

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO**

**BIBLIOTECA DIGITAL: UMA PROPOSTA PARA
PUBLICAÇÃO E DISSEMINAÇÃO DO
CONHECIMENTO PRODUZIDO ATRAVÉS DAS
TESES E DISSERTAÇÕES**

DISSERTAÇÃO DE MESTRADO

Juçara Salete Gubiani

**Santa Maria, RS, Brasil
2005**

**BIBLIOTECA DIGITAL: UMA PROPOSTA PARA
PUBLICAÇÃO E DISSEMINAÇÃO DO CONHECIMENTO
PRODUZIDO ATRAVÉS DAS TESES E DISSERTAÇÕES**

por

Juçara Salete Gubiani

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de
Produção, Área de Tecnologia da Informação, do Centro de Tecnologia, da
Universidade Federal de Santa Maria (RS), como requisito parcial para obtenção
do grau de
Mestre em Engenharia de Produção

Orientador: Dr. Prof. Rafael Port da Rocha

**Santa Maria, RS, Brasil
2005**

**Universidade Federal de Santa Maria
Centro de Tecnologia
Programa de Pós-Graduação em Engenharia de Produção**

A Comissão Organizadora, abaixo assinada,
aprova a Dissertação

**BIBLIOTECA DIGITAL: UMA PROPOSTA PARA
PUBLICAÇÃO E DISSEMINAÇÃO DO CONHECIMENTO
PRODUZIDO ATRAVÉS DAS TESES E DISSERTAÇÕES**

Elaborada por
Juçara Salete Gubiani

Como requisito parcial para a obtenção do grau de
Mestre em Engenharia de Produção

COMISSÃO EXAMINADORA

Prof. Rafael Port da Rocha, Dr.
(Presidente e Orientador - UFRGS)

Prof. João Hélio Righi de Oliveira, Dr. (UFSM)

Prof. Luis Felipe Dias Lopes, Dr. (UFSM)

Santa Maria, RS, 14 de Julho de 2005

AGRADECIMENTOS

Ao professor, amigo e orientador Rafael Port da Rocha, pelo incentivo, amizade, competência, paciência, compreensão e dedicação na condução do desenvolvimento e realização desta dissertação.

Aos professores João Hélio Righi de Oliveira e Luis Felipe Dias Lopes, pela contribuição, capacidade, motivação e amizade.

Aos demais professores e colegas do PPGEF pela amizade, apoio e colaboração.

A Universidade Federal de Santa Maria que como Instituição pública cumpre seu papel na sociedade formando profissionais com qualidade para um mercado de trabalho cada vez mais especializado.

Aos meus amigos que acima de tudo sabem o verdadeiro valor da palavra amizade.

E finalmente a minha família que soube compreender todas as vezes que fui ausente em nome do estudo e da profissão.

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia de Produção
Universidade Federal de Santa Maria

BIBLIOTECA DIGITAL: UMA PROPOSTA PARA PUBLICAÇÃO E DISSEMINAÇÃO DO CONHECIMENTO PRODUZIDO ATRAVÉS DAS TESES E DISSERTAÇÕES

Autor: Juçara Salete Gubiani
Orientador: Rafael Port da Rocha
Local e Data da Defesa: Santa Maria, 14 de Julho de 2005

Com o surgimento da *World Wide Web (Web)*, houve uma grande mudança na forma de comunicação entre as pessoas e as organizações. A ligação entre documentos que podem estar situados em qualquer lugar na rede de computadores, através da *Internet* e do conceito de *Universal Resource Locator*, possibilitou a recuperação rápida de informações, antes materializadas em ambientes de acesso restrito. Publicar na *Internet* passou a ser uma tarefa simples, porém era preciso buscar qualidade nas publicações e um portal único de acesso e busca de informações relevantes. Surgem propostas com padrões para publicação propiciando a interoperabilidade de informações. Neste contexto temos as bibliotecas digitais oferecendo condições para a formação de uma grande rede de dados e conhecimento. Este trabalho tem como objetivo analisar tecnologias para publicação na *Web* e sugerir uma proposta para publicar o conhecimento produzido através das Teses e Dissertações dos Programas de Pós-Graduação da Universidade Federal de Santa Maria (UFSM), bem como medir a aceitabilidade na comunidade acadêmica da nova forma de divulgação e disseminação do conhecimento. A ferramenta proposta deve usar padrões e tecnologias que ofereçam total integração com as demais bibliotecas do país e fazer parte do projeto da Biblioteca Digital de Teses e Dissertações (BDTD), bem como estar ligada com a *Networked Digital Library of Theses and Dissertations* da *Virginia Tech University*. Após análise de tecnologias, protocolos de publicações e a visão da comunidade acadêmica, o sistema TEDE desenvolvido pelo Instituto de Ciência e Tecnologia (IBICT), foi considerado como solução para publicação. Sua escolha de deu, em função do mesmo incorporar tecnologias com protocolos capazes de coletar e possibilitar a transferência de informações garantindo a interoperabilidade necessária entre bibliotecas e por conseqüência a disseminação do conhecimento produzido pelos programas de pós-graduação através da publicação na *Web* das teses e dissertações.

Palavras-chave: bibliotecas digitais; publicações eletrônicas; arquivos abertos; interoperabilidade; metadados; tecnologia da informação

ABSTRACT

Master Degree Dissertation
Pos-Graduate Program in Production Engineering
Federal University of Santa Maria

BIBLIOTECA DIGITAL: UMA PROPOSTA PARA PUBLICAÇÃO E DISSEMINAÇÃO DO CONHECIMENTO PRODUZIDO ATRAVÉS DAS TESES E DISSERTAÇÕES

(DIGITAL LIBRARY: A PROPOSAL FOR PUBLICATION AND DISSEMINATION OF
THE KNOWLEDGE PRODUCED THROUGH THESES AND DISSERTATIONS)

Author: Juçara Salete Gubiani

Adviser: Rafael Port da Rocha

Place and Date of the Defense: Santa Maria, July 14th, 2005

With the coming of the World Wide Web (Web), there has been a great change in the form of communication between the people and the organizations. The link among documents that can be situated in any place on the computer network, through the Internet and the concept of Universal Resource Locator, made the fast recovery of information possible, before materialized in environments of restricted access. Publishing on Internet has become a simple task, however it was necessary to search for quality in publications and an only vestibule of access and search of excellent information. Proposals with standards for publication have appeared propitiating the interoperabilidade of information. In this context we have the digital libraries offering conditions to make a great net of data and knowledge. This work has the aim to analyze technologies for publication on the Web and to suggest a proposal to publish the knowledge produced through the Theses and Dissertation of the Programs of Pos-Graduation of the Federal University of Santa Maria (UFSM), as well as measuring the acceptability in the academic community of the new form of spreading and dissemination of the knowledge. The proposal tool must use standards and technologies that offer total integration with the other libraries in the country and be part of the project of the Digital Library of Theses and Dissertations (BDTD), as well as being linked with the Digital Networked Library of Theses and Dissertations of Virginia Tech University. After analysis of technologies, publication protocols and the view of the academic community, the system TEDE developed by the Institute of Science and Technology (IBICT), was considered as solution for publication. It was chosen because it incorporates technologies with protocols capable to collect and to make the transference of information possible being guaranteed the necessary interoperabilidade among libraries and for consequence the dissemination of the knowledge produced by the programs of pos-graduation through the publication of the theses and dissertations on the Web.

Keywords: *digital libraries; electronic publications; open archives; interoperability; metadata; technology of the information.*

LISTA DE FIGURAS

Figura 1 – Número de novas submissões por mês no <i>arXiv</i> desde 1991 até abril/2005.	40
Figura 2 – Interoperabilidade Z39.50 X OAI-PMH.....	44
Figura 3 – Colheita de metadados / <i>Harvesting</i>	45
Figura 4 – Iteração entre <i>Data Provider</i> e <i>Service Provider</i>	46
Figura 5 – Arquitetura de OAI do NDLTD.....	56
Figura 6 – Sistema MARIAN em uma rede semântica de recuperação de ETD-MS	57
Figura 7 – Estratégia de interoperabilidade da BDB – Busca distribuída.....	61
Figura 8 – Estratégia de interoperabilidade da BDB – Coleta Automática.....	62
Figura 9 – <i>Link</i> das Instituições de ensino com o IBICT e com o NDLTD	63
Figura 11 – <i>Interface</i> padrão do TEDE	106
Figura 12 – Submissão da T&D	107
Figura 13 – Pesquisa no Portal da DBTD	110

LISTA DE TABELAS

Tabela 1 – Metadados já importados na BDTD.	64
Tabela 2 – Metadados coletados por <i>Harvesting</i> na BDTD.....	64
Tabela 3 – Consulta realizada entre março 2004 e março 2005.....	103
Tabela 4 – Comparativo entre instituições em relação à bibliotecas digitais de T&D's.....	105

LISTA DE ACRÔNIMOS

ARC	<i>Cross Archive Searching Service</i>
ARL	<i>Association of Research Libraries</i>
BDB	Biblioteca digital Brasileira
BTD	Banco de Teses e Dissertações
BDTD	Biblioteca digital Brasileira de Teses e Dissertações
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CERN	Centro Europeu de Pesquisas Nucleares em Genebra
CLIR	<i>Council on Library and Information Resources</i>
CGS	<i>Council of Graduate Schools</i>
CNI	<i>Coalition for Networked Information</i>
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CTC	Comitê Técnico Consultivo
DC	<i>Dublin Core</i>
DLF	<i>Digital Library Federation</i>
DMCI	<i>Dublin Core Metadata Initiative</i>
DTD	<i>Document Type Definition</i>
EBNF	<i>Extended Backus-Naur Form</i>
ETD	<i>Electronic Thesis and Dissertation</i>
ETD-MS	<i>Electronic Thesis and Dissertation Metadata Standard</i>
FINEP	Financiadora de Estudos e Projetos
FGDC	Federal Geographic Data Committee
HTML	<i>Hypertext Markup Language</i>
IES	Instituições de Ensino Superior
ISO	<i>International Organization for Standardization</i>
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
LANL	<i>Los Alamos National Laboratory</i>

MCT	Ministério da Ciência e Tecnologia
MEC	Ministério da Educação e Cultura
NCSTR	<i>Networked Computer Science Technical Reference Library</i>
NDLTD	<i>Networked Digital Library of Theses and Dissertation</i>
INPI	Instituto Nacional de Propriedade Industrial
OAI	<i>Open Archives Initiative</i>
OAI-PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
OAMS	<i>Open Archives Metadata Set</i>
OPAC	<i>Open Public Access Catalog</i>
RDF	<i>Resource Description Framework</i>
SESU	Secretaria de Ensino Superior
SIBi	Sistema Integrado de Bibliotecas da USP
SIE	Sistema de Informações para o Ensino
SGDB	Sistema de Gerência de Banco de Dados
SGML	<i>Standard Generalized Markup Language</i>
SPARC	<i>Scholarly Publishing & Academic Resources Coalition</i>
SURA	<i>Southeastern Universities Research Association</i>
TEDE	Sistema de Publicação Eletrônica de Teses e Dissertações
TDE	Tese e Dissertação Eletrônica
T&D	Teses e Dissertações
UDK	<i>Umwelt-Datenkatalog</i>
URL	<i>Universal Resource Locator</i>
WEB	<i>World Wide Web</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

RESUMO.....	3
ABSTRACT	4
LISTA DE FIGURAS	5
LISTA DE TABELAS.....	6
LISTA DE ACRÔNIMOS.....	7
1 INTRODUÇÃO	12
1.1 Problema de pesquisa.....	14
1.2 Objetivos	15
1.2.1 Objetivo geral	15
1.2.1.1 Objetivos específicos	15
1.3 Justificativa	16
1.4 Delimitação do tema	17
1.5 Método e técnica	17
1.6 Estrutura do trabalho.....	17
2 PRODUÇÃO DO CONHECIMENTO.....	19
2.1 Publicação do conhecimento.....	19
2.2 Documentos e arquivos digitais	22
2.3 Cooperação entre sistemas de informação na era dos arquivos digitais	25
3 PUBLICAÇÃO NA WEB.....	27
3.1 Tecnologias para publicação na <i>Web</i>	27
3.1.1 Metadados.....	28
3.1.1.1 Vocabulário <i>Dublin Core</i>	29
3.1.1.2 Representação de metadados através de linguagens de marcas.....	31
3.1.1.2.1 SGML – Standard Generalized Markup Language.....	32

3.1.1.2.2	HTML – Hypertext Markup Language	33
3.1.1.2.3	XML – Extensible Markup Language	34
4	ARQUIVOS ABERTOS	36
4.1	A Publicação Científica na <i>Web</i>	36
4.2	O <i>software</i> arXiv	38
4.3	<i>Open Archives Initiative</i>	40
4.4	Interoperabilidade entre Documentos Digitais	42
4.5	Protocolo OAI-PMH	43
4.6	Interoperabilidade	46
4.6.1	Interoperabilidade Semântica	47
4.6.2	Interoperabilidade “Linkagem”	47
4.6.3	Serviço de Busca a Repositório de Arquivos Abertos.....	48
5	BIBLIOTECA DIGITAL DE TESES E DISSERTAÇÕES – BDTD	50
5.1	NDLTD – <i>A Global Digital Library Of Electronic Theses And Dissertations</i>	52
5.1.1	Construindo uma biblioteca digital.....	53
5.1.2	NDLTD como uma federação de bibliotecas digitais.....	55
5.1.3	O Sistema MARIAN de armazenamento e pesquisa.....	56
5.2	Projetos institucionais no Brasil de bibliotecas digitais de T&D’s	57
5.2.1	Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).....	58
5.2.2	O sistema TEDE	65
5.2.2.1	Procedimento e uso do sistema TEDE.....	67
5.2.3	A biblioteca digital de T&D’s da Universidade de São Paulo	69
5.2.4	Biblioteca digital de T&D’s do PPGEP/UFSC	72
5.2.5	Líber – Universidade Federal de Pernambuco.....	74
5.2.6	Biblioteca digital da UNICAMP	75
5.2.7	Biblioteca digital do IMPA.....	76
5.3	Projetos de Instituições Internacionais.....	77
6	METODOLOGIA	79
6.1	Estrutura metodológica da pesquisa.....	80
6.1.1	Em relação ao objetivo proposto e a finalidade da pesquisa	80
6.1.2	Em relação à forma de abordagem	81
6.1.3	Em relação à natureza da pesquisa	82
6.1.4	Em relação aos procedimentos adotados	82
6.2	O Delineamento da pesquisa.....	84

6.3	A Descrição sucinta das atividades.....	85
6.3.1	A formação do referencial bibliográfico.....	85
6.3.2	A percepção do problema e a definição do método proposto para a análise.....	86
6.3.3	A escolha da atividade produtiva e a estratégia de ação.....	87
6.3.4	A Entrevista e os participantes do universo da pesquisa.....	87
6.3.5	Participantes do universo da pesquisa.....	88
6.3.6	A análise documental e análise dos dados.....	88
7	ANÁLISE DAS ENTREVISTAS.....	89
7.1	Análise das respostas.....	90
8	PROPOSTA PARA PUBLICAÇÃO DE T&D's NA UFSM.....	102
8.1	Implantando o TEDE Modular.....	105
8.1.1	Papel da pós-graduação.....	107
8.1.2	Papel do autor.....	108
8.1.3	Papel da biblioteca.....	108
8.1.4	Publicação, colheita de metadados e consulta.....	109
8.1.5	Infra-estrutura exigida para o <i>software</i>	110
8.1.5.1	Treinamento para uso do <i>software</i>	110
8.1.5.2	Sensibilização da comunidade universitária.....	111
8.1.5.3	Ambiente operacional.....	111
8.1.5.4	Grupos de trabalho.....	112
8.1.5.5	Projeto-piloto.....	112
9	CONCLUSÃO.....	114
	REFERÊNCIAS.....	120

1 INTRODUÇÃO

Historicamente o desenvolvimento de uma sociedade acontece quando o país obtém conhecimentos, recursos e tecnologias suficientes para promover investimentos em todas as áreas da economia. Neste sentido, existe uma relação direta entre desenvolvimento econômico e social com o desenvolvimento da ciência e da tecnologia (C&T). À ciência e a tecnologia está reservado um papel fundamental no desenvolvimento da sociedade. Para tal, a disponibilidade da informação é de fundamental importância para a promoção e o crescimento da pesquisa no país (MARCONDES & SAYÃO, 2001).

A busca e o uso integrado das tecnologias de comunicação, de computação e de conteúdos em formato digital, cujo fator determinante hoje é a *Internet*, tem contribuído nos últimos anos para criar um novo ambiente de acesso, disseminação, cooperação e promoção do conhecimento em uma escala global. Muitas questões relativas a essa nova forma de armazenamento e disseminação do conhecimento, que não guardam similares como os tradicionais materiais impressos em papel, mas que a cada dia promovem um processo evolutivo de disseminação do conhecimento, ainda não foi objeto de uma completa avaliação pelos profissionais responsáveis pela catalogação e distribuição da informação.

A *Internet* é atualmente o meio de socialização do conhecimento mais utilizado, propiciando facilidades que extrapolam o conceito tradicional de informação bibliográfica baseada em documentos, como artigos e periódico, trabalhos em congressos, teses etc. Além das tradicionais formas de publicação, atualmente as tecnologias de informação oferecem a comunidade de pesquisa novos recursos para publicação na *World Wide Web (Web)*, como documentos multimídia, listas de discussão, fóruns eletrônicos, conferências em on-line, imagens (de satélites, de microscópios, em tempo real), modelos animados, bancos de arquivos eletrônicos de *e-prints*¹. Estes recursos servem de subsídio à pesquisa melhorando a comunicação e cooperação entre pesquisadores, diminuindo o tempo de publicação dos resultados e principalmente garantindo a originalidade intelectual.

¹ Arquivos do tipo *e-print* são textos em formato digital que correspondem a artigos de pesquisa revisados pelos pares.

As ferramentas disponíveis para publicação na *Internet* são capazes de abrir novas possibilidades intelectuais, propiciando recursos que extrapolam em muito aqueles oferecidos por documentos em papel, de leitura linear. Para muitos autores, a *Internet* representa, neste sentido, uma mudança de paradigma comparável à invenção da imprensa por Gutemberg. Essa mudança de paradigma se faz sentir também no aspecto da comunicação científica. A *Internet* tornou-se um mecanismo de comunicação de alcance mundial, imediato, interativo e multidirecional, ou seja, qualquer pessoa pode publicar nela. A publicação é automaticamente acessível, o autor pode receber um retorno imediato sobre o que publicou em qualquer lugar e de qualquer lugar. Um pesquisador espera a máxima divulgação de seus trabalhos, para que os resultados de sua pesquisa tenham o maior impacto possível sobre as pesquisas de seus pares e sobre outras publicações que possam advir resultante de seu trabalho.

Recentes estudos confirmam que as publicações eletrônicas são muito mais citadas que as publicações em papel: "*The mean number of citations to offline articles is 2,74, and the mean number of citations of online articles is 7,03, an increase of 157%*"², (LAWRENCE, 2001 *apud* MARCONDES & SAYÃO, 2001). A informação sobre os percentuais de publicação na *Web* mostrado acima e supondo um crescimento gradual após 2001, entende-se que para o desenvolvimento e maturidade da pesquisa no Brasil, é necessário, recursos para o desenvolvimento de mecanismos de publicação eletrônica e a sua disponibilidade para a comunidade de pesquisa.

A publicação na *Web*, tornou-se a forma mais rápida de divulgar os resultados das pesquisas, a comunicação científica, que no passado recente, tinham por base o periódico científico em papel, com seu esquema de revisão por pares, juntamente com o monopólio das grandes editoras científicas na publicação do resultado final, sofre um grande impacto frente à *Internet*. A técnica de revisão por pares, onde pesquisadores trocam idéias sobre um determinado assunto, em um sistema tradicional de comunicação demanda um tempo maior. Associado a isso, o controle da publicação nas mãos das grandes editoras, elevando o custo, fez com que a comunidade científica se mobilizasse na busca de uma alternativa. Dentro da comunidade de física, é criado o sistema *arXiv*, um repositório de documentos de arquivos *e-prints* baseado nos padrões dos arquivos abertos e com interoperabilidade das informações publicadas. A partir deste repositório, a comunidade científica internacional passa a oferecer uma alternativa prática e a um custo muito menor para a publicação do resultado de suas

² O número médio das citações aos artigos impressos é 2,74, e o número médio das citações de artigos digitais é 7,03, um aumento de 157%.

pesquisas. Junto com os arquivos automatizados de *e-prints*, surge a iniciativa chamada *Open Archives Initiative* definindo padrões para publicação na *Web*.

Essas transformações têm exercido profunda influência na concepção e funcionamento dos sistemas de informação automatizados, especialmente aqueles voltados para as atividades de pesquisa. A popularização da *Internet*, a partir de 1990, permitiu o surgimento de uma nova realidade para esses sistemas, antes orientados basicamente para recuperação de referências bibliográficas catalogas em bases de dados isoladas e textos em papel, voltam-se hoje para a recuperação distribuída de objetos digitais, textos completos, imagens em movimento, som etc., estabelecendo como palavras de ordem a publicação na *Internet* e a busca constante pela interoperabilidade entre fontes de informação heterogêneas e globalmente distribuídas.

No Brasil, algumas instituições de ensino e pesquisa, com a intenção de fomentar e fornecer meios para que a comunidade brasileira de pesquisa possa publicar sua produção acadêmica de teses e dissertações (T&D) diretamente na *Web*, estão agregando esforços em direção ao desenvolvimento e uso de bibliotecas digitais com padrões abertos e tecnologias que propicie a interoperabilidade das informações publicadas nesses ambientes independente do local e instituição depositária da informação. Essa iniciativa possibilita a criação de um ambiente colaborativo, onde o pesquisador pode interagir com outros pesquisadores sem nenhuma interferência e publicar seus trabalhos sem dificuldades diretamente na rede aumentando a visibilidade da pesquisa, otimizando o fluxo de comunicação científica e reduzindo o ciclo de geração de novos conhecimentos.

1.1 Problema de pesquisa

No sentido disseminar o resultado de suas pesquisa, o Programa de Pós-Graduação em Engenharia de Produção considera de extrema importância para a Universidade Federal de Santa Maria e sua comunidade universitária, a existência de um ambiente onde os programas de pós-graduação possam publicar as T&D's produzidas pelos alunos. Esse ambiente requer a implantação de uma ferramenta de *software* com interface *Web* onde todos os resultados das pesquisas passam a ficar totalmente transparente. Para que o processo tenha sucesso e aceitação dentro da comunidade acadêmica, a seguinte questão foi levantada neste trabalho: **Qual a melhor forma de disseminar o conhecimento produzido através das T&D's desenvolvidas nos programas de pós-graduação no contexto da UFSM?**

Para responder a esse questionamento os seguintes aspectos precisam ser considerados:

- a) As tecnologias de comunicação, atualmente, propiciam a publicação e disseminação do conhecimento produzido através das T&D's, na *Web*?
- b) Existe uma tecnologia padrão e portanto recomendada para publicar e disponibilizar T&D's com segurança na *Web* e que ofereça interoperabilidade das informações?
- c) As pessoas usam T&D's como base de conhecimento para realização de seus trabalhos e ou pesquisas?
- d) Como é tratado o direito autoral sobre o trabalho produzido? Ele pode ser um impedimento para a publicação da T&D na *Web* caso o autor não repasse o direito para a pós-graduação?
- e) Qual a visão das coordenações dos programas de pós-graduação em relação à disseminação do conhecimento produzido através das T&D's na *Web*?

1.2 Objetivos

Considerando o problema definido anteriormente, foram formulados os seguintes objetivos

1.2.1 Objetivo geral

Analisar tecnologias disponíveis, aspectos legais em relação ao direito autoral bem como a aceitação da comunidade universitária em relação à questão cultural, social e as praticas relacionadas ao assunto abordado para elaboração de uma proposta de implantação de uma biblioteca digital visando à publicação e disseminação do conhecimento gerado através das T&D's dos programas de pós-graduação da UFSM.

1.2.1.1 Objetivos específicos

Como objetivos específicos foram considerados os seguintes itens:

- a) Analisar o atual uso, como base de conhecimento, de publicações que estão disponibilizadas de forma tradicional nas bibliotecas da UFSM;
- b) Estudar e analisar tecnologias disponíveis para publicação e disseminação do conhecimento de T&D's na *Web*;
- c) Analisar a viabilidade da implantação de uma biblioteca digital de T&D's nos programas de pós-graduação no sentido de oferecer à comunidade acadêmica um ambiente para submissão, busca e recuperação de textos completos em meio digital que utilize a metodologia dos arquivos abertos oferecendo um padrão de interoperabilidade entre bibliotecas a nível nacional e internacional;
- d) Verificar, através de entrevistas nos programas de pós-graduação dentro da comunidade da UFSM, a aceitação de uma biblioteca digital para publicar e disseminar os resultados de suas T&D's.

1.3 Justificativa

A produção científica e a comunicação dos resultados para a sociedade acontece de forma dispersa e muitas vezes em um tempo longo demais considerando o avanço tecnológico. Pesquisadores, em centros de pesquisas e universidades, produzem e comunicam seus resultados de pesquisa através de diferentes mecanismos de comunicação, de formatos, de idiomas e em momentos diversos.

Atualmente existem tecnologias e padrões que permitem a interoperabilidade dessas informações tanto a nível nacional quanto internacional. Nesse sentido as bibliotecas digitais oferecem mecanismos para a disponibilidade de trabalhos completos na *Web*. Em uma universidade, um mestre ou doutor em formação tem nas bibliotecas digitais um recurso inestimável e de rápido acesso. Enquanto nas bibliotecas tradicionais o pesquisador se defronta com limitações de tempo, espaço e outros recursos, as bibliotecas digitais oferecem baixo custo e facilidades de pesquisa aumentando a produtividade através da diminuição do tempo de produção e da divulgação dos resultados.

Considerando, que a disseminação do conhecimento é um fator preponderante para o desenvolvimento de uma sociedade, este trabalho, através do estudo de tecnologias para publicação na *Web* e análise de projetos já implantados de bibliotecas digitais em outras instituições de ensino, proporcionará uma proposta de um modelo de biblioteca digital para publicar as T&D's produzidas pelos programas de pós-graduação da UFSM.

1.4 Delimitação do tema

No trabalho, são analisadas tecnologias para publicação de T&D's na *Web* que possibilitem a interoperabilidade de informação entre bibliotecas digitais promovendo a cooperação na comunidade acadêmica através de um serviço de comunicação inovador baseado em padrões abertos, com ênfase na iniciativa dos arquivos abertos que define um modelo para publicação na *Web*. A ferramenta para tal serviço deve oferecer interatividade entre o aluno, o programa de pós-graduação e a biblioteca depositária das T&D's. Deve também ser de fácil compreensão e oferecer total gerenciamento ao programa de pós-graduação, o qual, a partir da submissão da T&D possa controlar o fluxo de tramitação da pesquisa até o fechamento e envio definitivo à biblioteca depositária do documento digital. A partir da publicação a comunidade deve ter total acesso para consultas.

A revisão da literatura, a análise de ferramentas de publicação na *Web*, buscou viabilizar uma proposta de implantação nos programas de pós-graduação da Universidade Federal de Santa Maria, de uma biblioteca digital com o objetivo de publicar e disseminar as pesquisas produzidas pelos alunos através das T&D's. Num primeiro momento será considerado o Programa de Pós-Graduação em Engenharia da Produção (PPGEP) como objeto de ensaio e teste da ferramenta de publicação em um ambiente experimental.

1.5 Método e técnica

Para a elaboração deste trabalho foram analisados conceitos teóricos e opiniões de diversos autores sobre o tema. Empregou-se a abordagem qualitativa, que oferece ao pesquisador informações de natureza mais subjetiva e latente, implicando não só numa análise do discurso do entrevistado, como também em sua postura mais global, diante das questões que lhe são colocadas.

1.6 Estrutura do trabalho

Considerando os objetivos definidos na introdução, o trabalho está estruturado em 9 capítulos a saber:

Este capítulo abordou o tema do trabalho e sua organização. A definição e a natureza do problema em estudo, os objetivos da pesquisa e a relevância do estudo para o segmento pesquisado, métodos e técnicas utilizados na elaboração do trabalho bem como a delimitação do estudo.

No segundo capítulo, encontram-se a revisão sobre a produção do conhecimento que teve suas origens no Renascimento acentuando-se fortemente na Revolução Industrial até o advento da *Internet*.

Na seqüência, no terceiro capítulo, são apresentadas as tecnologias existentes para publicação na *Web* e as novas formas de disseminação do conhecimento, tanto nas comunidades de pesquisa quanto no público em geral.

No quarto capítulo discute-se tecnologias para *Web* que auxiliam o processo de publicação científica e disseminação de informação através dos arquivos abertos compreendendo temas como *e-prints*, auto-arquivamento, revisão por pares, vocabulários controlados entre outros.

O quinto capítulo é apresentado o objeto da pesquisa que trata de bibliotecas de teses e dissertações (BDT) e as tecnologias para divulgação desse conhecimento na *Web* através do uso de bibliotecas digitais. Nesta seção também é analisado o estado da arte em relação às publicações de T&D's em nível nacional e internacional descrevendo algumas iniciativas institucionais.

A metodologia usada para a elaboração da pesquisa é apresentada no sexto capítulo.

Os resultados obtidos a partir das entrevistas realizadas nos programas de pós-graduação são discutidos no sétimo capítulo.

No oitavo capítulo, é apresentada uma proposta para publicação de T&D's na UFSM e a partir do objetivo geral e das discussões do tema durante a elaboração do trabalho. No nono capítulo encontra-se a conclusão do trabalho, considerando os objetivos propostos na introdução, o que efetivamente foi realizado durante a execução do trabalho e as recomendações para trabalhos futuros.

2 PRODUÇÃO DO CONHECIMENTO

A sociedade, hoje, contextualizada em uma economia globalizada, historicamente vem passando por transformações ao longo do tempo. Estudos e pesquisas concluem que o desenvolvimento econômico e social de uma nação está baseado em três fatores básicos: o trabalho, o capital e o conhecimento. A importância relativa entre eles, ao longo dos anos, vem se deslocando do primeiro, passando pelo segundo e se concentrando, nos últimos tempos, no terceiro, onde o conhecimento é apoiado em uma infra-estrutura de informações amplas e instantâneas. Os avanços das tecnologias da informação reforçam as mudanças econômicas e sociais que promovem uma revolução nos negócios e na sociedade. A nova economia baseada na informação está emergindo onde o comércio e o investimento são globalizados e empresas competem com o conhecimento, redes eletrônicas e agilidade na tomada de decisão nos negócios.

Neste capítulo é abordado a publicação do conhecimento na sua forma tradicional através da publicação em papel e a sua disponibilidade em bibliotecas tradicionais, como também, em meio digital, a partir do surgimento de tecnologias de informação capazes de promover a publicação e a disseminação do conhecimento através da *Internet*.

2.1 Publicação do conhecimento

A partir da revolução industrial, o conhecimento passou a ter um papel fundamental como insumo produtivo, como um bem capital dentro das empresas e organizações em um processo que teve suas origens no renascimento. Sua gestão, a utilização e principalmente o acesso se tornam cada vez mais atividades críticas em um sistema de economia globalizada.

Segundo Marcondes & Sayão (2001a), historicamente, a forma de disponibilizar o conhecimento é através da sua publicação em papel e posterior exposição para acesso público em bibliotecas tradicionais. Porém, a preservação desse material em papel é uma constante preocupação de bibliotecários, administradores de universidades e centros de pesquisa. Como ficará no futuro o conhecimento hoje armazenado em papel? Atualmente, já temos coleções

deteriorando-se em função de ambientes sem condições necessárias para acomodação do material impresso. A cada dia novas publicações são adquiridas e a exigência de mais espaço é uma realidade. A saída certamente passará por grandes mudanças em relação à infraestrutura de tecnologias e pessoas, provavelmente deverá ocorrer uma reciclagem dos profissionais, novas funções serão atribuídas aos bibliotecários, coleções de valor histórico deverão ser encaminhadas a museus e o resto provavelmente será reciclado. Alguns autores relacionam alguns problemas em relação ao material impresso, como por exemplo:

- Os altos preços de materiais impressos e respectivo armazenamento em estantes e similares;
- contínuo número de itens disponíveis para compra, com o agravo de repetição de conteúdos;
- duplicações de operações catalográficas e de coleções em bibliotecas do mesmo tipo e áreas de conhecimento;
- necessidade contínua de espaço e aquisição de estantes para armazenar materiais e para acomodar usuários;
- redução de orçamentos para aquisição de novas publicações;
- elevação de custos com assinaturas de revistas especializadas;
- despesas mais altas com materiais de encadernação, declínio nas compras.

Além das considerações a respeito da tradicional forma de publicação do conhecimento, é constante a dificuldade de identificar a informação que está sendo publicada, gera-se mais rapidamente informação acadêmica do que a capacidade dos pesquisadores em administrá-la, dos editores em publicá-la, do bibliotecário em coletá-la e dos estudiosos em consultá-la. Grande parte da produção acadêmica ocorre por força de suas necessidades de ascensão à carreira acadêmica e busca de promoções. Repetições de conteúdo, geração de assuntos repetidos e fileiras de estantes abarrotadas de publicações não significa que haja adição de novos conteúdos para benefício da sociedade ou para o aumento de informação.

Existe uma grande quantidade de informações produzidas e disponibilizadas em meio eletrônico nas diferentes atividades sociais, a chamada “explosão informacional”. À medida que essas informações são produzidas e disponibilizadas sem uma gerência em relação aos padrões de publicações, cresce a dificuldade de identificação, acesso e utilização destes conteúdos pelas comunidades de pesquisa.

A partir da metade do século XX, com o surgimento de tecnologias para publicação em meio digital, a saída encontrada pelas bibliotecas foi a cooperação: a publicação das fichas do catálogo da *Library of Congress* dos Estados Unidos data de 1905, o projeto MARC³ é do fim da década de 60. Segundo Marcondes & Sayão (2001a), com o surgimento do computador e o desenvolvimento de tecnologias de informação oferecendo condições para a publicação do conhecimento, as bibliotecas passam a ter a possibilidade de prover acesso não só a documentos dos seus próprios acervos, mas também aos armazenados em acervos de outras bibliotecas.

A partir de 1990, com a melhoria nos sistemas de comunicação e a popularização da *Internet*, as bibliotecas têm publicado documentos diretamente na rede, tornando desta forma o acesso a um documento digital uma questão de conhecer o seu endereço eletrônico, a URL⁴. No entanto, a grande maioria dos documentos publicados na *Web* não é de responsabilidade das bibliotecas, e quase sempre estão fora dos padrões de publicação, desta forma, prejudicando a busca da informação de qualidade na *Internet*. Tem-se facilidade de acesso na *Web*, mas, em contrapartida, uma grande dificuldade de encontrar informações relevantes e que possam ser utilizadas como fonte de pesquisa. São bilhões de páginas publicadas ao longo do tempo, sobre diferentes assuntos, na maioria das vezes sem o uso de qualquer padrão de publicação, de forma caótica e sem a preocupação com a recuperação da informação.

Na *Internet* não existe ordem e controle, a informação é disponibilizada de maneira caótica, diferente de uma biblioteca tradicional, onde a informação segue padrões de publicação usando técnicas de catalogação e vocabulários controlados. Para piorar a situação, a informação é disponibilizada em diferentes idiomas agravando o problema do controle do vocabulário a ser usado na recuperação. Os mecanismos ou motores de busca indexam a *Internet* periodicamente, de forma automática, cegamente, sem compreender o tema de uma publicação em uma página, simplesmente extraíndo palavras do texto em *Hypertext Markup Language* (HTML) da página pesquisada armazenando estas palavras isoladas junto ao endereço da página, em uma base de dados para exibição na consulta. Além disso, a indexação é feita por páginas HTML isoladas, não considerando que diversas páginas estão inter-relacionadas, formando um *site*.

³ MARC - *Machine Readable Cataloging*, padrão internacional para publicação de documentos. O formato utiliza números, letras e sinais gráficos para marcar os campos e informações no registro bibliográfico em bibliotecas.

⁴ *Universal Resource Locator*

Os programas-robôs dos mecanismos de busca só "enxergam" páginas HTML estáticas quando fazem sua rotina de indexação, deixando de considerar grande quantidade de informações sob a forma de registros contidos em bases de dados disponíveis na *Internet* – a chamada *Deep Web* – (BERGMAN, 2001), informações que ficam "invisíveis". Estes registros são acessados somente por meio das interfaces destas bases de dados proprietárias, que pressupõem uma interação entre um usuário humano e a base de dados e, portanto, ficam inacessíveis aos programas-robôs. Segundo Bergman (2001), a *Surface Web* é estimada em cerca de 2,5 bilhões de páginas, enquanto a *Deep Web* seria cerca de 500 vezes maior, dificultando o estabelecimento de filtros de qualidade para as informações encontradas na *Internet*.

2.2 Documentos e arquivos digitais

Atualmente, no mundo globalizado onde a economia de um país é totalmente dependente dos demais países e parceiros no comércio, a produção científica também passou a ser institucionalizada, ou seja, não existe pesquisa isolada e sim uma interdependência entre as comunidades de pesquisa. Para a ciência é fundamental o trabalho coletivo, onde pesquisadores e grupos de pesquisa trabalham sobre resultados já obtidos pelos seus pares. Neste contexto, a disponibilidade de tecnologias de comunicação passa a ser um fator determinante na disseminação do conhecimento. A ciência precisa de mecanismos eficientes de comunicação científica para avançar e cumprir o ciclo da produção de conhecimento. Mecanismos que possam propiciar a melhoria do registro dos resultados, da coleta e estocagem destes registros, a disseminação dos resultados e o reuso em outras pesquisas, tanto em atividades da produção, quanto também como fonte para gerar novos conhecimentos, independente de fronteiras de países.

Até o final do século XX, os periódicos científicos tradicionais eram considerados a melhor forma de publicação e comunicação da comunidade científica. Atualmente a comunidade científica troca idéias através de conversas pessoais, cartas, telefonemas, arquivos *preprints*⁵ distribuídos aos pares, estes últimos bastante ágeis e atualizados em termos de velocidade, mas ainda sem filtros de qualidade. Entretanto, a cooperação entre os pares, usando canais de comunicações formais para a revisão dos trabalhos de congressos e

⁵ Texto de propriedade do autor, ainda não apreciado por um corpo de referência e antes da publicação.

artigos de periódico, acarreta uma maior lentidão da publicação dos resultados de pesquisas porem com um diferencial em relação à melhoria da qualidade do resultado final.

Marcondes & Sayão (2002), destacam que, a partir de 1665, com a publicação do *Journal des Savans* e da *Philosophical Transactions da Royal Society*, o periódico acadêmico vem tendo um papel destacado como veículo por excelência de comunicação científica. Com a instituição do sistema de revisão por pares, a partir do fim do século XIX, o uso de periódicos passou a ser considerado como a melhor forma de produzir pesquisa institucionalizada, juntamente com todo aparato acadêmico e científico, atingindo a configuração atual, onde as comunidades de pesquisa trocam informações e dividem o esforço na produção do conhecimento. No mundo dos documentos impressos, os periódicos científicos tinham papel destacado como coroamento de um sistema de comunicação científica institucionalizada e reconhecido pela comunidade acadêmica, que fazia um compromisso entre velocidade e filtros de qualidade. Dentre os papéis por eles cumpridos, destacam-se os seguintes: disseminação ampla dos resultados de pesquisa; controle de qualidade, através do mecanismo de revisão por pares; homologação de prioridade nas descobertas científicas; reconhecimento dos autores; criação de um arquivo público de conhecimentos com cópias armazenadas em bibliotecas de centros de documentação.

O surgimento de novas áreas de pesquisa e a moderna visão integrada das ciências da vida leva a identificação de pontos de contato entre diferentes ramos da ciência ou a novos problemas de caráter interdisciplinar. Desta forma, a comunicação científica através de periódicos torna-se mais restrita e especializada, levando à diminuição das edições e ao conseqüente aumento do custo final dos periódicos científicos, bem como das edições de anais de eventos para seus usuários. Os custos das assinaturas de periódicos, tornaram-se proibitivos para as comunidades de pesquisa, bibliotecas e centros de documentação, trazendo sérias conseqüências para o desenvolvimento da ciência, especialmente nos países em desenvolvimento, cujos recursos para investimento em ciência e tecnologia são escassos, como é o caso do Brasil.

Com a popularização da *Internet* e dos mecanismos de publicação diretamente na rede, a partir de 1990 a comunidade acadêmica passa a ter uma nova alternativa de publicação, a nível internacional, dos resultados de suas pesquisas. A publicação na *Web*, além da diminuição dos custos propicia a imediata disponibilidade dos resultados, os quais passam automaticamente a servir de bases a outras pesquisas e serem citados por outros trabalhos. A comunidade científica vê as publicações eletrônicas na *Web* como um meio de aumentar sua

visibilidade e qualidade, acelerando o avanço da ciência e disseminando os resultados das pesquisas a nível global. Segundo Stevan Harnard⁶ o resultado das pesquisas é como patrimônio da humanidade. Na busca de uma alternativa, surge dentro da comunidade científica mundial, os arquivos eletrônicos informais e auto-geridos, os arquivos eletrônicos de *e-print*.

Em 1991, no *Los Alamos National Laboratory*, no Novo México, o físico Paul Ginsparg cria o *arXiv*, sistema de submissão e recuperação de dados baseado na filosofia dos arquivos abertos, onde os usuários podem recuperar ou submeter os artigos, tanto por uma *interface on-line* na *Web*, como via correio eletrônico, marcando claramente uma transição do modelo de comunicação tradicional, baseado em publicações periódicas formalmente estabelecidas, para um novo e surpreendente paradigma (GINSPARG, 1996).

No Reino Unido, na Universidade de Southampton, é criado o *software e-print*, o qual foi construído de forma bastante flexível exigindo o mínimo de esforço de configuração e adaptação. A versão genérica do *software e-print* é interoperacional com todos os padrões de arquivos abertos, podendo, independente de onde o artigo estiver arquivado, ser recuperado através de um protocolo padrão.

As condições tecnológicas oferecidas pela *Internet* possibilitaram a troca de informações e a publicação de documentos eletrônicos, mas existem outros fatores que determinaram o surgimento dos arquivos de *e-prints*, podemos relacionar:

- a) a lentidão do ciclo de edição das revistas comparado à rapidez da geração de novos conhecimentos de algumas áreas;
- b) a renúncia ao direito sobre a obra imposto pelas revistas, que impede a ampla disseminação pelo autor de suas idéias;
- c) a perspectiva extremamente rígida e conservadora dos esquemas de revisão por pares, adotados pelas revistas, que não raro são impeditivos ao surgimento de idéias inovadoras;
- d) o alto custo da subscrição dos periódicos, seja em papel ou em meio eletrônico.

Na busca de um padrão para troca de informação entre os arquivos *e-prints* publicados, a comunidade científica internacional se organizou no sentido de torná-los interoperáveis, isto é, passíveis de serem consultados simultaneamente. A iniciativa conhecida como *Open Archive Initiative*, foi criada objetivando a integração de arquivos *e-prints*

⁶ <http://www.nature.com/nature/debates/e-access/Articles/hamad.html>

promovendo a interoperabilidade e intercâmbio automático de metadados entre bibliotecas digitais. Esse assunto será abordado em um capítulo a parte.

2.3 Cooperação entre sistemas de informação na era dos arquivos digitais

Não existem fronteiras na produção e publicação do conhecimento, entretanto, a produção da ciência e a comunicação dos resultados das pesquisas na nossa sociedade acontecem de forma espacialmente dispersa. Pesquisadores, em diversos locais geográficos, produzem e comunicam os resultados de suas pesquisas através de mecanismos de comunicação científica, porém, sem um padrão de publicação que possa indexar as informações publicadas e assim oferecer à comunidade de pesquisa qualidade na recuperação das mesmas.

Os sistemas de informação, hoje, são considerados os grandes aliados das comunidades de pesquisa, desempenhando um papel fundamental na economia globalizada, agregando valor ao buscar a interoperabilidade das informações produzidas, por natureza, de forma dispersa e oferecer a possibilidade da recuperação dessas informações independente do local e forma de armazenamento.

Durante a execução de uma pesquisa científica, muitos documentos são produzidos porém nem sempre publicados e portanto não são encontrados no circuito editorial convencional, como trabalhos apresentados em eventos, *preprints*, teses e dissertações, que noticiam com grande atualidade os resultados de pesquisa. O papel dos sistemas de informação sempre foi o de se contrapor a esta dispersão, provendo um ponto de concentração para a comunicação dos resultados das pesquisas. Os sistemas especializados organizaram-se para prover acesso principalmente a este tipo de documentação não convencional, chamada de "literatura cinzenta", podemos citar o *software* NTIS, o *Dissertation Abstracts*, o *Chemical Abstracts*, ERIC e muitos outros.⁷ Coletar essa literatura sempre foi caro e extremamente trabalhoso para os sistemas especialistas, devido à dispersão e heterogeneidade das informações.

⁷NTIS: www.ntis.gov/search.htm;

Dissertation Abstracts: <http://library.dialog.com/bluesheets/html/bl0035.html>;

Chemical Abstracts: <http://library.dialog.com/bluesheets/html/bl0399.html>;

ERIC: www.eric.ed.gov

Com a publicação diretamente na *Web* esse quadro começa a mudar radicalmente, mas surge outro problema, a quantidade de informações disponível na rede é tão grande que identificar, localizar, descobrir a existência e acessar informações relevantes torna-se crítico, demandando um tempo proibitivo aos usuários. Hoje em dia não é mais suficiente, para garantir o máximo de visibilidade de seu acervo, que bibliotecas digitais simplesmente disponibilizem seus dados na *Internet*. Torna-se necessário a definição de padrões para publicação e interoperabilidade entre as bibliotecas, para que o usuário possa buscar a informação relevante em um único ambiente, um portal de acesso padrão.

Como aconteceu no passado, onde as bibliotecas convencionais eram responsáveis pela catalogação, guarda e disponibilidade dos documentos, cabe agora as bibliotecas digitais a questão de cooperarem, sob novas bases organizacionais e tecnológicas, para garantir o máximo de visibilidade a seus acervos. Atingir esta visibilidade não significa mais necessariamente que alguém buscando informações terá de acessar o *site* da biblioteca digital ou arquivo eletrônico para ter acesso aos documentos digitais nele depositados. A possibilidade que seus acervos possam ser consultados simultaneamente, sem que um usuário acesse cada site individualmente, a chamada interoperabilidade, tem sido buscada como forma de viabilizar a tecnologia dos arquivos abertos. Atingir a interoperabilidade entre repositórios de arquivos *e-prints* ou bibliotecas digitais, distintos e heterogêneos, possibilitando que possam ser consultados simultaneamente, envolve um aporte intenso em termos de tecnologias, protocolos e padronização.

Neste capítulo foi relatado a situação da publicação do conhecimento considerando o modelo tradicional de publicação e as possibilidades futuras em consequência do surgimento da *Internet*. No próximo capítulo veremos como é realizada atualmente a publicação na *Web* considerando as tecnologias disponíveis e a busca pela interoperabilidade das informações.

3 PUBLICAÇÃO NA WEB

Hoje, as tecnologias de informática, computação e comunicação desempenham um papel diferenciado na produção e publicação de conhecimento. (MARCONDES & SAYÃO, 2002). Em se tratando de ciência e tecnologia, este papel é mais acentuado, na medida em que a ciência institucionalizada está centrada em mecanismos de comunicação rápida, propiciando um menor tempo nos resultados de pesquisa, que por sua vez estão baseados fortemente nas tecnologias de informação. No ciclo de comunicação científica, as bibliotecas têm um papel fundamental. A elas cabem, neste ciclo, os papéis de coleta, registro, estocagem e disseminação do conhecimento. A evolução das tecnologias de informação, no entanto, vem alterando substancialmente este papel e junto com isto o próprio conceito de biblioteca.

Este capítulo trata de tecnologias para publicação na *Web*, metadados e o vocabulário *Dublin Core*, como instrumentos no sentido de promover a interoperabilidade de informações publicadas na *Internet*.

3.1 Tecnologias para publicação na *Web*

O surgimento da *Internet* e das novas tecnologias de comunicação disponíveis na rede têm provocado profundas mudanças em relação à publicação científica. (COSTA, 1996, p.231 *apud* VIDOTI, 2004), “... as novas tecnologias de informação estão modificando o modo como os cientistas disseminam informações a respeito de seus trabalhos, tanto na comunidade científica quanto para o público em geral”. No contexto atual, tornou-se evidente a morosidade do processo da comunicação científica tradicional frente à rapidez com que algumas áreas do conhecimento se desenvolvem e promovem a divulgação dos seus trabalhos. Soma-se a isso a questão da transferência dos direitos autorais para os editores, o que nem sempre corresponde aos interesses dos autores.

A importância do processo de revisão feita pelos pares e o tempo que isso requer é muitas vezes um limitador do processo de disseminação de novas idéias, favorecendo a formação de um círculo restrito de editores e autores. Nesse sentido os arquivos abertos e

automatizados de *e-prints* oriundos e referendados pela *Open Archives Initiative*, aparecem como um modelo mais eficiente para a disseminação dos resultados de pesquisa, disponibilizando uma base tecnológica que oferece uma estrutura de mediação entre as instituições científicas e os usuários, de uma forma ampla e com custos relativamente baixos para as comunidades de pesquisa.

3.1.1 Metadados

Para propiciar a interoperabilidade entre bases de dados que apresentam um alto grau de distribuição e heterogeneidade, como é o caso do assunto focado neste trabalho, o uso de metadados é considerado de fundamental importância e o fator determinante na recuperação das informações catalogadas e disponíveis em um repositório comum. Quando se tem uma vasta coleção de dados como, por exemplo, um livro, uma biblioteca, ou até mesmo toda a *Web*, faz-se necessário mecanismos que busquem e indexem informações solicitadas com qualidade. Na recuperação de informações são utilizados índices e tabelas que descrevem seu conteúdo a fim de classificar e organizar a informação e agilizar a busca, nesse caso, um índice é um metadado definido durante a publicação da informação.

Metadado pode ser definido como sendo “dado sobre o dado”. Na definição, “Meta” é um prefixo de auto-referência sobre o conteúdo ou dado considerado, de forma que ao juntarmos meta mais dado, temos “metadados” qual seja, “dados sobre dados” ou dados associados a objetos que auxiliam seus usuários potenciais proporcionando conhecimentos sobre suas existências e características, descrevendo suas unidades de medidas, qualidade e objetivos.

Metadados auxiliam na padronização da descrição, no processamento e na integração de dados heterogêneos, facilitando o acesso e utilização dos mesmos (GUNTHER, 1997). Metadados são utilizados em páginas HTML, nas marcações (*meta*), para descrever configurações que são interpretadas e utilizadas pelo navegador (*browser*). Portanto, metadados são necessários quando se quer organizar uma grande quantidade de informação.

Profissionais da área de sistemas de informações juntamente com a comunidade científica, consideram os metadados como um fator crucial na busca pela interoperabilidade entre informações heterogêneas e distribuídas. A definição de “dado estruturado sobre dado”, é concisa e é empregada em diferentes áreas com um objetivo similar, ou seja, “permitir uma

melhor integração, troca, acesso e interpretação dos dados” (GUNTHER, 1997). Para isso torna-se necessário a busca de um padrão de definição dos metadados.

Atualmente existem diversos padrões voltados para domínios particulares do conhecimento, podemos citar o modelo *Umwelt-Datenkatalog* (UDK), o padrão *Federal Geographic Data Committee* (FGDC), ambos usados em aplicações de georeferenciamento. Simon & Tomasic (1998), concluem, após análise dos padrões existentes, que não existe e nunca existirá um padrão de metadado único devido à natureza heterogênea das aplicações.

O surgimento de diversos padrões de metadados originou um problema de incompatibilidade entre os padrões. O padrão *Dublin Core* (WEIBEL, 1999), no contexto das bibliotecas digitais, foi uma das primeiras tentativas de se gerar um padrão de metadados que fosse comum a todos os demais padrões.

3.1.1.1 Vocabulário *Dublin Core*

Dublin Core (DC), é um conjunto de termos usados para definir metadados e vocabulários. Ele foi desenvolvido pela *Dublin Core Metadata Initiative* (DMCI)⁸, um fórum cujo objetivo é desenvolver padrões de metadados que suportem uma grande variedade de aplicações, comunidades e modelos de negócio. O objetivo principal da DCMI é proporcionar integração de metadados, permitindo que recursos de diversas áreas do conhecimento possam ser descritos e amplamente reutilizados.

O *Dublin Core* é constituído de um conjunto padrão de 15 elementos utilizados para descrever recursos, propostos na versão 1.1⁹. Cada elemento é opcional e pode ser repetido. Esses elementos foram definidos tendo em vista a descrição de documentos, artigos e páginas *Web*. Segundo Hillmann (2003), as principais características do padrão *Dublin Core* são:

- Simplicidade na descrição dos recursos;
- Entendimento semântico universal dos elementos;
- Escopo internacional e extensibilidade;
- Entensibilidade – O *Dublin Core* permite que seja realizado adaptações à necessidades adicionais de descrição, ou seja, permite que comunidades diferentes adicionarem outros elementos específicos dentro de uma área de atuação.

⁸ www.dublincore.org/

⁹ <http://www.niso.org/standards/resources/Z39-85.pdf>

Elementos que compõem o *Dublin Core*:

- *dc:title* – (*Título*) Um nome dado ao recurso pelo criador/autor ou publicador/editor.
- *dc:creator* – (*Criador*) As pessoas ou organizações principais responsáveis pela criação do conteúdo intelectual do recurso.
- *dc:subject* – (*Assunto*) A essência do conteúdo do recurso. O assunto deve ser expresso por palavras-chaves ou frases que descrevam o assunto ou conteúdo do recurso. O uso de vocabulários controlados é encorajado.
- *dc:description* – (*Descrição*) Uma descrição textual do conteúdo do recurso, incluindo abstracts no caso de documentos textuais, ou descrição de conteúdo no caso de recursos visuais.
- *dc:publisher* – (*Publicador*) A entidade responsável por tornar o recurso disponível na presente forma, tais como uma editora/publicadora, um departamento de uma universidade, uma entidade.
- *dc:contributor* – (*Colaborador*) Pessoa ou organização não-especificada no elemento criador, que tenha dado contribuição intelectual significativa para o recurso, mas cuja contribuição é considerada secundária para a pessoa ou instituição especificada no elemento Criador, como por exemplo, editor, tradutor, ilustrador.
- *dc:date* – (*Data*) A data em que o recurso tornou-se disponível na presente forma. Uma data associada com a criação ou disponibilidade do recurso. Recomenda-se adotar a ISSO 8601, formatos de datas e tempo, *World Wide Web Consortium (W3C¹⁰) Technical Note¹¹*, inclui diferentes formatos de datas.
- *dc:type* – (*Tipo*) A natureza ou gênero do conteúdo do recurso, tais como texto, som, dados, *software*, interativo, evento e objeto físico.
- *dc:format* – (*Formato*) O formato de armazenamento físico ou digital do recurso, ele pode ser selecionado de uma lista em desenvolvimento pelo *Dublin Core Workshop Series*.
- *dc:identifier* – (*Identificador*) Um identificador não-ambíguo (único) de um recurso em um dado contexto.

¹⁰ Consórcio de empresas, profissionais, cientistas e instituições acadêmicas que é responsável pela criação de padrões tecnológicos que regulam a *World Wide Web*.

¹¹ <http://www.w3.org/TR/NOTE-datetime>

- *dc:source* – (*Origem*) Informação sobre um segundo recurso do qual o presente recurso é derivado.
- *dc:language* – (*Linguagem*) Idioma em que se encontra o conteúdo do recurso.
- *dc:relation* – (*Relação*) Possibilita relacionamento com outros recursos. A especificação desse elemento visa fornecer um meio de expressar relacionamentos entre recursos que tem relação normal com outros, mas que existem por si mesmos.
- *dc:coverage* – (*Coverage*) As características espaciais ou temporais do conteúdo intelectual do recurso. Cobertura espacial refere-se a região física, longitude e latitude ou ainda a nomes de lugares de uma lista controlada. Cobertura temporal refere-se sobre o que é o recurso, e não quando foi criado ou tornado disponível.
- *dc:rights* – (*Direitos*) Informações sobre direitos autorais do recurso. Um identificador que vincula a uma declaração de direitos sobre propriedade, ou um identificador que vincula a um serviço que fornece informações sobre o direito de propriedade do recurso.

O conjunto de metadados *Dublin Core* suporta qualificadores para especificar o significado de um elemento (HILLMAN, 2000) e pode ser codificado em formatos como HTML e *Extensible Markup Language* (XML), (BECKETT, 2004). Como é explicitado na própria proposta da *Dublin Core*, o conjunto de metadados deve ser simples e intuitivo a ponto de permitir que o próprio autor descreva seu trabalho. Os ambientes de submissão de trabalhos, no caso as bibliotecas digitais, adotam este padrão na sua arquitetura. Ao submeter seu trabalho, o autor preenche um formulário com os metadados pertinentes sobre o assunto abordado como se fosse fazer um índice de palavras chaves para posterior coleta dessas informações por outros serviços disponíveis na rede. Os elementos do vocabulário são bastante abrangentes, podendo ser utilizados em várias aplicações, e não apenas na descrição de documentos.

3.1.1.2 Representação de metadados através de linguagens de marcas

Historicamente, usa-se a palavra "marcação" para descrever anotações ou marcas em um texto, que tem por objetivo dar instruções para alguém ou uma máquina sobre a maneira como uma parte do texto deveria ser representada. Pode-se definir que um sublinhado ondulado indica negrito, símbolos especiais para passagens do texto a serem omitidas ou impressas com uma fonte especial, dentre outras. Como a formatação e a impressão de textos

tornaram-se automatizadas, o termo foi estendido para todos os tipos de códigos de marcação em textos eletrônicos. Todos os textos impressos são codificados com sinais de pontuação, uso de letras maiúsculas e minúsculas, regras para a disposição do texto na página, espaço entre as palavras etc. Estes elementos são visto como um tipo de "marcação", cujo objetivo é ajudar o leitor na determinação de onde uma palavra termina e onde outra começa, ou identificar características estruturais, por exemplo, cabeçalhos ou simples unidades sintáticas como parágrafos e sentenças. Codificar ou "marcar" um texto para processamento por computadores é também um processo de tornar explícito o que é conjetural. Indica como o conteúdo do texto deve ser interpretado no momento do processamento do texto pelo computador.

Linguagens ditas “linguagens de marcação” ou *markup languages*, são utilizadas para a transferência e representação de dados na *Internet*, neste trabalho será abordado aspectos gerais de linguagens de marcação *Standard Generalized Markup Language* (SGML), XML e HTML, sendo esta última considerada hoje como padrão para publicação e transferência de informações na *Internet*. As linguagens de marcação são as bases conceituais de uma nova geração de aplicações nas áreas da informação e da tecnologia da informação, possibilitando a democratização do acesso à informação organizada na *Internet* permitindo a evolução das pesquisas em direção à chamada *Web* semântica, e neste caso, o desenvolvimento de ontologias a respeito do assunto a ser pesquisado.

A seguir é descrito a origem das linguagens de marcação e sua utilização na *Internet*, os conceitos complementares necessários ao entendimento das linguagens de marcação que possibilitam ao autor especificar a forma dos dados no documento, além de permitir definições semânticas ao dado.

3.1.1.2.1 SGML – *Standard Generalized Markup Language*

Standard Generalized Markup Language definido pela ISO 8879 em 1986, representa um padrão internacional para definição de estrutura e conteúdo de diferentes tipos de documentos eletrônicos. A SGML pode ser chamada de "língua mãe", um padrão internacional, não proprietário e de código aberto, utilizado já há bastante tempo para troca eletrônica de dados podendo ser utilizada por diferentes sistemas informatizados para descrever tipos diferentes de documentos em áreas da atividade humana, desde transcrições

de antigos até documentação técnica para aviões de guerra, registros de pacientes em unidades médicas até notação musical.

Um dos objetivos do SGML é garantir que documentos codificados de acordo com suas regras possam ser transportados de um ambiente de *hardware* e *software* para diferentes ambiente, sem perda de informação. Cada documento SGML carrega consigo sua própria especificação formal, o *Data Type Document* (DTD).

O DTD é uma espécie de gramática formal criada a partir da notação *Extended Backus-Naur Form* (EBNF), que define como as marcas devem ser interpretadas, quais as regras que restringem o uso de cada marca nos diferentes contextos do documento e, até mesmo, quando for relevante, a ordem em que as marcas devem aparecer no documento. SGML é uma linguagem para definir outras linguagens, ou ainda, uma linguagem para construir diferentes tipos de documentos, independente de padrões, ou seja, a definição do padrão do documento e todas as informações necessárias para o tratamento da informação fazem parte da própria codificação definida pela DTD.

3.1.1.2.2 HTML – *Hypertext Markup Language*

No início dos anos 80, SGML passou a ser usada em várias organizações, entre as quais o Centro Europeu de Pesquisas Nucleares em Genebra (CERN), onde um pesquisador chamado Dan Connolly usou a linguagem em seu programa de edição de hipertextos. Em seguida, Tim Berners-Lee juntamente com Dan Connolly e Ralph R. Swick inventam o *World Wide Web*, graças a uma idéia revolucionária na época: o *link* ou ligação entre documentos que poderiam estar situados em qualquer lugar na rede de computadores de seu laboratório ou do mundo, através da *Internet* e do conceito da URL¹².

O *Hypertext Markup Language* foi definido pela IETF-RFC1866¹³ e consiste de uma aplicação específica do SGML utilizada na *Internet*. O HTML define um tipo de documento simples, com marcações fixas projetadas para uma classe de relatórios técnicos de uso comum em escritórios, como, por exemplo, cabeçalhos, parágrafos, listas, ilustrações e algumas possibilidades para hipertexto e multimídia. HTML possui um grupo de *tags* predefinidos, concebidos com a função de organizar a informação a ser transferida por meio de páginas

¹²<http://www.w3.org/1999/04/WebData>

¹³<http://www.ietf.org> - *Internet Engineering Task Force*

Web sendo usado em milhares de aplicações, incluindo navegadores, editores, *softwares* de *e-mail*, servidores de base de dados entre outros.

3.1.1.2.3 XML – *Extensible Markup Language*

Extended Markup Language é uma versão abreviada do SGML, que possibilita ao autor especificar a forma dos dados no documento, além de permitir definições semânticas do dado especificado na linguagem. Um arquivo eletrônico XML, pode conter, simultaneamente, dados e descrição da estrutura do documento, através do DTD. O XML obtém benefícios omitindo as partes mais complexas e menos utilizadas do SGML. Linguagem derivada do SGML e idealizada por Jon Bosak, engenheiro da Sun Microsystems, o qual apresentou ao W3C sua idéia de explorar o SGML em aplicações voltadas para *Internet*. Criada em 1996, inicialmente como uma versão simplificada do SGML, a linguagem teve total aceitação que em fevereiro de 1998, tornou-se uma especificação formal, reconhecida pelo W3C.

De acordo com o W3C, entre os objetivos estabelecidos na especificação da linguagem XML, estão as seguintes características:

- ser diretamente utilizável na *Internet*;
- ser legível por humanos;
- possibilitar um meio independente para publicação eletrônica;
- permitir a definição de protocolos para troca de dados pelas empresas (independentemente da plataforma de *hardware* e *software*);
- facilitar às pessoas o processamento de dados pelo uso de *softwares* de baixo custo;
- facilitar a utilização de metadados que auxiliam na busca de informações;
- aproximar "produtores" e "consumidores" de informação.

Neste capítulo foi apresentado as tecnologias disponíveis para publicação na *Web* e a importância da definição de padrões de vocabulários no sentido de organizar o conhecimento privilegiando a recuperação das informações. A facilidade propiciada pelas linguagens de marcas, que, possibilitam publicar a estrutura da informação através de definição de marcas identificadoras da informação na própria estrutura do código, em especial a linguagem XML,

considerada como padrão para troca de informações entre sistemas diferentes. No próximo capítulo veremos o padrão de troca de informações permitido pelo uso de tecnologias de arquivos abertos o qual define um padrão de interoperabilidade de informações entre bibliotecas.

4 ARQUIVOS ABERTOS

A *Internet* e as novas tecnologias de comunicação disponíveis na rede têm provocado profundas mudanças no mundo da publicação científica. Neste contexto, surgem os arquivos abertos desenvolvidos para prover um padrão de publicação na *Web*, visando à interoperabilidade entre arquivos disponibilizados na rede com o objetivo de facilitar a recuperação e uso das informações como fonte de pesquisa.

Neste capítulo é apresentado o padrão de interoperabilidade denominado Arquivos Abertos, sua origem, motivação e uso pela comunidade científica na publicação e troca de informações na *Web*.

4.1 A Publicação Científica na *Web*

O conceito de arquivos abertos tem como referencia, em relação a um padrão de publicação, outubro de 1999. A convenção de Santa Fé realizada no Novo México, a qual foi organizada por Paul Ginsparg, Rick Luce e Herber Van Sompel, e de responsabilidade do *Council on Library and Information Resources (CLIR)*, da *Digital Library Federation (DLF)*, da *Scholarly Publishing & Academic Resources Coalition (SPARC)*, da *Association of Research Libraries (ARL)* e do *Los Alamos National Laboratory (LANL)*¹⁴.

Durante esta Convenção foram definidos os princípios básicos de uma nova era para a publicação científica. Foi definido um padrão para troca de informação entre arquivos eletrônicos, a iniciativa ficou conhecida como *Open Archives Initiative (OAI)*. Os arquivos abertos são considerados como o mais democrático e eficiente modelo para disseminação de resultados de pesquisas. Uma referência importante sobre o histórico, os desenvolvimentos recentes e a dimensão da (OAI), inclusive com uma entrevista com Paul Ginsparg, pode ser encontrada em Sena (2000).

Na convenção, discutiu-se o funcionamento dos serviços de auto-arquivamento e o estabelecimento de normas de adesão à iniciativa, definição de metadados, especificações de

protocolo e padronização de linguagens. Mais tarde, em 2001, uma reunião dos membros da comunidade definiu os seguintes conceitos em relação padronização: o auto-arquivamento, interatividade entre usuários através da revisão entre os pares e a interoperabilidade entre os arquivos abertos através de padronização de metadados e protocolos.

O auto-arquivamento (*self archiving*) refere-se ao direito de o próprio autor enviar o seu texto para publicação em um repositório de arquivos do tipo *e-print*, garantindo visibilidade aos seus trabalhos de pesquisa sem a interferência de terceiros. Um conceito inovador cujos objetivos são tornar o texto disponível para colheita dos dados no menor tempo possível oferecendo acesso democrático e gratuito das publicações eletrônicas, com isso diminuindo o monopólio das grandes editoras científicas, que até recentemente detinham em seu poder os direitos de publicação

Os arquivos de textos eletrônicos são inteiramente dirigidos pelos cientistas e flexíveis o bastante, tanto para coexistir com os sistemas de publicação tradicional, como para ajudar estes sistemas a se envolverem com algo mais próximo das necessidades dos pesquisadores (TRISKA, 2001). Um arquivo *e-print*, pode ser classificado como: *preprint*, texto de propriedade do autor, ainda não apreciado por um corpo de referência e antes da publicação; *reprints*, texto publicado e na maioria das vezes de propriedade do editor em função da proteção da cópia, *copyright*. Segundo Stevan Harnard, “antes de haver a arbitragem e publicação, eles são *preprints*, propriedade do autor. Depois da arbitragem, aceitação e publicação, eles são em função do acordo de *copyright*, usualmente, de propriedade do editor e chamados de *reprints*”.

Em relação à revisão entre os pares, a filosofia dos arquivos abertos tem como propósito a transparência das críticas e sugestões feitas aos textos eletrônicos depositados no repositório de arquivos de *e-prints*. Desta forma, o ambiente, possibilita que toda a comunidade tenha acesso ao processo de revisão e de versões do texto geradas com base nas sugestões. Acaba o conceito de revisão sigilosa feita por um comitê científico, caracterizando mais uma vez o aspecto democrático desta iniciativa. Nesse sentido, Ginsparg (2000) *apud* Triska (2001), menciona que a disseminação rápida oferecida pelos arquivos abertos não é incompatível com a tradicional revisão entre os pares. Para ele, a longo prazo, os arquivos abertos oferecem uma estrutura mais funcional de organização da literatura do que a atual organização existente no processo de revisão entre os pares.

¹⁴ http://www.openarchives.org/meetings/SantaFe1999/sfc_entry.htm.

A interoperabilidade de informações, ou seja, a possibilidade de troca de informações entre sistemas de informação na comunidade científica, passa a ser determinante para a criação dos arquivos abertos. Na perspectiva deste movimento a nível internacional de publicação eletrônica, a interoperabilidade envolve uma série de aspectos, tais como um conjunto mínimo de metadados, tipo de arquitetura subjacente do sistema, abertura para a criação de serviços de bibliotecas digitais de terceiros, integração com o mecanismo de comunicação já existente no meio científico, possibilidade de uso em contextos interdisciplinares e a contribuição para criação de um sistema de medida de uso e de citação, Triska (2001).

A Convenção de Santa Fé estabeleceu os seguintes aspectos como mecanismos para se atingir a interoperabilidade:

- a) definição de um conjunto mínimo de metadados;
- b) concordância no uso de uma sintaxe comum XML para representar e transportar tanto o *Open Archives Metadata Set* (OAMS) como os conjuntos de metadados específicos de cada repositório;
- c) definição de um protocolo comum *Open Archives Dienst Subset* para possibilitar a extração do OAMS e dos metadados específicos dos repositórios participantes.

No que se refere à criação de repositórios de documentos eletrônicos, a Convenção recomenda que sejam observados os seguintes aspectos:

- a) mecanismo de submissão;
- b) armazenamento a longo prazo;
- c) política de gerenciamento da submissão e da preservação dos documentos inseridos no repositório;
- d) interface aberta que permita que terceiros possam coletar dados do repositório.

4.2 O software arXiv

Como já citado anteriormente, antes da Convenção de Santa Fé, em agosto de 1991, o físico e pesquisador Paul Ginsparg do laboratório de Los Alamos no Novo México cria o primeiro repositório de documentos eletrônicos de *e-print* baseado na filosofia dos arquivos abertos, o *arXiv*¹⁵, o qual tornou-se um repositório global de artigos não-revisados pelos pares

¹⁵ <http://arxiv.org/>

nas áreas de física e disciplinas correlatas, matemática, ciências não-lineares, lingüística computacional e neurociências.

Inicialmente, o repositório continha arquivos *e-pints* de uma comunidade de 200 físicos, nos anos seguintes esse número passou para 35 mil usuários, localizados em mais de 70 países, processando mais de 70 mil transações eletrônicas por dia, conseguindo suplantar alguns periódicos tradicionais em certas áreas da física como meio de disseminação de novas pesquisas. Seus usuários podem recuperar ou submeter os artigos tanto por uma *interface on-line* na *Web* como via correio eletrônico. Neste ambiente é permitida a atualização de artigos pelos autores, além de serem oferecidos serviços de alerta e mecanismos de busca adaptados a diferentes perfis de usuários¹⁶.

A principal característica do sistema *arXiv*, no entender dos seus mentores, diz respeito à publicação acadêmica sem interesse comercial com foco principal na maior distribuição possível do trabalho, sem qualquer cobrança de *royalties* ou *pay-per-view*, resguardado, porém, o uso justo (*fair use*) da obra. O autor tem o direito de publicar o documento eletrônico onde quiser, permitindo que se estabeleçam apontadores que indiquem a URL onde está o texto, sem admitir, contudo, a reprodução do trabalho em outro servidor, nem a venda sem seu consentimento.

Apesar do grande sucesso do *arXiv*, existe uma preocupação entre os bibliotecários em relação ao arquivamento diante do grande número de publicações e ausência de padrões nas tecnologias de informação. A resistência à nova tecnologia é natural considerando que não se pode jogar fora um sistema comprovadamente bom como é o caso das bibliotecas tradicionais que se mantém há séculos, em função de suas falhas e limitações, sem um embasamento seguro de que a substituição de mídia é confiável e apresenta um substituto estável para o armazenamento seguro no presente e no futuro.

Na figura 1 é apresentado uma consulta em tempo real na *Web* do *software arXiv*, as barras indicam o número de novas submissões recebidas, a cada mês, desde 1991. “*arXiv monthly submission rate statistics*¹⁷”.

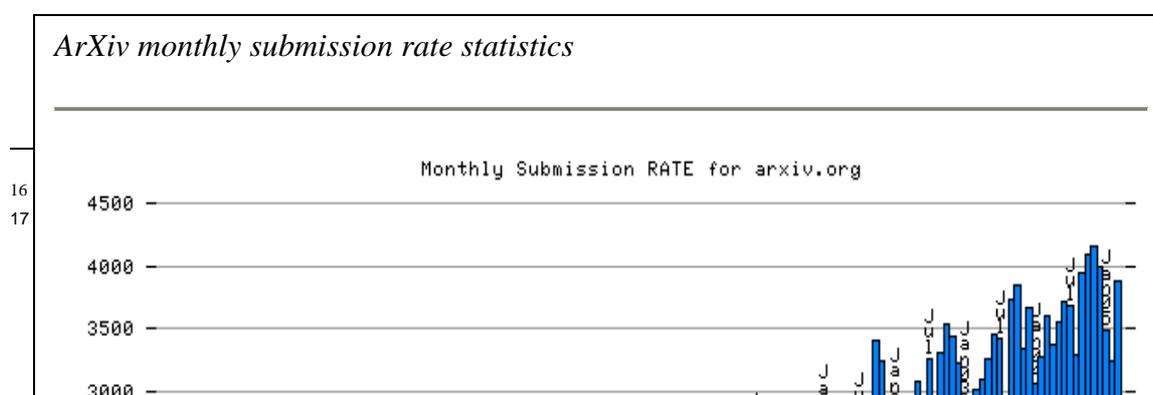


Figura 1 – Número de novas submissões por mês no *arXiv* desde 1991 até abril/2005.

4.3 *Open Archives Initiative*

A interoperabilidade entre arquivos eletrônicos de *e-prints*, foi alcançada mediante adoção de um conjunto de especificações técnicas e princípios organizacionais bastante simples, porém potencialmente poderosos e de grande alcance conhecida como *Open Archive Initiative*¹⁸, e tem como objetivos básicos apoiar o desenvolvimento de arquivos de *e-prints* e criar uma arquitetura tecnológica padronizada que sustente a interoperabilidade entre eles.

Na arquitetura da *Open Archives Initiative*, foram desenvolvidas tecnologias, padrões e metodologias para publicação, disponibilização, metadados e intercâmbio automático de metadados entre bibliotecas digitais. A dimensão da iniciativa dos *Open Archives* a nível mundial e sua cobertura regional e temática podem ser consultadas na lista dos arquivos eletrônicos existentes em <http://www.osti.gov/eprints/>.

Um repositório de arquivos de *e-prints* baseado nos padrões dos arquivos abertos, apresenta facilidades que permitem a um autor submeter diretamente seus trabalhos,

¹⁸ <http://www.openarchives.org/>

armazená-los em forma digital permanentemente, editá-los, substituí-los e receber críticas e contribuições. Para isso o autor necessita somente conhecer o endereço eletrônico na rede. Ao submeter um trabalho para armazenar e disponibilizar em meio eletrônico, autor deve descrever, em um formulário de catalogação, os metadados tais como autor, título, idioma, assunto, ou seja, definir um vocabulário para o arquivo objetivando a posterior extração de informações permitindo assim a recuperação do documento na *Web*.

Na arquitetura da OAI, o *site* que hospeda o arquivo permite a consulta e o acesso direto aos arquivos eletrônicos nele armazenado. Um servidor de *e-prints* compatível com o protocolo padrão de OAI, deve permitir a exposição de metadados dos trabalhos nele armazenados para colheita automática e reuso por provedores de serviços de informação disponíveis na rede, possibilitando a criação de outros serviços a partir dessa exposição.

No Brasil, a publicação na rede de textos completos de interesse em C&T começa a ser uma realidade. Algumas Instituições adotaram o padrão de arquivos abertos para publicar suas pesquisas, em um capítulo será descrito a situação atual de algumas iniciativas pioneiras, podemos citar:

- portal [Scientific electronic library online](http://www.scielo.br) SciELO (<http://www.scielo.br>), que abrange dezenas de periódicos, associado a uma metodologia para publicar e prover acesso a periódicos eletrônicos em texto completo na *Web*. A SciELO é o resultado de um projeto de pesquisa da [FAPESP](#) - Fundação de Amparo à Pesquisa do Estado de São Paulo, em parceria com a [BIREME](#) - Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde. A partir de 2002, o Projeto passou a ter o apoio do [CNPq](#) - Conselho Nacional de Desenvolvimento Científico e Tecnológico;
- proposta de desenvolvimento de um ambiente *Web* para edição de anais de congresso do Centro de Informações Nucleares (CIN) e da Comissão Nacional de Energia Nuclear (CNEN), através do resumo de artigos, periódicos, teses, capítulos de livros, trabalhos de congressos;
- arquivos de *e-prints* do [Instituto Nacional de Matemática Pura e Aplicada](#) (IMPA);
- publicações digitais do Laboratório de Automação de Museus, bibliotecas digitais e Arquivos do Departamento de Engenharia Elétrica da PUC-Rio (LAMBDA) (<http://www.lambda.maxwell.ele.puc-rio.br>);
- Banco de Teses e Dissertações (BDT) do Programa de Pós-graduação em Engenharia de Produção da Universidade Federal de Santa Catarina PPGE/UFSC (<http://teses.eps.ufsc.br/>);

- repositório de teses da USP (<http://www.teses.usp.br>) e a iniciativa do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) que desenvolveu um *software* para armazenar T&D's produzidas pelas universidades do país.
- portal de Periódicos da Capes/MEC (<http://www.periodicos.capes.gov.br>), criado em novembro de 2000, para universalizar o acesso à informação no meio acadêmico-científico, o Portal é atualmente o maior banco de informações científicas da América Latina, atendendo 152 instituições.

Algumas dessas iniciativas serão melhor detalhadas em um capítulo à parte.

4.4 Interoperabilidade entre Documentos Digitais

A interoperabilidade entre documentos digitais pede ser alcançada através de três formas de consulta, ou seja, três tecnologias de publicação na *Web*. A primeira forma é através de um esquema chamado de Federação, onde a consulta é distribuída entre as instituições cooperantes; outro esquema chamado de colheita ou *harvesting*, que monta um repositório único colhendo metadados das instituições cooperantes e o terceiro esquema chamado de reunião ou *gathering*, que acontece através do uso de motores de busca na *Web*.

Na federação, uma consulta é submetida a um componente global, que a distribui a uma federação de bases documentais, que as executam e enviam seus resultados ao componente global. Nesse tipo de consulta a interoperabilidade é imediata e exige a participação síncrona de todas as bases de dados da federação. Muitas bibliotecas digitais usam esse tipo de consulta para prover interoperabilidade de informação, entretanto todos os membros da federação devem, obrigatoriamente, concordar, implementar e manter os serviços acordados, requerendo cooperação entre os participantes. O protocolo para prover a interoperabilidade é o Z39.50, e deve rodar localmente nos clientes.

Os participantes da *Open Archives Initiative* rejeitam este protocolo de comunicação em função da complexidade da execução das consultas, devido a exigência do sincronismo entre as bases documentais dos participantes.

Na colheita, os metadados de catalogação das bases documentais são expostos, para serem coletados por outros serviços, num sistema de *Data Provider* e *Service Provider*, ou seja, servidores de dados e servidores de serviço, proporcionando uma implementação e implantação mais fácil, à medida que não exige interoperabilidade imediata. O protocolo

utilizado é o *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), que será discutido logo abaixo, o qual promove a interoperabilidade de informação entre os arquivos de *e-prints*.

Na reunião, a interoperabilidade de dá através de motores de busca, que indexam documentos HTML sem nenhuma padronização, são link's de documentos disponibilizados em sites sem qualquer gerenciamento, ou seja, qualquer pessoa ou organização pode publicar na *Web* e sua informação pode ser automaticamente relacionada em uma pesquisa. Não requer que as organizações estejam preparadas para cooperar e proporciona um serviço de baixa qualidade.

4.5 Protocolo OAI-PMH

No protocolo OAI-PMH a interoperabilidade não é imediata, não é portanto, um protocolo para busca *on-line* entre repositórios de arquivos *e-prints*, bibliotecas digitais ou qualquer servidor na rede que queira expor metadados, ou seja, tornar visíveis metadados de documentos nele armazenados para um serviço externo. O protocolo de busca das novas informações adicionadas extrai essas novas informações através dos metadados definidos no momento do arquivamento do arquivo depois da submissão.

Esse procedimento propicia a interoperabilidade de informações entre arquivos de *e-prints* e é chamado de “Colheita”. Além de colheitas de metadados, o protocolo fornece outros serviços tais como citação, redes cooperativas de informação, redes de bibliotecas de teses e dissertações, edição de revistas, submissão de contribuições, avaliação, comentários, catálogo coletivo de revistas eletrônicas, descoberta de recursos, mineração de dados, educação à distância.

O protocolo de recuperação de informações Z39.50, utilizado na federação distribui uma busca imediata e simultânea por vários servidores que hospedam catálogos de bibliotecas exigindo de que todos os membros da federação concordem, implementem e mantenham os serviços acordados, requerendo total cooperação entre os participantes. Entretanto, no protocolo OAI-PMH a solução é mais simples e menos onerosa em termos de recursos computacionais consumidos conforme veremos neste capítulo. As diferenças na interoperabilidade com o uso do protocolo Z39.50 X e o OAI-PMH pode ser visualizada na figura 2.

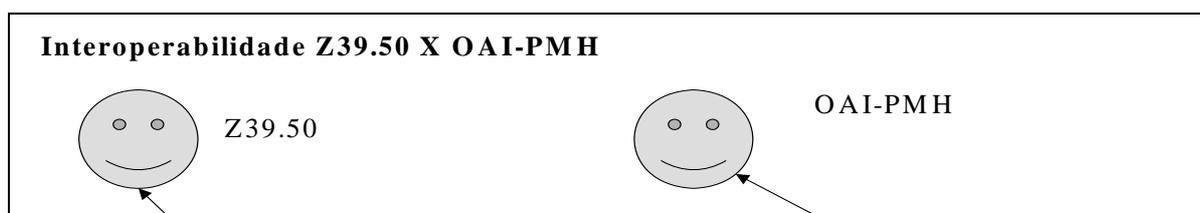
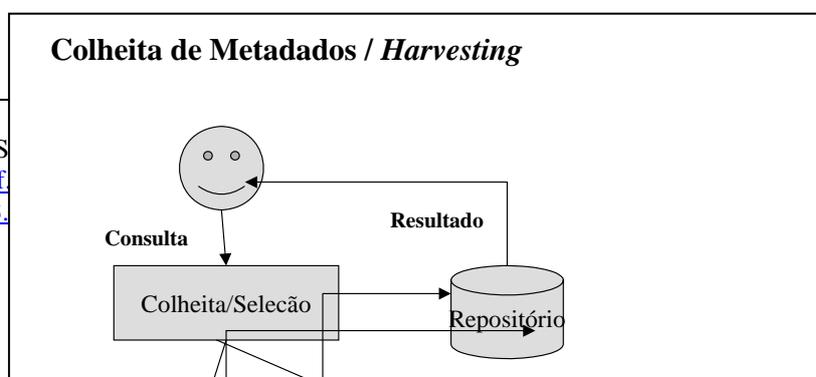


Figura 2 – Interoperabilidade Z39.50 X OAI-PMH¹⁹

O OAI-PMH propicia o acesso unificado a acervos de diferentes provedores de dados, por meio de um portal *Web* único ou a constituição de bases de dados qualificadas sobre temas específicos, ou um periódico eletrônico com textos avaliados e submetidos a um esquema de revisão entre os pares, desenvolvido a partir dos metadados coletados de diversos provedores de dados como uma colheita conforme mostrado na figura 3.

A troca de mensagens que ocorre entre o servidor do provedor de dados e o programa robô externo do provedor de serviços para a transferência de metadados é unidirecional, ou seja, o provedor de serviços faz solicitações ao provedor de dados, que responde enviando metadados na rede. As solicitações do provedor de serviço são feitas via protocolo HTTP²⁰, usando comandos da linguagem *Common Gateway Interface* (CGI), codificados por meio dos métodos específicos, GET ou POST, que fazem a interface entre os provedores. As solicitações são respondidas pelo provedor de dados com o envio de dados das respostas ou metadados dos documentos armazenados codificados na linguagem de marcação XML²¹.



¹⁹ Marcondes & S

²⁰ <http://www.ietf>

²¹ <http://www.w3>

Figura 3 – Colheita de metadados / *Harvesting*

O OAI-PMH estabelece como vocabulário padrão o *Dublin Core Metadata Element Set*²², para a organização das informações com o conjunto mínimo de metadados a ser suportado pelos provedores de dados em resposta a uma solicitação de um provedor de serviços. Entretanto, o provedor de serviços pode, a seu critério, oferecer outros formatos de metadados, mais amplos e complexos, com por exemplo o padrão MARC de dados.

Os metadados, formando um registro de cada documento armazenado no provedor de dados, têm um identificador único, formado pelo identificador do provedor de dados mais um identificador do registro. Cada registro apresenta também um selo temporal, denominado *datestamp*, que indica a data da criação ou última alteração do documento associado a este registro. O *datestamp* é a chave que permite a coleta automática dos metadados do provedor de dados a partir de uma determinada data, possibilitando, portanto, a sincronização entre os registros do provedor de dados e de um provedor de serviços que forneça um serviço de acesso simultâneo a metadados de documentos armazenados em diversos provedores de serviço. A interação entre *Service Providers* e *Data Providers* usando o protocolo OAI-PMH pode ser visualizada a seguir na figura 4.

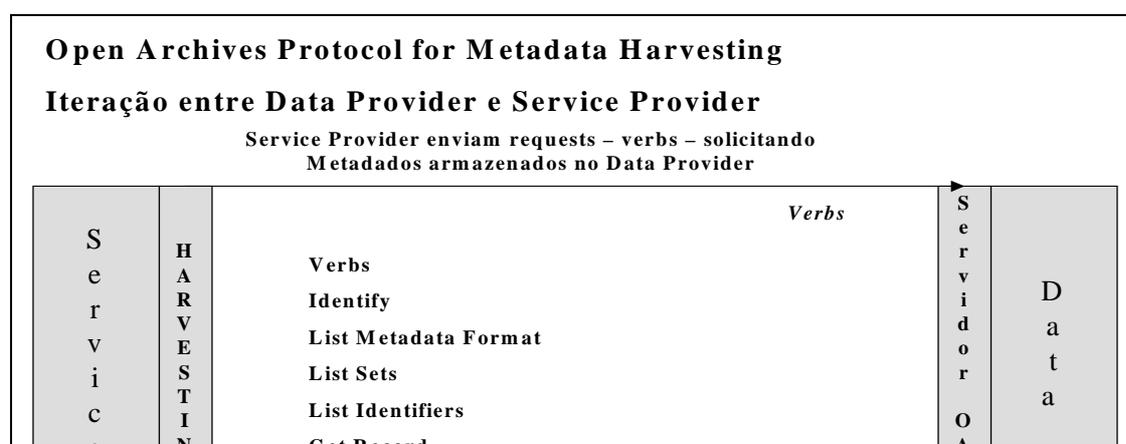


Figura 4 – Iteração entre *Data Provider* e *Service Provider*²³

O protocolo de consulta prevê ao todo seis comandos que um programa robô de provedor de serviços pode solicitar a um provedor de dados, para colher metadados de documentos armazenados:

- *identify* – obtém dados administrativos sobre o provedor de dados, sua política de publicação de documentos, seu escopo etc.;
- *ListSets* lista as classificações sob as quais os documentos são organizados no provedor de dados;
- *ListMetadataFormats* – lista os formatos de metadados por meio dos quais os metadados dos documentos armazenados no provedor de dados podem ser apresentados;
- *ListIdentifiers* lista os identificadores de registros armazenados no provedor de dados, podendo opcionalmente limitar estes registros a partir de uma data, ou pertencentes a um *set*;
- *ListRecords* lista os metadados dos registros armazenados no provedor de dados segundo um formato de metadados, especificando todos que pertencem a um *set* ou todos a partir de uma data;
- *GetRecords* obtém os metadados dos registros armazenados segundo um formato de metadados, dado um identificador de registro.

4.6 Interoperabilidade

²² <http://purl.org/DC/documents/rec-dces-19990702.htm>

²³ <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>

A interoperabilidade entre os arquivos abertos tem como objetivo transformar cada um dos arquivos em parte de um arquivo global para a realização de pesquisas *on-line* permitindo que sistemas distintos e heterogêneos possam buscar e agregar valor à informação criada por outro, gerando novos serviços e novas visões para a mesma informação. Além da questão técnica e tecnológica, o sucesso da interoperabilidade depende da gestão, articulação e cooperação mútua entre sistemas com padrões definidos. Sendo necessário uma estrutura que suporte fóruns de discussões adequados no sentido de estabelecer e tomar decisões endossadas pelo grau de representatividade dessas organizações envolvidas. Este é o caso da DLF - *Digital Library Federation*, da PILA - *Publisher International Linking Association* e do próprio OAI - *Open Archives Initiative*.

4.6.1 Interoperabilidade Semântica

A interoperabilidade semântica diz respeito ao uso generalizado de instrumentos comuns de descrição temática com um conjunto mínimo de metadados padronizados, uma linguagem de descrição temática de cobertura ampla, tal como a tabela de Áreas do Conhecimento da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), que predefine uma estrutura hierárquica de conhecimento e a criação de servidores de autoridades cooperativos. A interoperabilidade semântica tem como reflexo imediato a melhoria na qualidade da recuperação das informações e a otimização da consulta a sistemas interligados. Esses resultados são bastante perceptíveis pelo usuário final.

4.6.2 Interoperabilidade “Linkagem”

A *Internet* nos traz outra forma de interoperabilidade que é a *linkagem* ou enlaces entre sistemas. A *linkagem* permite a navegação via *hiperlinks* entre as várias manifestações do trabalho acadêmico de um indivíduo, normalmente dispersas em vários sistemas, seja como autor, orientador ou membro de banca de teses ou dissertações eletrônicas, seja como autor de artigos de periódico, de trabalhos em congressos ou acessando seu currículo em um sistema de currículos. Mas é de fundamental importância a adoção de padrões e metodologias que garantam a persistência dos endereços eletrônicos dos recursos informacionais, como o

PURL - Persistent URL²⁴ e o DOI²⁵ – *Digital Object Identifier* no sentido de preservar o investimento na *linkagem* entre sistemas.

4.6.3 Serviço de Busca a Repositório de Arquivos Abertos

Com objetivo de recuperar os documentos eletrônicos existentes nos diversos repositórios de arquivos abertos criados no mundo, surgiu o primeiro serviço de busca denominado *Cross Archive Searching Service* (ARC)²⁶. Este serviço utiliza arquitetura em Java e tem plataforma independente, podendo trabalhar com qualquer servidor *Web*. Atualmente, a busca abrange os seguintes repositórios:

- *A Celebration of Women Writers*
- *Software arXiv* - Laboratório Nacional de Los Alamos – *Open Archives*
- *California International and Area Studies Digital Repository*
- *Cogprints - Cognitive Sciences Eprint Archive*
- *Dermatology Digital Repository*
- *Humbolt University of Berlin, Germany, Document Server*
- *Library of Congress Open Archive Initiative Repository 1*
- *LTRS - Langley Technical Reports Server (LTRS)*
- *M.I.T. Theses*
- *National Advisory Committee for Aeronautics*
- *NCSTRL - Networked Computer Science Technical Reference Library*
- *NDLTD - Network Digital Library Theses and Dissertations*
- *NSDL - Open Archives Server at Cornell University*
- *OCLC - Online Computer Library Center Theses and Dissertation Repository*
- *Open Video*
- *Perseus Digital Library*
- *PhsyNet, Oldenburg, Germany, Document Server*
- *Resource Discovery Network*
- *The Oxford Text Archive*
- *Tobacco Control Digital Repository*
- *University of Illinois at Urbana Champaign, Digital Library Initiative*

²⁴ <http://www.purl.org/>

²⁵ <http://www.doi.org>

5 BIBLIOTECA DIGITAL DE TESES E DISSERTAÇÕES – BDTD

Historicamente as bibliotecas foram criadas para armazenar e disponibilizar o resultado de pesquisas e documentos de interesse de uma comunidade ou de uma nação. Frequentemente elas são descritas como o coração das universidades e centros de pesquisas. Um bom acervo bibliográfico, em geral, reflete o nível da pesquisa da comunidade onde ele está inserido, e a eficiência com a qual esse acervo é acessado e utilizado pelos leitores é de grande importância para o crescimento científico e tecnológico de uma nação. Hoje, com as tecnologias de informação que propiciam o desenvolvimento de bibliotecas digitais, o acesso ao conhecimento deixou de ter limitações quanto ao local do acervo bibliográfico, outrora geograficamente localizado e focado no documento impresso, agora disponível em meio digital na *Internet* para ser consultado simultaneamente, sem restrições em relação ao tempo ou local.

As bibliotecas digitais são consideradas sistemas de informação extremamente complexos suportando a criação, gestão, distribuição e preservação de fontes de informação, permitindo uma interação eficaz e eficiente entre as diversas sociedades que se beneficiam do conhecimento armazenado nesses ambientes. O nível de complexidade atribuído na construção de uma biblioteca digital acontece à medida que a mesma deve ter uma estrutura que possa receber e fornecer informações para outras bibliotecas como se fossem construídas sobre a mesma base de dados, provendo total interoperabilidade entre elas. Academicamente, elas também são conhecidas como bibliotecas eletrônicas, principalmente no Reino Unido, bibliotecas virtuais quando utiliza recursos da realidade virtual, bibliotecas sem paredes e bibliotecas cibernéticas.

Mundialmente, temos várias iniciativas tratando do assunto, particularmente neste capítulo, discutiremos a iniciativa da biblioteca digital em rede de teses e dissertações da *Virginia Tech University*, o *Networked University Digital Library* (NUDL)²⁷, dos Estados Unidos.

²⁷ <http://www.nudl.org>,

No Brasil, temos o projeto da biblioteca digital de teses e dissertações coordenada pelo IBICT, que visa proporcionar a integração, a partir de uma interface padrão, de todos os repositórios da informação científica do país. O principal objetivo é disponibilizar os trabalhos produzidos nos programas de pós-graduação das universidades, além de inserir a produção científica da instituição no consórcio brasileiro de teses e dissertações, que, mantido pelo IBICT, dissemina os trabalhos defendidos no âmbito dos programas de pós-graduação do país. Esse objetivo somente poderá ser alcançado a partir da integração de todos os sistemas de informação que publicam as T&D's existentes nas Instituições de Ensino Superior (IES), nacionais.

O termo “biblioteca digital” surgiu em 1990 com a intenção de prover recursos em áreas-chaves de pesquisa. A partir de 1993 ficou claro que havia uma nova ordem em termos de armazenamento e publicação do conhecimento produzido nos grandes centros de pesquisa, (FOX & GONÇALVES, 2001). As bibliotecas digitais passam a oferecer benefícios e vantagens em relação às bibliotecas tradicionais:

- disponibilidade e alta qualidade de recursos de *multimedia* rompendo barreiras físicas e conceituais;
- conectividade na rede com tecnologias que permitem a interatividade possibilitando a criação de ricos espaços virtuais para a pesquisa;
- tecnologia digital que permite a criação de serviços avançados e inovadores que seriam extremamente difíceis de se executar em bibliotecas tradicionais ou praticamente impossíveis além do alto custo. Podemos citar como exemplo novos métodos de visualização e iteração de informações, eficientes mecanismos de pesquisa, complexa interconexão através de hipertextos e bibliografia referenciada e disseminação automática da informação.

Um excelente exemplo de uma biblioteca digital extremamente complexa é a biblioteca digital em rede da Universidade de Virginia, o projeto do NUDL foi desenvolvido para que a propriedade intelectual produzida nas universidades pudesse ser mais acessível, estimulando a transferência de tecnologia, a colaboração internacional e o conhecimento ser compartilhado através de todas as disciplinas.

O NUDL permite a configuração de uma rede de bibliotecas digitais de teses e dissertações através do *Network Digital Library of Theses and Dissertations (NDLTD)*²⁸, uma

²⁸<http://www.ndltd.org/>

iniciativa internacional para melhorar a educação, permitindo que os estudantes publiquem suas teses e dissertações em meio eletrônico. NUDL se propõe a atender os seguintes objetivos:

- 1) fazer com que o trabalho dos pesquisadores escritos em diferentes línguas possam ser publicados e disseminados como fonte de pesquisa para as demais comunidades de pesquisas;
- 2) estar interligado com diversos outros centros de pesquisa tais como NCSTRL <http://www.ncstrl.org> e CoRR <http://xxx.lanl.gov/archive/cs/intro.html> para arquivos de reprints/preprints na troca de informações; links com NSDL <http://www.nsdlnsf.org>, anteriormente SMETE-lib, para a ciência, a matemática, a engenharia, e as tecnologias na educação; e a *Open Archives Initiative* visando a interoperabilidade de dados e informações;
- 3) preparar as gerações futuras de alunos oferecendo conhecimento necessário para a realização de suas tarefas e que eles mesmos possam submeter pessoalmente seus próprios trabalhos de T&D's para publicação.

5.1 NDLTD – A *Global Digital Library Of Electronic Theses And Dissertations*

NDLTD, diferentemente de outras bibliotecas digitais, foi projetado para que as universidades e seus alunos possam publicar suas T&D's na *Web* através de um sistema de submissão, onde os próprios pesquisadores publicam suas pesquisas com um baixo custo e com garantias quanto a autenticidade e propriedade da pesquisa. Usando o protocolo OAI para arquivamento das informações, a Universidade da Virginia, através da NDLTD, estabelece uma rede de cooperação regional, nacional e internacional contribuindo para o avanço da tecnologia e da ciência do país.

A concepção do *electronic theses and dissertations* (ETDs) foi inicialmente discutida em 1987 em uma reunião com membros de universidades e centros de pesquisa com representantes da *Virginia Tech* (Edward Fox da Computer Science e Susan Bright da Computing Center), *University of Michigan*, *SoftQuad*, e *ArborText*. Em seguida a *Virginia*

Tech desenvolve, a partir da linguagem SGML, uma definição de documento com vistas para a publicação de ETD's, o *Document Type Definition*. Edward Fox e John Eaton têm colaborado desde o início no projeto, investigando problemas relativos à produção, arquivamento e acesso aos dados dentro da comunidade local. Desde 1992 o grupo tem trabalhado com o *Coalition for Networked Information* (CNI), com o *Council of Graduate Schools* (CGS), UMI e outras organizações, colaborando nos projetos e participando das reuniões sobre o assunto.

Adicionalmente, a *University Library's Scholarly Communications Project*, desenvolveu procedimentos e sistemas para processamento e arquivamento fornecendo acesso público para pesquisa aos alunos de graduação da *Virginia Tech*. Para ajudar na implementação do projeto, em 1996 foi definido que o *Southeastern Universities Research Association* (SURA)²⁹, com base na *Virginia Tech*, seria responsável pela pesquisa, desenvolvimento e a disseminação na *Web*.

Objetivos considerados pelo projeto NDLTD:

- os estudantes devem aprender sobre bibliotecas digitais, aplicando esse conhecimento para buscar informações para sua pesquisa, construindo e submetendo sua própria ETD para a publicação;
- as universidades devem aprender sobre bibliotecas digitais, coletando, catalogando, arquivando e disponibilizando as ETD para pesquisa;
- as universidades possam aprender e desenvolver o potencial de suas propriedade intelectual e de produções;
- a melhoria do nível do ensino com compartilhamento mais eficaz do conhecimento e,
- a disponibilidade imediata do resultado das pesquisas e o conhecimento poder ser compartilhado por todos os alunos e a comunidade.

5.1.1 Construindo uma biblioteca digital

²⁹<http://www.ndltd.org/news/sura.en.html>

A *Virginia Tech University*, através do NUDL, desenvolveu uma biblioteca digital para armazenar as ETD's e as atividades de pesquisas desenvolvidas pelos membros do NDLTD, além das demais instituições consorciadas. O *software* desenvolvido pelo NUDL, usa o banco de dados MySQL (servidor de banco de dados SQL multiusuário e *multi-threaded*), para armazenar as informações catalogadas. Consiste em uma série de *Web Pages* e de scripts em PERL (linguagem de programação estável e multiplataforma, usada em aplicações críticas em diversos setores, com destaque em administração de sistemas, manipulação de textos, processamento de linguagem natural, programação *Web* e biotecnologia), que fornecem uma interface padrão para usuários da *Web*, pesquisadores, autores de ETD's, equipes de programadores e os bibliotecários para catalogar e gerenciar arquivos e metadados relacionados a uma coleção eletrônica de ETD. O banco de dados da BD inclui interfaces e funcionalidades básicas para buscas de texto completo baseado em palavras chaves bem como navegação local pelo departamento da universidade ou pelo autor do trabalho.

Este *software* e sua documentação correspondente, em versões melhoradas, são empacotados periodicamente e disponibilizados sem custo para outras instituições cooperantes do NDLTD. Recentemente o *software* foi atualizado para suportar funções do *Open Archives Initiative* disponibilizando bibliotecas de *software* para integração com o protocolo OAI-PMH permitindo colher automaticamente os metadados de T&D's dos novos membros integrantes da rede NDLTD. Outras ferramentas relacionadas OAI estão também disponíveis para aquelas que usam pacotes de *software* alternativos.

As submissões de T&D's são controladas por um sistema do *workflow*, que guiam e direcionam o autor em todas as etapas do processo. Envolve a gravação do arquivo da T&D em um banco de dados, fornecendo metadados on-line e em um formato padrão para publicação do trabalho na BD da universidade. O NDLTD suporta e permite a submissão de trabalhos em um pequeno conjunto de formatos padrão, incluindo XML, SGML, pdf, e latex bem como formatos padrão de *multimedia*. São usados formatos de *Document Types Definitions* de linguagens de marcação para troca de informação como o XML e SGML, e um esquema de XML para T&D's também está disponível, além de instruções para a conversão dos formatos não suportados pelo OAI.

O padrão de metadados desenvolvido pelo NDLTD para descrever T&D's é baseado no elemento do núcleo de *Dublin Core* incluindo elementos adicionais específicos para T&D's. Incorporando campos especiais para representar a informação como por exemplo,

membros de conselho e de comitê, o grau e o nível de instrução associados com o trabalho, instituição que concede o grau e a área de estudo da pesquisa. Foi projetado para, também, assegurar que metadados de diferentes línguas seja incluído a respeito do mesmo trabalho

5.1.2 NDLTD como uma federação de bibliotecas digitais

NDLTD é uma federação de Bibliotecas digitais com esforços relacionados à publicação de T&D's. Um dos objetivos principais da federação é fornecer serviços transparentes e integrados aos membros do NDLTD os quais devem manter suas próprias coleções de T&D's individualizadas, porém integradas na federação. A procura e a navegação no ambiente, constituem alguns destes serviços básicos. Um dos primeiros projetos construídos com estes objetivos foi o sistema federado da busca que está disponível no *site* do NDLTD³⁰.

As perguntas submetidas aos membros locais do NDLTD são transformadas em argumentos de busca para pesquisa nos vários sistemas da federação. Os resultados de cada local visitado são tratados, porém, nem todos são devidamente tratados devido a complexidade de fundir resultados da busca sem conhecimento da apresentação subjacente. O sistema sofre também da latência elevada da rede e da disponibilidade incerta dos usuários.

Para superar esses problemas, foi adotado uma nova solução baseada "em uma arquitetura do catálogo união", usando protocolo da iniciativa dos arquivos abertos que colhe os metadados em formatos específicos tornando acessível em portais espelhados. Esta solução tem a vantagem que em uma única base de dados onde podendo oferecer vários serviços os quais podem ser construídos sobre a mesma base, porém requerendo um papel mais ativo e mais colaborativo dos membros de NDLTD.

Os portais atuais oferecem uma relação na *Web* ao catálogo de união de *Electronic Thesis and Dissertation* incluindo o sistema de *Virtua* trabalhando em uma modalidade de produção e o sistema de biblioteca digital de MARIAN para a investigação de tendências de novas pesquisas e serviços de T&D's como é mostrado na figura 5.

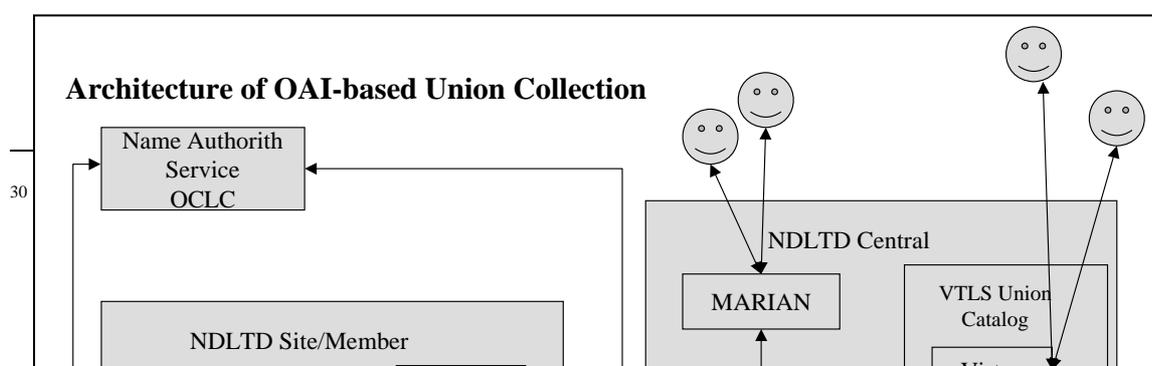


Figura 5 – Arquitetura de OAI do NDLTD³¹

5.1.3 O Sistema MARIAN de armazenamento e pesquisa

O MARIAN é um sistema de biblioteca digital projetado e construído para armazenar, pesquisar e recuperar um grande número de objetos digitais complexos. Originalmente foi desenvolvido para catálogos de bibliotecas para recuperação on-line, obteve sucesso para coleções de tamanhos e de estruturas de dados diferentes sendo posteriormente melhorado para dar suporte a bibliotecas digitais e *Web* semântica onde a informação é dada com um significado bem definido, permitindo melhor interação entre os computadores e as pessoas (FOX & GONÇALVES, 2001).

O modelo dos dados de MARIAN combina três conceitos fortes. Primeiramente, a estrutura e os relacionamentos das coleções de MARIAN são capturados em formulário de uma rede de informação explícita de nós e ligações. Os modelos gráfico-baseados similares são eficientes em recuperar dados originais e semi-estruturados de documentos da *Web* e entre sistemas de diferentes bibliotecas digitais. Em segundo, MARIAN expande este modelo insistindo que os nós e as ligações de uma coleção representada graficamente podem ser membros de classes de objetos.

³¹Fox & Gonçalves, 2001

As classes são organizadas em métodos similares e ligadas por pontos em grafos semi-estruturados, sendo estritamente mais poderosas porque dão forma a uma rede completa de subconjuntos, suportando herança de comportamento. Além disso, desde que os nós no gráfico da coleção são exemplos de classes do objeto da informação, podem também suportar comportamentos de objetos complexos, entende-se neste caso como objetos contendo imagens e sons. Em terceiro lugar os nós ou as ligações podem representar descrições mais complexas de busca. MARIAM é especialista em pesquisas distribuídas buscando informações sobre grandes objetos de informação. A figura 6, acima, mostra um exemplo da representação semântica da rede de MARIAN para o padrão de metadados de uma biblioteca digital de *Electronic Thesis and Dissertation Metadata Standard* (ETD-MS).

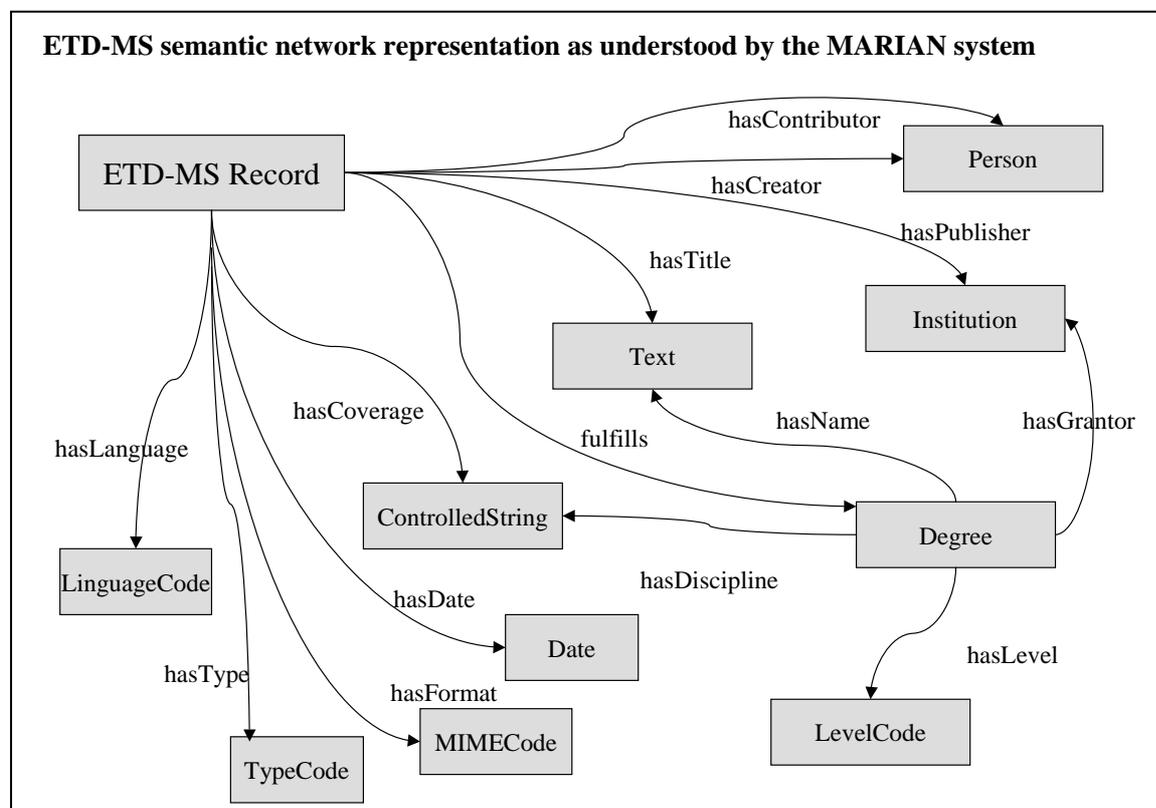


Figura 6 – Sistema MARIAN em uma rede semântica de recuperação de ETD-MS³²

5.2 Projetos institucionais no Brasil de bibliotecas digitais de T&D's

No Brasil, algumas Universidade e Instituições de Pesquisa estão concentrando esforços no sentido de disponibilizar na *Web* o conhecimento produzido internamente para

³²Fox & Gonçalves, 2001

que o mesmo possa ser usado por outras instituições formando uma rede de pesquisa. Podemos relacionar a iniciativa do IBICT, que, coordena o projeto de implantação da Biblioteca Digital Brasileira (BDB), cujo foco está na integração dos mais importantes provedores de dados e de serviços de informações em C&T do país. A integração desses provedores de dados acontece em torno das questões prioritárias para a BDB, que são a publicação e a disponibilidade de textos completos e outros objetos digitais na *Internet* e a interoperabilidade entre os diversos sistemas/serviços de informação participantes através de um portal único de acesso, preservando-se a independência e peculiaridades de cada sistema/serviço participante.

Para conseguir a cooperação dos provedores de dados, a especificação do conjunto de metadados, a configuração do ambiente, os padrões e procedimentos por parte dos provedores de dados visando à interoperabilidade com a BDB deverão ser os mais simples e menos onerosos e ter a garantia de total independência em relação aos demais provedores.

5.2.1 Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

O IBICT coordena o projeto da Biblioteca Digital de Teses e Dissertações (BDTD), cujo objetivo é integrar os sistemas de informação de teses e dissertações existentes nas Instituições de Ensino Superior (IES) brasileiras, além de estimular o registro e a publicação de teses e dissertações em meio eletrônico. A BDTD é uma das ações do programa da BDB, com o apoio financeiro da Financiadora de Estudos e Projetos (FINEP).

O projeto teve início em 2001 quando foi instalado um comitê técnico consultivo (CTC), constituído por representantes do IBICT, CNPq, Ministério da Educação e Cultura (MEC) através da CAPES e da Secretaria de Ensino Superior (SESU), a FINEP e três universidades: USP, PUC-RIO e UFSC. O CTC é um colegiado que tem por objetivo referendar o desenvolvimento da BDTD, assim como, atuar na especificação de padrões a serem adotados no âmbito do sistema da BDTD. Em particular, o comitê apoiou e aprovou o padrão brasileiro de Metadados para Teses e Dissertações (MTD-BR), (MAFFIA,2003).

Em 2002, é apresentado o modelo final do projeto da BDTD, com dois objetivos a saber: O registro bibliográfico e a publicação eletrônica de T&D's existentes nos acervos das IES brasileiras. Ao desenvolver o projeto, o IBICT amplia a abrangência da BDTD e disponibiliza para os usuários um catálogo coletivo nacional de T&D's através de um portal único de busca de informações.

O papel do IBICT na rede é de agregador, participando tanto como provedor de dados e provedor de serviço. Como provedor de serviços, o IBICT mantém a BDTD nacional, a qual é composta pela integração das iniciativas nacionais de registro bibliográfico e publicação T&D's. Para a integração dessas iniciativas a BDTD utiliza padrões de metadados e tecnologias dos arquivos abertos.

Mesmo considerando o avanço acelerado das tecnologias de informação e comunicação, o corpo técnico do IBICT preocupou-se em usar no projeto tecnologias consolidadas, cujo grau de estabilidade e confiabilidade tenham sido comprovados no país e no exterior, e que, prioritariamente, tenham sido aplicados nas principais experiências internacionais levando em conta a facilidade de uso dessas tecnologias, a facilidade de implantação e manutenção pelas comunidades de pesquisa.

Em função da diversidade de *software* livre, sempre que possível foi usado *software* de domínio público, considerando a qualidade, a documentação sobre o *software* e a facilidade de manutenção dos mesmos. Após a definição do uso dos arquivos abertos no ambiente e do estabelecimento de contatos com especialistas envolvidos nesta área, o IBICT testou alguns *softwares* para implementação da solução final de comunicação científica. Dentre os programas analisados, o *software e-print* caracterizou-se como o mais adequado. O *e-print* é um *software* desenvolvido pela Universidade de Southampton no Reino Unido e é possível verificar o seu funcionamento do repositório *CogPrints Cognitive Sciences Eprint Archive* e *California Digital Library*.

O *software* da BDTD foi desenvolvido com flexibilidade e exigiu o mínimo de esforço de configuração e adaptação. Conforme descrito anteriormente, a versão genérica do *software e-print* é interoperacional com todos os protocolos OAI, isso significa que a informação pode ser recuperada através do protocolo OAI via um "arquivo virtual" global e para tal deve-se usar um provedor de serviço, no caso um *Open Archives Service Providers* como o serviço ARC, por exemplo. Os serviços de recuperação de documentos oferecidos pelo *software e-print* são:

- serviço de alerta (*subscription*);
- serviço de busca por área de conhecimento e o;
- serviço de busca (*search*) simples e avançada e serviço de busca pelo número do documento no repositório.

O modelo de interoperabilidade proposto para a BDB aproxima-se bastante dos modelos do portal da ND LTD (SULEMAN, 2001) e do Arc - *Cross Archive Searching*

*Service*³³. Ambos os sistemas fazem colheita de metadados de provedores de dados, alimentando uma base de dados central de metadados usando o esquema de provedor de serviço e provedor de dados. A proposta do projeto é tornar o portal da BDB na *Internet* a materialização da BDB em C&T a nível nacional. Um *site* que, através de diferentes mecanismos de interoperabilidade, deverá oferecer ao pesquisador acesso unificado e integrado a diferentes recursos de interesse para C&T, mesmo que estes recursos sejam heterogêneos e se encontrem globalmente distribuídos, sem a necessidade de navegar e consultar cada recurso individualmente.

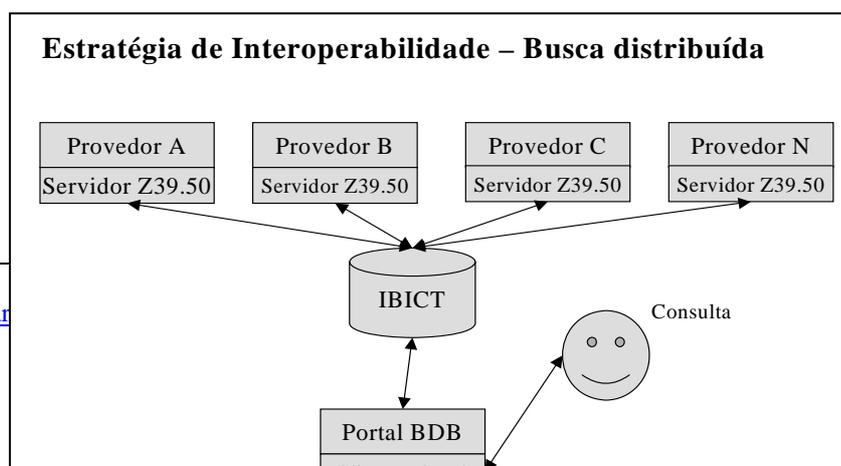
Entre estes recursos estão incluídos, na versão inicial do portal, periódicos eletrônicos brasileiros que fazem parte do portal SciELO, periódicos mantidos pela Bireme, anais eletrônicos de eventos brasileiros em C&T a serem disponibilizados pela CNEN/CIN, bancos de T&D's publicados na USP, UNICAMP, UFSC, PUC-Rio, ENSP/Fiocruz, repositórios de arquivos *e-prints* brasileiros que começam a ser disseminados na *Internet*, como o do IMPA.

O modelo de interoperabilidade da BDB, está baseado em dois elementos:

- mecanismo de submissão de consultas, que propicia buscas a partir de uma interface única do portal acessando diferentes recursos informacionais que compõem a BDB;
- conjunto de metadados que descrevem e fornecem uma visão unificada dos diferentes conjuntos de documentos.

Através dos **mecanismos de submissão de consultas**, a BDB incorpora as principais alternativas tecnológicas analisadas anteriormente, buscas distribuídas e busca em uma base centralizada de metadados obtida mediante coleta automática de dados através de *harvesting*.

Para isso, o portal da BDB dispõe de um programa cliente Z39.50, que permite acesso integrado por meio da distribuição de consultas aos catálogos na *Internet* das principais bibliotecas universitárias do país e estrangeiras servidas pelo protocolo Z39.50. Qualquer outro recurso informacional na *Internet* que seja servido por este protocolo também poderá ser acessado do portal da BDB, como, por exemplo, um arquivo de *e-prints* ou do portal Scielo, o qual planeja implementar um servidor Z39.50 para sua base. Na figura 7 é mostrado o modelo geral de interoperabilidade da BDB.



³³<http://www.ar>

Figura 7 – Estratégia de interoperabilidade da BDB – Busca distribuída³⁴

Apesar da BDB recuperar recursos informacionais, via consultas distribuídas, usando o serviço local o Z39.50, a BDB mantém em seu *site* uma base comum de metadados, obtida pelo processo de *harvesting* dos metadados. Estes recursos ou serviços são periodicamente coletados através do OAI-PMH e dos provedores de dados compatíveis com este protocolo. Os metadados coletados via FTP, arquivos HTML ou arquivos texto também fazem parte da solução, esses formatos não são compatíveis com o Z39.50 ou com o OAI.

No caso dos arquivos coletados via FTP e HTML, os metadados são processados e armazenados na base de metadados diretamente no *site* da BDB onde, com um servidor do protocolo Z39.50 local as informações ficam acessíveis a partir do portal da BDB como qualquer outro recurso servido por este protocolo.

Com esta possibilidade, arquivos acadêmicos de *preprints* eletrônicos podem ser implantados em departamentos de instituições de ensino superior, institutos de pesquisa, sociedades científicas, periódicos eletrônicos, as publicações on-line da revista Ciência da Informação do IBICT, ou projetos específicos interinstitucionais, como Projeto Genoma, poderão se integrar à BDB. A figura 8 mostra a integração desses recursos.

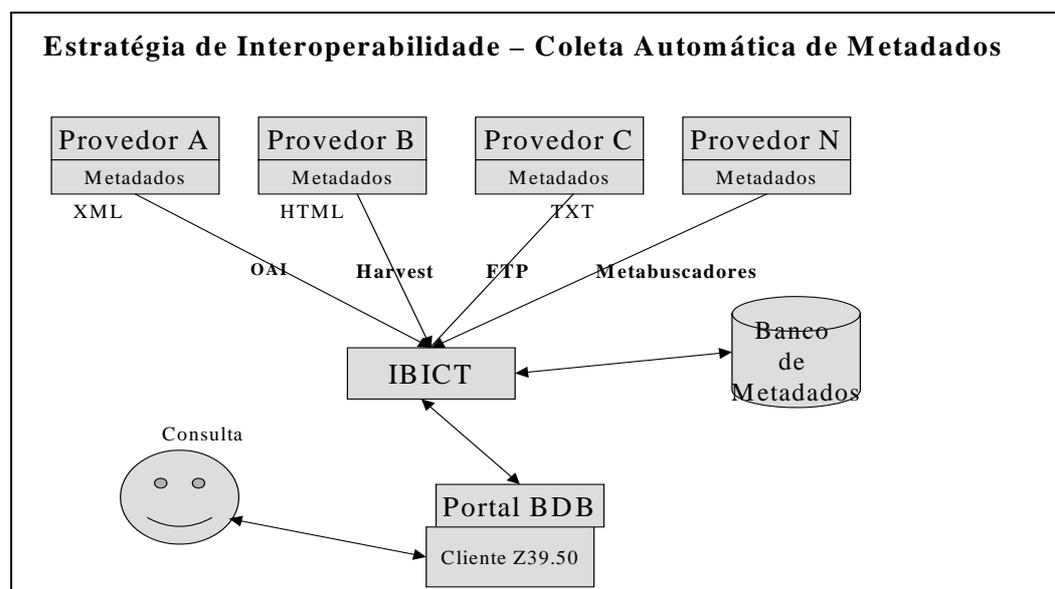


Figura 8 – Estratégia de interoperabilidade da BDB – Coleta Automática³⁵

O segundo elemento do esquema de interoperabilidade da BDB é o **conjunto de metadados** para a descrição das diferentes tipologias de documentos originários dos diferentes recursos informacionais que compõem a BDB. O padrão utilizado é o vocabulário *Dublin Core* que suporta qualificadores para especificar o significado de um elemento e pode ser codificado em formatos como HTML e XML. O vocabulário *Dublin Core* é considerado padrão internacional para descrição de recursos, além da simplicidade na definição do conjunto de metadados que permitindo que o próprio autor ao submeter seu trabalho preencha um formulário com os metadados relativos a pesquisa.

O principal serviço atualmente disponível na BDTD é o de busca e recuperação de documentos de T&D's produzidos no Brasil ou produzidos por brasileiros no exterior. Usuários interessados em encontrar T&D's relevantes aos seus problemas poderão, a partir do portal da BDTD, realizar buscas nessas em bibliotecas digitais sem que para tal tenham que visitar cada uma delas individualmente nos repositórios locais dessas instituições provedoras de dados.

A BDTD trabalha com entidades cooperantes que produzem as informações nos seus provedores de dados. O IBICT colhe os metadados desses provedores de dados e insere na base de dados da BDTD nacional a qual contém os dados de todos os provedores de dados que estão na rede ao mesmo tempo as informações ficam disponíveis para a colheita pelo NDLTD na *Virgínia Tech University*. Na figura 9 onde é mostrado o link das Instituições com o IBICT e com o NDLTD.

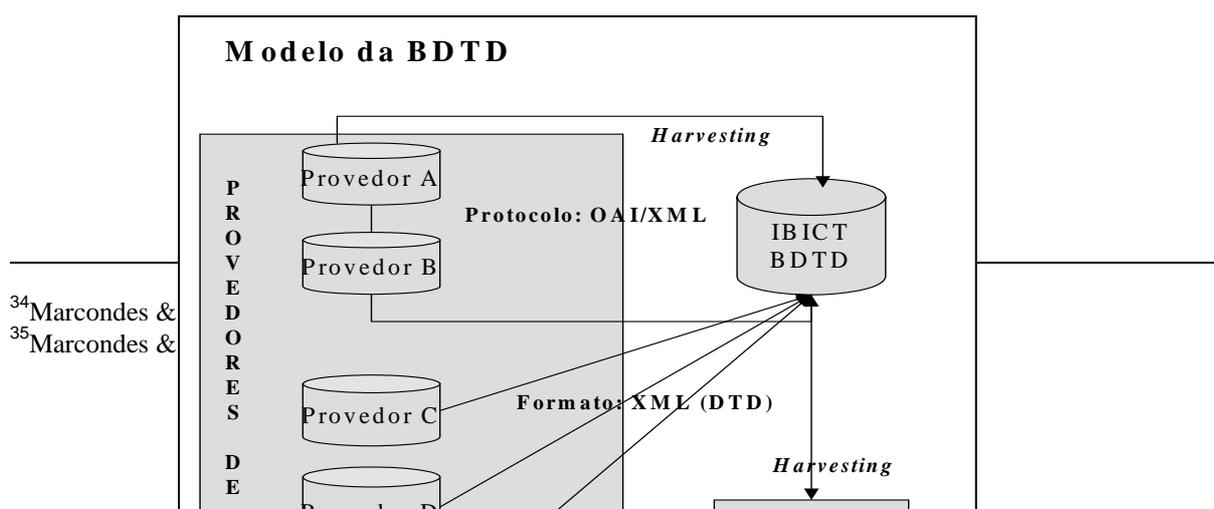


Figura 9 – *Link* das Instituições de ensino com o IBICT e com o NDLTD

A maioria dessas instituições cooperantes da BDTD, não usa o protocolo OAI e sim padrões próprios de catalogação e publicação na *Web*. Abaixo as Instituições de ensino superior que já publicam suas T&D's em BD's e são integrantes do sistema BDTD gerenciado pelo IBICT:

- Biblioteca digital da Universidade Federal do Rio Grande do Sul - <http://www.biblioteca.ufrgs.br/bibliotecadigital//>
- Biblioteca digital da USP - <http://www.teses.usp.br//>
- Biblioteca digital da PUC-Rio - http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/db2www/PRG_0490.D2W/INPUT?CdLinPrg=pt/
- Instituto Nacional de Pesquisas Espaciais INPE – <http://www.inpe.br//>
- Pontifícia Universidade Católica de São Paulo PUCSP - <http://www.pucsp.br//>
- Pontifícia Universidade Católica do Rio Grande do Sul PUCRS - <http://www.pucrs.br//>
- Pontifícia Universidade Católica do Rio de Janeiro PUC-Rio - <http://www.puc-rio.br//>
- Universidade Católica de Brasília UCB - <http://www.ucb.br//>
- Universidade Católica de Pernambuco UNICAP - <http://www.unicap.br//>
- Universidade Estadual de Campinas UNICAMP - <http://www.unicamp.br//>
- Universidade Federal de Lavras UFLA - <http://www.ufla.br//>
- Universidade Federal de Pernambuco UFPE - <http://www.ufpe.br//>
- Universidade Federal de São Paulo Escola Paulista de Medicina UNIFESP/EPM - <http://www.unifesp.br//>

- Universidade Federal do Paraná UFPR - <http://www.ufpr.br//>
- Universidade Federal do Pará UFPA - <http://www.ufpa.br//>
- Universidade Federal do Rio Grande do Sul UFRGS - <http://www.ufrgs.br/universidadeviva//>
- Universidade Federal do Rio de Janeiro UFRJ - <http://www.ufrj.br//>
- Universidade de Brasília UnB - <http://www.unb.br//>
- Universidade de São Paulo USP - <http://www2.usp.br/publishing/insite.cgi/>

Na tabela 1 podemos ver a situação atual da publicação e coleta de metadados de T&D's no acervo da BDTD do Brasil, separados por repositório em 10 de abril de 2005.

Na tabela 2 podemos ver a contagem dos registros de metadados em processo de colheita antes da importação para a BDTD. Metadados brutos coletados em 10 de abril de 2005.

Tabela 1 – Metadados já importados na BDTD.

Instituição	T&D
Teses Exterior (IBICT)	95
Inatel	4
PUC Rio	1594
Uni. Católica de Brasília	89
Uni. Católica de Pelotas	5
Uni. Fed. Fluminense	56
Uni. Fed. De Sergipe	12
Uni. Fed. De Santa Catarina	1894
Uni. Fed. De São Carlos	273
UNICAMP	4223
Uni. Católica de Pernambuco	26
Uni. Do Vale do Rio dos Sinos	19
USP	2474

Fonte: IBICT

Tabela 2 – Metadados coletados por *Harvesting* na BDTD

Instituição	T&D
Teses Exterior (IBICT)	122
Inatel	3
INPE	176
PUC Rio	1594
Uni. Católica de Brasília	91
Uni. Católica de Pelotas	5

Uni. Fed. Fluminense	58
Uni. Fed. De Sergipe	12
Uni. Fed. De Santa Catarina	1894
UNICAMP	4226
Uni. Católica de Pernambuco	26
Uni. Do Vale do Rio dos Sinos	19
USP	2512

Fonte: IBICT

5.2.2 O sistema TEDE

Atualmente, com o objetivo de apoiar as IES que ainda não implementaram bibliotecas digitais de teses e dissertações localmente, o IBICT coordena a implantação do Sistema de Publicação Eletrônica de Teses e Dissertações, TEDE, desenvolvido para que as IES possam através dele efetuar a publicação eletrônica de suas T&D's produzidas localmente e gerar suas bibliotecas digitais locais.

O Sistema TEDE e a camada do protocolo OAI-PMH, que o acompanha, permite às IES utilizar os padrões e tecnologias necessárias para a sua imediata integração à BDTD, cujo objetivo, conforme descrito anteriormente, a integração de todas as iniciativas brasileiras de publicação eletrônica na *Web* e registro bibliográfico de T&D's e a exposição dos metadados destas T&D's na NDLTD da *Virginia Tech University*.

O ambiente de submissão foi desenvolvido para ser compatível com o *Open Public Access Catalog* (OPACs). Com essa ferramenta as IES poderão efetuar a publicação eletrônica de suas T&D's pelo Sistema TEDE formando uma grande rede de publicação e de disseminação do conhecimento, e, posteriormente, por meio de um módulo específico do sistema, exportar os metadados para o OPAC.

Usuários interessados em encontrar T&D's relevantes aos seus problemas de informação poderão, a partir de um portal único, no caso podemos considerar o portal da BDTD para realizar consultas nessas diversas iniciativas sem que para tal tenham que visitar cada uma delas individualmente e realizar buscas nos repositórios locais dessas instituições provedoras de dados. O sistema possui um módulo específico para exportar os metadados para o OPAC.

Na sua concepção, é utilizado o padrão de metadados MTD-BR compatível com os padrões ETD-MS específico para T&D's e o vocabulário *Dublin Core* para a descrição de

qualquer tipo de objeto digital usado no trabalho. O padrão MTD-BR foi escolhido por propiciar aproveitamento de sistemas legados e permitir interligação com outros *Webservices* como por exemplo a plataforma Lattes do CNPq.

O sistema consiste de uma aplicação desenvolvida para *Web*, implementado em HTML usando banco de dados e apresentando uma interface amigável com um sistema de navegação orientado por menus. A navegação entre os módulos de operação do sistema é bastante simples e conta com ajuda automática na opção de operação de cada módulo.

A solução desenvolvida é composta por dois sistemas, TEDE Modular e TEDE Simplificado. O Sistema TEDE Modular requer a interação entre o curso de pós-graduação, o autor da tese e a biblioteca central da IES durante o processo de publicação eletrônica da T&D. Construído em módulos, possui funções específicas para cada etapa do processo de publicação do trabalho. O Sistema TEDE Simplificado, possibilita a publicação eletrônica da T&D pela IES diretamente pela biblioteca central e neste caso a biblioteca central faz a catalogação da T&D e publica na *Web* sem as informações de responsabilidade da pós-graduação e do autor, como banca, fomento e os dados completos da pesquisa.

O TEDE modular e o simplificado atendem ao mesmo objetivo proposto pelo IBICT, qual seja, a integração das iniciativas brasileiras de publicação eletrônica e registro bibliográfico de T&D's.

O sistema TEDE foi desenvolvido baseado nos seguintes objetivos:

- incrementar a disponibilidade de conteúdos de teses e dissertações, ampliando o acesso às mesmas;
- apoiar as universidades no aperfeiçoamento da educação superior;
- educar os futuros profissionais / pesquisadores no uso de ferramentas para publicação eletrônica e no uso efetivo de bibliotecas digitais;
- melhoria dos serviços oferecidos;
- informação mais rapidamente disponível para os usuários;
- informação sempre disponível;
- redução de trabalho;
- catalogação de documentos eletrônicos;
- eliminação do manuseio do documento
- eliminar a necessidade de espaço nas estantes para o armazenamento de teses e dissertações.

5.2.2.1 Procedimento e uso do sistema TEDE

No Sistema TEDE, a publicação da T&D é automatizada de forma cooperativa envolvendo autor, curso de pós-graduação e a biblioteca. Para instalação do TEDE, a instituição deve disponibilizar um computador com o sistema operacional baseado na plataforma *UNIX* além dos seguintes *softwares* básicos:

- Servidor Apache: servidor *Web* que funciona em qualquer sistema operacional e que permite que linguagens de programação sejam agregadas a ele como módulos, o que gera mais segurança, estabilidade e performance.
- Servidor de banco de dados MySQL: servidor de banco de dados SQL multiusuário e multi-*threaded*. Sua implementação consiste em um *daemon* (programa residente em memória) servidor chamado "mysqld" e diversos programas clientes e bibliotecas. Suas principais metas são velocidade, robustez e facilidade de uso. A base sob a qual o MySQL foi construído é formada por um conjunto de rotinas que foram utilizadas em ambiente de produção com alta demanda por muitos anos. O banco de dados MySQL esta sempre em desenvolvimento, e oferece um conjunto de funções que atende as necessidades do TEDE.
- Linguagem PHP: *Hypertext Preprocessor* - É uma linguagem de *server-side* e *open-source* para criação de páginas *Web* de *e-commerce* dinâmicas e outros aplicativos para a *Web*. Uma página *Web* dinâmica é aquela que apresenta conteúdo de banco de dados sob demanda.

O TEDE Modular é composto pelos módulos: administração, pós-graduação, autor, biblioteca e busca de documentos. A preparação do *software* para utilização é sempre através do módulo administração, a customização envolve o cadastro dos dados específicos da instituição receptora do sistema assim como de suas unidades, tais como, programas de pós-graduação, das bibliotecas depositárias e dos operadores do sistema.

A customização também envolve escolha de interfaces para o sistema local, inclusão da logomarca da instituição, habilitação dos operadores dos sistemas e a identificação dos servidores de arquivos e de *e-mail* geral utilizados pelo sistema TEDE. Após a customização, o sistema está preparado para iniciar os procedimentos de publicação das teses e dissertações da IES.

Cada módulo tem funções específicas para cada etapa do processo de publicação de uma T&D e eles devem ser executados na seqüência a saber: administração; pós-graduação;

autor; biblioteca e busca. Com exceção dos módulos autor e busca, os módulos administração, pós-graduação e biblioteca devem ter operadores habilitados pela instituição para executá-los.

Após o ambiente liberado, a publicação de uma T&D sempre deve ser iniciada no módulo da pós-graduação. Neste módulo, o operador habilitado para tal inclui dados do autor, título do trabalho, membros da banca e o papel dos colaboradores.

Depois do cadastro da T&D pela pós-graduação, o operador do sistema cria uma conta para o autor. Somente após executar esse procedimento é que o autor tem acesso ao sistema para entrar ou alterar os dados de seu trabalho e também para anexar os arquivos de texto, imagem, animação, etc. que o compõem. No final deste processo, o autor submete a T&D à pós-graduação para revisão de dados. Caso seja detectado algum erro nos dados ou arquivos inseridos pelo autor, o operador do módulo de pós-graduação insere um texto no campo especificado para tal, indicando o erro detectado. Esta observação é encaminhada via e-mail para o autor para as devidas correções. O autor volta a operar o sistema, efetua as correções e novamente submete seu trabalho para a pós-graduação.

Essa iteração entre pós-graduação e autor ocorre até que a pós-graduação considere os dados e arquivos da T&D como corretos. Depois do fechamento, o trabalho pode ser liberado para a biblioteca depositária onde o operador do módulo insere metadados bibliográficos para complementar a catalogação fazendo uma revisão geral dos metadados já inseridos nas fases anteriores. Caso ainda detecte erros, a biblioteca poderá devolver a T&D para as devidas correções.

No momento em que uma T&D é publicada pela instituição provedora de dados, o IBICT através da BDTD realiza a colheita dos metadados e a mesma passa a estar disponível na *Web* para busca, por meio do módulo correspondente ou pelo portal da BDTD.

Na figura 10 temos uma representação da iteração do *workflow* definido entre o autor do trabalho com o curso de pós-graduação, o curso com a biblioteca central da instituição e a biblioteca com o repositório local de trabalhos publicados.

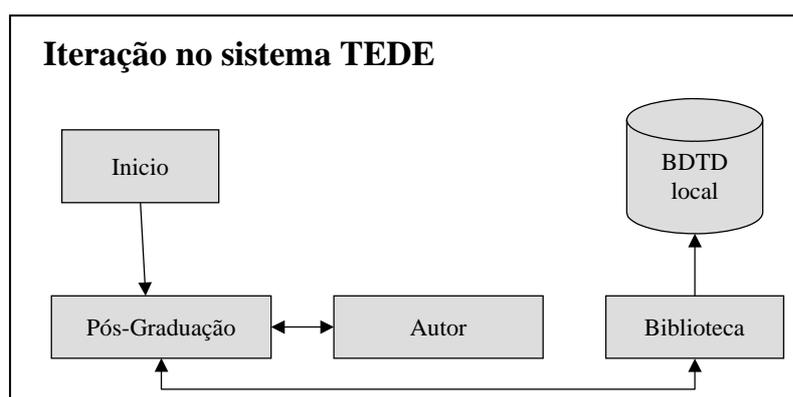


Figura 10 – Iteração no sistema TEDE

O *workflow* necessário para a publicação de T&D's no TEDE Modular, determina o papel de cada módulo dentro do sistema: o autor, antes de iniciar o processo de submissão, deve preencher o formulário de autorização para a publicação eletrônica de sua T&D e contatar a secretaria de seu curso de pós-graduação; a pós-graduação gerência os dados dos alunos, suas contas de acesso ao sistema e gerência dos dados dos contribuidores, além de revisão e aprovação final das T&D's; a biblioteca realiza a catalogação e liberação das T&D's, previamente aprovadas pelos cursos de pós-graduação, para o acesso público; a administração do sistema, customiza e habilita as contas de acesso para os cursos de pós-graduação e bibliotecas.

5.2.3 A biblioteca digital de T&D's da Universidade de São Paulo

Em junho de 2001 a USP implantou a sua biblioteca digital de teses e dissertações objetivando facilitar o acesso remoto a base de conhecimento produzida pela Instituição. Hoje, a USP possui um dos maiores sistemas de pós-graduação do país produzindo anualmente em média 1.500 teses de doutorado e 2.600 dissertações de mestrado nos diversos programas de pós-graduação³⁶. A biblioteca digital engloba T&D's nas áreas de humanas, exatas e biológicas, com diferentes estruturas e conteúdos, das mais simples, somente texto até aquelas mais complexas com vídeos e imagens (MASIERO, BREMER & COLLETA, 2001).

O projeto da biblioteca digital foi motivado a nível institucional, através da publicação de uma portaria do Reitor constituído a uma comissão de implementação da biblioteca digital de T&D's da USP com o envolvimento de várias pessoas no processo. O projeto teve a duração de um ano onde a Pró-Reitoria de Pós-Graduação atuou como intermediária entre a comissão e o conselho de pós-graduação contando com total apoio jurídico para a definição quanto à legislação, principalmente em relação à questão dos direitos autorais (MASIERO, BREMER & COLLETA, 2001).

³⁶<http://www.teses.usp.br/>

Os arquivos armazenados na biblioteca digital são de formato PDF e o *software* aplicativo utilizado é do NDLDT da *Virginia Tech University*, iniciativa descrita no capítulo anterior. A USP instalou e adaptou o código fonte de módulos de *software* desenvolvidos pelo NDLDT através de uma associação com a *Virginia Tech University*.

Como o *software* não disponibiliza um robô de busca, a USP utiliza o *software* Insearch, da INSITE³⁷, que concedeu, em convênio, o direito de uso gratuito para a Universidade. Nesse processo, muitos componentes, inicialmente programados na linguagem PEARL, foram refeitos na linguagem PHP, novos componentes foram desenvolvidos pela equipe da USP, uma nova interface como também a interface com os demais sistemas existentes dentro da Universidade.

Enquanto o *software* era customizado pela equipe técnica, a USP se preocupava com os seguintes pontos:

- a definição de diretrizes para a elaboração de teses e dissertações, baseadas nos manuais já existentes na Universidade;
- definição do conjunto de metadados para a descrição bibliográfica das teses e dissertações, segundo os padrões do *Dublin Core*;
- aquisição e instalação, em todas as bibliotecas, de *softwares* necessários ao funcionamento da biblioteca digital.

A USP, visando à operação, evolução e garantia do sistema, definiu responsabilidades para os centros de informática, as bibliotecas e os programas de pós-graduação após a implantação.

O direito autoral teve ampla discussão e ocorreu entre a comissão, o conselho de pós-graduação e o jurídico da USP. Ficou definido que o autor da tese é o aluno e que, como autor, só ele pode autorizar a publicação da T&D na biblioteca digital. Para tal, foi elaborado um documento que deve ser analisado e preenchido pelo aluno antes de submeter o seu trabalho para a publicação na *Web*.

A comissão analisou a possibilidade dos usuários efetuarem pagamento para consultar ou fazer cópia das teses. Na medida, que, para a execução de uma tese ou dissertação o autor, no caso o aluno, a Universidade e, muitas vezes, órgãos públicos de apoio à pesquisa, principalmente a CAPES, o CNPq, a Fapesp, além do custo da Universidade para manter o sistema em operação, essa questão não teve relevância no processo de implantação da BD.

³⁷<http://www.insite.com.br>

Foram ouvidos alunos e orientadores sobre o processo de implantação da BD, na visão deste segmento, existem dúvidas em relação à tornar disponível o trabalho científico para um público potencialmente tão grande como o propiciado pela *Internet*, são elas:

- maior facilidade para que as teses sejam copiadas, aumentando os casos de plágio;
- muitos autores pensam em publicar suas teses em forma de livros, principalmente na área de ciências humanas, e isso poderia causar problemas de direito autoral, perda de receita etc.;
- muitos autores, principalmente nas áreas de ciências básicas, têm a preocupação de que se possa perder a originalidade do trabalho, prejudicando a publicação de artigos sobre a tese em periódicos, pedidos de patentes, registro de espécimes (biologia) etc.;
- de forma mais velada, alguns deixaram transparecer uma preocupação quanto à qualidade do trabalho, principalmente em relação à redação, o que ficaria mais exposto na biblioteca digital.

Além dos itens acima, os orientadores consideraram dois pontos que deveriam ser melhor trabalhados quanto ao conteúdo do texto digital armazenado pelo autor:

- quanto à mudança do texto após a defesa, isto é, sobre a possibilidade de que o aluno corrija a tese na versão digital, levando em conta as observações dos examinadores;
- o aluno submeta a tese com mudanças em relação ao exemplar defendido, sem autorização do orientador, ou submeta, sem perceber, um arquivo com erros.

O sistema é neutro em relação ao primeiro caso, que deve ser resolvido por uma decisão administrativa. Há várias alternativas: manter as versões digital e impressa da mesma forma, manter o exemplar impresso como foi defendido e inserir na biblioteca digital o exemplar corrigido etc. Para o segundo caso, o sistema fornece dois pontos de controle. O primeiro é efetuado pela seção de pós-graduação, que insere e confere os dados relativos à tese provenientes do sistema de pós-graduação (banca examinadora, data da defesa, título etc.) e nesse momento abre os arquivos submetidos pelo aluno e faz uma conferência superficial. O segundo é efetuado nas bibliotecas, que também abrem os arquivos quando inserem os dados bibliográficos e fazem uma conferência geral, antes de liberar a tese para consulta (MASIERO, BREMER & COLLETA, 2001).

5.2.4 Biblioteca digital de T&D's do PPGEP/UFSC

O Banco de Teses e Dissertações (BTD) do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina (PPGEP/UFSC), desenvolvido pelo Grupo Stela do PPGEP/UFSC, foi criado para atender além dos alunos pós-graduandos, professores e pesquisadores, também aos gestores do processo de construção do conhecimento, ou seja, os coordenadores de área, gestores de turmas, líderes de grupos de pesquisa e a coordenação do programa de pós-graduação da UFSC (PACHECO & KERN, 2001).

O PPGEP foi criado em 1970 e contempla até o presente momento aproximadamente 4900 trabalhos em texto completo de conclusão de mestrado e doutorado já defendidos. Todos os trabalhos concluídos já estão lançados na BTD³⁸, onde pelo site pode-se verificar a estatística por ano e grau de trabalho defendido. A BTD entrou em produção a partir de 1995 e foi elaborada para:

- interessados em assistir as defesas marcadas no PPGEP, para obtenção do título de mestre, doutor, ou para aprovação no exame de qualificação do doutorado: <http://teses.eps.ufsc.br/defesas.asp>
- interessados em contatar autores de trabalhos defendidos nos últimos quatro meses e em fase de preparação final pelos autores: <http://teses.eps.ufsc.br/defesas.asp>
- interessados em conhecer as teses e dissertações do PPGEP: <http://teses.eps.ufsc.br/tese.asp>
- alunos em fase final encaminham o pedido de defesa: <http://teses.eps.ufsc.br/pedidodefesa.asp>
- interessados em conhecer o perfil das defesas do Programa nos últimos anos <http://teses.eps.ufsc.br/estatisticas.asp>
- Interessados em outros links com T&D <http://teses.eps.ufsc.br/links.htm>

O BTD foi constituído com a finalidade de disponibilizar a memória das T&D's do PPGEP, destacando a transparência no processo de produção de conhecimento do programa. Devido a total falta de qualidade no serviço de rede, o governo do estado, elevou a velocidade

³⁸ <http://teses.eps.ufsc.br/IndicadoresTotais.asp>

de tráfego da rede beneficiando e possibilitando, além do projeto do BTD a implantação do programa de ensino a distância do PPGE/UFSC.

A digitalização dos arquivos de T&D's entregue pelos alunos no passado armazenadas em papel foram digitalizadas e convertidas para o formato de texto com marcação HTML. A primeira dissertação do PPGE/UFSC foi colocada no *site* do programa no início de 1996. Entre 1996 e 1999, foram manualmente transformadas e disponibilizadas no BTD 357 T&D's. Nesse período, lançaram-se projetos como o do *Massachusetts Institute of Technology*, ProQuest-UMI, *Virginia Tech* e TB-IBICT, entre outros (PACHECO & KERN, 2001).

O Grupo Stela, em 1999, desenvolveu o primeiro banco digital de teses em engenharia química, com informações de todos os programas da área no país sendo que, a manutenção do banco digital é de responsabilidade do Curso de Pós-Graduação em Engenharia Química da Universidade Federal de Santa Catarina (CPGEnq/UFSC). Este projeto serviu de base para a conformação do atual BTD do PPGE/UFSC. Em dezembro de 2000, o Grupo Stela reuniu o conjunto de arquivos digitalizados no BTD, unindo-os à Plataforma de Gestão Acadêmica do PPGE/UFSC por meio do enlace entre os arquivos e os correspondentes registros administrativos das T&D's defendidas no programa.

O *site* da biblioteca digital foi construído em sete áreas de recursos oferecidos aos usuários: informações gerais, defesas marcadas ou recentemente realizadas, acesso aos documentos e informações sobre trabalhos defendidos: teses, dissertações e exames de qualificação de doutorado no PPGE/UFSC, serviço de marcação de defesas por parte de alunos concluintes, estatísticas sobre os trabalhos de conclusão de curso no PPGE/UFSC, *links* relacionados aos assuntos do BTD e indicadores de acesso segundo diversos critérios de consultas.

A biblioteca oferece aos gestores do programa e aos interessados, diversos indicadores referentes às T&D's defendidas no PPGE/UFSC. Seu acionamento gera cálculos diretos e *on-line* sobre o sistema acadêmico do PPGE/UFSC. As estatísticas incluem tempo de titulação por ano e a distribuição de defesas por área de concentração. A consulta a dados estatísticos pode ser geral, para todo o programa, ou específicas de um orientador.

Posteriormente, o CNPq e o PPGE/UFSC firmaram acordo visando à conexão *Web* com as bases de currículos Lattes e o BTD. O acordo possibilitou os seguintes recursos adicionais aos seus usuários *Web*:

- para o CNPq, a partir da consulta aos currículos Lattes de professores do PPGEP/UFSC, o usuário pode verificar a atuação de orientação do professor no Programa e ler textos completos das T&D's orientadas;
- para o PPGEP/UFSC, a partir da consulta ao detalhamento da atividade de orientação do professor e dos textos completos das T&D's orientadas por este, o usuário pode consultar o currículo Lattes do orientador.

5.2.5 Líber – Universidade Federal de Pernambuco

O projeto Narrativas Digitais do Virtus - Laboratório de *Hipermídia* da Universidade Federal de Pernambuco (UFPE), foi criado com o objetivo de estudar e viabilizar a publicação eletrônica na *Internet*. Neste sentido, o laboratório desenvolveu o Líber, que tem como objetivo permitir que o usuário possa publicar e acessar gratuitamente textos em formato digital.

Liber é uma palavra de origem latina que significa livro. O projeto Narrativas Digitais retoma a concepção mais genérica da palavra livro e apresenta o Liber como um espaço para organização do pensamento com características *hipermídicas*. Toma como referência pesquisas recentes em ferramentas narrativas, que apontam igualmente para a criação de um novo padrão em que textos, imagens e sons são acoplados em novas estruturas narrativas. Um produto totalmente digital e diferenciado da tecnologia impressa (MOUTINHO, CUNHA FILHO & LIMA, 2002)

No Liber, a inserção de documentos depende do preenchimento de formulário simples e a publicação se faz de forma rápida e fácil. O usuário conta ainda com as seções de edição, nas quais poderá excluir ou alterar informações relativas ao texto, e de acervo, onde qualquer usuário da *Internet* poderá, mediante consulta, acessar e comentar todos os documentos inseridos no Liber. Publicar no Liber implica assumir a responsabilidade pelo processo de redação, de edição, disponibilidade e divulgação dos textos que o usuário inserir. Algumas características do Líber:

- a) *rapidez no processo de divulgação de informações*: há apenas um formulário e nele poucos campos a serem preenchidos;
- b) *facilidade*: instruções claras tornam o *site* acessível a autores de diferentes idades;
- c) *gratuidade*: o Liber não exige qualquer tipo de pagamento pela publicação de obras;

- d) *interatividade*: o usuário poderá contatar o autor e comentar a obra, enviando mensagens que poderão ser lidas por todos os usuários do Liber;
- e) *edição*: o autor poderá atualizar sua obra ou excluí-la;
- f) *navegabilidade*: com facilidade o usuário do Liber encontra um sistema de navegação em que se identificam facilmente os procedimentos possíveis de serem realizados;
- g) *nacionalidade*: no Brasil, poucos são os espaços abertos para disponibilização de textos eletrônicos com as características acima mencionadas. O Liber pretende ampliar o espaço para publicações eletrônicas no Brasil, caracterizando-se como um produto de qualidade na língua portuguesa.

As primeiras obras disponibilizadas, a partir da parceria com a editora universitária foram, Naufrágio e Prosopopea, de Afonso Luiz Piloto e Bento Teyxeyra, respectivamente. Essas obras foram publicadas e impresso por volta do início do século XVII e reeditadas pela editora da UFPE em 1969. Três séculos depois, no início de 2001, ganharam no Liber sua versão eletrônica.

Por pertencer á um ambiente acadêmico, o projeto Liber espera auxiliar, especialmente, estudantes e professores a superar os constantes obstáculos à publicação, incompatíveis com a contínua produção de documentos produzidos por professores, estudantes, funcionários e outros. A parceria com a Editora da UFPE, permitirá ao autor publicar sua obra nos formatos impresso e digital em formato PDF. Na próxima etapa do projeto, pretende-se incorporar novas tecnologias narrativas, com imagens e sons. A intenção é desenvolver ferramentas para publicação usando tecnologias que permitam interoperabilidade que possam contribuir para a ampliação do universo de leitura, autoria e principalmente para a pesquisa.

5.2.6 Biblioteca digital da UNICAMP

A biblioteca digital da UNICAMP usa o sistema Nou-Rau, implementando um serviço on-line para armazenamento e obtenção de documentos, provendo acesso controlado e mecanismos eficientes para busca. Foi desenvolvida objetivando disponibilizar e difundir a produção científica, acadêmica e intelectual da Universidade em formato eletrônico: artigos, fotografias, ilustrações, teses, obras de arte, registros sonoros, revistas, vídeos e outros documentos de interesse ao desenvolvimento científico, tecnológico e sócio-cultural. A partir

da implantação da BD com a disponibilidade do serviço de publicação e consulta aos textos completos, a comunidade acadêmica e de pesquisa passou a ter:

- maior agilidade na divulgação e obtenção da informação;
- a produção da UNICAMP publicada para a comunidade acadêmica interna e de outras instituições de pesquisa nacionais e internacionais;
- o uso simultâneo do documento por vários pesquisadores, nos seus próprios ambientes de trabalho;
- o acesso na *Web* da produção acadêmica de T&D;
- biblioteca distribuída e acessível por várias classes de usuários da *Internet* e a preservação dos textos originais conforme regulamentação interna.

A Universidade definiu alguns procedimentos para o envio dos arquivos contendo o conteúdo das T&D's em relação aos formatos com a extensão ponto doc (Word), ponto ps (Post-Script) ou ponto pdf (Adobe). Não serão aceitos outros tipos de formatos, cabendo ao autor a conversão dos documentos que não estejam nos formatos estabelecidos, para o aceite da publicação na biblioteca digital. A autorização para publicação na *Internet* deve ser entregue junto com o documento digital conforme modelo da universidade definido através de portaria do reitor.

O sistema oferece estatísticas por área: Biomédicas, Humanas, Exatas e Tecnológicas com aproximadamente 4500 T&D's disponibilizadas no sistema³⁹.

5.2.7 Biblioteca digital do IMPA

A Biblioteca do Instituto de Matemática Pura e Aplicada (IMPA) possui um acervo de cerca de 30 mil volumes em livros e 32 mil volumes em periódicos, correspondentes a mais de 600 títulos em assinaturas correntes, em matemática e áreas mais afins, tendo sido escolhida como biblioteca-base na área de matemática pura e aplicada. Além desse acervo, a biblioteca possui um volume considerável de T&D's, publicações do próprio instituto, como também de outras instituições, apostilas e atas bem como outras pré-publicações (CHATAIGNIER & SILVA, 2001).

O projeto de Informatização da biblioteca do IMPA tem como principal objetivo disponibilizar o acervo da biblioteca via *Internet*, não só para a comunidade do IMPA, como

³⁹ http://www.unicamp.br/unicamp/servicos/servicos_arqbibl.html

também para outras comunidades acadêmicas nacionais e internacionais. Esse projeto teve início por volta de 1994, a partir de um trabalho em que foram analisados vários sistemas para bibliotecas disponíveis no mercado.

O IMPA optou, como solução, pela compra do *software* Horizon, desenvolvido pela *Ameritech Library Services*, hoje Epixtech⁴⁰, utilizado como solução de automação de bibliotecas nos EUA e em outros países. O Horizon é um sistema em ambiente gráfico com arquitetura cliente/servidor que utiliza um banco de dados relacional padrão SQL. O formato de catalogação utilizado pelo Horizon é o padrão internacional MARC. A arquitetura desse sistema consiste em um servidor de banco de dados, destinado ao catálogo, funciona em plataforma *Windows* com um servidor *Web* em ambiente UNIX. Os clientes que servem aos administradores do sistema e bibliotecários, operam em ambiente *Windows*. A interface *Web* usa o protocolo Z39.50 para prover a interoperabilidade e comunicação com o banco de dados.

A biblioteca mantém, uma página na *Internet* com informações de interesse para a comunidade de matemática, assim como *links* à portais, como o Periódicos CAPES, bases de dados, como o *MathSciNet*, o *MATH Database* e o *Web of Science*, além de periódicos assinados pela instituição, alguns desses acessos são restritos ao domínio IMPA. A biblioteca é composta, basicamente, de livros, periódicos e *preprints* (dissertações de mestrado, teses de doutorado e artigos ainda não publicados). No caso de *preprints*, é importante que o conteúdo integral, e não apenas uma referência ao mesmo, esteja disponível à comunidade acadêmica.

A solução adotada para disponibilizar o seu acervo de *preprints* foi participar do projeto Math-Net⁴¹, iniciativa internacional de criação de uma rede distribuída de repositórios de *preprints* de matemática na *Internet*, acessáveis por um serviço de busca e navegação.

5.3 Projetos de Instituições Internacionais

NDLTD <http://www.ndltd.org/> (*Network Digital Library of Theses and Dissertations*) biblioteca eletrônica de teses e dissertações autorizadas por estudantes das instituições membros da rede. As pesquisas desenvolvidas para a criação desta rede envolvem assuntos

⁴⁰ <http://www.epixtech.com>

⁴¹ <http://www.math-net.org>

como a criação de um *workflow* para submissão de ETD o desenvolvimento de XML e de DTD para ETDs e o suporte para a biblioteca digital de teses e dissertações eletrônicas.

arXiv.org, <http://www.arxiv.org/>, do Laboratório Nacional de Los Alamos, é o pioneiro na construção de *open archives*. Foi iniciado em 1991 por Paul Ginsparg. Nas últimas décadas, tornou-se um repositório global de artigos não-revisados pelos pares nas áreas da física, matemática, ciência da computação e ciências não-lineares.

CogPrints (*Cognitive Sciences Eprint Archive*) da Universidade de Southampton, no Reino Unido, inspira-se na arquitetura do *arXiv* e aceita trabalhos escritos em qualquer língua, porém há preferência ao inglês, especialmente como língua para ingresso dos resumos dos trabalhos. Abrange psicologia, lingüística e neurociências e subáreas da ciência da computação, filosofia e biologia. O célebre Stevan Harnard é quem faz uma avaliação preliminar dos trabalhos enviados.

NCSTRL (pronuncia-se "ancestral", em inglês, *Networked Computer Science Technical Reference Library*) é uma coleção internacional de relatórios de pesquisa na área de ciências da computação que são disponibilizados sem fins comerciais por várias instituições e arquivos *on-line*.

RePEc, (*Research Papers in Economics*) uma iniciativa dos pesquisadores da área de economia. Citando apenas um dos três significados que Krichel atribui ao termo "RePEc", trata-se de uma coleção de arquivos na *Web* ou em *ftp* que provê dados estruturados sobre documentos impressos e eletrônicos na área de economia. Existem 100 voluntários, em 25 países, contribuindo para a existência do RePEc.

6 METODOLOGIA

A pesquisa investiga o mundo em que o homem vive e o próprio homem. O estudo sobre um determinado problema depende das fontes de informações sobre o mesmo. Essas informações podem ser obtidas através de observações, de reflexões pessoais, de pessoas com experiências no assunto, de participação em eventos ou em acervos de conhecimentos que possua o registro do dado. (CHIZZOTTI, 1998).

Na busca do conhecimento, toda ciência utiliza-se de métodos, que são técnicas ou procedimentos usados para coletar e analisar dados relacionados a uma necessidade de pesquisa ou hipótese. Esses métodos abrangem, entre outros, a participação de pessoas em entrevistas nas suas diversas formas, a aplicação de questionários, a observação de comportamento e o exame de documentos ou registros da atividade produtiva ou humana.

Segundo o Código de Ética da Associação Nacional de Empresas de Pesquisa (ANEP), pesquisa significa “coleta sistemática e o registro, classificação, análise e apresentação objetiva de dados sobre hábitos, comportamentos, atitudes, valores, necessidades, opiniões e motivações de indivíduos e organizações dentro do contexto de suas atividades econômicas, sociais, políticas e cotidianas”.

Neste capítulo é apresentada a metodologia da pesquisa utilizada no trabalho, iniciando pela sua caracterização e estrutura metodológica, a seguir o método de procedimento adotado: pesquisa documental com abordagem qualitativa baseada em uma ampla revisão bibliográfica conjuntamente com a realização de entrevistas com alguns coordenadores dos programas de pós-graduação da UFSM para ouvir os representantes da comunidade e a definição de uma proposta para publicação de T&D's na *Web* através de uma biblioteca digital e a apresentação dos resultados obtidos na conclusão do trabalho.

6.1 Estrutura metodológica da pesquisa

A elaboração da pesquisa científica se deu com base em quatro classificações: segundo os objetivos e finalidade propostos, a forma de abordagem, a natureza e os procedimentos adotados pelo pesquisador.

6.1.1 Em relação ao objetivo proposto e a finalidade da pesquisa

Quanto aos objetivos e finalidade das pesquisas, podem ser classificadas em três grandes grupos, conforme Triviños (1987, p.109), que faz a seguinte diferenciação: pesquisa exploratória, a descritiva e a experimental. Este trabalho está situado no grupo de pesquisa exploratória qualitativa, pelas características do mesmo em relação ao grau de novidade e da recente exploração do tema de forma científica.

Segundo Triviños (1987, p.104), pesquisa exploratória permite ao investigador aumentar sua experiência em torno de um determinado problema. O pesquisador parte de uma hipótese e aprofunda seu estudo nos limites de uma realidade específica buscando antecedentes, maior conhecimentos para, em um segundo momento planejar uma pesquisa descritiva ou experimental.

A pesquisa de caráter exploratória, segundo Gil (1989, p.39) envolve o levantamento do estado da arte de uma tecnologia, uma situação ou mesmo uma área do conhecimento através da revisão bibliográfica relacionada com o processo pesquisado, para que a vivência com o problema possa estimular e facilitar a compreensão do fato. Este tipo de pesquisa busca basicamente desenvolver, esclarecer e modificar conceitos e idéias para a formulação de novas abordagens *a posteriori*. Portanto, esta forma de estudo objetiva proporcionar o maior conhecimento possível para o pesquisador sobre o objeto de análise, para que o mesmo possa formular ou reformular problemas mais precisos e específicos, bem como a criação de hipóteses para estudos posteriores no sentido da continuidade da pesquisa pela Instituição.

6.1.2 Em relação à forma de abordagem

Quanto à forma de abordagem o trabalho se caracteriza como pesquisa qualitativa, que segundo Chizzotti (1998, p.78) “... é uma designação que abriga correntes de pesquisa muito diferentes, que se fundamentam em alguns pressupostos contrários ao modelo experimental”. A pesquisa qualitativa tem como objetivo, via de regra “provocar o esclarecimento de uma situação para a tomada de consciência pelos próprios pesquisadores dos seus problemas e das condições que os geram, a fim de elaborar os meios e estratégias de resolvê-los”, Chizzotti (1998, p.104) e que adota métodos e técnicas de pesquisa diferentes dos estudos experimentais.

Godoy, (1995, p.58) recomenda que uma pesquisa qualitativa deve apresentar as seguintes características:

- considerar o ambiente como fonte direta dos dados e o pesquisador como instrumento chave;
- possuir caráter descritivo;
- o processo ser o foco principal de abordagem e não o resultado ou o produto;
- a análise dos dados ser realizada de forma intuitiva e indutivamente pelo pesquisador;
- não requerer o uso de técnicas e métodos estatísticos; e por fim,
- ter como preocupação maior a interpretação de fenômenos e a atribuição de resultados.

Ainda segundo Godoy (1995, p.58), a pesquisa qualitativa não busca enumerar e/ou medir os eventos pesquisados, nem utiliza instrumental estatístico na análise dos dados. O ponto de partida são questões ou focos de interesses amplos, que vão se definindo na medida que o estudo se desenvolve. Um estudo desta categoria necessariamente envolve dados descritivos sobre organizações, pessoas, lugares e as formas de interação que se dão pelo contato do agente pesquisador com o objeto em análise, procurando compreender os fenômenos segundo a vivência e visão dos atores que convivem diuturnamente com o caso fático.

Sendo a pesquisa qualitativa, é adotado o método histórico-antropológico, que “captam os aspectos específicos dos dados e acontecimentos no contexto em que acontecem” (CHIZZOTTI, 1998, p. 79).

Uma pesquisa qualitativa tem como ponto de partida a existência de uma relação dinâmica entre o sujeito/organização e o mundo real, o que obriga ao pesquisador a despojar-

se de preconceitos e adotar uma posição aberta em relação a tudo que é observado, uma vez que todos os atores que interagem com a pesquisa são “reconhecidos como sujeitos que elaboram conhecimentos e produzem práticas adequadas para intervir nos problemas que intensificam” (CHIZZOTTI, 1998, p.83).

6.1.3 Em relação à natureza da pesquisa

Em relação à natureza, a pesquisa classifica-se em aplicada, uma vez que uma de suas intenções é a mudança de paradigma das instituições na forma de seu processo de publicação e disseminação dos trabalhos de conclusão de T&D's dos programas de pós-graduação da UFSM. A pesquisa aplicada, além de envolver verdades e interesses localizados, “objetiva gerar conhecimento para a aplicação prática em soluções de problemas específicos”.

6.1.4 Em relação aos procedimentos adotados

Este trabalho desenvolveu-se a partir de um referencial bibliográfico existente, principalmente livros, artigos, pesquisas na *Internet* e a análise de viabilidade da implantação de uma biblioteca digital de T&D's no PPGEP a UFSM. Para tal, foi analisado tecnologias disponíveis no mercado buscando uma ferramenta que possibilite total interoperabilidade com as demais bibliotecas digitais no país e exterior, bem como uma entrevista com os coordenadores de pós-graduação da UFSM no sentido de avaliar a expectativa em relação à disseminação do conhecimento produzido através das T&D's na *Web*.

Como o trabalho analisa uma proposta de implantação de uma biblioteca digital de T&D's no PPGEP a UFSM, o mesmo classifica-se à categoria de pesquisa cujo objeto é uma unidade que se avalia (sociedade, empresa, pessoa, comunidade, ferramentas, etc.), não admite visões isoladas, parceladas e estanques. Segundo Triviños (1987, p.133), a análise de uma situação ou de um caso é uma categoria de pesquisa cujo objeto é uma unidade, onde um conjunto de dados descreve uma fase ou a totalidade do processo social, em suas várias relações internas e nas suas fixações culturais.

Segundo Chizzotti (1998, p. 102), o desenvolvimento do um estudo de caso pré-supõe três fases:

- a) **a seleção e delimitação do caso a ser estudado:** o caso deve ser uma referência significativa para merecer a investigação e, por comparações aproximativas, apto para fazer generalização a situações similares ou autorizar inferência em relação ao contexto da situação analisada. Este é precisamente o momento de definir os aspectos e os limites do trabalho, a fim de reunir informações sobre um campo específico e fazer análises sobre uma dada organização, a partir dos quais se possa compreender uma determinada realidade;
- b) **o trabalho de campo:** busca reunir e organizar um conjunto probatório de informações. Pressupõe uma negociação prévia para que se tenha acesso a documentos e pessoas necessários a concretização do estudo;
- c) **a organização e redação do relatório:** que poderá apresentar um estilo narrativo, descritivo ou analítico. Esta última etapa também pode ser de registro de caso, isto é, o produto final do qual consta uma descrição do objeto de estudo.

Este trabalho foi desenvolvido em três fases: a fase exploratória, a fase de coleta de dados e a fase de análise e interpretação dos dados.

Fase exploratória, onde ocorreu a revisão bibliográfica pertinente ao tema: revisão sobre a produção do conhecimento que teve suas origens no renascimento acentuando-se fortemente na revolução industrial até o advento da *Internet*; as tecnologias existentes para publicação e disseminação do conhecimento de T&D's na *Web*; Arquivos Abertos, compreendo temas como *e-prints (preprint e posprint)*, auto-arquivamento, revisão por pares, vocabulários controlados; tecnologias para divulgação do conhecimento armazenado nas T&D's na *Web* através do uso de bibliotecas digitais.

Fase de coleta de dados: esta fase foi desenvolvida em duas etapas:

Primeira etapa: avaliação documental da situação atual da busca pelo conhecimento armazenado nas T&D's das bibliotecas convencionais da UFSM e análise de tecnologias para publicação na *Web*.

Segundo Godoy (1995, p.60), uma das vantagens básicas da análise documental é que os documentos constituem uma fonte não-reativa, uma vez que as informações neles contidas são imutáveis, sendo uma fonte natural de informações à medida que, por se originarem num determinado contexto histórico, econômico e social, refletem a inserção da organização neste mesmo contexto.

Segunda Etapa: realização de entrevistas semi-estruturada, aplicada aos coordenadores dos programas de pós-graduação da UFSM.

Fase de análise e interpretação dos dados: abordagem de caráter qualitativo, cujo objetivo foi revelar a situação atual da disseminação do conhecimento armazenado nas T&D's no âmbito da UFSM, a utilização da *Web* como busca de conhecimento pela comunidade de pesquisa e quais as expectativas quanto à publicação na *Web* das T&D's usando como meio uma biblioteca digital. Além da análise dos dados obtidos através das entrevistas, nesta fase também temos a definição de uma proposta de ferramenta para publicação na *Web* das T&D's produzidas pelos programas de pós-graduação da UFSM.

Como enfoque teórico do significado de inter-relações no desenvolvimento do trabalho, utilizou-se do enfoque sistêmico. Segundo Triviños (1987, p. 81) o enfoque sistêmico, que tem suas raízes na Teoria Geral dos Sistemas elaborada por Ludwig von Bertalanffy, parte da idéia de que existem numerosas relações no interior do objeto que se estuda, mas que este também está ligado ao meio externo por um conjunto de inter-relações. O enfoque sistêmico dirige sua atenção especialmente ao estudo dos sistemas altamente complexos, sistemas técnicos altamente desenvolvidos.

Este trabalho é de caráter qualitativo, onde abordou-se, a disseminação do conhecimento, sua evolução segundo as novas tecnologias de publicação na *Web* procurando revelar a situação atual da disseminação do conhecimento de T&D's na *Web* através de uma biblioteca digital.

A abordagem qualitativa da pesquisa não se apresenta como uma proposta rígida, ela permite que a imaginação e a criatividade levem o investigador a explorar novos enfoques.

Nesta busca foi utilizado o delineamento da pesquisa a seguir apresentado.

6.2 O Delineamento da pesquisa

Num primeiro momento, a pesquisa passou pelo o QUÊ deve ser buscado para permitir uma maior disseminação do conhecimento produzido através das T&D's nos programas de pós-graduação da UFSM. Uma verificação das tecnologias para publicação de T&D's na *Web* que promovam a interoperabilidade entre aplicações e a cooperação na comunidade acadêmica provendo um serviço de informação e comunicação inovador melhorando o tempo de publicação das pesquisas.

Num segundo momento, ONDE, será realizada a pesquisa. O que leva a busca de uma unidade do setor produtivo que seja representativa e que tenha ações concretas para validar a proposta.

Em terceiro lugar, na busca de uma definição espaço temporal, o QUANDO, isto é, qual a dimensão temporal da pesquisa. Neste caso foi utilizado, como referencial, o momento atual, em função do grande avanço das tecnologias .

E finalmente, o COMO, isto é, qual o método a ser utilizado na implementação do trabalho proposto.

- formação do referencial bibliográfico;
- percepção do problema, a formulação da hipótese e a definição do método proposto;
- escolha da atividade produtiva e estratégia de ação;
- análise documental;
- coleta dos dados;
- análise e interpretação dos dados

6.3 A Descrição sucinta das atividades

Em relação às atividades desenvolvidas durante a execução do trabalho, comentaremos o referencial bibliográfico, o problema e o método usado para análise, a escolha da atividade produtiva e a estratégia de ação, como foi feita a análise e coleta dos dados para o caso da entrevista realizada e o universo da pesquisa.

6.3.1 A formação do referencial bibliográfico

Durante a realização dos créditos necessários no PPGEP-UFSM, iniciou-se a fundamentação teórica através de pesquisa a bibliografia existente, de forma contínua e sistematizada procurando-se atingir obras nacionais e internacionais. Os portais de excelência como SCIELO e o IBICT foram usados para busca das informações disponibilizadas na *Internet* que versassem sobre biblioteca digital, organização do conhecimento, disseminação do conhecimento, tecnologias para publicação onde citamos a iniciativa dos arquivos abertos, linguagens de marcação entre outros.

Todo este referencial foi agrupado e sistematizado segundo a lógica da contextualização histórica e evolução dos instrumentos e formas de publicação do conhecimento buscando ter uma visão ampla e sistêmica da produção e disseminação do conhecimento produzido através de T&D's e seu uso em C&T.

6.3.2 A percepção do problema e a definição do método proposto para a análise

Atualmente, a *Internet* é o meio de socialização do conhecimento mais utilizado, propiciando facilidades e rompendo barreira outrora existente em função da forma e meio de armazenamento do conhecimento. A busca e o uso integrado das tecnologias de comunicação, de computação e de conteúdos em formato digital, vem contribuindo nos últimos anos para criar um novo ambiente de acesso, disseminação, cooperação e promoção do conhecimento em uma escala global.

Constatou-se que hoje não é mais possível uma instituição de ensino ou um centro de pesquisa pensar e agir de forma isolada. Espera-se que os resultados de suas pesquisas possam ser publicados em um curto espaço de tempo, passível de recuperação podendo assim servir de subsídio para novas pesquisas na comunidade ou por outras comunidades de pesquisa.

Atualmente, os programas de pós-graduação publicam o resultado de suas pesquisas através das T&D's, as quais ficam armazenadas nas diversas bibliotecas da UFSM. No sentido de melhorar a disseminação do conhecimento produzido, o Programa de Pós-Graduação em Engenharia de Produção sugere a implantação a nível institucional de uma ferramenta para publicar na *Web* as T&D's produzidas pelos alunos da UFSM.

A partir desta realidade reuniu-se referencial para responder a seguinte pergunta: Qual a melhor forma de publicar e disseminar na *Web* o conhecimento produzido através das T&D's nos programas de pós-graduação das instituições de ensino superior, especificamente na UFSM?

A formulação do modelo teve como ponto de partida um estudo detalhado de tecnologias de publicação na *Web* no sentido de estruturar o referencial teórico mostrando uma visão global da situação atual em relação à publicação do conhecimento e a impressão dos representantes da comunidade de acadêmica de pesquisa da UFSM representados pelos coordenadores de programas de pós-graduação. Neste sentido, foi realizado entrevistas com os coordenadores de programas de pós-graduação da UFSM, por grande área do conhecimento, os quais esperam obter o máximo possível das pesquisas produzidas nos seus

programas com o objetivo de disseminar o conhecimento e contribuir para o desenvolvimento de C&T do país.

6.3.3 A escolha da atividade produtiva e a estratégia de ação

Na escolha da atividade produtiva a ser pesquisada, foram levados em consideração alguns requisitos tais como a Instituição envolvida:

- ser uma Instituição de Ensino com vários centros independentes de pesquisa;
- ter catalogado em seu acervo as publicações que permitissem uma análise do seu processo de evolução em relação a parâmetros de qualidade do seu produto final;
- ter o maior interesse em modificar ou implementar uma nova forma de publicação de suas pesquisas oferecendo a comunidade maior transparência ao produto produzido, no caso a publicação de suas pesquisas;
- ter a intenção de ser um referencial de qualidade para a comunidade acadêmica de pesquisa.

A estratégia de ação foi apresentar o projeto de pesquisa para a comunidade acadêmica principalmente ressaltando a importância da disseminação do conhecimento desenvolvido pelos programas de pós-graduação através da publicação de suas T&D's na *Web* através de uma biblioteca digital cujo ambiente oferece facilidade de acesso e segurança quanto ao material disponibilizado para consulta à comunidade acadêmica.

6.3.4 A Entrevista e os participantes do universo da pesquisa

Consistiu em entrevistar pessoas da comunidade acadêmica que participam diariamente e desenvolvem suas atividades de ensino e pesquisa dentro da Instituição. Utilizou-se um questionário o qual tinha como objetivo analisar as diferentes visões dos coordenadores dos programas de pós-graduação. As perguntas foram elaboradas com a finalidade de verificar o nível de conhecimento dos entrevistados sobre a atual situação da publicação e disseminação do conhecimento produzido através das T&D's na *Web* em termos de Brasil e quais suas expectativas considerando a atual realidade brasileira em termos de acesso às tecnologias além do direito autoral sobre o trabalho.

6.3.5 Participantes do universo da pesquisa

Participaram deste estudo, no item referente à entrevista, os coordenadores de programas de pós-graduação da UFSM, um coordenador por grande área de conhecimento. A escolha dos entrevistados, se deu em função da atuação do grupo na comunidade acadêmica de pesquisa da UFSM cujo entendimento é que eles representam o grupo de pesquisadores vinculado ao programa e seu trabalho junto à comunidade acadêmica é de manter o programa em um elevado grau de qualidade.

6.3.6 A análise documental e análise dos dados

Inicialmente foi realizada uma pesquisa documental sobre o assunto basicamente usando a *Internet* como fonte como é mostrado nos primeiros capítulos. Posteriormente, buscou-se no banco de dados da Instituição, verificar a situação atual em relação à busca do conhecimento armazenado nas bibliotecas da UFSM das T&D's catalogadas nas bibliotecas. O capítulo 7 apresenta o resultado de cada pergunta da entrevista realizada sintetizando a visão dos coordenadores dos programas de pós-graduação, que, neste trabalho representam a comunidade acadêmica de pesquisa, na seqüência, o capítulo 8, temos o detalhado a proposta da ferramenta para publicação na *Web* das T&D's.

7 ANÁLISE DAS ENTREVISTAS

Na análise das informações obtidas nas entrevistas realizadas, procurou-se sintetizar em uma única visão o pensamento do segmento entrevistado, ou seja, a visão dos coordenadores de programas de pós-graduação da UFSM considerando os seguintes aspectos em relação ao objeto de estudo:

- As coordenações, hoje, orientam seus alunos quanto à utilização da *Web* para realizar pesquisas em portais de excelências;
- Atualmente, adota-se alguma estratégia para propiciar o acesso e o uso das T&D's publicadas e armazenadas na Biblioteca Central da UFSM;
- As T&D's são consideradas pelos Programas como um referencial bibliográfico e base de conhecimento para a continuidade e o desenvolvimento de novas pesquisas dentro dos programas de pós-graduação;
- A aceitabilidade por parte das coordenações quanto à publicação de T&D's em uma biblioteca digital como forma de disseminação e recuperação do conhecimento produzido pelos programas de pós-graduação;
- Como é atualmente tratado o direito autoral da T&D pelo programa de pós-graduação e se existe alguma restrição sobre o assunto.

Buscando a opinião dos coordenadores de programas de pós-graduação em relação ao assunto acima descritos, elaborou-se 10 perguntas que foram aplicadas e sintetizadas no decorrer deste capítulo. A decisão de escolher os coordenadores de programas de pós-graduação foi por entender que os mesmos representam o segmento que define a atuação, a qualidade, a abrangência e os rumos da pesquisa de cada programa dentro da UFSM.

Foram entrevistados, no mínimo um coordenador de Programa por área de conhecimento, a saber: Ciências Exatas e da Terra; Ciências Biológicas; Engenharias; Ciências da Saúde; Ciências Agrárias; Ciências Sociais e Aplicadas; Ciências Humanas; Lingüística, Letras e Artes.

Periodicamente, a CAPES realiza a avaliação dos programas de pós-graduação de todo o país, essa ação faz com que os Programas estejam constantemente procurando melhorar os índices de qualidade, os quais, são traduzidos em parâmetros considerados no momento da

avaliação. O tempo de produção de uma T&D juntamente com as publicações nacionais e internacionais de artigos resultantes do trabalho original também são considerados na avaliação. Neste sentido, a disponibilidade de informações durante a execução do trabalho, principalmente na elaboração do referencial teórico, passa a ser um determinante da garantia da qualidade, da diminuição do tempo de desenvolvimento da pesquisa.

Durante as entrevistas, inicialmente procurou-se situar o coordenador de cada programa de pós-graduação dentro do objeto de estudo, qual seja bibliotecas digitais e a publicação de T&D's na *Web*. Cada pergunta elaborada foi sintetizada em uma visão geral dos coordenadores destacando as diferenças de opiniões ou contribuições relevantes. No final de cada pergunta tem-se a conclusão da idéia em função do objetivo da questão considerada. A análise das entrevistas é relatada logo abaixo.

7.1 Análise das respostas

Pergunta 01

A coordenação deste programa de pós-graduação adota alguma estratégia no sentido de disseminar o conhecimento produzido através das T&D's aos seus alunos? De que forma?

Nesta pergunta, procurou-se identificar se as coordenações dos programas de pós-graduação adotam alguma estratégia para disseminar o conhecimento produzido pelo programa através das T&D's no sentido de divulgar para a comunidade acadêmica de pesquisa e dar continuidade as pesquisas do programa realizadas pelos alunos egressos.

Todos foram unânimes em afirmar que após a avaliação da pesquisa por uma banca qualificada, o resultado sendo considerado bom, a T&D passa a ser indicada como um referencial bibliográfico e base de conhecimento para outros trabalhos na área. A qualidade da T&D determinará a sua divulgação pelos docentes do programa como fonte de pesquisa no meio acadêmico.

Alguns Programas trabalham com grupos de pesquisa e neste caso são os professores do grupo que definem a continuidade dos trabalhos após o término de uma pesquisa ou publicação de uma T&D.

Os coordenadores consideram que o referencial bibliográfico tem qualidade a partir do momento que a T&D gerou publicação de artigos e periódico em congressos ou revistas da área. O coordenador do Programa de Pós-Graduação em Química salienta que no seu programa 99% dos artigos e periódicos são escritos em inglês e submetidos para publicação

internacional, onde, segundo ele o trabalho passa a ser efetivamente divulgado mas considera a T&D como base de referência para a pesquisa.

Considerando as opiniões dos coordenadores, percebe-se que ainda existe uma carência em relação à divulgação do conhecimento armazenado na T&D no âmbito dos programas. Eles consideram os artigos e periódicos resultantes das T&D's como fonte do referencial bibliográfico em função do acesso, ou seja, esse material é submetido à publicação em revistas e circulam no meio acadêmico independente da Instituição de origem, isso não acontece nas T&D's materializadas em bibliotecas convencionais.

Pergunta 02

Todos os programas de pós-graduação, historicamente, publicam e disponibilizam suas T&D's na Biblioteca Central e nas Bibliotecas Setoriais. Qual a sua opinião sobre o uso de T&D's produzidas na UFSM e em outras Instituições como base de conhecimento nas pesquisas do seu programa?

O objetivo desta questão foi verificar se os programas de pós-graduação da UFSM, atualmente usam T&D's da UFSM e de outras Instituições como base de conhecimento nas pesquisas do programa.

“Critica Fortuna”, quer dizer avaliar o conhecimento produzido por outra pessoa. O estado da arte de um determinado assunto é fruto de uma fundamentação teórica baseada na visão da atualidade, ou seja, o que há de novo em termos de pesquisa sobre o assunto.

Segundo o coordenador do Programa de Pós-Graduação em Engenharia Elétrica, o acesso deve ser irrestrito a todas as publicações. Entretanto a Biblioteca Central da UFSM tem dificuldade de catalogar o atual acervo e isso dificulta o acesso ao conhecimento. Ele comenta, que desde 1997 possui um acervo de livros de alta relevância na área e engenharia elétrica, adquiridos em através de um projeto, e o mesmo está disponível somente para uso dos alunos do Programa por orientação da Biblioteca Central em detrimento da falta de infraestrutura para acervo e catalogação.

O coordenador Programa de Pós-Graduação em Química coloca que busca em outros programas as T&D's produzidas e publicadas mas afirma ter dificuldade de localização do material completo na biblioteca da CAPES. Segundo ele, a CAPES possui todas as T&D's

dos programas de pós-graduação credenciados do país, porém, o *site* não tem filtros de qualidade e muitas vezes o pesquisador desiste e solicita o material por outras vias.

Percebe-se claramente a necessidade do uso de T&D's como fonte de pesquisa independente da instituição que a publicou. Por outro lado, tem-se dificuldades de acesso ao material publicado nas bibliotecas digitais e disponibilizados em portais, como por exemplo da CAPES. Recentemente a CAPES disponibilizou novas bases de dados e patentes, plataformas e serviços de pesquisa dos conteúdos com o objetivo de universalizar o acesso à informação no meio acadêmico-científico. Segundo a CAPES, o portal é atualmente o maior banco de informações científicas da América Latina e atende 152 Instituições.

Pergunta 03

Considerando, que a UFSM ainda não possui uma Biblioteca digital de T&D's, a coordenação deste programa de pós-graduação atualmente tem alguma forma de medir o uso de T&D's como referencial bibliográfico de pesquisa pelos docentes e alunos do programa? Existe alguma estatística em relação à quantidade de consultas por área de pesquisa? Da frequência do uso de T&D's nos trabalhos de pesquisa dos alunos e professores? O programa acha relevante trabalhar esses ou outros indicadores relativos ao assunto?

Neste questionamento procurou-se verificar se as coordenações dos programas de pós-graduação possuem alguma forma de medir o uso de T&D's, como referencial bibliográfico de pesquisa, pelos docentes e alunos do programa considerando que a UFSM ainda não tem uma biblioteca digital de T&D's. A estatística em relação à quantidade de consultas por área de pesquisa, a frequência do uso de T&D's nos trabalhos de pesquisa dos alunos e professores, se existe forma de quantificar este uso e se informações estatísticas são relevantes para o programa.

Conforme relato, nenhum programa de pós-graduação da UFSM consegue medir índices em relação às publicações de T&D's mas todos afirmam ser de alta relevância a obtenção de índices que possam ajudar na gestão da qualidade do Programa.

A coordenadora do Programa de Pós-Graduação em Educação colocou que na última reunião do comitê a nível nacional em março de 2005, ficou claro para os Programas a necessidade de medir índices de qualidade. A CAPES, num primeiro momento, sugere aos programas que identifiquem as melhores pesquisas para que se possa classificar os 5 melhores

trabalhos. Como medir o melhor sem uma ferramenta que auxilie na busca de indicadores de qualidade.

O coordenador Programa de Pós-Graduação em Química coloca que em um futuro próximo o número de vezes que um autor é citado em publicações de artigos vai ser um balizador da qualidade de publicação do pesquisador. Esse índice já é considerado por algumas revistas de publicação na área de química.

O coordenador Programa de Pós-Graduação em Medicina Veterinária comenta que o portal do SciELO oferece dados estatísticos sobre publicações de artigos e periódicos mas não das T&D's. Para o autor é o retorno do seu esforço, sua pesquisa ser considerada por outros pesquisadores como base de conhecimento e acima de tudo a questão da transparência na produção.

A coordenadora do Programa de Pós-Graduação em Letras procura assistir as defesas para poder medir e saber o que está sendo produzido pelo programa.

Ficou claro que para a gestão administrativa dos programas de pós-graduação, a possibilidade de obter-se índices que demonstrem o nível de qualidade da produção acadêmica do programa, é uma questão de sobrevivência frente a competitividade dentro da própria instituição e as exigências da CAPES, órgão que fiscaliza a produção individual a nível nacional.

Pergunta 04

A *Internet* e as tecnologias de comunicação disponíveis na *Web* têm provocado profundas mudanças em relação à publicação científica. De que forma a coordenação deste programa de pós-graduação orienta e incentiva seu corpo docente e discente a usar a *Web* como meio prático e seguro de pesquisa na produção de suas T&D's?

Considerando o surgimento da *Internet* e das novas tecnologias de comunicação disponíveis na *Web*, as quais têm provocado profundas mudanças no mundo da publicação científica, procurou-se verificar nesta questão de que forma as coordenações dos programas de pós-graduação tem orientado e incentivado seu corpo docente e discente a usar a *Web* como meio prático e seguro de pesquisa na produção de suas T&D's.

Segundo alguns coordenadores, é difícil medir segurança em relação ao material publicado. Os programas aconselham seus alunos a buscar informações em Institutos de pesquisa na área de atuação mas nem sempre se tem a garantia de qualidade. Segundo o

coordenador do Programa de Pós-Graduação em Engenharia Elétrica, na área de engenharia, existem dois centros de excelência, o *IEEE* americano e o *IEE* inglês que garantem a qualidade do material publicado. O portal da CAPES e o SciELO também são citados como referência de qualidade no Brasil.

Alguns coordenadores dizem que nem sempre uma Instituição de renome oferece somente qualidade, muitas vezes à busca por melhores índices sem um controle do que está sendo produzido ou algum tipo de balizador externo pode resultar na ineficiência.

Neste item questionado, percebeu-se que todos os coordenadores estão preocupados com a busca da informação na *Web* em função da massificação provocada pela explosão da *Internet* onde não existe critério para a publicação. O conselho de todos os programas é no sentido da consulta buscar informações a partir de portais de excelência. A publicação em bibliotecas digitais integradas em um consórcio nacional ajudará a referendar a qualidade do material publicado localmente e disponibilizado em um portal único de busca.

Pergunta 05

Até pouco tempo todas as publicações científicas eram impressas e armazenadas em bibliotecas tradicionais. Em 1991, no Los Alamos National Laboratory no novo México, o físico Paul Ginsparg cria o *software arXiv* onde os usuários podem recuperar ou submeter os artigos tanto por uma *interface on-line* na *Web* como via correio eletrônico marcando claramente uma transição do modelo de comunicação tradicional, baseado em publicações periódicas formalmente estabelecidas para uma nova realidade.

Em 1999, durante a Convenção de Santa Fé no Novo México organizada também por Ginsparg, é definido um padrão para troca de informação entre arquivos eletrônicos, a iniciativa ficou conhecida como *Open Archives*, arquivos abertos que define padrões para publicação na *Web*. Hoje, com os mecanismos eficientes de comunicação na *Web* e tecnologias para construção de bibliotecas digitais podemos ter o registro dos resultados, a coleta e estocagem destes registros, a disseminação e o reuso, tanto em atividades produtivas quanto como fonte para gerar novos conhecimentos. A coordenação deste programa de pós-graduação tem interesse de usar esse tipo de tecnologia para publicação de sua produção intelectual de T&D's independente de consulta à comunidade acadêmica? Considerando o contexto, dê a sua opinião sobre esta realidade?

A intenção desta pergunta é verificar se coordenação dos programas de pós-graduação tem interesse de usar esse tipo de tecnologia para publicação de sua produção intelectual de T&D's independente de consulta à comunidade acadêmica e a opinião sobre a interoperabilidade de informações propiciada por essas tecnologias em detrimento do crescimento da pesquisa.

Os Programas entendem que preservado o direito do autor quanto à autenticidade da pesquisa à publicação da produção acadêmica na *Web* é bem vinda. A disponibilidade da produção de T&D's na *Web* fará com que tanto o autor quanto o orientador do trabalho passem a ter uma maior preocupação com a qualidade da publicação e em contra-partida o Programa se qualifica.

Para os coordenadores do Programa de Pós-Graduação em Química e da Pós-Graduação em Medicina Veterinária a publicação da produção acadêmica em termos de artigos e periódico em revistas é a garantida de que a pesquisa está sendo usada por outros pesquisadores. A disseminação das informações, neste caso, corre independente de qualquer tecnologia salientando que as editoras assumem a responsabilidade da publicação da pesquisa, neste caso, cabe à elas tornam público o conhecimento.

Quanto à publicação de T&D's na *Web*, todos concordam que o resultado a curto e médio tempo deverá ser satisfatório principalmente em função do uso de um portal como base de busca e a disponibilidade de tecnologia local para publicar e consultar.

Percebeu-se que até o presente momento, os coordenadores de programas de pós-graduação da UFSM, não tem a visão ampla da existência de tecnologias e ferramentas disponíveis no mercado capazes de publicar e disponibilizar informações com qualidade na *Web*. O sentimento da comunidade é de que o papel impresso e assinado por uma editora ainda referenda e qualifica a informação. Novamente, acredita-se que, com a divulgação de portais de busca de informação, como a BDTD, a CAPES e o SciELO essa realidade pode ser modificada e as T&D's passem a ser uma fonte de pesquisa rápida e segura para os alunos dos programas de pós-graduação.

Pergunta 06

Biblioteca digital é um ambiente onde os programas de pós-graduação podem divulgar suas T&D's na *Web* conforme um padrão que permite a total integração entre outras bibliotecas a nível nacional e internacional. De um lado as tecnologias emergentes e de outro

o Brasil, um país em crescimento com várias dificuldades econômicas e sociais onde o cesso a *Internet* ainda é bastante precário, mesmo no âmbito das Universidades. Qual é a opinião desta coordenação em relação à publicação de T&D's na *Web* para o desenvolvimento da C&T no país?

Esta pergunta tem como objetivo verificar a visão das coordenações em relação ao desenvolvimento da C&T diante das possibilidades oferecidas pelas tecnologias que propiciam a disseminação imediata do conhecimento na *Web* das T&D's produzidas nas Universidades, levando-se em conta as dificuldades em relação a financiamentos para pesquisa e o acesso a *Internet*.

Na visão dos coordenadores, a divulgação do conhecimento na *Web* realmente socializa o conhecimento mas não contribui para o desenvolvimento do país se não houver uma ação institucional, ou seja, o governo preocupado em organizar a pesquisa a nível nacional. Ações isoladas demoram, esforço em conjunto propícia o crescimento, une as iniciativas nivelando e aumentando a qualidade da pesquisa.

Para a coordenadora do Programa de Pós-Graduação em Letras, a comunidade científica é altamente rigorosa e seletiva com a divulgação de suas pesquisas, em geral o acesso é restrito a alguns pesquisadores da área e neste caso o acesso independe de tecnologias e sim da qualificação do grupo de pesquisa. Na comunidade acadêmica, algumas áreas protegem suas pesquisas principalmente no item segredo de patente.

Para a maioria dos coordenadores de Programa, o acesso ao conhecimento produzido integra e aproxima as comunidades. Algumas coordenações comentam sobre a inexistência de um portal único. A CAPES poderia oferecer um portal com links para todas as instituições que publicam com qualidade. Um grande projeto do governo integrando ações, a pesquisa ganha, o pesquisador passa a ter um local seguro e prático para buscar informações. Segundo o coordenador da Pós-Graduação em Química, as T&D's são enviadas para a CAPES, neste caso é uma questão de tornar esta informação publica para toda a comunidade de pesquisa e isso só vai acontecer se o site oferecer eficiência nos recursos de busca.

Percebeu-se nesta questão, que a maioria dos coordenadores dos Programas, preocupam-se com a melhoria da produção acadêmica, com a clara intenção de contribuir para o desenvolvimento do país. Entretanto, a divulgação dos resultados das pesquisas, na maioria das vezes, fica no âmbito da Biblioteca Central em material impresso e sem acesso irrestrito ao publico. Percebe-se também a falta de ações do governo no sentido de direcionar pesquisas para determinadas áreas. Essas ações podem ser realizadas através da criação de

grupos de pesquisas e financiamentos especiais em áreas carentes e com potencial de desenvolvimento.

Pergunta 07

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) coordena um projeto de implantação dos Arquivos Abertos que é composto de um conjunto de ações que integram o projeto da biblioteca digital brasileira. É uma proposta inovadora que possibilita não somente a publicação de textos eletrônicos diretamente na rede, mas também o envio de comentários e sugestões aos documentos eletrônicos disponíveis em um repositório virtual. Esta iniciativa do IBICT propõe ampliar a visibilidade nacional e internacional da produção intelectual brasileira em C&T, melhorar o fluxo de comunicação científica e tecnológica e incrementar o ciclo de geração de novos conhecimentos.

A partir deste projeto o IBICT desenvolveu um ambiente para publicação de T&D's chamado TEDE, o qual é oferecido às IFES, sem ônus para a Instituição, para que os programas de pós-graduação possam publicar suas T&D's em uma biblioteca digital com total integração com a BDTD. O *software* oferece o gerenciamento da defesa da T&D com o cadastro completo do trabalho, da data da defesa, da banca, do órgão financiador além de consultas. Quando o autor terminar o trabalho o Programa submete a T&D para a Biblioteca Central para a catalogação final e publicação na *Web*. A coordenação deste programa possui infra-estrutura necessária para absorver o *software* TEDE e qual sua visão quanto aos objetivos propostos pelo IBICT?

A intenção desta pergunta é verificar se as coordenações tem infra-estrutura para absorver o sistema e a visão das mesmas sobre os objetivos proposto pelo IBICT através do uso do TEDE como ferramenta de disseminação do conhecimento.

Quanto ao item infra-estrutura necessária para absorver a *software*, a maioria dos programas diz que é uma questão de adaptação e que o resultado será muito bom para a comunidade acadêmica. Um dos programas sugere que a Pró-Reitoria de Pós-Graduação e Pesquisa disponibilize em um local onde todos os Programas possam encaminhar seus alunos para fazer o gerenciamento e submissão de T&D's para a biblioteca central.

Quanto aos objetivos do *software* proposto pelo IBICT, visibilidade no Brasil sim, a nível internacional somente será alcançado se tivermos as publicações traduzidas para a

língua inglesa. O coordenador do Programa de Pós-Graduação em Química comenta sobre a edição em inglês do jornal JBC, o qual é editado dentro a USP. Nas áreas chamadas de “ciências duras”, fica difícil oferecer transparência internacional se as publicações não foram na língua inglesa. A Alemanha é citada como exemplo, para buscar a transparência internacional os pesquisadores germânicos tem publicado seus trabalhos na língua inglesa e não mais sua língua original.

Na análise da questão ficou clara a importância da publicação do trabalho na *Web* e a questão da infra-estrutura passa pela organização da própria Instituição na definição de uma meta e ação, “publicar as T&D’s do programas de pós-graduação da UFSM em uma biblioteca digital e tornar transparente a pesquisa a nível nacional e internacional”. Essa ação deverá ser institucional e envolver todos os segmentos e as unidades relacionadas com a pesquisa dentro da universidade.

Pergunta 08

Bibliotecas digitais, nos remete a um futuro sem documentos impressos, armazenados em locais identificados por um endereço na *Web*, porém de acesso imediato. Para o desenvolvimento da C&T de um país, quando menor o tempo gasto na busca da informação, menor será o custo e maior o ganho da sociedade. Qual a sua opinião sobre este futuro sem o papel impresso?

Esta pergunta procurou identificar qual a opinião dos coordenadores de pós-graduação sobre o mundo virtual, o conhecimento em meio digital num futuro sem o papel impresso.

Na visão dos coordenadores, uma forma de publicação não exclui a outra. O papel facilita a leitura e não fica fora do ar, disponibilidade imediata. Falta recurso para o pesquisador adquirir tecnologia que ofereça condições de total acesso, tem-se velocidade de rede mas falta o incentivo para aquisição de computadores que possam ser transportados com segurança e praticidade para qualquer lugar.

Admirável mundo novo! A coordenadora do Programa de Pós-Graduação em Letras questiona como ficará no futuro a base de conhecimento das ciências que independem de novas tecnologias. As primeiras publicações registradas e catalogadas são as bases de conhecimento para as ciências humanas, não existe futuro sem o entendimento do passado. A *Web* para a pesquisa e para a otimização do tempo, o papel para a leitura.

Neste item, percebeu-se que existe uma preocupação em relação às publicações em meio digital. Isso acontece porque ainda existe uma certa insegurança em relação a durabilidade e manutenção de informações disponibilizadas em meio magnético. As publicações tradicionais, em papel, as pessoas podem visualizar, porém, também estão sujeitas aos desgastes como ações do tempo, da umidade, incêndios, etc. Em 11/09/2001, tivemos um exemplo do uso seguro das informações em meio magnético. Após a queda das duas torres gêmeas onde funcionava o “coração financeiro” dos americanos em New York, dois dias depois do desastre, todas as operações foram restabelecidas em função das cópias de segurança das bases de dados. A segurança das informações em meio magnético é uma questão de política, local, nacional e internacional.

Pergunta 09

Quando uma Instituição publica suas pesquisas em uma biblioteca digital a informação é imediatamente liberada e todas as comunidades de pesquisa podem usá-la como fonte para suas pesquisas. Desta forma, a qualidade da publicação deve ser objeto de preocupação pois torna globalmente transparente a produção e por consequência o nível intelectual do corpo docente. Como este programa de pós-graduação pretende tratar este assunto internamente?

Procurou-se nesta pergunta, verificar como os programas de pós-graduação pretendem tratar a questão da qualidade da produção do seu Programa.

Todos os coordenadores foram unânimes em afirmar que este problema não existe e que a transparência faz parte da política da Instituição. O acesso deve ser irrestrito a todas as publicações e se possível o sistema deve disponibilizar espaço para críticas, resenhas e opiniões. A responsabilidade pela qualidade deve ser acima de tudo do professor orientador. Ao coordenador cabe a tarefa de fiscalizar e observar o que ocorre no dia-a-dia, sempre atento aos índices que demonstram qualidade para evitar a falência do Programa a exemplo de situação ocorrida dentro da UFSM, onde o melhor Programa de Pós-Graduação em Educação Física do país foi considerado pela CAPES sem qualidade suficiente para continuar funcionando.

Hoje, as publicações de artigos e periódicos em revistas de qualidade elevam o nível do Programa, a publicação da T&D deve ser tratada da mesma forma. No Programa de Pós-Graduação em Química, o docente para ser aceito como professor orientador, deve ter acima de tudo, uma boa publicação de artigos e periódicos a nível internacional. O seu ingresso no

Programa deve agregar valores, seu currículo é avaliado no sentido de medir o interesse do Programa sobre sua produção acadêmica.

Percebeu-se que a qualidade das produções não é um fator preocupante, a busca pela qualidade é uma necessidade. Não existe nenhum impedimento para a divulgação e disseminação das T&D's produzidas em uma biblioteca digital no âmbito da UFSM.

Pergunta 10

A questão dos direitos do autor, de reprodução das obras e do direito à informação ganha hoje uma importância considerável, à medida que se populariza o uso da informação digitalizada. As Instituições que adotam bibliotecas digitais para publicar suas T&D's exigem do autor um documento de repasse do direito autoral para o Programa. Na UFSM, conforme determinação da Pró-Reitoria de Pós-Graduação e Pesquisa, o Programa solicita ao aluno uma autorização de repasse do direito autoral para a Universidade publicar na Biblioteca Central sua T&D em meio digital. O programa tem encontrado alguma dificuldade em relação à obtenção desta autorização? Tem alguma consideração a fazer a respeito do assunto?

Nesta pergunta procurou-se identificar quais as dificuldades que o Programa enfrenta em relação à obtenção desta autorização ou se existe alguma consideração a fazer a respeito do assunto.

Considerando que existe uma orientação da Pró-Reitoria de Pós-Graduação e Pesquisa, a maioria dos coordenadores não tem encontrado nenhum problema em relação à obtenção da assinatura da transferência do direito autoral da publicação para a UFSM.

O coordenador do Programa de Pós-Graduação em Engenharia Elétrica comenta que em função de tecnologias trabalhadas, pode haver casos em que não é possível liberar os resultados da pesquisa em detrimento de marcas e patentes do produto pesquisado. Muitas vezes parte do trabalho pode ser retirado do contexto no sentido de proteger um resultado. Existem situações em que o órgão de financiamento pode ser um impedimento para a publicação em função do recurso financeiro investido na pesquisa.

O coordenador do Programa de Pós-Graduação em Química comenta que algumas pesquisas geram registro de patente do produto. O registro é feito no Instituto Nacional de Propriedade Industrial (INPI), e a partir do registro do produto ou da pesquisa, as dificuldades deixam de existir em relação ao uso da tecnologia por terceiros.

No caso do registro de um produto desenvolvido no âmbito da UFSM, o direito à patente é distribuída entre o autor e a Universidade, ou seja, 30% com o autor e 70% com a Instituição.

Quando o direito a publicação é transferido para uma editora, o autor transfere também seu direito ao lucro sobre as vendas. O coordenador da Pós-Graduação em Química exemplifica a transferência de direitos de publicação de um livro escrito em conjunto entre três pesquisadoras da Química para uma editora internacional. A elas é reservado o direito a um desconto de 20% na compra dos exemplares e o brinde de somente um exemplar do livro que custa hoje na Europa aproximadamente R\$ 500,00.

Neste item, percebeu-se que o assunto não é objeto de discórdia entre autor e Programa e, em princípio, não existe nenhuma dificuldade quanto ao autor transferir o direito autoral de sua T&D para a UFSM publicar a produção em uma biblioteca digital mantida pela Universidade.

Neste capítulo apresentou-se a análise das respostas dos questionamentos feitos aos coordenadores dos programas de pós-graduação da UFSM. No próximo capítulo apresenta-se a proposta de uma ferramenta para publicação na *Web* das T&D's produzidas pela UFSM.

8 PROPOSTA PARA PUBLICAÇÃO DE T&D'S NA UFSM

Este trabalho, além dos objetivos analisados no capítulo anterior, onde, procurou-se medir a aceitabilidade do uso de bibliotecas digitais para publicação e recuperação de informações nos programas de pós-graduação da UFSM, também se propôs a estudar tecnologias disponíveis para publicação e disseminação do conhecimento produzido através de T&D's na *Web* bem como, analisar a viabilidade da implementação de uma biblioteca digital para oferecer à comunidade acadêmica um ambiente para submissão, busca e recuperação de textos completos em meio digital e que utilize a metodologia de arquivos abertos.

Para a definição de uma proposta, foi estudado tecnologias de comunicação e de armazenamento de informações, juntamente com uma análise de instituições de ensino e comunidades de pesquisa que já usam bibliotecas digitais para publicar suas T&D's bem como elas estão organizadas em torno do tema a nível nacional e internacional. Foram analisadas tecnologias e ferramentas, implementadas em instituições nacionais e internacionais, que propiciam a publicação e disseminação do conhecimento na *Web*. Algumas universidades brasileiras disponibilizam sua produção acadêmica de T&D's na *Web* através de projetos institucionais próprios porém usando padrões já estabelecidos por outras comunidades de pesquisa em relação às tecnologias de informação garantindo desta forma a interoperabilidade das informações entre diferentes bibliotecas digitais. Entretanto, o desenvolvimento de projetos próprios exige da instituição de ensino o envolvimento de pessoal de diferentes áreas e especializações, desta forma muitas vezes, o custo de implementação, implantação e manutenção inviabiliza a execução do projeto.

Atualmente, a UFSM dispõe de um *software* próprio para catalogação de T&D's desenvolvido pelo Centro de Processamento de Dados (CPD), o qual está em uso pela biblioteca central. O *software* que realiza a catalogação e gerenciamento do acervo da UFSM, faz parte da solução de gestão educacional SIE, que propicia à instituição um sistema de gestão administrativa mas não tem nenhuma forma de integração com as demais instituições, portanto, não existe um protocolo de troca de informação dele com outros sistemas. O *software* foi desenvolvido em uma arquitetura proprietária, não possuindo um protocolo que possa propiciar a interoperabilidade de dados com outras bibliotecas no Brasil e no exterior.

Atualmente, as bibliotecas da UFSM catalogaram no SIE aproximadamente 15296 T&D's desenvolvidas no âmbito da Universidade e aproximadamente 6271 de outras universidades. Na tabela 3, podemos verificar o número de empréstimos e renovações nas T&D's armazenadas em papel nas bibliotecas setoriais e na Biblioteca Central catalogas no sistema SIE no período de um ano. O sistema oferece consulta do catálogo bibliográfico na *Web* e mediante senha de acesso permite a reserva do exemplar e sua renovação caso o mesmo esteja emprestado ao usuário.

Tabela 3 – Consulta realizada entre março 2004 e março 2005

BIBLIOTECA	MATERIAL	EMPRÉSTIMO	RENOVAÇÃO
Biblioteca Setorial CT	T&D outras IES	34	77
Biblioteca Setorial CEFD	T&D outras IES	347	493
Biblioteca Setorial CE	T&D outras IES	139	178
Biblioteca Setorial CCSH	T&D outras IES	34	50
Biblioteca Setorial CCR	T&D outras IES	252	402
Biblioteca Central	T&D outras IES	472	1054
Biblioteca Setorial CT	T&D - UFSM	260	496
Biblioteca Setorial CEFD	T&D - UFSM	383	476
Biblioteca Setorial CE	T&D - UFSM	553	721
Biblioteca Setorial CCSH	T&D - UFSM	179	121
Biblioteca Setorial CCR	T&D - UFSM	1858	3275
Biblioteca Central	T&D - UFSM	2480	5629
Total		6991	12972

Fonte: Empréstimos de T&D no sistema SIE da UFSM.

A partir da sociabilidade e abrangência resultante da *Internet* associada às tecnologias de informática que promovem a integração de informações, definiu-se que o modelo para a publicação e disseminação na *Web* do conhecimento produzido pelos programas de pós-graduação da UFSM através das T&D, deve ser baseado em padrões que ofereça a total interoperabilidade entre as demais instituições que já publicam suas pesquisas na *Web* através de bibliotecas digitais.

A ferramenta escolhida como mais adequada para a proposta deve obrigatoriamente possuir as seguintes características:

- usar o protocolo OAI-PMH que promove a interoperabilidade entre arquivos *e-print*, através da extração de metadados dos repositórios das bibliotecas digitais de T&D's associadas;

- usar o padrão de metadados MTD-BR compatível com os padrões ETD-MS específico para T&D's para que possa oferecer o aproveitamento de sistemas legados permitindo a interligação com outros *Webservices* como por exemplo o Lattes usado pelo CNPQ;
- usar a linguagem de marcação XML para representar e transportar o conjunto de metadados de cada repositório de T&D;
- usar o vocabulário *Dublin Core* como padrão para definição de metadados descrevendo recursos de qualquer tipo de objeto digital fornecendo um modelo padrão de representação da base de conhecimento envolvida;
- estar interligada às demais bibliotecas nacionais e internacionais publicando os metadados das T&D's com isso fazendo parte de um portal único para pesquisa, no caso interligada à biblioteca digital brasileira de teses e dissertações;
- ser de fácil instalação e uso por parte dos programas de pós-graduação;

Para atender as condições necessárias, após análise, considerou-se que no momento atual o projeto coordenado pelo IBICT, cuja intenção é viabilizar a implantação de bibliotecas digitais nas IES integrando-as à BDTD e com a NDLTD da *Virgínia Tech University*, é o mais adequado. O *software* foi desenvolvido institucionalmente e oferecido as IES sem custo, de fácil implantação e gerenciamento pelos programas de pós-graduação e bibliotecas das instituições.

A adoção do sistema TEDE desenvolvido pelo IBICT e da camada do protocolo OAI-PMH, que o acompanha, permita à UFSM integrar-se imediatamente à BDTD e a exposição dos metadados das suas T&D's na NDLTD.

O Sistema TEDE foi desenvolvido para ser compatível com o OPAC que possibilita a instituição efetuar a publicação eletrônica de T&D's pelo sistema TEDE e, posteriormente, por meio de um módulo específico do sistema, exportar os metadados para o OPAC. O TEDE utiliza o padrão de metadados MTD-BR, compatível com os padrões ETD-MS, específico para T&D's, e o vocabulário *Dublin Core*, para a descrição de qualquer tipo de objeto digital usado no trabalho. O padrão MTD-BR foi escolhido por propiciar aproveitamento de sistemas legados e permitir interligação com outros *Webservices* por exemplo, a plataforma Lattes do CNPq citado anteriormente.

Na tabela 4 pode-se verificar a atual situação de algumas instituições em relação ao uso de bibliotecas digitais considerando o protocolo utilizado, o tratamento do direito autoral, a data de início, a quantidade de acervo já disponibilizado e o tipo de serviço oferecido.

Na UFSM é feito somente a catalogação da T&D's e não é disponibilizado o texto eletrônico.

Tabela 4 – Comparativo entre instituições em relação à bibliotecas digitais de T&D's

Implementação	USP	UFSC	UNICAMP	UFPE	IMPA	UFSM
Protocolo OAI	Sim	-	Não	Não	Z39.50	Não
Interoperabilidade com BD	Sim	Sim	Sim	Não	Sim	Não
Direito Autoral repassado às IES	Sim	Sim	Sim	-	-	Sim
Data de início	2001	1995	2001	2001	1994	2002
Quantidade no acervo de T&D's	-	4900	4500	-	-	21567*
Dowload do arquivo eletrônico	Sim	Sim	Sim	Não	Sim	Não

Fonte: Site das IES.

8.1 Implantando o TEDE Modular

Objetivando a análise do uso do *software* de publicação eletrônica TEDE por parte dos programas de pós-graduação, foi instalado pelo CPD da UFSM o TEDE Modular e disponibilizado para uso experimental sem a possibilidade de publicar os dados na BDTD. Na figura 11 podemos verificar a tela principal do sistema, que, conforme comentado anteriormente foi construído orientado por menus, onde temos a parte referente ao autor, a pós-graduação, a biblioteca, a administração do sistema e a busca de informações na biblioteca digital.

Para realizar os testes de uso da ferramenta optou-se por usar a atual infra-estrutura de redes, servidor hospedeiro do *software* e de dados disponível no CPD da Universidade sem custo para a Instituição. Seguindo as orientações dos manuais do produto, não houve nenhuma dificuldade na instalação. O tempo e a complexidade foi considerado baixo pela equipe de suporte técnico do CPD. O produto foi instalado provisoriamente e no momento que a Universidade decidir pela implantação e uso do produto a nível institucional, o CPD passa a exercer o controle e manutenção dos dados conforme a política de backup no sentido de manter a integridade das informações disponibilizadas pelo sistema TEDE.

Inicialmente, o administrador do sistema deve proceder a catalogação da biblioteca central e das bibliotecas setoriais da instituição que serão as depositárias das T&D's. Posteriormente, deve-se cadastrar todos os programas de pós-graduação da universidade, criando uma conta de acesso com senha para cada uma e indicando em qual biblioteca depositária ficará armazenado as T&D's daquele programa.

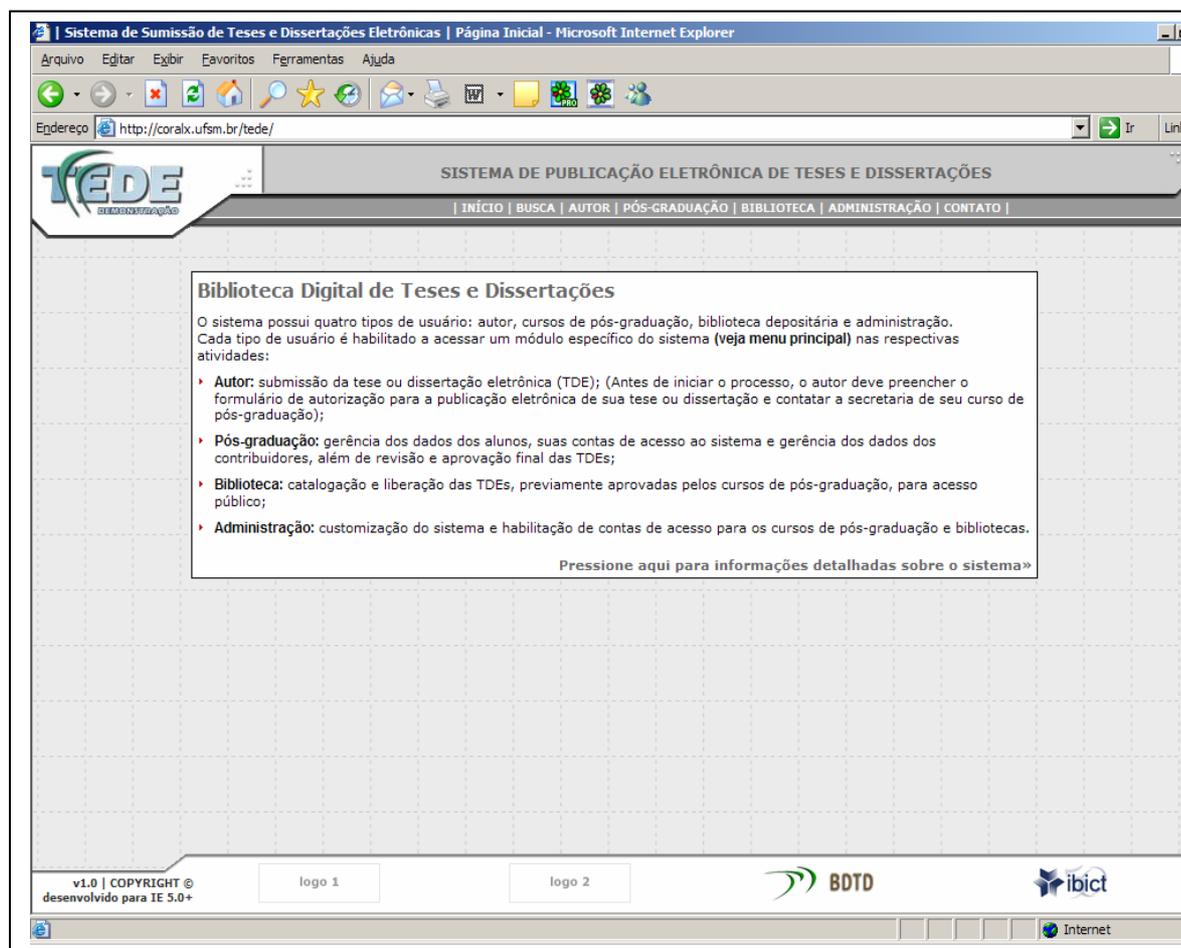


Figura 11 – Interface padrão do TEDE

No sentido de demonstrar o processo de publicação de uma T&D, descrevemos o papel de cada usuário no sistema e suas atividades. O processo sempre é iniciado pela pós-graduação com o cadastro dos dados da T&D e do autor juntamente com a assinatura da ficha de liberação do direito autoral para publicação. Finalizado o cadastro da T&D no sistema, a pós-graduação cria e libera um usuário e senha para o autor completar as informações sobre o trabalho e posterior retorno para a pós-graduação. Quando a pós-graduação aprova a T&D, o trabalho é submetido para a biblioteca central para que o processo seja finalizado. A biblioteca central revisa os metadados e realiza a catalogação da T&D objetivando a publicação liberando para a colheita de metadados pela BDTD controlada pelo IBICT. Após a

colheita pelo IBICT, os metadados são liberados para que o NDLTD também realize a colheita disponibilizando a T&D no seu portal. Em cada um dos módulos seus operadores informam metadados necessários à publicação da T&D mas apenas o autor é responsável pela entrada do texto completo em formato PDF de seu trabalho.

8.1.1 Papel da pós-graduação

A pós-graduação cadastra os dados do autor no sistema, o título da T&D, os membros da banca, o papel dos colaboradores, o órgão financiador caso exista. Terminado este procedimento, o operador do sistema cria uma conta para o autor acessar o sistema consultando ou alterando os dados de sua T&D e também para anexar os arquivos de texto, imagem, animação, etc. que compõem seu trabalho.

The screenshot shows a web browser window with the address bar displaying the URL: `http://coralx.ufsm.br/tede_pos/alterar_tde_inicial.php?tsIdentificador=4&pfCod=5`. The page title is 'Sistema de Publicação Eletrônica de Teses e Dissertações | Pós-Graduação | Alteração de TDE - Microsoft Internet Explorer'. The browser's menu bar includes 'Arquivo', 'Editar', 'Exibir', 'Eavoritos', 'Ferramentas', and 'Ajuda'. The address bar has 'Ir' and 'Links' buttons.

The main content area is titled 'TDE' and contains the following form fields:

- Autor**
 - Nome: Juçara Salete Gubiani
 - Lattes: [empty]
 - CPF: 428.452.890-49
 - E-mail: jucara@smail.ufsm.br
- Afiliação**
 - Nome: Universidade Federal de Santa Maria
 - Sigla: UFSM
 - Pais: Brasil
 - UF: Rio Grande do Sul
 - CNPJ: 31.231.231/0001-81
 - URL: http://www.ufsm.br
- Titulação**
 - Grau: Mestre (dropdown menu)
 - Titulação: Mestre em Engenharia de Produção (text input)
- Local de Defesa**
 - Pais: Brasil (dropdown menu)
 - UF: Rio Grande do Sul (dropdown menu)
 - Cidade: Santa Maria (text input)
- Data da defesa**
 - 20 (dropdown menu) - Junho (dropdown menu) - 2005 (text input) (AAAA)

On the left side of the page, there is a navigation menu with the following items: 'Programa: 1 Programa de Pós-Graduação em Engenharia de Produção - TI', '» TDE', '» Alterar', '» Contribuidor', '» Inserir', '« Voltar', and '« Menu Principal'.

Figura 12 – Submissão da T&D

Como o objetivo é ilustrar e mostrar as facilidades e dificuldades do uso do *software*, o tempo despendido para cadastrar uma T&D pelo usuário de pós-graduação foi de 5 minutos. Na figura 12 podemos verificar a interface de uma T&D já cadastrada no sistema e pronta para ser enviada ao autor para completar os dados relativos ao seu trabalho.

8.1.2 Papel do autor

A pós-graduação cadastra o trabalho e o sistema TEDE envia por e-mail a conta e senha para o autor acessar o sistema. O autor recebe o e-mail e pode acessar sua T&D para inserir ou alterar os dados, anexar os arquivos de texto, imagem, animação etc. que compõem seu trabalho. Ao terminar o processo, que sempre deve ser depois da aprovação do trabalho, o autor submete a T&D à pós-graduação para revisão e encaminhamento para a biblioteca central. No teste realizado não houve dificuldade quanto à execução deste procedimento no sistema.

Durante a revisão por parte da pós-graduação, ao serem detectados erros nos dados inseridos pelo autor ou nos arquivos da sua T&D, a pós-graduação manterá um diálogo com o autor por meio da opção de operação observação. Este diálogo é administrado pelo sistema, que emite *e-mails* automáticos tanto ao autor, quanto aos operadores habilitados da pós-graduação, informando as ações realizadas. Este fluxo entre autor e pós-graduação acontece até que esta considere os dados e arquivos da T&D corretos, para enviar a tese ou dissertação à biblioteca para a catalogação. Antes de enviar para a pós-graduação o autor deve escolher uma das seguintes opções:

1. Liberar o conteúdo dos arquivos para acesso público;
2. Liberar o conteúdo dos arquivos somente para a comunidade da universidade;
3. Reter o conteúdo dos arquivos por motivos de patente, publicação ou direitos autorais ou;
4. Liberar o conteúdo de alguns arquivos para acesso público, restringir o conteúdo de outros arquivos para acesso somente da comunidade da universidade e/ou reter o conteúdo de alguns arquivos por motivos de patente, publicação e/ou direitos autorais.

8.1.3 Papel da biblioteca

Quando, finalmente pronta a T&D é submetida para a biblioteca central, o usuário autorizado insere os metadados bibliográficos para a catalogação e realiza uma revisão geral dos metadados inseridos na T&D, tanto pelo autor, quanto pela pós-graduação. Esta ação pode dar origem a um novo diálogo entre a biblioteca e a pós-graduação e se necessário entre a pós-graduação e o autor. Este diálogo também será realizado por meio da opção observação

e administrado por meio de *e-mail* emitido automaticamente pelo sistema aos operadores habilitados destes módulos e o autor.

A catalogação da T&D é realizada pela biblioteca, somente na fase final do processo de submissão. Após o encerramento do fluxo da T&D entre os interessados e os arquivos forem considerados corretos para a publicação eletrônica é liberado o acesso público. Durante a catalogação é realizada a citação do autor conforme as normas da ABNT e o número de chamada da T&D dentro das normas bibliográficas.

Esse procedimento não foi executado devido a natureza do teste de submissão da T&D. Para operacionalizar esta atividade, a Instituição deve estar oficialmente cadastrada no IBICT para que o mesmo proceda a colheita e liberação dos metadados das T&D's dos programas de pós-graduação da UFSM.

8.1.4 Publicação, colheita de metadados e consulta

Após a catalogação e liberação da T&D pela biblioteca central, o IBICT exerce seu papel de provedor de serviço através da BDTD nacional, que integra as iniciativas nacionais de registro bibliográfico e publicação de T&D. A BDTD através do protocolo *harvesting* colhe os metadados informados e disponibiliza o serviço de busca e recuperação no seu portal na *Web*⁴². Como provedor de dados o IBICT expõe os metadados que foram incorporados na BDTD nacional para serem coletados pelo provedor de serviços internacional NDLTD.

O IBICT faz o papel de agregador de dados e serviços no âmbito do sistema TEDE, sua arquitetura assim como as tecnologias adotadas no projeto garantem que todas as T&D's nas IES brasileiras estejam disponíveis em nível local, em nível nacional e em nível internacional.

Para enriquecimento do trabalho, seria ideal que pudéssemos visualizar a consulta de uma T&D no portal da BDTD. Em se tratando de um teste, não foi possível liberarmos a T&D para colheita pelo IBICT através da BDTD.

Todas as T&D's produzidas no Brasil ou produzidas por brasileiros no exterior ficam disponíveis para recuperação conforme é mostrado na figura 13, onde é podemos verificar o resultado de uma consulta no portal da BDTD. A consulta realizada foi através da informação do nome de um pesquisador, o retorno da consulta foram todos os trabalhos publicados pelo

⁴² bdt.d.ibict.br/bdt.d/

pesquisador na Universidade Federal de Santa Catarina. Os dados foram catalogados pelo sistema implantado no PPGEP/UFSC e descrito no capítulo 5 deste trabalho.

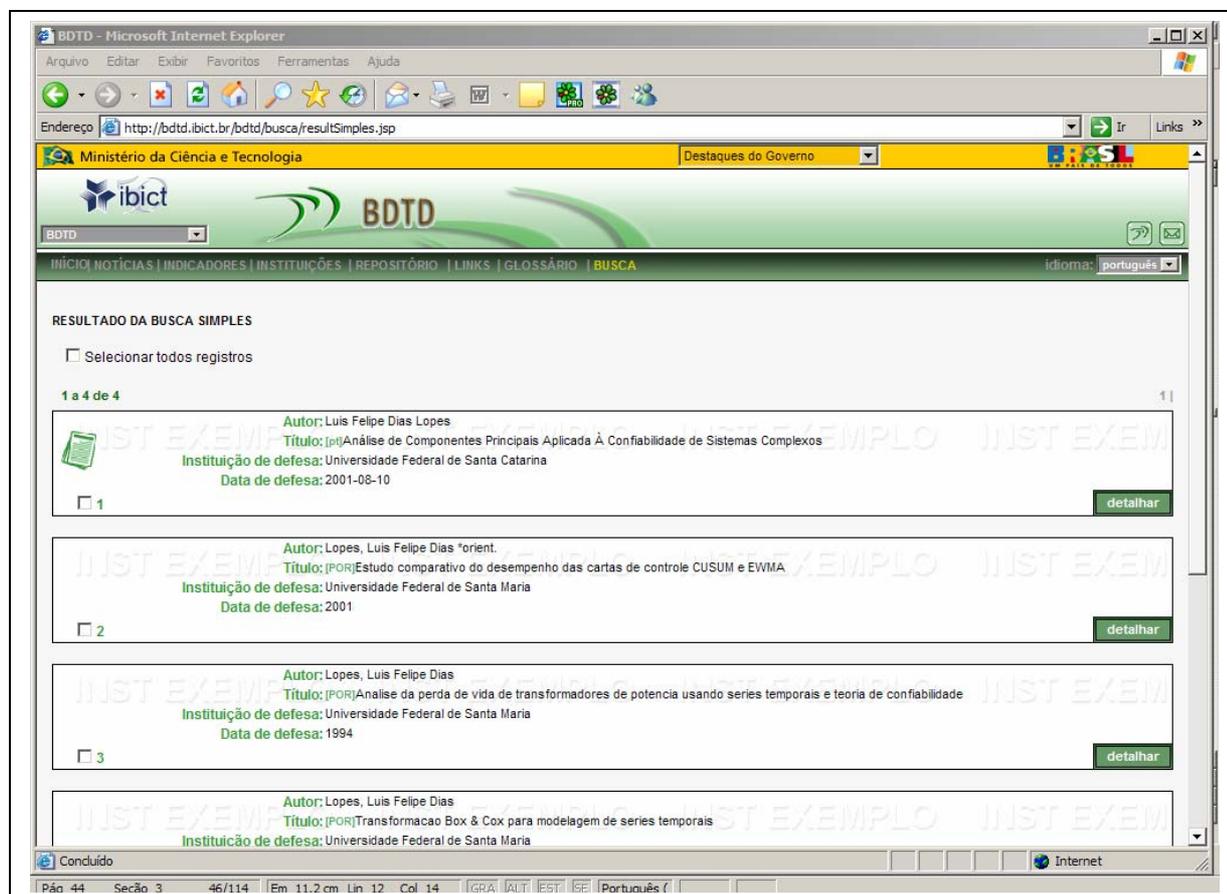


Figura 13 – Pesquisa no Portal da BDTD

8.1.5 Infra-estrutura exigida para o *software*

Para colocar em produção, o IBICT sugere que instituição recebedora do sistema TEDE organize um plano de implantação com algumas atividades bem pontuais buscando o comprometimento de vários segmentos da comunidade universitária. As atividades serão descritas logo abaixo.

8.1.5.1 Treinamento para uso do *software*

No material distribuído pelo IBICT, sugere-se a capacitação de pelo menos dois técnicos da IES para a implantação do TEDE na instituição. O treinamento de utilização do sistema deve ser realizado dentro do IBICT ser repassado a outros técnicos no âmbito da IES.

A capacitação pelo IBICT de técnicos das IES, ocorre durante os workshops realizados para este fim. Sugere-se a adoção da metodologia e conteúdos do workshop ministrado pelo IBICT para o repasse do treinamento às equipes que participarão da implantação, seja nos projetos piloto a serem implantados nas IES, seja na implantação definitiva das bibliotecas digitais de T&D's nas IES.

Em 2004, dois técnicos da UFSM, um analista de sistema pertencente ao CPD e uma bibliotecária do quadro de pessoal da Biblioteca Central da Universidade receberam treinamento nas dependências do IBICT em Brasília.

8.1.5.2 Sensibilização da comunidade universitária

Para o sucesso da implantação dentro da IES, o IBICT sugere que a Instituição realize ações de sensibilização no âmbito da universidade nos níveis gerenciais, reitor, pró-reitor de pesquisa, coordenadores de cursos de pós-graduação, diretor de biblioteca e alunos de pós-graduação, uma vez que estes são os agentes envolvidos na publicação da T&D. Estas atividades têm o objetivo de difundir o projeto da biblioteca digital de T&D's local e seu impacto na visibilidade da produção científica local, nos fluxos operacionais e no estabelecimento de uma interação mais estreita entre o programa de pós-graduação, o aluno e as bibliotecas. Abaixo segue algumas atividades aconselhadas para promover a sensibilização da comunidade acadêmica receptora do sistema:

- realização de palestras e workshops para divulgação da BDTD;
- produzir material de divulgação tais como folder, banner e outros que esclareçam sobre a finalidade do projeto e sobre a importância sobre a necessidade da interação entre as unidades operacionais do sistema TEDE;
- utilizar a rede interna da Universidade para veiculação de textos de sensibilização;
- utilizar a página da BDTD nacional para acesso a metadados e/ou a textos completos provenientes dos diversos provedores e sua integração internacional.

8.1.5.3 Ambiente operacional

O IBICT orienta, conforme manual de instalação e operação do sistema TEDE, que a IES prepare o ambiente computacional nos moldes da especificação contida no manual. É oferecido suporte técnico durante a instalação e implantação do *software*. Caso haja

necessidade, técnicos do IBICT podem se deslocar até a IES para ajudar na implantação além da disponibilidade de listas de discussões sobre o assunto.

8.1.5.4 Grupos de trabalho

O IBICT sugere que a instituição receptora do sistema crie dois grupos de trabalho. Um comitê para a definição de estratégia de implantação do *software* na universidade e uma equipe técnica que se responsabilizará pela operação do pacote e treinamento local, quando novos cursos de pós-graduação e bibliotecas associadas a estes se incorporarem ao projeto. O comitê de implantação tem como objetivo estratégico a difusão do projeto no âmbito da universidade. Este comitê deve contar com representantes de cursos de pós-graduação, bibliotecas, pró-reitorias. Este comitê deve desenvolver um plano estratégico de implantação da BDTD na universidade e acompanhar seu desenvolvimento. É recomendável que o Comitê de implantação providencie junto ao reitor, um documento que formalize a criação da biblioteca digital de teses e dissertações da universidade. Sugere-se que o coordenador do comitê de implantação seja a pessoa de contato entre o projeto local e a coordenação nacional da BDTD. A equipe técnica da IES deve estar preparada para receber treinamento tão logo os técnicos que participaram do workshop oferecido pelo IBICT retornem às suas instituições de origem e deve ser formada com pelo menos:

- dois representantes de bibliotecas (técnicos de informação) para serem treinados na operação do pacote;
- dois representantes do curso de pós-graduação escolhida para serem treinados na operação do pacote;
- dois técnicos de informática para conhecer o pacote e dar apoio em eventuais problemas de infra-estrutura;
- coordenador do comitê de implantação.

8.1.5.5 Projeto-piloto

Devido à mudança no processo de publicação de T&D's dentro da IES, o IBICT sugere a implantação de um projeto piloto no âmbito da universidade com no máximo dois cursos de pós-graduação com o compromisso de fornecer o acompanhamento e o apoio técnico necessário na adoção do sistema TEDE dentro da instituição. Este acompanhamento

pode ser realizado por meio de lista de discussão dos participantes dos workshops organizados pelo IBICT onde todos os problemas e dúvidas devem ser reportados para que a equipe do IBICT possa apoiar na solução dos mesmos.

Depois do teste do projeto piloto do sistema TEDE, o comitê de implantação e a equipe técnica devem iniciar ações para a expansão da abrangência da BDTD no âmbito da IES. Neste sentido o IBICT sugere que a instituição analise a implantação do projeto piloto e com base nessa experiência planeje a implantação do sistema nos demais programas de pós-graduação com a seguinte definição:

- nomear um responsável pela administração do sistema;
- definir as metas para o crescimento da BDTD local;
- definir e planejar a incorporação gradativa de outros cursos de pós-graduação. A implantação, de uma só vez, de um número muito grande de cursos pode congestionar o processo, torná-lo moroso e comprometer o sucesso na implantação;
- definir um cronograma detalhando quais os cursos que serão contemplados em cada fase da implantação;
- definir uma estratégia de treinamento para os novos participantes da BDTD local;
- definir o texto relativo à autorização do autor para a publicação eletrônica de sua T&D e criando infra-estrutura de apoio ao aluno/autor na conversão de sua tese ou dissertação em formato a ser publicado (PDF);
- discutir junto ao IBICT, através da lista de discussão a ser criada, o apoio à IES na implantação da BDTD local.

Neste capítulo foi apresentado a ferramenta para publicação de T&D's dos programas de pós-graduação da UFSM. O ambiente considerado como proposta é o Sistema TEDE de publicação de T&D's desenvolvido pelo IBICT e oferecido às IES sem custo.

9 CONCLUSÃO

Historicamente, a responsabilidade de guardar e disseminar o conhecimento produzido por uma nação é repassado para as bibliotecas. Vimos durante o desenvolvimento do trabalho que a economia esta centrada em três fatores, o trabalho, o capital e por último, em franca ascensão, o conhecimento. Percebe-se a importância e a necessidade de se ter a disponibilidade do conhecimento, para se obter informações imediatas em todas as áreas da economia, principalmente no desenvolvimento de novas pesquisas. Nesse sentido, este trabalho procurou responder o seguinte questionamento: Qual a melhor forma de disseminar o conhecimento produzido através das T&D's desenvolvidas nos programas de pós-graduação no contexto da UFSM?

Para responder a questão considerada na introdução da pesquisa, procurou-se estudar e analisar tecnologias disponíveis no mercado e os aspectos legais para a formulação de uma proposta de implementação de uma biblioteca digital para publicação e disseminação na *Web* do conhecimento gerado através das T&D's dos programas de pós-graduação da UFSM.

Levantou-se um amplo referencial bibliográfico sobre o assunto, onde, foram analisadas tecnologias e padrões para divulgação e disseminação de informações na *Web* considerando como pré-requisito na análise a interoperabilidade dessas informações, objetivando buscar subsídios para fundamentar o problema.

Para a conclusão sobre uma proposta de ferramenta a ser utilizada na publicação de T&D's produzidas na UFSM, a análise de tecnologias e interfaces de algumas bibliotecas digitais em instituições no Brasil e no exterior foi de extrema relevância. Procurou-se analisar instituições que oferecem para a comunidade acadêmica um ambiente para submissão, busca e recuperação de textos completos, em meio digital, com o uso de padrões, de interoperabilidade e troca de informações.

Para avaliar se as pessoas usam T&D's como base de conhecimento e qual a visão da comunidade acadêmica sobre a disseminação de T&D's na *Web*, foram realizadas entrevistas com os coordenadores de programas de pós-graduação da Universidade considerando as áreas de conhecimento do CNPq. O direito à propriedade intelectual também foi considerado, porém com menor ênfase conforme veremos neste capítulo.

Todas as questões colocadas acima deram embasamento e foram consideradas como pré-requisito para uma proposta de ferramenta para publicar na *Web* as T&D's produzidas pelos programas de pós-graduação da UFSM, objetivando a disseminação do conhecimento.

No que diz respeito às tecnologias de comunicação e de informação, não se tem dúvidas quanto à existência de ferramentas que oferecem condições de publicação na *Web* e a total disponibilidade dos dados nela publicados. A ferramenta definida para a proposta, e que atende aos objetivos definidos, usa tecnologias discutidas durante a pesquisa bibliográfica e possui padrões que oferecem total interoperabilidade das informações publicadas. Desenvolvido pelo IBICT, o sistema TEDE de publicação foi construído usando os padrões de arquivos abertos, discutido durante o trabalho, cujo protocolo propicia a interoperabilidade das informações e está disponível para todas as IES, para que as mesmas publiquem sua produção de T&D's em bibliotecas digitais locais e, posteriormente, essas informações possam ser disponibilizadas na BDTD.

O sistema TEDE usa o protocolo OAI-PMH para colheita dos metadados, a linguagem de marcação XML para representar e transportar o conjunto de metadados dos repositórios locais de T&D's e o vocabulário *Dublin Core* como padrão para a descrição e definição dos metadados. Com o uso do vocabulário *Dublin Core*, é possível descrever recursos das T&D's independente do tipo de objeto digital, ou seja, textos, sons, imagens, animações, entre outros. A interoperabilidade das informações entre as bibliotecas digitais, que atualmente é objeto de discussão e necessidade em termos de sistemas de informações, é garantida com o uso de *software e-print* usado na ferramenta e o protocolo OAI-PMH via extração de metadados dos repositórios usando o padrão de metadados MTD-BR compatível com os padrões ETD-MS específicos para publicação de T&D's. A tecnologia de construção do TEDE, permite também o aproveitamento de sistemas legados disponíveis dentro das instituições possibilitando a interligação com outros *Webservices* como por exemplo a plataforma *Lattes* usado pelo CNPq.

No sistema TEDE, após a publicação da T&D pela instituição através da sua biblioteca central, o IBICT exerce seu papel de provedor serviço, entra na rede e colhe os metadados nas bibliotecas digitais locais e expõe a T&D na BDTD tornando disponível o trabalho para consultas no portal. De imediato, o IBICT faz seu papel de provedor de dados expondo os metadados que foram incorporados na BDTD Nacional para serem coletados pelo provedor de serviços internacional do NDLTD. A partir da exposição dos metadados da

T&D's e a colheita pela NDLTD, o trabalho é disponibilizado no portal da *Virgínia Tech University*.

Com o uso do sistema TEDE, a UFSM passa a fazer parte da rede internacional de publicação, hoje a nível de Brasil, controlada pelo IBICT. Além dessa possibilidade, a Instituição recebe todo o suporte ao produto desde a instalação, o treinamento e acompanhamento por pessoas qualificadas com atendimento local se houver necessidade. A Universidade terá os custos de aquisição de infra-estrutura necessária em termos de *hardware e software* conforme descrito no capítulo referente à especificação técnica, lembrando que o TEDE, usa *software* livre em sua construção. Considerando a facilidade do uso da ferramenta, acredita-se que as secretarias dos programas de pós-graduação, com o devido treinamento das pessoas, não terão dificuldades na sua operação e o programa passa a ter visão de sua produção de T&D's disponibilizada na *Web* com todos os dados do trabalho, dados da banca e do fomento caso exista.

A questão da segurança e a disponibilidade das T&D's publicados na *Web* é assegurada em função da escolha da ferramenta. O ambiente TEDE é um projeto do IBICT o qual propicia as IES a total integração da ferramenta à BDTD oferecendo um portal único para pesquisa conforme mostrado no capítulo anterior. Por se tratar de um projeto Institucional é de interesse da nação, manter e tornar disponível na *Web* publicações de T&D's em benefício da C&T. Um dos objetivos do portal da BDTD é reunir publicações de T&D's de todas as IES e disponibilizar o acesso na *Web* sem restrições. Desta forma, as instituições participantes do projeto possuem garantia da continuidade e manutenção da ferramenta por parte do IBICT.

No Brasil, algumas IES desenvolveram ambientes próprios para divulgação de sua produção acadêmica de T&D's, podemos citar a USP, a UNICAMP, a UFSC/PPGEP entre outras. Porém, o desenvolvimento de projetos proprietários demanda altos investimentos e o envolvimento de pessoal qualificado. Com o projeto desenvolvido pelo IBICT cuja intenção é de buscar a produção acadêmica das IES e disponibilizar para a comunidade de pesquisa em um portal de acesso livre, a UFSM passa a ter a oportunidade de divulgar sua produção acadêmica de T&D's na *Web* com total interoperabilidade das informações e segurança quanto à autenticidade das pesquisas somada a garantia da continuidade do projeto sem a preocupação com custos de desenvolvimento e manutenção do *software*.

Na questão do uso de T&D's como referencial bibliográfico no desenvolvimento das pesquisas dentro dos programas de pós-graduação da UFSM, buscou-se, através de uma

pesquisa no banco de dados do SIE da Universidade analisar o uso através da informação de empréstimos e renovações conforme mostrado na tabela 3 do capítulo 8. Analisando a informação dos empréstimos, percebe-se que no período de um ano foi emprestado aproximadamente 7000 T&D's disponibilizadas em papel, o que nos fornece uma média de 600 empréstimos no mês. Considerando um universo de 2250 alunos ativos nos programas de pós-graduação, obtemos uma média de 0,25 retirada de T&D's no mês por aluno. A relação é de 4 alunos para uma pesquisa realizada em T&D's armazenadas nas bibliotecas da UFSM por mês.

Para as coordenações, nas T&D's está contido o conhecimento atualizado e de relevância, na maioria das vezes a base para a publicação de artigos e periódicos os quais hoje são medidas de qualidade dos programas. A disponibilidade imediata dessa base de conhecimento, trará um ganho no tempo de produção e disseminação de novos conhecimentos e a promoção do pesquisador em conjunto com a instituição, não somente no ambiente em que ele está inserido como também a nível mundial. Espera-se que após a disponibilidade na *Web* das T&D's da UFSM em bibliotecas digitais, o índice de consulta possa ser melhorado, o aluno não precisará deslocar-se até uma biblioteca para obter a informação, bastará acessar a *Internet* e o portal da BDTD.

A preocupação com a disseminação e a busca do conhecimento dentro da instituição, deve ser uma ação institucional, a *Virgínia Tech University* é um exemplo. O projeto NUDL permite o desenvolvimento de redes de publicações do conhecimento através do NDLTD, no qual o sistema TEDE está integrado, com um objetivo claro de melhorar a educação e preparar as gerações futuras de alunos oferecendo conhecimento necessário para a realização de suas tarefas e que eles mesmos possam submeter pessoalmente seus próprios trabalhos de T&D's para publicação. Uma mudança cultural, não só em relação às práticas dos alunos como também ao comportamento dos orientadores e da própria instituição, que deve oferecer condições de acesso às tecnológicas em todas as áreas de pesquisa.

Durante as entrevistas com os coordenadores dos programas de pós-graduação, ficou claro que parte do referencial bibliográfico está nas T&D's publicadas recentemente, o estado da arte do objeto em estudo geralmente encontra-se nas últimas publicações do assunto. Geralmente o uso de T&D's como referência durante a execução da pesquisa se dá em função do assunto abordado, da área de concentração e por consequência do grupo de trabalho onde o autor está inserido ou do próprio orientador da pesquisa. Algumas áreas do conhecimento como por exemplo *Linguística, Letras e Artes*, buscam o referencial bibliográfico na literatura

que deu origem a própria área de conhecimento, a atualidade serve de base de comparação no presente.

Percebeu-se também, que apesar de todas as tecnologias disponíveis na *Web*, muitas vezes o acesso ao conhecimento encontrado nas T&D's fica restrito a ambientes de difícil busca, com filtros que não oferecem facilidade ao usuário pesquisador. As coordenações dos programas de pós-graduação da UFSM entrevistados demonstraram interesse em disseminar o conhecimento produzido através das T&D's em bibliotecas digitais e conseqüentemente a liberação na *Web*. Essa afirmação procede em função da atual restrição de acesso ao conhecimento publicado e armazenado em papel na Biblioteca Central e em função do acesso restrito ao local físico. As coordenações afirmam que a disseminação do conhecimento, hoje, acontece em função da publicação de artigos e periódicos publicados em revistas científicas, mesmo considerando que a base para essas publicações, muitas vezes, está na T&D. A iniciativa da BDTD coordenada pelo IBICT, cujo objetivo é a integração das bibliotecas digitais brasileiras de T&D's das IES é uma solução imediata para a disseminação do conhecimento em um portal único de busca.

O direito autoral não é considerado um impedimento para publicar na *Web* das T&D's. Essa questão não foi objeto discutido durante o trabalho, considerando a lei nº 9.610, de 19 de fevereiro de 1998 que altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. A UFSM desde 2003 tem solicitado aos programas de pós-graduação a autorização para que a T&D possa ser publicada em meio digital.

Durante as entrevistas, uma das questões consideradas diz respeito obtenção da autorização conforme determinação da Pró-Reitoria de Pós-Graduação e Pesquisa. Todos os coordenadores de Programas afirmaram não ter encontrado nenhuma dificuldade quanto à obtenção da autorização e em caso da necessidade de proteção dos direitos deve-se registrar a patente da T&D no INPI. Após o registro o autor ou autores tem a proteção contra cópias e qualquer publicação advinda da T&D deve informar o registro da patente.

Analisando todos os aspectos relacionados, tais como a disponibilidade de tecnologias, infra-estrutura de comunicação, questões culturais em relação ao uso e armazenamento de informações, a relevância da qualidade das publicações, a troca de informação entre as comunidades de pesquisa e a disseminação do conhecimento nos permite concluir que a publicação na *Web* das T&D's tornará transparente a produção acadêmica dos programas de pós-graduação da UFSM, no contexto nacional e internacional através do uso do *software*

TEDE com a disponibilidade para consulta dos trabalhos no portal da BDTD e do NDLTD na *Virgínia Tech University*.

Este trabalho propôs a implantação de uma ferramenta para publicação de T&D's na *Web*. Fica como sugestão para outros trabalhos, após a implantação em nível Institucional, a avaliação da produção do conhecimento em todos os programas de pós-graduação da UFSM considerando a disponibilidade de acesso às T&D's.

REFERÊNCIAS

DAY, Michael. **The scholarly journal in transition and the PubMed Central proposal.** *Ariadne*, v. 21, Sept.1999. Disponível em: <<http://www.ariadne.ac.uk/issue21/pubmed/>>. Acesso em: 10/2004.

DRABENSTOTT, Karen M.; BURMAN, Celeste M. **Revisão analítica da biblioteca do futuro.** *Ci. Inf.*, Maio/Ago. 1997, vol.26, no.2. ISSN 0100-1965.

DRABENSTOTT, Karen M.; BURMAN, Celeste M. **Revisão analítica da biblioteca do futuro.** *Ci. Inf.* [online]. Maio/Ago. 1997, vol.26, no.2. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651997000200012&lng=pt&nrm=iso>. ISSN 0100-1965. Acesso em: 10/2004.

BECKETT, Dave [RDF/XML Syntax Specification \(Revised\)](#) Dave Beckett (editor), W3C Recommendation, 10 February 2004. Acesso em: 10/2004.

BERGMAN, Michael K. The deep *Web*: surface hidden value. *Journal of Electronic Publishing*, v. 7, n. 1, Aug. 2001. Disponível em: <<http://www.press.umich.edu/jep/07-01/bergman.html>>. Acesso em: 08/2004.

BERNERS-LEE, Tim; CONNOLLY, Dan; SWICK, Ralph R. **Web Architecture: Describing and Exchanging Data**, < <http://www.w3.org/1999/04/WebData>>. Acessado em: 11/2004.

CHATAIGNIER, Maria Cecilia Pragana; SILVA, Margareth Prevot da;. **Biblioteca Digital: a experiência do Impa.** *Ci. Inf.* v.30 n.3 Brasília set./dez. 2001. ISSN 0100-1965 .

CHIZZOTTI, A. **Pesquisa em Ciências Humanas e Sociais.** São Paulo. Cortez Editora, 3^a ed. 1998.

COSTA, S.M. de S. **Controle de qualidade em periódicos científicos eletrônicos disponibilizados na Internet: a questão do julgamento pelos pares.** *Revista de Biblioteconomia de Brasília*, v. 20, n.2, p.227-236, jul./dez.1996.

FOX, Edward A., GONÇALVES, Marcos André, **Technology and Research in a Global Networked University Digital Library (NUDL)**. *Ciência da Informação*, Brasília, Vol. 30, No. 3, 2001.

GIL, A .C., **Métodos e técnicas de pesquisa social**. 2. ed. São Paulo: Atlas, 1989.

GINSPARG, Paul. Creating a global knowledge network. *In: FREEDOM OF INFORMATION CONFERENCE: the impact os open access on biomedical research*, 2000, New York Academy of Medicine, BioMed Central. New York, 2000. Disponível em: <<http://www.biomediccentral.com/info/ginsparg-ed.asp>>. Acessado em: 10/2004.

GINSPARG, Paul. *Winners and Losers in the Global Research Village*. Disponível em: <http://arxiv.org/blurp/pg96unesco.html> . Acessado em: 10/2004.

GODOY, A. S., **Introdução a pesquisa qualitativa e suas possibilidades**. ERA, São Paulo. 1995. v. 35, n. 3, p. 21-29.

GOMEZ, Maria Nélide González de. **O papel do conhecimento e da informação nas formações políticas ocidentais**. *Ciência da Informação*, Brasília, v.16, n. 2, p. 157-167. jul./dez. 1987.

GUNTHER, O.; VOISARD, A. **Metadata in Geographic and Environmental Data Menagement**. In: W. Klas, e A Sheth, editors. *Managing Multimedia Data: Using Metadata to Integrate and Apply Digital Data*. McGraw Hill, 1997

GRUBER, Tom .R. **A Translation Approach to Portable Ontologies**. *Knowledge Acquisition*, v.5, n.2, p.199-200, 1993.

HARNARD, Stevan. *The self-archiving initiative: nature Web debates*. Disponível em:< <http://www.nature.com/nature/debates/e-access/Articles/harnad.html>> Acesso em: 11/ 2004.

HILLMANN, Diane. **Using Dublin Core, 2000. Dublin Core Metadata Initiative**. Disponível na World Wide Web: < <http://dublincore.org/documents/2000/07/16/usageguide/>>. Acesso em 10/2004.

HILLMANN, Diane. **Using Dublin Core, 2003. Dublin Core Metadata Initiative**. Disponível na World Wide Web: < <http://dublincore.org/documents/2003/08/26/usageguide/>> . Acesso em 10/2004.

ISO. Disponível em: <www.iso.org/cate/d16387.html>.

MAFFIA, Sueli Amélia Melhim; SOUTHWICK, Silvia Barcellos; ROCHA, Eliana da Conceição. **Metodologia para implantação das Bibliotecas Digitais de Teses e Dissertações locais das Instituições de Ensino Superior – IES**. IBICT, Brasília, 2003.

MARCONDES, Carlos Henrique; SAYAO, Luis Fernando. **Documentos digitais e novas formas de cooperação entre sistemas de informação em c&t**. *Ci. Inf.* [online]. set./dez. 2002, vol.31, no.3 [citado 04 Setembro 2004], p.42-54. Disponível na World Wide Web: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000300005&lng=pt&nrm=iso>. ISSN 0100-1965.

MARCONDES, Carlos Henrique; SAYAO, Luís Fernando. **Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira**. *Ci. Inf.* [online]. set./dez. 2001, vol.30, no.3 [citado 04 Setembro 2004], p.24-33. Disponível na World Wide Web: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000300004&lng=pt&nrm=iso>. ISSN 0100-1965.

MASIERO, Paulo César; BREMER, Carlos Frederico; COLETTA, Teresinha das Graças *et al.* **A Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo**. *Ci. Inf.*, set./dez. 2001, vol.30, no.3, p.34-41. ISSN 0100-1965.

MOUTINHO, Karina, CUNHA FILHO; Paulo C.; LIMA, Alessandra Marques de *et al.* **Liber: alternativa para publicação eletrônica**. *Ci. Inf.*, maio/ago. 2002, vol.31, no.2, p.80-85. ISSN 0100-1965.

OPEN archives initiative. Disponível em: <<http://www.openarchives.org/>>.

Open archives: **caminho alternativo para a comunicação científica**. *Ciência da Informação*, Brasília, v. 29, n. 3, set./dez. 2000

PACHECO, Roberto Carlos dos Santos; KERN, Vinícius Medina. **Transparência e gestão do conhecimento por meio de um banco de teses e dissertações: a experiência do PPGEP/UFSC**. *Ci. Inf.*, set./dez. 2001, vol.30, no.3, p.64-72. ISSN 0100-1965.

PINHEIRO, Lena Vania Ribeiro. **Comunidades científicas e infra-estrutura tecnológica no Brasil para uso de recursos eletrônicos de comunicação e informação na pesquisa**. *Ci. Inf.* [online]. set./dez. 2003, vol.32, no.3 [citado 04 Setembro 2004], p.62-73. Disponível na

World Wide Web: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652003000300008&lng=pt&nrm=iso>. ISSN 0100-1965.

THE SANTA FÉ CONVENTION FOR THE OPEN ARCHIVES INITIATIVE.

Disponível em: <<http://www.openarchives.org/>>.

TRISKA, Ricardo; CAFÉ, Ligia. **Arquivos abertos: subprojeto da Biblioteca Digital Brasileira.** *Ci. Inf.*, set./dez. 2001, vol.30, no.3, p.92-96. ISSN 0100-1965.

TRIVIÑOS, Augusto N. S. Introdução à pesquisa em Ciências Sociais: A pesquisa qualitativa em educação. 1. ed, São Paulo: Atlas, 1987. SENA, Nathália Kneipp.

SIMON, E., TOMASIC, A.; GALHARDAS, H. A **Framework for classifying Scientific Metadata.** American Association for Artificial Intelligence, 1998.

VIDOTTI, Silvana Aparecida Borsetti Gregorio; OLIVEIRA, Gabriela Pereira de; SOUZA, Maria Fernanda Sarmiento. **A iniciativa dos arquivos abertos como alternativa a publicações científicas in:** II Simpósio Internacional de Bibliotecas digitais, 2004, Campinas, 2004.

WEIBEL, Stuart. **The State of the Dublin Core Metadata: Initiative.** D-Lib Magazine, April, 1999.