.edu/wya/papers/NSDL-Interop.doc>>. Acesso em: 5 jan. 2010.

ARMS, W. Y. et al. A spectrum of interoperability: the *site* for science for prototype for the NSDL. **D-Lib Magazine**, Reston, v. 8, n. 1, p. 2002. Disponível em: <<http://vivo.cornell.edu/individual/vivo/n10897>>. Acesso em: 12 jan. 2010.

BAILEY JR., C. W. **Institutional repositories, tout de suite**. 2005. Disponível em: <<http://www.digital-scholarship.org/ts/irtoutsuite.pdf>>. Acesso em: 15 jan. 2010.

BAPTISTA, A. A. A falar nos entendemos: a interoperabilidade entre repositórios digitais. In: GOMES, M. J. G.; ROSA, F. (Org.). **Repositórios institucionais: democratizando o acesso ao conhecimento**. Salvador: EDUFBA, 2010. p. 71-90. Disponível em: <http://repositorium.sdum.uminho.pt/bitstream/1822/11517/1/RI_Ana%20Alice%20Baptista.pdf>. Acesso em: 4 jan. 2010.

BAPTISTA, A. L.; MACHADO, A. B. Um gato preto num quarto escuro: falando sobre metadados. **Rev. Bibliotec. Brasília**, Brasília, v. 25, n. 1, p. 77-90, 2001

BARITÉ, M. Organización del conocimiento: um nuevo marco teórico-conceptual em bibliotecología y documentación. N: CARRARA, K. (rg.). **Educação, universidade e pesquisa**. Marília: UNESP, 2001. p. 35-60.

BARITÉ ROQUETA M. Sistemas de organização do conhecimento: uma tipologia atualizada. **Inf. Inf.**, Londrina, v. 16, n. 3, p. 122-139, jan./jun. 2011.

BAX, M. P. Introdução às linguagens de marcas. **Ciênc. Inf.**, Brasília, v. 30, n. 1, p. 32-38, 2001.

BERNERS-LEE, T. The semantic Web. **Sci. Am.**, 17 May 2001. Disponível em: <<http://www.scientificamerican.com/article.cfm?id=the-semantic-Web>>. Acesso em: 27 nov. 2010.

BHAT, M. H. Interoperability of *open access* repositories in computer science and IT: an evaluation. **Library Hi Tech.**, v. 28, n. 1, p. 107-118, 2010.

BIRREL, D.; DUNSIRE, G.; MENZIES, K. **Online catalogue and repository interoperability study (OCRIS)**. Appendice 2 to OCRIS Final Report. Case Studies in 2 higher education institution libraries. Strathclyde: Centre for Digital Library Research, 2009.

BORBINHA, J.; FREIRE, N. **Metadados**. 23 out. 2002. Disponível em: <<http://metadados.bn.pt/index.html>>. Acesso em: 20 set. 2011.

BOTERAM, F. "Content architecture". Semantic interoperability in an international comprehensive knowledge organization system. **Aslib Proc.**, London, v. 62, n. 4/5, p. 406-414, 2010.

BRAGA, K. S. Aspectos relevantes para a seleção de metodologia adequada à pesquisa social em Ciência da Informação. In: MUELLER, S. P. M. (Org.). **Métodos para a pesquisa em ciência da informação**. Brasília: Thesaurus, 2007. cap. 1, p. 17-38.

BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, São Paulo. **Anais...** São Paulo: ANCIB, 2008. Disponível em: <<http://www.slideshare.net/doritchka/brascher-e-caf-organizacao-da-informao-ou-do-conhecimento>>. Acesso em: 20 dez. 2011.

BRASCHER, M.; CARLAN, E. Sistemas de organização do conhecimento: Antigas e novas linguagens. In: ROBREDO, J.; BRASCHER, M. (Org.). **Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento**. Brasília: IBICT, 2010. Cap. 8, p. 147-176.

BROUGHTON, V. et al. Knowledge organization. In: KAJBERG, L.; LØRRING, I. (Org.) **European curriculum reflections on library and information science education**. Copenhagen: Royal School of Library and Information Science, 2005. cap. 7, p. 133-148. Disponível em: <<http://biblis.db.dk/uhtbin/hyperion.exe/db.leikaj05>> . Acesso em: 2 jul. 2011.

CALAZANS, A. T. S. Estudo de caso: uma estratégia de pesquisa. In: MUELLER, S. P. M. (Org.). **Métodos para a pesquisa em ciência da informação**. Brasília: Thesaurus, 2007. cap. 2, p. 39-62.

CAMPOS, L. F. B. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Enc. Bibli.: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, v. 12, n. 23, p. 16-45, 2007. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12n23p16/390>>. Acesso em: 2 maio 2011.

CAMPOS, M. L. A.; CAMPOS, L. M.; MEDEIROS, J. S. A representação de domínios de conhecimento e uma teoria de representação: a ontologia de fundamentação. **Inf. Inf.**, Londrina, v. 16, n. 3, p. 140-164, 2011.

CAMPOS, M. L. A.; GOMES, H. E. Taxonomia e classificação: o princípio de categorização. **DataGramaZero**, v. 9, n. 4, p. ago. 2008. Disponível em: <http://www.dgz.org.br/ago08/Art_01.htm>. Acesso em: 5 jan. 2012.

CARLAN, E. **Sistemas de organização do conhecimento**: uma reflexão no contexto da Ciência da Informação. 100 f. 2010. Dissertação (Mestrado em Ciência da Informação) – Departamento de Ciência da Informação, Universidade de Brasília, Brasília, 2010.

CARVALHO, M. C. R. Bibliotecas universitárias brasileiras e a implantação de repositórios institucionais. **Rev. Inf. Univers**, v.1, jul./dez. 2009. Disponível em: <<http://www.siglinux.nce.ufrj.br/~gtbib/site/2009/06/implantacao-de-repositorios/>>. Acesso em: 07 set. 2009

CASTRO, F. F.; SANTOS, P. L. V. A. C. A consistência da informação descritiva em repositórios digitais: caminhos para a interoperabilidade. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 24., Maceió, 2011. **Anais...** Disponível em: <<http://www.febab.org.br/congressos/index.php/cbbd/xxiv/paper/view/327>>. Acesso em: 20 set. 2011.

CASTRO, F. F.; SANTOS, P. L. V. C. Os metadados como instrumentos tecnológicos na padronização e potencialização dos recursos informacionais no âmbito das bibliotecas digitais na era da Web Semântica. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 22., 2007, Brasília. **Anais...** Brasília: ABDF/UNB, 2007.

CATARINO, M. E. **Integração das folksonomias nos metadados**: identificação de novos elementos como contributo para a descrição de recursos em repositórios. 2009. Tese (Doutorado em Tecnologias e Sistemas de Informação) – Departamento de Sistemas de Informação, Universidade do Minho, 2009.

CHAN, L. M.; ZENG, M. L. Ensuring interoperability among subject vocabularies and knowledge organization schemes: a methodological analysis. In: IFLA COUNCIL AND GENERAL CONFERENCE, Glasgow, 18-24 Aug. 2002. Disponível em: <<http://www-db.stanford.edu/~kevin/publications/int>>. Acesso em: 2 maio 2011.

CHAN, L. M.; ZENG, M. L. Metadata interoperability and standardization: a study of methodology Part I. Achieving interoperability at the schema level. **D-Lib Magazine**, v. 12, n. 6, June 2006. Disponível em: <<http://www.dlib.org/dlib/june06/zeng/06zeng.html>>. Acesso em: 2 maio 2011.

CLARKE, S. G. D. **ISO 25964**: a standard in support of KOS interoperability. Disponível em: <<http://www.iskouk.org/conf2011>>. Acesso em: out. 2011

CINTRA, A. M. M. et al. **Para entender as linguagens documentárias**. 2. ed., rev. e ampl. São Paulo: Polis, 2002.

COSTA, S. Abordagens, estratégias e ferramentas para o acesso aberto via periódicos e repositórios institucionais em instituições acadêmicas brasileiras. **Liinc em Rev.**, Rio de Janeiro, v. 4, n. 2, p. 218-232, 2008.

COSTA, S. M. S.; LEITE, F. C. L. Insumos conceituais e práticos para iniciativas de repositórios institucionais de acesso aberto à informação científica em bibliotecas de pesquisa. In: SAYÃO, L. et al. (Org.). **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009. p. 163-202.

CROW, R. **Sparc**. Institutional repository checklist & resource guide. Washington: SPARC, 2002.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.

CURRÁS, E. **Ontologias, taxonomia e thesaurus em teoria de sistemas e sistemática**. Brasília: Thesaurus, 2010.

DARBY, R. M. et al. Increasing the productivity of interactions between subject and institutional repositories. **New Rev. Inf. Network.**, v. 14, p. 117-135, 2008.

DIGITAL REPOSITORY INFRASTRUCTURE VISION FOR EUROPEAN RESEARCH. **Directrizes DRIVER 2.0**: directrizes para fornecedores de conteúdos: exposição de recursos textuais com o protocolo OAI-PMH. (Versão portuguesa Serviços Documentação da Universidade do Minho, abr. 2009). Disponível em: <http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final__PT.pdf>. Acesso em: out. 2010.

DODEBEI, V. Repositórios institucionais: por uma memória criativa no ciberespaço. In: SAYÃO, L. et al. (Org.). **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009. p. 83-106.

DODEBEI, V. L. D. **Tesauro**: linguagem de representação e memória documentária. Niterói: Intertexto, 2002.

DOER, M. Increasing the power of semantic interoperability for the European Library. **ERICIM**, n. 66, July 2006. Disponível em: <<http://eprints.sics.se/299/1/doerr.html>>. Acesso em: 13 out. 2010.

DRAKE, M. A. Institutional repositories: hidden treasures. **Searcher**, v. 12, n. 5, May 2004. Disponível em: <<http://www.infotoday.com/searcher/may04/drake.shtml>>. Acesso em: 13 maio 2011.

DSPACE@MIT. **MIT DSpace policies and guidelines**. Disponível em: <<http://libraries.mit.edu/dspace-mit/build/policies/index.html>>. Acesso em: 2 dez. 2010.

DUBLIN CORE METADADA INICIATIVE. DCMI. **Metadada Terms**: Section 4: Vocabulary encoding schemes. Disponível em: <<http://dublincore.org/documents/dcmi-terms/#H4>>. Acesso em: 2 fev. 2011.

DUVAL, E. et al. Metadata principles and practicalities. **D-Lib Magazine**, v. 8, n. 4, p. 1-15, abr. 2002. Disponível em: <<http://www.dlib.org/dlib/april02/weibel/04weibel.html>>. Acesso em: 15 nov. 2002.

EIJNDHOVEN, K.; GRAAF, M. **Inventory study into the present type and level of OAI compliant Digital Repository activities in the EU**. [S.l.]: SurfFoundation, 2007. Disponível em: <<http://www.pleiade.nl/rapport/DRIVERInventoryStudy.pdf>>. Acesso em: 5 nov. 2010.

ENABLING OPEN SCHOLARSHIP. **Formulating an institutional Open access policy**. Liège: Université of Liège, [2010].

ESPÍRITO SANTO, J. **Delineamentos de metodologia científica**. São Paulo: Loyola, 1992.

FACHIN, G. R. B. et al. Gestão do conhecimento e a visão cognitiva dos repositórios institucionais. **Perspect. Cienc. Inf.**, Belo Horizonte, v. 14, n. 2, p. 220-236, maio/ago. 2009.

FERREIRA, M.; BAPTISTA, A. A. The use of taxonomies as a way to achieve interoperability and improved resource discovery in DSpace-based repositories. In: RAMALHO, J.C.; SIMÕES, A.; LOPES, J.C. (Ed.). XATA 2005 : XML : aplicações e tecnologias associadas. **Actas da III Conferência Nacional**, Braga, 2005. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/873/1/paper-25.pdf>>. Acesso em: 3 jun. 2011.

FERNANDES, L. A.; GOMES, J. M. M. Relatórios de pesquisa nas ciências sociais: características e modalidades de investigação. **ConTexto**, Porto Alegre, v. 3, n 4, p. 1-23, 2002.

FERREIRA, S. M. S. P.; SOUTO, L. F. Dos sistemas de informação federados à federação de bibliotecas digitais. **Rev. Bras. Biblio. Doc.**, Nova Série, São Paulo, v. 2, n. 1, p.23-40, jan./jun. 2006.

FONSECA, F., EGENHOFER, M.; BORGES, K. Ontologias e interoperabilidade semântica entre SIGs. In: WORKSHOP BRASILEIRO EM GEOINFORMÁTICA - GEOINFO2000, 2., 2000. **Anais...** Disponível em: <<http://mtc-m18.sid.inpe.br/col/dpi.inpe.br/vagner/2000/07.04.15.32/doc/011.pdf>>. Acesso em: 5 jan. 2010.

FOULONNEAU, M.; ANDRÉ, F. **Investigative study of standards for digital repositories and related services**. Amsterdam: Amsterdam University Press, 2008.

FUJITA, M. S. L. Organização do conhecimento: algumas considerações para o tratamento temático da informação. In: CARRARA, K. (org.). **Educação, universidade e pesquisa**. São Paulo: FAPESP, 2001. p. 29-34.

GALHO, T. S.; MORAES, S. M. W. **Categorização automática de documentos de texto utilizando lógica difusa**. Gravataí: Universidade Luterana do Brasil. Disponível em: <<http://inf.unisul.br/~ines/workcomp/cd/pdfs/2405.pdf>>. Acesso em: 20 maio 2012.

GALINDO, M.; PEREIRA, M.; LIMA, C.M. V. Bibliotecas digitais e metadados: uma abordagem integradora. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS., 2., 2004. **Anais...** Disponível em: <<http://cutter.unicamp.br/document/?code=8283>>. Acesso em: 2 mar. 2011.

GARCIA, S. S.; MOURA, A. M. C.; CAMPOS, M. L. M. **Metadados para documentação e recuperação de imagens**. Relatório Técnico n.º 040/DE9/99. São Paulo: Instituto Militar de Engenharia, 1999. Disponível em: <<http://hermes.dpi.inpe.br:1910/col/dpi.inpe.br/banon/2004/04.21.13.08/doc/TeseMetadadosImagens.pdf>>. Acesso em: 13 maio 2011.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2008.

GILLILAND-SWETLAND, A. J. **Introduction to metadata: setting the stage**. 2000. Disponível em: <<http://www.slis.kent.edu/~mzeng/metadata/Gilland.pdf>>. Acesso em: 4 nov. 2010.

GODIN, B.; GINGRAS, Y. The place of universities in the system of knowledge production. **Res. Policy**, Amsterdam, v. 29, p. 273-278, 2000.

GOMES, H. E. Tendências da pesquisa em organização do conhecimento. **Pesq. bras. Ci. Inf.**, Brasília, v. 2, n. 1, p. 60-68, jan./dez. 2009.

GOMES, H.E.; GUIMARAES, L. S. Organização do conhecimento e recuperação da informação na era das tecnologias digitais. **Bibliotextos**, 15 mar. 2012. Disponível em: <<http://bibliotextos.files.wordpress.com/2012/03/organizac3a7c3a3o-do-conhecimento-e-recuperac3a7c3a3o-da-informac3a7c3a3o-na-era-das-tecnologias.pdf>>. Acesso em: 10 mar. 2012.

GÓMEZ DUEÑAS, L. F. **Modelos de interoperabilidad en bibliotecas digitales y repositorios documentales**: caso Biblioteca Digital Colombiana BDCOL. Disponível em: <http://eprints.rclis.org/handle/10760/14878#.T57tRLOm_NU>. Acesso em: 2 mar. 2012.

GRUBER, T. **What is an ontology?** 1996. Disponível em: <<http://ksl.stanford.edu/kst/what-is-an-ontology.html>> Acesso em: abril 2007.

GUIMARÃES, F. J. Z. **Utilização de ontologias no domínio B2C**. 2002. 195 f. Dissertação (Mestrado em Informática) - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2002.

GUY, M. **Interoperability focus**: about. Bath: UKOLN, 2006. Disponível em: <<http://www.ukoln.ac.uk/interop-focus/about/>>. Acesso em: 5 jan. 2010.

HARNARD, S. Entrevista com Stevan Harnard. **Ci. Inf.**, Florianópolis, n. esp., p. x-xv, 2007.

HEERY, R. Introduction. In. HEERY, R. et al. (Org.). **An evaluation study on the development and implementation of community repositories to support research and learning and teaching**. DELOS Project nº 507618: Report on digital repositories: an avaluation study on the development and implementation of community repositories which support research (and learning and teaching). 2006a. p. 1-11. Disponível em: <<http://delos-wp5.ukoln.ac.uk/project-outcomes/WP-5-0-1/d-5-1-1-report.pdf>>. Acesso em: 12 mar. 2011.

HEERY, R. Summary and recommendations. In. HEERY, R. et al. (Org.). **An evaluation study on the development and implementation of community repositories to support research and learning and teaching**. DELOS Project nº 507618: Report on digital repositories: an avaluation study on the development and implementation of community repositories which support research (and learning and teaching). 2006b. p. 50-51. Disponível em: <<http://delos-wp5.ukoln.ac.uk/project-outcomes/WP-5-0-1/d-5-1-1-report.pdf>>. Acesso em: 12 mar. 2011.

HILL, L. L. et al. Collection metadata solutions for digital library applications. **J. Am. Soc. Inf. Sci.**, v. 50, n. 13, p. 1169-1181, 1999.

HJØRLAND, B. Fundamentals of knowledge organization. **Knowl. Org.**, Bonn, v. 30, n. 2, p. 87-11, 2003. Disponível em: <<http://www.clir.org/pubs/reports/pub91/contents.htm>>. Acesso em: 25 jan. 2011.

HJØRLAND, B. **Knowledge organization systems**. 2007. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm>. Acesso em: 13 maio 2011.

HJØRLAND, B. What is knowledge organization (KO). **Knowl. Org.**, Bonn, v. 35, n. 3/3, p. 86-100, 2008.

HJØRLAND, B. Library and information science and philosophy of science. **J. Doc.**, London, v. 61, n. 1, p. 5-10, 2005.

HODGE, G. **Systems of knowledge organization for digital libraries**: beyond traditional authorities files. Washington, DC: CLIR, 2000. Disponível em: <<http://www.clir.org/pubs/reports/pub91/contents.htm>> Acesso em: 24 maio 2008.

IBICT. DSPACE - glossário. Disponível em: <http://dspace.ibict.br/index.php?option=com_content&task=view&id=43&Itemid=77>. Acesso em: 5 nov. 2010.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **O manifesto da IFLA sobre a Internet**. Genève: IFLA, 2002. Disponível em: <<http://www.ifla.org/III/misc/im-pt.htm>>. Acesso em: jun. 2009.

OS INVESTIGADORES e o *open access*: apresentação pra investigadores e docentes. **RCAAP**. Disponível em: <http://projecto-devel.rcaap.pt/index.php?option=com_content&view=article&id=26&Itemid=38&lang=pt>. Acesso em: 20 out. 2011.

JARDIM, J. M.; SILVA, S. C. A.; NHARRELUGA, R. S. Análise de políticas públicas: uma abordagem em direção às políticas públicas de informação. **Perspect. Ciênc. Inf.**, Belo Horizonte, v. 14, n. 1, p. 2-22, Apr. 2009

KOCH, T. Subject access and semantic interoperability in repositories. In: HEERY, R. et al. (Org.). **An evaluation study on the development and implementation of community repositories to support research and learning and teaching**. DELOS Project nº 507618: Report on digital repositories: an avaluation study on the development and implementation of community repositories which support research (and learning and teaching). 2006. p. 38-49. Disponível em: <

wp5.ukoln.ac.uk/project-outcomes/WP-5-0-1/d-5-1-1-report.pdf>. Acesso em: 12 mar. 2011.

KURAMOTO, H. Peter Suber fala sobre o estágio atual das iniciativas OA. **Blog do Kuramoto**, postado em 7 jul. 2011.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Brasília: Briquet de Lemos, 1986.

LAVILLE, C.; DIONNE, J. **Construção do saber: manual de metodologia da pesquisa em ciências humanas**. Belo Horizonte: UFMG, 1999.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto**. Brasília: Ibict, 2009.

LEITE, F. C. L. Comunicação científica e gestão do conhecimento: enlaces conceituais para a fundamentação da gestão do conhecimento científico no contexto da universidade. **Transinformação**, Campinas, v. 19, n. 2, p. 139-151, maio/ago 2007.

LEITE, F. C. L.; COSTA, S. Repositórios institucionais como ferramentas de gestão do conhecimento científico no ambiente acadêmico. **Persp. Cienc. Inf.**, Belo Horizonte, v. 11, n. 2, p. 206-219, maio/ago. 2006.

LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. **ARL Bimonthly Report**, Washington, n. 226, 2003. Disponível em: <<http://www.arl.org/resources/pubs/br/br226/br226ir~prints.html>>. Acesso em: 17 dez. 2008.

LYNCH, C. A.; GARCIA-MOLINA, H. Interoperability, scaling, and the digital libraries research agenda: a report on the may 18-19, 1995. IITA Digital Libraries Workshop, 1995. Disponível em: <<http://www.sis.pitt.edu/~repwkshop/papers/DL1995.pdf>>. Acesso em: maio 2011.

LYNCH, C. A.; LIPPINCOTT, J. K. Institutional repository deployment in the United States as of early. **D-Lib Magazine**, v. 11, n. 9, Sept. 2005. Disponível em: <<http://www.dlib.org/dlib/september05/lynch/09lynch.html>>. Acesso em: maio 2011.

MAI, J. E. **The future of general classification**. 2003. Disponível em: <<http://www.haworthpress.com/store/product.asp?sku=J104>>. Acesso em: 12 jul. 2011.

MARCONDES, C. H.; SAYÃO, L. F. Documentos digitais e novas formas de cooperação entre sistemas de informação em C& T. **Ci. Inf.**, Brasília, v. 3, n. 31, p. 42-53, set./dez. 2002.

MARCONDES, C. H.; SAYÃO, L. F. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. **Ci. Inf.**, Brasília, v. 30, n. 3, p. 24-33, set./dez. 2001.

MARCONDES, C. H.; SAYÃO, L. F. Introdução: repositórios institucionais e livre acesso. In: SAYÃO, L. et al. (Org.). **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009.

MCCULLOCH, E. Multiple terminologies: an obstacle to information retrieval. **Library Rev.**, Bradford, v. 53, n. 6, p. 1723-1737, 2004.

MELERO, R. et al. The situation of *open access* institutional repositories in Spain: 2009 report. **Inf. Res.**, v. 14, n. 4, p. 1-21, Dec. 2009.

METADATA basics. **Dublin Core Metadata Initiative**. Disponível em: <<http://dublincore.org/metadata-basics/>>. Acesso em: 5 jan. 2010.

MILLER, P. I say what I mean, but do I mean what I say? **Ariadne**, v. 23, p. 1-11, 2000a. Disponível em: <<http://www.ariadne.ac.uk/issue23/metadata/>>.

MILLER, P. Interoperability. What is it and why should I want it? **Ariadne**, v. 24, p. 1-7, June 2000b. Disponível em: <<http://www.ariadne.ac.uk/issue24/interoperability/intro.html>>. Acesso em: 31 jan. 2011.

MIRANDA, M. L. C. A organização do conhecimento e seus paradigmas científicos: algumas questões epistemológicas. **Informare: Cadernos do Programa de Pós-Graduação em Ciência da Informação**, Rio de Janeiro, v. 5, n. 2, p. 64-77, 1999.

MIRANDA, M. L. C. **Organização e representação do conhecimento: fundamentos teórico-metodológicos na busca e recuperação da informação em ambientes virtuais**. 2005. 354 f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

MODESTO, F. A cara da biblioteca na próxima geração do OPAC. **INFOhome**. 2010. Disponível em: <http://www.ofaj.com.br/colunas_conteudo.php?cod=564>. Acesso em: 13 jan. 2012.

MOEN, W. E. Mapping the interoperability landscape for networked information retrieval. In: ACM/IEEE-CS JOINT CONFERENCE ON DIGITAL LIBRARIES, 1., 2001. **Proceedings**.... New York: ACM, 2001. Disponível em: <www.unt.edu/wmoen/publications/MapInteropJCDFinal.pdf>. Acesso em: 5 jan. 2011.

MONTEIRO, F. S. **Organização da informação em repositórios digitais institucionais com ênfase na descrição física e descrição temática**. 2008. 199f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, 2008.

MOURA, M. A. Interoperabilidade semântica e ontologia semiótica: a construção e o compartilhamento de conceitos científicos em ambientes colaborativos *online*. **Inf. Inf.**, Londrina, v. 16, n. esp., p. 165-179, jan./jun. 2011.

MOYANO GRIMALDO, W. A. La interoperabilidade n los lenguajes documentales y el caso del proyecto Armarc para autoridades de matéria. Disponível em: <http://eprints.rclis.org/bitstream/10760/13723/1/La_interoperabilidad_en_los_lenguajes_documentales.pdf>. Acesso em: 20 jan. 2011.

NAGAMORI, M.; SUGIMOTO, S. Metadata schema registry as a tool to enhance metadata interoperability. **TCDL Bulletin**, v. 3, n. 1, 2006. Disponível em: <<http://www.ieee-tcdl.org/Bulletin/v3n1/nagamori/nagamori.html>>. Acesso em: 5 jan. 2011.

NATIONAL INFORMATION STANDARDS ORGANIZATION. ANSI/NISO Z39.19 – Diretrizes para a construção formatar, e gestão de monolíngues vocabulários controlados. 2005. Disponível em: <www.niso.org/standards/resources/Z39-19.html>. Acesso em: 2 nov. 2011.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **A framework of guidance for building good digital collections**. 3. ed. Bethesda: NISO, 2007.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **The Dublin Core metadata element set**. Bethesda: NISO Press, 2001. Disponível em: <http://www.niso.org/apps/group_public/download.php/6578/The%20Dublin%20Core%20Metadata%20Element%20Set.pdf>. Acesso em 9 set. 2011).

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding metadata**. Bethesda: NISO, 2004.

NEEDLEMAN, M. Z39.50 – a review, analysis and some thoughts on the future. **Library Hi Tech**, Ann Arbor, v. 18, n. 2, p. 158-165, 2000.

NEVES, D. A. Information science and human cognition: an information processing approach. **Ci. Inf.**, Brasília, v. 35, n. 1, p. 39-44, 2006.

NICHOLSON, D. et al. **HILT**: High-level Thesaurus Project: final report to RSLP & JISCH. 2001. Disponível em: <<http://cdlr.strath.ac.uk/pubs/nicholsond/HILTfinalreport.pdf>>. Acesso em: 20 ago. 2011.

NUNES, P. Conceito de organização. **Portal do administrador**. 2008 Disponível em: <<http://www.htmlstaff.org/xkurt/projetos/portaldoadmin/modules/news/article.php?storyid=995>>. Acesso em: 20 ago. 2011.

OPEN DOAR. Directory of *Open access* Repositories. **Search or browse for repository**. Disponível em: <<http://www.opendoar.org/find.php>>. Acesso em: 20 ago. 2012.

PAEPCKE, A. et al. Interoperability for digital libraries: problems and directions. 2010. **CiteSeer^xbeta**. Disponível em: <<http://www.db.stanford.edu/~kevin/publications/int>>. Acesso em: 2 mar. 2011.

PANTOJA, A. O. **O ensino da odontologia no Espírito Santo**. [Vitória, 1971].

PARK, J.R. Metadata quality in digital repositories: a survey of the current state of the art. **Cat. & Classif. Q.**, New York, v. 47, p. 213-228, 2009.

PARK, J.R.; TOSAKA, Y. Metadata creation practices in digital repositories and collections: schemata, selection criteria, and interoperability. **Inf. Technol. Libr.**, Chicago, v. 29, n. 3, p. 104-116, 2010. Disponível em: <<http://www.ala.org/ala/mgrps/divs/lita/ital/292010/2903sep/park.cfm>>. Disponível em: jun. 2011.

PATEL, M. et al. **Semantic interoperability in digital library systems**. Bath: Network of Excellence on Digital Libraries, 2005. (Project n.º 507618 DELOS).

PAYETTE, S. et al. Interoperability for digital objects and repositories: the Cornell/CNRI experiences. **D-Lib Magazine**, Reston, v. 5, n. 5, p. 1-23, May 1999. Disponível em: <<http://www.dlib.org/dlib/may99/payette/05payette.html>>. Acesso em: 12 dez. 2010.

PEREIRA, A. M.; RIBEIRO JUNIOR, D. I.; NEVES, G. L. C. Metadados para a descrição de recursos da internet: as novas tecnologias desenvolvidas para o padrão Dublin Core e sua utilização. **Rev. ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 10, n. 1, p. 241-249, jan./dez. 2005.

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

RALYTÉ, J. et al. A knowledge-based approach to manage information systems interoperability. **Information Systems**, Elmsford, v. 33, p. 754-784, 2008.

RAMALHO, R. A. S. **Desenvolvimento e utilização de ontologias em bibliotecas digitais: uma proposta de aplicação**. 2010. 145 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista, Marília, 2010.

RELVÃO, R. **Elementos de metadados para a aprendizagem à distância**. 2003. 47 f. Relatório de Projeto (Mestrado em Sistemas de Informação) – Departamento de Sistemas de Informação, Universidade do Minho, Guimarães, 2003.

RIEH, S. Y. et al. Perceptions and experiences of staff in the planning and implementation of institutional repositories. **Libr.Trends**, Champaign, v. 57, n. 2, p. 168-190, Fall 2008.

RODRIGUEZ MARTINEZ, F. J. Agents and ontologies working together to federate digital libraries. **TCDL Bul.**, v. 2, n. 2, 2006. Disponível em: <<http://www.ieee-tcdl.org/Bulletin/v2n2/rodriguez-martinez/rodriguez-martinez.html>>. Acesso em: 2 maio 2011.

ROSA, F.; GOMES, M. J. Dos promotores aos utilizadores: estudos sobre o RepositóriUM. In: GOMES, M. J. G.; ROSA, F. (Org.). **Repositórios institucionais: democratizando o acesso ao conhecimento**. Salvador: EDUFBA, 2010. p.153-203.

ROWLEY, J. BS 8723 Structured vocabularies for information retrieval: Part 1: Definitions, symbols and abbreviations, and Part 2: Thesauri. **J. Doc.**, London, v. 63, n. 3, p. 428-431, 2007.

SAN SEGUNDO MANUEL, R. **Sistemas de organización del conocimiento: Ila organización del conocimiento en las bibliotecas españolas**. Madrid: Universidad Carlos III de Madrid; Boletín Oficial del Estado, 1996.

SAYÃO, L. F. Padrões para bibliotecas digitais abertas e interoperáveis. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, v. 2, n. 2, 2007. Disponível em: <<http://revista.ibict.br/pbcib/index.php/pbcib/article/view/794>>. Acesso em: 20 ago. 2011.

SAYÃO, L. F.; MARCONDES, C. H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, Campinas, v. 20, n. 2, p. 133-148, maio/ago. 2008.

SELVAGE, M. Y et al. Achieve semantic interoperability in a SOA: patterns and best practices. **DeveloperWorks**, 23 June 2006. Disponível em: <www.ibm.com/.../library/ws-soa-seminterop.html>. Acesso em: 25 jan. 2011.

SHEARER, K. Institutional repositories: towards the identification of critical success factors. **Can. J. Inf. Libr. Sci.**, Ontario, 2003. Disponível em: <http://cais-acsi.ca/proceedings/2003/Shearer_2003.pdf>. Acesso em: 10 dez. 2010.

SHERPA ROMEO. **Políticas de Copyright e de auto-arquivo de editores**. 2010. Disponível em: <<http://www.sherpa.ac.uk/romeo/index.php?flDnum=&la=pt>>. Acesso em: 2 dez. 2010.

SHINTAKU, M.; BRÄSCHER, M. **DSpace versão 1.4** : uma análise das facilidades relacionadas ao assunto. In: Seminário Internacional de Bibliotecas Digitais Brasil, 2007. Disponível em: <<http://cutter.unicamp.br/document/?code=23471>>. Acesso em: 2 jul. 2011.

SHINTAKU, M.; FERREIRA, S. M. M. S. P.; ROBREDO, J. Repositórios brasileiros implementados com DSpace. In: ENCONTRO NACIONAL DE ENSINO E PESQUISA EM INFORMAÇÃO., 10., Salvador, 2011. Disponível em: <<http://www.cinform2011.ici.ufba.br/modulos/submissao/Upload/37558.pdf>>. Acesso em: 2 mar. 2012.

SHIRI, A. Digital library research: current developments and trends. **Library Rev.**, Bradford, v. 52, n. 5, p. 198-202, 2003.

SLAVIC, A.; BROUGHTON, V. **Facet analytical theory in managing knowledge structures for humanities**. Outcomes and prototype: Project FAKTS. 2003. Disponível em: <<http://ucl.ac.uk/fatks/prototype.htm>>. Acesso em: 15 ago. 2011.

SOUZA, S. **CDU**: como entender e utilizar a edição-padrão internacional em língua portuguesa. 3. ed. rev. e ampl. Brasília: Thesaurus, 2004.

SOUZA, M. I. F.; ALVES, M. D. R. Representação descritiva e temática de recursos de informação no sistema agência Embrapa: uso do padrão Dublin Core. **Rev. Dig. Bibliotec. Ci. Inf.**, Campinas, v. 7, n. 1, p. 190-205, 2009.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ci. Inf.**,

Brasília, v. 29, n. 1, p. 93-102, abr. 2000. Disponível em: <<http://www.scielo.br/pdf/ci/v29n1/v29n1a10.pdf>>. Acesso em: 20 maio 2011.

SOUZA, T. B.; CATARINO, M. E.; SANTOS, P. C. Metadados: catalogando dados na Internet. **Transinformação**, Campinas, v. 9, n. 2, 1997. Disponível em: <<http://www.puccamp.br/~biblio/tbsouza92.html>>. Acesso em: 9 fev. 2000.

SUBER, P. **Open access overview**: focusing on *open access* to peer-reviewed research articles and their preprints. Disponível em: <<http://www.earlham.edu/~peters/fos/overview.htm>>. Acesso em: 10 dez. 2010.

SUBER, P. Victory in the Senate: update on the bill to mandate access at the NIH. **SPARC Access Newsletter**, v. 115, Nov. 2, 2007. Disponível em: <http://dash.harvard.edu/bitstream/handle/1/4322591/suber_nihsenate.html?sequence=1>. Acesso em: 10 dez. 2010.

SWAN, A. Institutional repositories: now and next. In: DALE, P.; BEARD, J.; HOLLAND, M. (Ed.). **University libraries and digital learning environments**. Surrey: Ashgate Publishing, 2011. Cap. 8, p. 119-134.

SWAN, A. **Open access self-archiving**: an introduction. 2005. Disponível em: <http://www.openscholarship.org/upload/docs/application/pdf/2009-01/open_access_self-archiving_-_an_introduction.pdf>. Acesso em: 2 dez. 2010.

TEIXEIRA, P. R. O que é SEO? Marketing de busca. Disponível em: <<http://www.marketingdebusca.com.br/seo/>>. Acesso em: 20 ago. 2011.

TENNIS, J. T. Epistemology, theory, and methodology in knowledge organization: toward a classification, metatheory, and research framework. **Know. Org.**, v. 35, n. 2/3, p. 102-112, 2008.

TENNIS, J. T. Layers of meaning: disentangling subject access interoperability. **Advances in Classification Res.**, Medford, v. 12, p. 113-122, 2004. Disponível em: <http://faculty.washington.edu/jtennis/Tennis_layers_2001.pdf>. Acesso em: 2 set. 2011.

TOLOSANA-CALASANZ, R. et al. On the problem of identifying the quality of geographic metadata. **Res. Adv. Tech. Dig. Lib.**, vol. 4172, pp. 232-243, 2006.

TOMAÉL, M. I.; SILVA, T. E. Repositórios institucionais: diretrizes para políticas de informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 28 a 31 de outubro de 2007. **Anais...** Salvador, 2007. Disponível

em: <<http://www.enancib.ppgci.ufba.br/artigos/GT5--142.pdf>>. Acesso em: 10 dez. 2010.

TORINO, E.; TORINO, L. P.; SILVA, T. E. Reflexões sobre política de informação em instituições acadêmicas: a via verde em foco. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 3., 2009. Disponível em: <<http://riut.utfpr.edu.br/jspui/handle/1/49>>. Acesso em: 10 dez. 2010.

TORINO, L. P. **Organização da produção científica em repositórios institucionais: um parâmetro para a UTFPR**. 2010. 150 f. Dissertação (Mestrado em Gestão da Informação) – Universidade Estadual de Londrina, Programa de Pós-Graduação em Gestão da Informação (Profissional), Londrina, 2010.

TRISKA, R.; CAFÉ, L. Arquivos abertos: subprojeto da biblioteca digital brasileira. **Ci. Inf.**, Brasília, v. 30, n. 3, p. 92-96, set./dez. 2001.

TRISTÃO, A. M. D.; FACHIN, G. R. B.; ALARCON, O. E. Sistemas de classificação facetada e tesauros: instrumentos para a organização do conhecimento. **Ci. Inf.**, v. 33, n. 2, 2004. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/viewArticle/88/81>>. Acesso em: 20 ago. 2011.

TRIVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 2007.

TUDHOPE, D.; KOCH, T.; HEERY, R. **Terminology services and technology. JISC state of the art review**. 2006. Disponível em: <http://www.jisc.ac.uk/media/documents/programmes/capital/terminology_services_and_technology_review_sep_06.pdf>. Acesso em: 2 fev. 2011.

TUDHOPE, D.; NIELSEN, M. L. Introduction to knowledge organization systems and services. **New Rev. Hypermed. Multimed.**, London, v. 12, n. 1, p. 3-9, June 2006.

UKOLN. interoperability focus: looking at interoperability. 2005. Disponível em: <<http://www.ukoln.ac.uk/interop-focus/about/leaflet.html>>. Acesso em: 20 jan. 2011.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. Sobre a UFES. 2010. Disponível em: <<http://portal.Ufes.br>>. Acesso em: 22 maio 2010.

VICKERY, B. C. **A note on knowledge organization**. Disponível em: <http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/Vickery_a_note_on_knowledge_organisation.htm>. Acesso em: 20 set. 2011.

W3C. SKOS Simple Knowledge Organization System – Home page. Disponível em: <<http://www.w3.org/2004/02/skos/>>. Acesso em: 12 mar. 2012.

WEIBEL, S. The Dublin core: a simple content description model for electronic resources. **Bul. Am. Soc. Inf. Sci.**, Silver Spring, v. 24, n. 1, p. 9-11, p.9-11, Oct./Nov. 1997.

WHITE, W. Institutional repositories: contributing to institutional knowledge management and the global research commons. In: INTERNATIONAL OPEN REPOSITORIES CONFERENCE, 4., 18-21 May 2009, Atlanta. Disponível em: <<http://eprints.soton.ac.uk/48552/>>. Acesso em: 22 maio 2010.

ZENG, M. L. Knowledge organization systems (KOS). **Knowl. Org.**, Bonn, v. 35, n. 3/3, p. 160-182, 2008.

ZENG, M. L. Linguagens de marcação específicas por domínio e metadados descritivos: funções para a descoberta de recursos científicos. **Enc. Bibli. R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n. esp. 2º sem. 2010. Disponível em: <<http://br.vlex.com/vid/linguagens-especificas-descritivos-descoberta-264811162>>. Acesso em: 2 maio 2011.

ZENG, M. L.; CHAN, L. M. Metadata interoperability and standardization: a study of methodology Part 2. Achieving interoperability at the record and repository levels. **D-Lib Magazine**, Reston, v. 12, n. 6, June 2006. Disponível em: <<http://www.dlib.org/dlib/june06/zeng/06zeng.html> >. Acesso em: 2 maio 2011

ZENG, M. L.; CHAN, L. M. Trends and issues in establishing interoperability among knowledge organization systems. **J. Am. Soc. Inf. Sci. Tech.**, v. 55, n. 5, p. 377-395, 2004.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 3. ed. Porto Alegre: Bookman, 2005.

GLOSSÁRIO

AUTOARQUIVAMENTO - Depósito de um documento digital pelo seu autor em *Website* publicamente acessível (IBICT, 2011).

BIBLIOTECAS DIGITAIS - “[...] conjunto de objetos digitais construídos a partir do uso de instrumentos eletrônicos, concebidos com o objetivo de registrar e comunicar pensamentos, ideias, imagens e sons, disponíveis a um contingente ilimitado de pessoas, dispersas onde quer que a plataforma WWW alcance” (ALVARENGA, 2001, p. 3), de forma que considerações em relação a esse tipo de biblioteca é extensivo aos RIs .

CATEGORIZAÇÃO - É realizada por meio da análise do domínio a partir de recortes que permitem identificação dos conceitos (categorias) que fazem parte do domínio (CAMPOS; GOMES, 2008, p. 4).

COLEÇÃO - Listagem de itens de alguma forma relacionados ou similares. Uma coleção pode ser caracterizada e classificada de acordo com algumas qualidades, como a extensão do arquivo, a natureza do documento, entre outros (em torno de um tópico ou tipo de informação). Para cada critério previamente estabelecido, poder-se-á dar origem a uma coleção específica. Não se pode criar uma coleção sem uma comunidade (IBICT, 2011).

COMUNIDADE - Ponto de entrada ao material depositado no repositório. Consiste numa *homepage* (ou página inicial) especificamente configurada para a comunidade, com o conjunto das coleções a ela pertencentes e o grupo de usuários a ela integrados, assim como suas informações, notícias e *links*, refletindo os interesses de tal comunidade. Em cada comunidade, podem existir inúmeras coleções, e em cada coleção podem existir inúmeros itens. Essa organização dá ao repositório a flexibilidade necessária para acomodar diferentes necessidades de cada comunidade, permitindo decidir sobre políticas (quem contribui com conteúdo, permissões de acesso, *workflow* e gerenciamento de coleções) (IBICT, 2011).

CONTAINER - Um item que armazena e gera dinamicamente metadados relacionados com um único recurso em múltiplos formatos, cada um dos quais pode ser recolhido como registros no protocolo OAI-PMH (DRIVER, 2009).

CROSSWALK - É um mapeamento semântico ou técnico, as vezes ambos, de uma estrutura de metadados para outro esquema de metadados” (PATEL et al., 2006, p. 21, tradução nossa).

ELEMENTO DE METADADO - Propriedade de um objeto que é definida em uma especificação. “Autor/*creator*”, Título/*“title”* e Autor/*“subject”* são propriedades definidas como elementos de metadados.

EXTENSÃO - Extensão é o conjunto de gêneros ou espécies possuidoras das mesmas qualidades (PIEIDADE, 1977, p. 28).

GRANULARIDADE - Refere-se ao tamanho, à decomponibilidade e ao grau em que um recurso é planejado para ser usado como parte de um recurso maior (NISO, 2007).

HARVESTING - Coleta automática de metadados que já é associada com recurso produzido automática ou manualmente.

INTENÇÃO - Intenção é o número de qualidades que distinguem as espécies, a soma de todas as propriedades comuns aos membros das classes (PIEIDADE, 1977, p. 28).

INTERONTOLOGIA - pode ser entendida como uma intersecção (sobreposição) a relação entre as entidades afins.

ITEM - É o documento individualmente considerado que compõe uma coleção. Para cada item, deve haver um metadado correspondente (IBICT, 2011).

LINGUAGEM DOCUMENTÁRIA - “[...] são linguagens construídas para indexação, armazenamento e recuperação da informação e correspondem a sistemas de símbolos destinados a traduzir os conteúdos dos documentos” (CINTRA, 2002, p. 39).

MAPEAMENTO - Processo de identificação de termos, conceitos e relações hierárquicas que são aproximadamente equivalentes (DOERR, 2001).

OBJETO - Sequência de bits digital (DRIVER, 2009).

OPEN ARCHIVES INITIATIVE PROTOCOL FOR METADADA HARVESTING (OAI-PMH) - é um protocolo desenvolvido pela Open Archives Initiative que define um mecanismo para coleta (*harvesting*) de registros de metadados em repositórios e proporciona a interoperabilidade.

PADRÃO DE CONTEÚDO - Padrão de autoridades ou conjunto de regras que determinam vocabulário, sintaxe ou formato em que o dado ou elemento de metadado está embutido (Tesauros, vocabulário controlado, AACR).

QUALIFICADORES - Informação que ajuda a definir o conteúdo do elemento do metadado. Os qualificadores identificam o atributo do esquema de valor do recurso (tesauro, sistemas de classificação), fornecendo um elemento de conteúdo de metadado.

RECURSOS - Qualquer sequência de bits, registrada como produção científica por uma pessoa autorizada (DRIVER, 2009).

REGISTRO - São metadados refletidos num formato único. Um registro é recuperado num XML como resposta a um pedido OAI-PHM de metadados de um item (DRIVER, 2009).

REGISTRO DE METADADOS - Conjunto de metadado de uma estrutura relevante, compreendendo todos os elementos relevantes, descrevendo um objeto.

REPOSITÓRIO DIGITAL - É um sistema de computador usado para armazenar coleções de uma biblioteca digital e disseminá-los aos usuários (ARMS apud KURAMOTO, 2009).

SERVIÇOS DE TERMINOLOGIA - “É um conjunto de serviços de extração, apresentação e aplicação de vocabulários, controlados e não controlados, incluindo os conceitos membros, termos e relacionamentos que descrevem o significado dos termos e facilitam a interoperabilidade semântica para fins de pesquisa, navegação, raciocínio semântico, indexação e classificação de assunto” (TUDHOPE; KOCH; HEERY, 2006, p. 42, tradução nossa).

SISTEMAS DE INFORMAÇÃO - Sistema automatizado ou manual que abrange pessoas, máquinas, e/ou métodos organizados para coletar, processar, transmitir e

disseminar dados que representam informação para o usuário e/ou cliente (SISTEMA DE INFORMAÇÃO, WIKIPEDIA).

TAXONOMIA - Pode referir-se tanto a uma classificação hierárquica das coisas, como aos princípios que baseiam a classificação (HODJE, 2000).

VOCABULÁRIO CONTROLADO - É um conjunto limitado de termos autorizados para uso na indexação e busca de documentos (LOPES, 2002).

APÊNDICE

APÊNDICE A ROTEIRO PARA PESQUISA

- a) Descrever a instituição
- b) Descrever o repositório
 - *software* utilizado
 - organização
 - características
 - materiais depositados
- c) Identificação e descrição dos instrumentos para recuperação e indexação de documentos

REPOSITÓRIO: DATA DA COLETA //				
IDENTIFICAÇÃO	ELEMENTO	QUALIFICADOR	INSTRUMENTO	FERRAMENTA
Subject	br-rjbn	Vocabulário controlado da Biblioteca Nacional (BR)	BN	
Subject	br-rjfgvb	Vocabulário da Fundação Getúlio Vargas	BIBLIODATA	
Subject	decs	Vocabulário controlado da Bireme	DECS	
Subject	lcsh	Vocabulário controlado da Library of Congress (EUA)	LCSH	
Subject	MeSH	Vocabulário controlado da Biblioteca Nacional de Medicina (EUA)	MeSH	
Subject	Udc	Classificação decimal de Dewey	CDU	
Subject	Udd	Classificação universal de Dewey	CDD	
Subject	LCC	Classificação da Library of Congress	LCC	
Subject	Termo livre			

- d) identificação de documentos com orientação para submissão;
- e) identificação de orientação para indexação e catalogação; e
- f) identificação de ocorrências relevantes.

APÊNDICE B

PLANILHA PARA IDENTIFICAÇÃO DOS INSTRUMENTOS E ELEMENTOS DE DESCRIÇÃO TEMÁTICA DOS RECURSOS UTILIZADOS PELOS REPOSITÓRIOS INSTITUCIONAIS

[illegible]

APÊNDICE C
QUESTIONÁRIO PARA IDENTIFICAÇÃO DE SERVIÇOS DOS REPOSITÓRIOS
INSTITUCIONAIS

QUESTIONÁRIO

A INTEROPERABILIDADE SEMÂNTICA NA PERSPECTIVA DA ORGANIZAÇÃO DO
CONHECIMENTO: UMA PROPOSTA PARA O REPOSITÓRIO INSTITUCIONAL DA
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

1. () Li e declaro que fui devidamente esclarecido(a) e concordo em participar
VOLUNTARIAMENTE da pesquisa coordenada pela mestrandia Morgana Carneiro de
Andrade.

DADOS SOBRE O REPOSITÓRIO INSTITUCIONAL (RI)

2. Nome do RI
3. URL do RI
4. *Software* utilizado pelo RI

UTILIZAÇÃO DE METADADOS

5. Esquema de Metadados utilizado

- () context_object
() DC qualified
() DC simple
() dcde2aleph
() DIDL
() imsmid
() MARC
() METS
() MODS
() primo_dc
() RDF
() rfc1807
() SWAP
() UK-LOM
() VRA4
() interno
() não sabe

Outra (Por favor, especifique) _____

ADMINISTRAÇÃO DO RI

6. Quem é o responsável pela administração do RI?

- () Biblioteca.
() Outro órgão em conjunto com a biblioteca.
() Outro departamento.
() Agente externo.

Outra (Por favor, especifique) _____

7. Quem realiza o depósito?

- ☐ Autoarquivamento pelos autores.
- ☐ Autoarquivamento pelos autores e controle de qualidade por profissionais da informação.
- ☐ Entregue pelos autores e depositado por profissionais da informação.
- ☐ Coletados pela administração e depositado por profissionais da informação.
- Outra (Por favor, especifique) _____

CONTEÚDO

8. Qual o tipo de material que pode ser inserido no RI?

- ☐ Livro/capítulo de livros.
- ☐ Revisões de livros.
- ☐ Papers apresentados em congressos.
- ☐ Pósteres apresentados em congressos.
- ☐ Artigo de periódicos.
- ☐ Artigos de periódicos submetidos (*pre-print*).
- ☐ Dissertações ou teses.
- ☐ Monografias (Especialização/Aperfeiçoamento).
- ☐ Trabalhos de graduação.
- ☐ Objeto de aprendizagem.
- ☐ Patentes.
- ☐ Documentos administrativos.
- ☐ Relatórios.
- ☐ Dados de pesquisa.
- ☐ Conteúdo de *site*.
- ☐ Documentos de discussão e trabalho.
- ☐ Documentos históricos.
- ☐ Materiais audiovisuais.
- ☐ *Softwares*.

Outra (Por favor, especifique) _____

METADADOS

9. Os metadados são mediados ou controlados pela administração do RI?

- ☐ Sim.
- ☐ Parcialmente.
- ☐ Não.
- ☐ Não sabe .

Outra (Por favor, especifique) _____

10. Que padrão de conteúdo de metadados é utilizado?

- ☐ AACR/RDA.
- ☐ EThOS/uketd_dc.
- ☐ ISAD(G).
- ☐ ISBD.
- ☐ DCAP.
- ☐ primo_dc.
- ☐ SWAP.
- ☐ UK-LOM.
- ☐ VR4A.
- ☐ Interno.
- ☐ NS/NC

Outra (Por favor, especifique) _____

INDEXAÇÃO DOS DOCUMENTOS

11. Qual das afirmações melhor descreve a indexação de assuntos

- ☐ Não é feita a indexação de assunto.
 - ☐ A atribuição de palavras-chave é livre, utilizando o idioma do País.
 - ☐ A atribuição de palavras-chave é livre, utilizando o idioma do País e em inglês .
 - ☐ Palavras-chave ou classificação são atribuídas de acordo com sistemas padronizados no idioma do País.
 - ☐ Palavras-chave ou classificação são atribuídas de acordo com sistemas padronizados no idioma do País e em inglês.
- Outra (Por favor, especifique) _____

12. A atribuição de palavras-chave é feita

- ☐ Pelos autores.
 - ☐ Por bibliotecários.
 - ☐ Por pessoal administrativo.
 - ☐ Não é obrigatório.
- Outra (Por favor, especifique) _____

13. Qual controle de autoridade de assunto é utilizado?

- ☐ DDC.
- ☐ CDU.
- ☐ LCC.
- ☐ LCSH.
- ☐ Bibliodata.
- ☐ Biblioteca Nacional (Brasil).
- ☐ MeSH
- ☐ DECS

Outra (Por favor, especifique) _____

14. Há algum instrumento ou procedimento para descrição dos recursos?

- ☐ Disponibilização de vocabulário controlado.
- ☐ Folksonomia.
- ☐ Tesauro.
- ☐ Esquema de categorização
- ☐ Taxonomia
- ☐ Redes semânticas
- ☐ Ontologia
- ☐ Sistema automático de indexação.
- ☐ Lista de descritores de assunto.
- ☐ Nenhum.

Outra (Por favor, especifique) _____

15. Os metadados são compartilhados com a biblioteca OPAC?

- ☐ Sim (100%).
- ☐ Mais 50%.
- ☐ Menos que 10%.
- ☐ Não.

INTEROPERABILIDADE

(compartilhamento de dados na máquina para nível de máquina)

16. O RI exporta metadado em OAI-PMH?

- ☐ Sim.
- ☐ Pendente.
- ☐ Não.
- ☐ Não sabe.

17. O seu repositório interopera com algum outro sistema interno?

- ☐ Sistema de gerenciamento de biblioteca/arquivo.

- ☐ Curso de administração.
- ☐ Gerenciamento de recursos eletrônicos.
- ☐ Empresa.
- ☐ Financeiro.
- ☐ Recursos humanos.
- ☐ Departamento de pesquisa administrativa.
- ☐ Ambiente de aprendizagem virtual.
- ☐ Ambiente virtual de pesquisa.

Outra (Por favor, especifique) _____

FUNÇÕES E SERVIÇOS

18. O conteúdo do repositório é pesquisável por algumas das opções

- ☐ BASE
- ☐ Citeseer
- ☐ motores de busca
- ☐ MEIND
- ☐ MetaGer
- ☐ OAI search
- ☐ OAIster
- ☐ OASE
- ☐ Open DOAR
- ☐ OPUS
- ☐ PLEIADI
- ☐ Scirus

Outra (Por favor, especifique) _____

19. Na sua concepção quais os metadados e instrumentos de organização do conhecimento que poderiam ser utilizados para a viabilização da interoperabilidade semântica

APÊNDICE D
TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

A interoperabilidade semântica na perspectiva da organização do conhecimento: uma proposta para o Repositório Institucional da Universidade Federal do Espírito Santo

Prezado(a) Senhor(a):

Gostaríamos de convidá-lo (a) a participar da pesquisa **“A interoperabilidade semântica na perspectiva da organização do conhecimento: uma proposta para o Repositório Institucional da Universidade Federal do Espírito Santo”**, realizada na Universidade Estadual de Londrina. O objetivo da pesquisa é “investigar as práticas na criação dos metadados e na utilização dos sistemas de organização do conhecimento em Repositórios Institucionais com vistas a fornecer subsídios para a potencialização da interoperabilidade semântica”. A sua participação é muito importante, pedimos que responda as perguntas do questionário no que diz respeito aos procedimentos adotados para a representação do conhecimento do seu Repositório Institucional. Gostaríamos de esclarecer que sua participação é totalmente voluntária, podendo você: recusar-se a participar, ou mesmo desistir a qualquer momento sem que isto acarrete qualquer ônus ou prejuízo à sua pessoa. Informamos ainda que as informações serão utilizadas somente para os fins desta pesquisa e serão tratadas com sigilo e confidencialidade, de modo a preservar a sua identidade.

A otimização da interoperabilidade semântica entre repositórios institucionais e demais sistemas de informações baseada nos princípios da organização do conhecimento reflete-se: no aumento do fator de impacto dos resultados de pesquisa por propiciar maior precisão e relevância nos processos de recuperação da informação; na socialização do conhecimento em nível global; na preservação e reuso da informação.

Caso você tenha dúvidas ou necessite de maiores esclarecimentos pode nos contactar Morgana Carneiro de Andrade, Rua Dr. Guilherme Serrano, n. 119 – apto. 2.403 – Barro Vermelho, Vitória, Espírito Santo – Email

morganaandrade@gmail.com), ou procurar o Comitê de Ética em Pesquisa Envolvendo Seres Humanos da Universidade Estadual de Londrina, na Avenida Robert Kock, nº 60, ou no telefone 33712490.

Li e declaro que fui devidamente esclarecido(a) e concordo em participar VOLUNTARIAMENTE da pesquisa coordenada pela mestrande Morgana Carneiro de Andrade.

☐ sim

☐ não

Atenciosamente,

Morgana Carneiro de Andrade

Mestranda do Programa de Pós-Graduação em Gestão da Informação

Departamento de Ciência da Informação

Centro de Educação, Comunicação e Arte

Universidade Estadual de Londrina

Email: morganaandrade@gmail.com

Skype: morganaandrade

ANEXO

ANEXO A APROVAÇÃO COMITÊ DE ÉTICA

Plataforma Brasil - Ministério da Saúde

Universidade Estadual de Londrina - UEL/ Hospital Regional do Norte do Paraná

PROJETO DE PESQUISA

Título: A interoperabilidade semântica na perspectiva da Organização do Conhecimento: uma proposta para o repositório institucional da Universidade Federal do Espírito Santo

Pesquisador: Morgana Carneiro de Andrade

Versão: 1

Instituição: UNIVERSIDADE ESTADUAL DE LONDRINA
(HOSPITAL UNIVER REGIONAL DO NORTE
PARANA))

CAAE: 02792612.9.0000.5231

PARECER CONSUBSTANCIADO DO CEP

Número do Parecer: 22511

Data da Relatoria: 14/05/2012

Apresentação do Projeto:

O projeto apresentado visa analisar os metadados descritivos e qualificadores utilizados para a organização do conhecimento com vistas a fornecer subsídios para a potencialização da interoperabilidade semântica dos repositórios institucionais. Trata-se de um estudo descritivo, com abordagem qualitativa. O método adotado é o estudo de casos múltiplos em que será utilizada a pesquisa documental. Serão empregados como instrumento de coleta de dados o roteiro de pesquisa e a planilha de dados, para identificação e coleta das informações públicas disponibilizadas nos sites dos repositórios, e o questionário estruturado eletrônico como instrumento auxiliar na complementação dessas informações. A amostra consta de 38 repositórios institucionais implantados por universidades/institutos de pesquisa nacionais com o incentivo do IbiCT/Finep.

Objetivo da Pesquisa:

Objetivo Geral: Analisar os metadados descritivos e qualificadores utilizados para a organização do conhecimento com vistas a fornecer subsídios para a potencialização da interoperabilidade semântica dos repositórios institucionais.

Objetivos específicos: 1) identificar metadados descritivos para a organização do conhecimento nos repositórios institucionais; 2) detectar elementos da representação temática que favoreçam a interoperabilidade semântica; 3) mapear instrumentos da representação temática que propiciem a interoperabilidade semântica; e 4) comparar os valores dos elementos de metadados utilizados para a representação temática identificados nos repositórios institucionais analisados.

Avaliação dos Riscos e Benefícios:

Riscos não há e como benefícios a autora relata que espera contribuir com a otimização da interoperabilidade semântica entre repositórios institucionais e demais sistemas de informações refletindo: no aumento do fator de impacto dos resultados de pesquisa por propiciar maior precisão e relevância nos processos de recuperação da informação; na socialização do conhecimento em nível global; na preservação e reúso da informação.

Comentários e Considerações sobre a Pesquisa:

O projeto de pesquisa justifica-se pela importância dos repositórios institucional, já que este apresenta-se como uma perspectiva de reunir, organizar, preservar, divulgar e garantir o acesso confiável e permanente da produção intelectual gerada na universidade em um único local. A busca de práticas que potencializem a interoperabilidade semântica nestes repositórios permitirá que eles atinjam um maior fator de impacto dos resultados de pesquisa por propiciar maior precisão e relevância nos processos de recuperação da informação, proporcionando a socialização do conhecimento em nível global.

Considerações sobre os Termos de apresentação obrigatória:

O projeto de pesquisa apresenta o TCLE em acordo com a resolução 196/96, apresenta o instrumento de coleta de dados, orçamento e financiamento e cronograma adequados.

Recomendações:

Não há recomendações.

Conclusões ou Pendências e Lista de Inadequações:

O projeto de pesquisa está aprovado.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

Projeto aprovado pelo colegiado

LONDRINA, 16 de Maio de 2012

Assinado por:
Alexandrina Aparecida Maciel Cardelli