

FUNDAÇÃO DE ENSINO “EURÍPIDES SOARES DA ROCHA”
CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA – UNIVEM
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

LUCAS SALVIANO ROMANI

**ANÁLISE E IMPLANTAÇÃO DE REPOSITÓRIO DIGITAL
UTILIZANDO SOFTWARE LIVRE DSPACE**

MARÍLIA
2009

LUCAS SALVIANO ROMANI

ANÁLISE E IMPLANTAÇÃO DE REPOSITÓRIO DIGITAL UTILIZANDO
SOFTWARE LIVRE DSPACE

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação da Fundação de Ensino “Eurípides Soares da Rocha”, mantenedora do Centro Universitário Eurípides de Marília – UNIVEM, como requisito para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Ms. ELVIS FUSCO

MARÍLIA
2009

Romani, Lucas Salviano.

Análise e Implantação de Repositório Digital utilizando Software Livre DSpace / Lucas Salviano Romani; orientador: Elvis Fusco. Marília, SP: [s.n.], 2009.

98 f.

Trabalho de Conclusão de Curso – Curso de Bacharelado em Ciência da Computação, Fundação de Ensino “Eurípides Soares da Rocha”, mantenedora do Centro Universitário Eurípides de Marília – UNIVEM, Marília, 2009.

1.Repositório Digital 2.Metadados 3.DSpace 4.Interoperabilidade
5.Software Livre

CDD: 004.21



CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

TRABALHO DE CONCLUSÃO DE CURSO – AVALIAÇÃO FINAL

Lucas Salviano Romani

**ANÁLISE E DESENVOLVIMENTO DE REPOSITÓRIO DIGITAL UTILIZANDO SOFTWARE
LIVRE DSPACE**

Banca examinadora da monografia apresentada ao Curso de Bacharelado em Ciência da Computação do UNIVEM/F.E.E.S.R., para obtenção do Título de Bacharel em Ciência da Computação.

Nota: 9,5 (nove e meio)

Orientador: Elvis Fusco

1º. Examinador: Fabio Lucio Meira

2º. Examinador: Fábio Dacêncio Pereira

Marília, 02 de dezembro de 2009.

A Deus, pois sem suas bênçãos nada se torna possível.

A meu pai Argemiro e minha mãe Sônia.

A minha família e verdadeiros amigos.

AGRADECIMENTOS

Agradeço a Deus, por me permitir e me abençoar na conquista de mais um objetivo da minha vida.

Agradeço a meus pais, Argemiro e Sônia, pelo apoio, pela presença, sacrifícios e paciência em todos os momentos da minha vida, e pelo amor e apoio nestes quatro duros anos e também em todos os outros da minha vida. Serei eternamente grato a vocês.

Agradeço ao amigo Mário Cleber, por ter me incentivado e apoiado tanto para que se tornasse possível a conclusão da minha graduação.

Agradeço ao meu irmão Thiago, por ter me apoiado durante estes anos de estudo de todas as formas possíveis e me orientado sempre com as palavras necessárias de se ouvir.

Agradeço ao meu orientador Elvis, pelo interesse e apoio no desenvolvimento deste projeto.

Agradeço aos verdadeiros amigos, pelo apoio e por entenderem as ausências devido ao comprometimento a este trabalho e aos meus estudos, em especial agradeço à Fabiana, Jenni, Eliana, Diego, Marília, Marcelle, Giovane, Vanessa, Omoto, Cyntia, Fogo, Mateus, Marina, Winkel, Tarsila e Cida.

Agradeço aos docentes do meu curso, pelos valiosos ensinamentos.

E agradeço a todos mais que estiveram presentes de forma positiva em minha vida durante estes anos de estudos, me apoiando e incentivando.

MUITO OBRIGADO A TODOS!

*"A vida é para quem topa qualquer parada,
e não para quem pára em qualquer topada."*

Bob Marley

ROMANI, Lucas Salviano. **Análise e Implantação de Repositório Digital utilizando Software Livre DSpace**. 2009. 98f. Trabalho de Conclusão de Curso de Bacharelado em Ciência da Computação – Centro Universitário Eurípides de Marília, Fundação de Ensino “Eurípides Soares da Rocha”, Marília, 2009.

RESUMO

O uso de meios computacionais para geração de documentos digitais, é o mais usado hoje em dia em todos os tipos de instituições, mas ocorre que grande parte destes documentos deixa de ser disponibilizado para aquelas pessoas as quais eles interessam. Este trabalho tem como objetivo o estudo dos repositórios digitais, com o intuito do desenvolvimento de um repositório digital capaz de fazer com que o conhecimento gerado pelo Grupo de Pesquisa – Novas Tecnologias em Informação do Programa de Pós-Graduação em Ciência da Informação da UNESP de Marília seja preservado no tempo e disponibilizado a todos os interessados de maneira eficiente e, a partir deste estudo, demonstrar o uso de repositórios digitais na gestão de documentos digitais em empresas e instituições de ensino e pesquisa.

Palavras-Chave: Repositório Digital. Metadados. Dspace. Interoperabilidade.

ROMANI, Lucas Salviano. **Análise e Implantação de Repositório Digital utilizando Software Livre DSpace**. 2009. 98f. Trabalho de Conclusão de Curso de Bacharelado em Ciência da Computação – Centro Universitário Eurípides de Marília, Fundação de Ensino “Eurípides Soares da Rocha”, Marília, 2009.

ABSTRACT

The use of computational methods to generate digital documents, is the most widely used today in all types of institutions, but it happens that great part of these documents is no longer available for those people whom it matters. This work has as purpose the study of the digital repositories, aiming the development of a digital repository able to ensure that the knowledge generated by the *Research Group – Novas Tecnologias em Informação* from the *Programa de Pós-Graduação em Ciência da Informação da Unesp de Marília* can be preserved in time and efficiently available to all interested, and after this study, demonstrate the use of digital repositories on the management of digital documents in companies and educational and research institutions.

Keywords: Digital Repository. Metadata. Dspace. Interoperability.

LISTA DE ILUSTRAÇÕES

FIGURA 1 - Gráfico com a quantidade de repositórios desenvolvidos por cada sistema	54
FIGURA 2 - Exemplo de representação dos metadados básicos no DSpace	60
FIGURA 3 - Processo de <i>workflow</i> no DSpace	62
FIGURA 4 - Funcionamento do Dspace	63
FIGURA 5 - Extraindo o arquivo “apache-maven-2.1.0-bin” para o disco local	73
FIGURA 6 - Alterando as variáveis de ambiente	74
FIGURA 7 - Criando as novas variáveis do sistema	74
FIGURA 8 - Abrindo o prompt de comando do Windows (cmd)	75
FIGURA 9 - Resultados esperados na checagem das variáveis de ambiente	75
FIGURA 10 - Abrindo o pgAdmin III do PostgreSQL	76
FIGURA 11 - Conectando ao banco de dados PostgreSQL	77
FIGURA 12 - Criando um novo “ <i>Login Role</i> ” no PostgreSQL	77
FIGURA 13 - Configurando o novo “ <i>Login Role</i> ” chamado dspace	77
FIGURA 14 - Criando um novo Banco de Dados	78
FIGURA 15 - Configurando o novo Banco de Dados chamado dspace	78
FIGURA 16 - Conferindo o novo <i>login role</i> e o novo banco de dados criados	79
FIGURA 17 - Ícone na área de notificação e opção de parar o serviço do servidor <i>web</i> Tomcat	79
FIGURA 18 - Pasta “dspace-1.5.2-src-release” extraída para disco local e pasta “dspace” criada pelo usuário no disco local (C:)	80
FIGURA 19 - Criando do pacote de instalação do DSpace	81
FIGURA 20 - Criação do pacote de instalação concluída com sucesso	81
FIGURA 21 - Realizando a instalação do DSpace	81
FIGURA 22 - Instalação do DSpace concluída com sucesso	81
FIGURA 23 - Criando a conta do administrador do repositório	82
FIGURA 24 - Copiando as pastas jspui e xmlui do Dspace para o Apache Tomcat	83
FIGURA 25 - Acessando http://localhost:8080/jspui	83
FIGURA 26 - Acessando http://localhost:8080/xmlui	84
FIGURA 27 - Divisão de uma página do DSpace	85
FIGURA 28 - Novo <i>banner</i> e o link “Sobre o software DSpace	85
FIGURA 29 - Linhas a serem eliminadas do arquivo “header-default.jsp” para que o link “Sobre o software DSpace” deixar de ser mostrado	85

FIGURA 30 - Alterando a cor de fundo dos quadros centrais da Page Content	86
FIGURA 31 - Resultado após as customizações realizadas no “styles.css”	87
FIGURA 32 - Link na <i>Navigation Bar</i> para acessar a interface de administrador e a interface de administrador	88
FIGURA 33 - Estrutura das Comunidades e Coleções do repositório	89
FIGURA 34 - Elementos do “input-forms.xml” traduzidos para o português	90
FIGURA 35 - Atualizando os arquivos do diretório “C:\dspace”	90

LISTA DE SIGLAS

AACR2: Anglo American Cataloguing Rules Edition 2
AIFF: Audio Interchange File Format
API: Application Programming Interface
ASCII: American Standard Code for Information Interchange
BSD: Berkeley Software Distribution
CDSWare: CERN Document Server Software
CERN: European Organization for Nuclear Research
CQL: Common Query Language
DCMI: Dublin Core Metadata Initiative
DOI: Digital Object Identifier
DP: Data Provider
DSpace: Digital Space – Institucional Digital Repository System
DTD: Data Type Document
EXIF: Exchangeable Image File
GDBM: GNU Database Manager
GPL: General Public Licence
GP-NTI: Grupo de Pesquisa - Novas Tecnologias em Informação
HP: Hewlett-Packard
HTML: Hypertext Markup Language
IBICT: Instituto Brasileiro de Informação em Ciência e Tecnologia
ID3: IDentify an MP3
IPR: Intellectual Property Rights
ISBN: International Standard Book Number
ISO: International Standard Organization
JPEG: Joint Photographic Experts Group
MARC: Machine Readable Cataloging
METS: Metadata Encoding and Transmission
MIT: Massachusetts Institute of Technology
MP3: MPEG - 1/2 Audio Layer 3
NGO: Non-governmental Organization
NISO: National Information Standards Organization

OAI: Open Archives Initiative

OAI-PMH: Open Archives Initiative – Protocol for Metadata Harvesting

OAMS: Open Archives Metadata Set

PDF: Portable Document Format

RI: Repositório Institucional

RTF: Rich Text Format

SGML: Standard Generalized Markup Language

SP: Service Provider

SRW: Search Retrieval Webservice

TIFF: Tagged Image File Format

UNESCO: United Nations Educational, Scientific and Cultural Organization

URI: Uniform Resource Identifier

URL: Uniform Resource Locator

W3C: World Wide Web Consortium

WWW: World Wide Web

XML: Extensible Markup Language

LISTA DE TABELAS

TABELA 1 - Exemplo de descrição de um recurso usando o padrão Dublin Core	41
TABELA 2 - Qualificadores do padrão Dublin Core.....	43
TABELA 3 - Atendimento dos requisitos obrigatórios para o repositório digital proposto nos aplicativos analisados	52
TABELA 4 - Atendimento dos requisitos desejáveis para o repositório digital proposto nos aplicativos analisados	53

SUMÁRIO

INTRODUÇÃO.....	14
CAPÍTULO 1 – REPOSITÓRIOS DIGITAIS	16
1.1 Funções Principais.....	18
1.2 Público Alvo e Benefícios	18
CAPÍTULO 2 – METADADOS	20
2.1 Definições.....	21
2.2 Características.....	22
2.3 Funções.....	24
2.4 Padrões de Metadados	24
2.4.1 Padrão MARC (<i>Machine Readable Cataloging</i>).....	27
2.4.2 Padrão METS (<i>Metadata Encoding and Transmission</i>)	29
2.4.3 Padrão Dublin Core	30
2.4.3.1 Especificações do Padrão Dublin Core	34
CAPÍTULO 3 – INTEROPERABILIDADE.....	44
3.1 Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)	46
3.2 Protocolo Z39.50	49
CAPÍTULO 4 – ANÁLISE DAS FERRAMENTAS DE CONTRUÇÃO DE REPOSITÓRIOS DIGITAIS	51
4.1 Comparação de Aplicativos.....	52
4.2 Análise Detalhada da Ferramenta DSpace	54
4.2.1 Metadados e Interoperabilidade no DSpace	58
4.2.2 Estrutura e Funcionamento do DSpace	58
4.2.3 Processo de Depósito e Workflow no DSpace	61
4.2.4 Buscas e Navegação no DSpace	63
4.3 Análise Detalhada da Ferramenta Greenstone.....	65
4.4 Análise Detalhada da Ferramenta CDSWare	66
4.5 Análise Detalhada da Ferramenta Nou-rau	69
4.6 Escolha do Sistema a ser usado para o Desenvolvimento do Repositório Digital Proposto	71
CAPÍTULO 5 – IMPLANTAÇÃO E CUSTOMIZAÇÃO DO REPOSITÓRIO DIGITAL ...	73
5.1 Instalação dos Componentes e do DSpace	73
5.2 Customização do Repositório Digital.....	84
5.3 Customização do Repositório Digital a partir da interface de administrador.....	88
5.4 Tradução dos componentes do DSpace	89
CONCLUSÃO.....	92
REFERÊNCIAS	94

INTRODUÇÃO

Com o uso intenso de meios computacionais para criação de produções científicas e documentos de uso geral, muita informação tem sido gerada, mas grande parte deixa de ser disseminada para aquelas pessoas e instituições as quais ela interessa e também documentos acabam sendo perdidos e precisam ser refeitos, sendo que poderiam ser facilmente reutilizados, provendo assim economia de tempo e pessoal nas organizações.

Atualmente muitas instituições de ensino e de pesquisa utilizam sistemas computacionais para gerenciar suas produções científicas, as quais são criadas nos mais diversos tipos de arquivos, como arquivos de texto (artigos, dissertações, relatórios, livros, teses, entre outros), planilhas eletrônicas, softwares, imagens, códigos fonte, arquivos de áudio, arquivos de vídeo, páginas *web*, entre outros, mas tais produções muitas vezes não são geridas adequadamente, o que leva a não serem preservadas no tempo e não serem disponibilizadas aqueles que possuem interesse em acessar tais produções, o que diminui as possibilidades de se ter maior proveito e consequentemente maior visibilidade desses documentos digitais.

Tais informações necessitam serem armazenadas para que não sejam perdidas no tempo, para que possam ser acessadas pelos interessados e para que sejam preservadas a fim de se permitir o reuso dos documentos criados pelas instituições. Desta maneira, faz-se necessário o uso de sistemas capazes de suprir estas necessidades, e para tais necessidades são apresentados os repositórios digitais, sistemas capazes de armazenar, gerir, disseminar e preservar as produções de uma instituição qualquer (LEWIS; YATES, 2008).

O Grupo de Pesquisa - Novas Tecnologias em Informação (GP-NTI) do Programa de Pós-Graduação em Ciências da Informação, da UNESP de Marília possui tais necessidades, ou seja, necessita de uma forma eficiente e segura de preservar e disseminar suas produções científicas para todos que se interessarem por elas, através de sistemas que sejam acessíveis a partir da internet e que sejam gratuitos.

Este trabalho tem como objetivo fazer uso de sistemas computacionais para que as produções científicas e os documentos gerados pelas instituições e organizações sejam disponibilizados ao público interessado e que este público tenha facilidade em acessá-los. Para isso pretende-se demonstrar o desenvolvimento e a customização de um repositório digital utilizando softwares livres para atender as necessidades de armazenamento e disseminação das produções científicas geradas pelo GP-NTI.

Além de permitirem que os conhecimentos gerados em instituições de ensino e pesquisa sejam de fato capazes de se tornarem conhecimentos explícitos para fins de pesquisas e de criação de novos conhecimentos, os repositórios digitais também encontram utilidade em empresas de modo geral, pois possuem a capacidade de transformar conhecimento tácito em conhecimento explícito, ou seja, empresas que utilizem repositórios digitais para armazenar, preservar e disponibilizar seus documentos digitais, deixarão de ter seus documentos não amplamente divulgados e pouco utilizados pelos interessados e por aqueles os quais tais documentos muito ajudariam, portanto os documentos digitais criados por seus membros, mas que não são socializados, poderão passar a serem socializados, ou seja, tornaram-se conhecimentos explícitos, pois irão permitir que outros membros da instituição além dos criadores dos documentos os usem e com isso aumentem o conhecimento geral da organização e consecutivamente melhorias em seus processos e maior ganho de tempo e agilidade nos cumprimentos de tarefas poderiam ser alcançados.

A pesquisa se apresenta como sendo exploratória, descritiva e explicativa, com base em revisão bibliográfica e desenvolvimento de um repositório digital utilizando software livre. A análise será feita de modo qualitativo e a redação se dará de modo interpretativo.

A revisão bibliográfica contará com estudo da teoria dos repositórios digitais, dos metadados e de interoperabilidade. Também irão ser estudados os softwares a serem utilizados no processo de desenvolvimento e customização do repositório digital.

CAPÍTULO 1 – REPOSITÓRIOS DIGITAIS

Com a evolução da tecnologia, o suporte às informações modificou-se muito e as produções científicas passaram a ser desenvolvidas em sua grande maioria através do uso de meios computacionais, surgiu então a necessidade de se desenvolver sistemas capazes de armazenar tais produções, estando elas em qualquer dos formatos existentes, tais como arquivos de texto, vídeos, imagens, sons, entre tantos outros, e também preservá-los no tempo e prover acesso a todos os interessados.

O grande aumento da capacidade de armazenamento dos computadores e a evolução da Internet e das redes, tornou possível pensar-se em repositórios digitais capazes de permitir o acesso ao conteúdo integral das obras em formato digital. Barton (2005, p.10) define repositórios institucionais como sendo “um banco de dados com um conjunto de serviços para capturar, armazenar, indexar, preservar e distribuir a pesquisa de uma instituição de ensino em formatos digitais”.

Repositórios digitais são conjuntos de obras em vários formatos diferentes, mas digitais, disponíveis para serem acessadas através de meios computacionais, podendo serem acessados a partir de redes de computadores locais ou pela internet. O conteúdo dos repositórios digitais não possui limitações, podendo existir repositórios institucionais, que agrupam projetos de pesquisas, teses de pós-graduação, trabalhos diversos desenvolvidos pelos membros da instituição, monografias, entre outros, e também repositórios para qualquer tipo de arquivos em formato digital que se deseje.

Tratando-se de repositórios institucionais de universidades Lynch (2003, p.02) os define como sendo um conjunto de serviços que uma instituição de ensino oferece aos membros de sua comunidade, objetivando o gerenciamento e disseminação eficiente dos materiais digitais gerados pela instituição e pelos membros de sua comunidade.

Também segundo Lynch (2003, p.02) é um compromisso da instituição a preservação a longo prazo dos materiais digitais armazenados em seus repositórios digitais, quando tal preservação for apropriada, bem como também sua organização adequada e o acesso ou distribuição deste materiais.

Os repositórios digitais institucionais, segundo definição do glossário elaborado pelo IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia) (REPOSITÓRIO..., 2007),

São sistemas de informação que armazenam, preservam, divulgam e dão acesso à produção intelectual de comunidades universitárias. Ao fazê-lo,

intervêm em duas questões estratégicas: contribuem para o aumento da visibilidade e o “valor” público das instituições, servindo como indicador tangível da sua qualidade; permitem a reforma do sistema de comunicação científica, expandindo o acesso aos resultados da investigação e reassumindo o controle acadêmico sobre a publicação científica.

Repositórios institucionais são conjuntos de serviços oferecidos pelas instituições para gerenciamento e disseminação dos materiais digitais criados pela instituição e pelos membros da sua comunidade, provendo acesso aos materiais digitais tanto para os membros internos da instituição quanto para os externos, através da internet, que também possuam interesse em tais materiais.

Rodrigues et al. (2006), Lynch (2003) e Crow (2002) afirmam que os repositórios digitais institucionais permitem que grande parte do conhecimento de uma instituição seja reunido, preservado e disseminado de forma eficiente, contribuindo desta maneira para o aumento da visibilidade das produções da instituição.

Afirmam que dentro dos últimos anos o número de repositórios digitais disponíveis tem crescido consideravelmente, especialmente na comunidade universitária, com objetivos focados no gerenciamento, na preservação, no armazenamento e também na divulgação das produções intelectuais desta comunidade. Instituições de ensino e pesquisa possuem interesse que seus membros disponibilizem suas produções científicas em tais repositórios digitais institucionais, uma vez que ao disponibilizarem tais produções, estão contribuindo diretamente para que as instituições tenham a possibilidade de ganhar mais prestígio e também credibilidade dentro da comunidade científica, além de contribuir efetivamente com o progresso científico.

Para Leite e Costa (2005, p.8):

Os repositórios institucionais (RI) possuem uma diversidade de tipologia de conteúdos e formatos que podem ser armazenados nos RI, tais como: artigos científicos, livros eletrônicos, capítulos de livros, *preprints*, *postprints*, relatórios técnicos, textos para discussão, teses, dissertações, trabalhos apresentados em conferências, palestras, material de ensino (slides, transparências, texto resumo, resenhas, trabalhos apresentados, entre outros), arquivos multimídia etc.

A necessidade de se preservar digitalmente documentos, que antes se encontravam apenas em papel, e permitir que os mesmos estejam acessíveis a qualquer pessoa, tem levado determinadas organizações, como universidades e bibliotecas, entre outras, a criar os seus próprios repositórios (SANTOS; TEIXEIRA; PINTO, 2005).

Segundo Barton (2005, p.16) um sistema de repositório institucional não é somente definido pelo *software* e pelo banco de dados que contém as coleções digitais. Um repositório institucional é um conjunto de serviços para aqueles que depositam material para as

comunidades de pesquisa e escolares envolvidas e para os usuários finais. Portanto um repositório institucional é constituído pelo *software*, pessoal responsável pelo gerenciamento do repositório e pelo seu conteúdo.

1.1 – Funções Principais

As principais funções de um repositório digital segundo identificação do autor são:

- Armazenamento de produções e documentos criados pelas instituições e seus membros;
- Preservação a longo prazo das produções;
- Publicação eletrônica eficiente das produções;
- Gerenciamento do conhecimento existente na instituição;
- Aumento da visibilidade das produções criadas pela instituição e seus membros;
- Prover a interoperabilidade entre repositórios através do uso de padrões de metadados e do protocolo OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*);
- Prover acesso livre a documentos e/ou arquivos de produções científicas;
- Armazenamento de documentos gerados dentro das instituições para gerenciamento próprio, como atas, relatórios, manuais, regulamentos, entre outros.

1.2 – Público Alvo e Benefícios

Os repositórios digitais possuem como público alvo qualquer instituição, sendo ela uma universidade, uma empresa de pesquisa ou qualquer outro tipo de instituição que deseje armazenar, gerenciar e preservar suas produções de maneira eficiente.

Segundo citado por diversos autores, a maioria dos repositórios digitais existentes são criados para instituições de ensino, mas também existem repositórios criados para diversos outros fins, como por exemplo, a Biblioteca Digital Jurídica (BDJur) do Superior

Tribunal de Justiça¹, a Biblioteca Digital do Senado Federal ²e a Biblioteca Digital sobre Corrupção³.

Entre os mais importantes motivos e benefícios que levam uma instituição a criar seu próprio repositório digital, segundo Noerr (2003) podem estar:

- Promoção a instituição perante seus concorrentes e potenciais clientes;
- Possibilidade de geração de renda, cobrando-se pelo acesso ao conteúdo do repositório;
- Participação em uma comunidade que produz conhecimento e o compartilha;

Outros benefícios identificados no uso de repositórios digitais pelas instituições são:

- Prover o acesso fácil e eficiente a toda à produção da instituição;
- Garantir a preservação digital dos arquivos presentes no repositório;
- Aumentar a visibilidade das produções da instituição, devido ao uso de padrões de metadados usados internacionalmente, o que facilita que estas produções sejam localizadas nos mais diversos mecanismos de busca existentes.
- Aumentar o interesse dos membros da instituição em criar novas produções, uma vez que saberão que as mesmas terão maior visibilidade tanto para os membros da instituição quanto para os interessados de fora da instituição.
- Aumentar o prestígio da instituição, uma vez que todo seu material de pesquisa estará sendo disponibilizado a todos interessados, fazendo com que esses saibam o potencial e o foco das pesquisas da instituição;
- Compartilhar coleções com outros repositórios institucionais através do uso de padrões de interoperabilidade, aumentando ainda mais a visibilidade da instituição e aumentando também o material disponível a ser usado pelos membros da instituição;
- Implantação de ambientes de gestão de conhecimento, persistindo e socializando os documentos digitais criados nas organizações, num ambiente centralizado.

Os repositórios digitais utilizam-se de metadados para permitir a busca e recuperação de seus objetos digitais e possibilitar a interoperabilidade com outras bases de dados.

¹ Ver <http://bdjur.stj.gov.br/xmlui/>

² Ver <http://www2.senado.gov.br/bdsf/>

³ Ver <http://bvc.cgu.gov.br/bvc.jsp>

CAPÍTULO 2 - METADADOS

A internet é o meio de maior utilização dos computadores e dos meios eletrônicos de armazenamento, e da busca e recuperação de informações armazenadas em meio eletrônico e com o seu desenvolvimento acelerado, capacidade de armazenamento e o aumento da quantidade de recursos disponibilizados levam à necessidade de ferramentas capazes de gerenciar, organizar e armazenar os recursos disponíveis em meios eletrônicos. Por recursos entende-se obras em quaisquer formatos, seja eles digitais ou não.

Pode-se compreender como recurso toda informação armazenada em meios eletrônicos que possa ser recuperada e acessada através da internet, não dependendo do seu formato, que pode ser texto, imagem, vídeo, som, entre muitos outros existentes.

Os mecanismos de busca disponíveis na internet, como Google¹, o Yahoo!², o Bing³ entre vários outros, recuperam muita informação, mas normalmente não com toda a precisão e qualidade que o usuário, a qual as informações disponibilizadas se destinam, deseja, ou seja, retornam resultados com pouca precisão.

A discussão teórica dos conceitos de precisão da informação não é o objetivo maior deste projeto, mas podem-se citar que o conceito de precisão para avaliação de fontes eletrônicas na internet fica muito próximo do conceito de avaliação de fontes impressas, definidos pelos bibliotecários. Então precisão seria a relação entre a quantidade de documentos relevantes recuperados e a quantidade total de documentos recuperados.

Para que uma busca tenha qualidade e aproxime-se do resultado esperado pelo usuário, os recursos disponíveis na internet necessitariam de possuir dados suficientes para descreverem seu conteúdo de forma mais adequada, mas esta eficiência na descrição dos recursos disponíveis em meios eletrônicos é pouco encontrada.

Por exemplo, pode-se usar as ferramentas de busca para localizar documentos que contenham referências a determinado autor, entretanto não é possível buscar somente os documentos que foram criados por este autor. Isto nos mostra que o resultado da busca traz muitos documentos que não são relevantes para o usuário, levando o mesmo a gastar tempo extra a procura dos documentos que realmente tenham interesse para sua pesquisa.

¹ Ver <http://www.google.com.br>

² Ver <http://www.yahoo.com.br>

³ Ver <http://www.bing.com>

O uso de metadados para descrição destes documentos permite realizar pesquisas de documentos escritos exatamente pelo autor determinado, do tipo autor = “nome_do_autor_determinado”.

“Metadados são a chave para garantir que os recursos vão sobreviver e continuar a ser acessíveis no futuro” (NISO, 2004, p.01).

2.1 - Definições

Normalmente chamados de dados sobre dados ou informações sobre informações, o termo metadados pode ser definido como conjuntos de dados chamados de elementos, que possuem número variável de acordo com o padrão, e que descrevem o conteúdo de um recurso, permitindo aos usuários ou ferramentas de busca recuperar este recurso. Estes elementos descrevem informações diversas a respeito do recurso, tais como nome, formato, data de criação, formato de arquivo, entre diversas outras, que aumentam ainda mais a eficiência da descrição de um recurso e aumentam o número de campos possíveis para pesquisas.

O termo metadados possui diferentes conceitos e significados dependendo da área que o profissional que o define trabalha, mas o foco principal dos metadados é a descrição da informação visando sua recuperação.

Segundo Niso (2004, p.01), “metadados são informações estruturadas que descrevem, explicam localizam ou senão facilitam a recuperação, uso ou o gerenciamento de um recurso informacional”.

Alvarenga (2003) refere-se aos metadados como sendo dados que descrevem a essência, atributos e contexto de emergência de um recurso (documento, fonte, etc) e caracteriza suas relações, visando ao acesso e ao uso potencial.

Já Souza; Vendrusculo e Melo (2000) os definem da seguinte maneira: “catalogação do dado ou descrição do recurso”.

Gill (1998) os descreve como “dados que servem para descrever grupos de dados que poderíamos chamar de objetos informatizados” ou “descrições estruturadas de um objeto informatizado”.

Como pode-se observar existem diversas definições para os metadados, mas após extensa pesquisa, metadados poderiam ser definidos como um conjunto de elementos que

descrevem um dado recurso, visando manter uma uniformidade nas descrições, com o objetivo principal de possibilitar sua busca e recuperação através dos mecanismos existentes.

Pode-se concluir que um conjunto de elementos que possua informações para identificar o conteúdo e de descrição de um dado recurso, possibilitando sua recuperação em meios eletrônicos pode ser chamado de metadados.

2.2 - Características

Tratando-se dos repositórios digitais, os elementos que formam os metadados devem conter dois tipos distintos de informações:

- elementos descritivos, que são aqueles que se referem as características explícitas do recurso, como autor, título, formato de arquivo, data de criação, entre outros;
- e elementos temáticos, que possuem conteúdo intelectual, ou seja, expressam o conteúdo do recurso, como as palavras-chave e as referências cruzadas.

Estes elementos de preferência devem usar esquemas de codificação que sejam comuns, como por exemplo, o uso de vocabulários controlados, esquemas de classificação e formatos de descrições formais, o que permite que esses metadados sejam trocados entre diferentes sistemas.

O uso de metadados facilita a busca e recuperação em bases de dados, pois a busca ganha eficiência e relevância, uma vez que se torna possível definir quais campos a serem pesquisado e quais valores procurar dentro de cada campo, o que não seria possível numa busca *full-text*, que procura exatamente pelo texto indicado na busca, sem se definir a que se refere tal texto, como título, autor ou outro elemento.

Na comparação entre ferramentas de busca utilizadas na internet, que utilizam indexação automática, com a busca por elementos de metadados descritos por seres humanos, diz que o uso de elementos de metadados trás resultados normalmente pequenos, mas bastante significativos quanto a relevância e detalhamento (KNIGHT, 2004, p.1-2).

De acordo com Niso (2004) o uso de metadados traz entre seus benefícios e funcionalidades permitir que os recursos sejam encontrados a partir de elementos relevantes, agrupamento de recursos similares e como conseqüência separar recursos com características distintas, além de também fornecer informações sobre a localização dos recursos.

Além de elementos descritivos de um recurso, os metadados podem possuir elementos que visem representar o contexto no qual ele está inserido, como localização de recursos, o tipo do recurso e sua relação com outros recursos.

Metadados podem ser usados para representar vários recursos, dependendo do domínio em que se encontram. Por exemplo, em um provedor *web* podem ser usados para localizar páginas na internet, podem ser usados para descrever as informações contidas em imagens oriundas de digitalizações, também em documentos criados em meio eletrônico, a fim de descrever a informação contida nos mesmos.

Gilliland-Swetland (1998) define que metadados podem ser divididos em cinco tipos distintos, relacionados com suas funcionalidades dentro do sistema que são usados:

- Metadados administrativos: são usados na administração de recursos;
- Metadados descritivos: são usados para descrever as informações sobre recursos;
- Metadados de conservação: estão diretamente relacionados com a conservação de recursos;
- Metadados técnicos: são relacionados com comportamento dos metadados e também com o funcionamento de sistemas;
- Metadados de uso: são relacionados ao tipo e ao nível do uso dos recursos.

Gilliland-Swetland (1998) também define atributos chave para os metadados, que são: fonte, que define que metadados podem ser criados no momento da criação do recurso ou posteriormente; métodos de criação, que definem como os metadados podem ser criados, por mecanismos computadorizados ou manualmente; caráter, que define que os metadados podem ser criados tanto por especialista quanto pelo responsável pelo dado recurso; situação, que define que os metadados podem ser de característica estáticos, ou seja, que nunca se alteram ou dinâmicos, que podem sofrer alterações e de larga e curta duração; estrutura, que definem que metadados podem ser estruturados, como por exemplo Dublin Core ou MARC (*Machine Readable Cataloging*) ou não estruturados; semântica, que definem que metadados podem ou não utilizar de vocabulários controlados para descrição dos seus elementos; e nível que define que metadados podem ser relacionados a objetos únicos ou a coleções inteiras de objetos.

Cole (2002, p.63-64) relata que os princípios a seguir ajudam na criação de metadados com boa qualidade:

- Devem ser apropriados aos materiais armazenados, usuários do acervo e uso pretendido para o acervo;

- Devem dar suporte para o gerenciamento a longo prazo dos objetos digitais;
- Devem ser confiáveis e
- Devem dar suporte a interoperabilidade.

Outra característica muito importante a cerca dos metadados é que os mesmos podem ser representados dentro do próprio recurso ou separadamente, como em uma base de dados de metadados.

2.3 - Funções

Segundo Niso (2004) os metadados possuem as mesmas funções na busca de recursos do que os bons catálogos fazem, que são:

- Permitir que os recursos sejam recuperados através de critérios relevantes;
- Identificar recursos;
- Agrupar recursos similares;
- Distinguir recursos não similares; e
- Oferecer informações sobre localização de recursos.

2.4 - Padrões de Metadados

Padrão de metadados pode ser entendido como um conjunto de elementos que descrevem um recurso seguindo um determinado conjunto de conceitos e requisitos, visando a descrição de recursos contidos em domínios específicos (FUSCO, 2009).

Sendo que cada obra de tipos diferentes pode possuir um esquema mais adequado para sua descrição, foram sendo desenvolvidos vários padrões, cada um mais específico para determinada áreas do que para outras. Um exemplo destas diferenças de descrições seria o elemento que especifica o “número de páginas” de um recurso, elemento que seria muito útil na descrição de livros e artigos, mas sem sentido algum para uma imagem, um som ou um vídeo. Em contrapartida o elemento “título” é útil para praticamente todas as obras.

Esta profundidade na descrição desejada por um profissional irá influenciar diretamente a escolha de um padrão de metadados, pois quanto maior for o detalhamento da descrição desejada, mais elemento o padrão escolhido deve possuir.

Assim como Cole (2002) relata, a capacidade de interoperabilidade depende do padrão de metadados. Caso duas bibliotecas distintas possuem a mesma obra em seu acervo, é de fato desejável e também possível que apenas uma delas faça o processo de descrever a obra. Partindo do pressuposto que os sistemas de informática das duas bibliotecas usem o mesmo padrão de metadados, ou caso não usem o mesmo padrão, ofereçam suporte para que tradução entre padrões seja realizada, a descrição realizada por uma das bibliotecas poderá ser exportada para a base de dados da outra, facilitando o trabalho da mesma.

Caso em um ambiente digital, como um repositório digital, dentro dos elementos do padrão de metadados encontre-se a localização eletrônica da obra, apenas importar os metadados já seria suficiente, e não haveria a necessidade de se importar a obra em si. Desta maneira quem procura pela obra poderá receber um endereço, que é a localização eletrônica da obra que estará sendo mantida em uma base de dados distinta da local onde o usuário está realizando a busca pela obras, desta maneira ambientes digitais como os repositórios digitais podem possuir uma qualidade e quantidade de conteúdo maior, sem a necessidade de armazenar a obra. Dessa maneira poderiam existir repositórios sem nenhuma obra armazenada, mas apenas com registros relacionados à determinada área de interesse e que seus metadados direcionem para a obra, que estará armazenada em outro local.

Os elementos dos esquemas de metadados recebem seus nomes de acordo com o contexto que estão inseridos, como compositor para uma música e autor para livros, e também recebem os valores, que são o conteúdo de cada elemento do esquema de metadados, como por exemplo o nome propriamente dito do autor. Portanto os elementos dos metadados são preenchidos com valores. Existem regras que tratam da sintaxe dos elementos e também é interessante que se hajam regras para especificação dos valores dos elementos, para que se possa ajudar a garantir a interoperabilidade, tópico que será tratado mais a frente.

A necessidade de padronização no preenchimento dos valores dos elementos dos metadados pode ser observada no seguinte exemplo. Caso uma obra possua dois títulos em sua capa, qual deles deverá ser usado para preencher o elemento *título* da obra? Caso não haja nenhuma regra neste preenchimento, duas pessoas que sejam responsáveis por preencher os valores dos elementos de metadados podem atribuir diferentes títulos para esta obra. A mesma observação vale para o tratamento do *assunto* da obra, elemento que é ainda mais fácil de ocorrer interpretações diferentes por pessoas diferentes.

Metadados, quando representados dentro do próprio recurso, para tratamento da sintaxe dos elementos dos metadados normalmente estão inseridos em páginas *Web* desenvolvidas em SGML (*Standard Generalized Markup Language*), em HTML (*Hypertext*

Markup Language) ou em XML (*Extensible Markup Language*), que são linguagens de marcação.

Tais linguagens de marcação foram desenvolvidas com a necessidade de manter uma maior interação e troca de informações entre os usuários de computadores e os computadores, que cresce cada vez mais com o aumento do número de computadores, com as novas tecnologias de comunicação e com o surgimento da WWW (*World Wide Web*) (BAX, 2001).

O SGML foi criado em torno de 1970, sendo uma linguagem que possibilita definir qualquer conjunto de marcas, portanto é auto-descritiva. Um documento SGML carrega consigo sua própria especificação, o que se chama de DTD (*Data Type Document*). Então podemos dizer que a linguagem SGML é usada para definir outras linguagens (BAX, 2001).

A partir do SGML surgiu o HTML, linguagem que possui um conjunto de marcas pré-definidas, criado com objetivo de organizar as informações que serão disponibilizadas através da internet. O tornou-se um dos padrões mais utilizados na internet, devido a existência das marcas pré-definidas.

Com o rápido avanço da internet e das necessidades dos seus usuários, foi criado o XML, resultado das pesquisas dos especialistas do W3C (*World Wide Web Consortium*), que no ano de 1996 disponibilizaram uma linguagem de marcação mais simples que o SGML e mais flexível que o HTML, criando assim uma linguagem de marcação voltada para as reais necessidades da internet. O XML tem como um de seus objetivos descrever o que os dados significam, além de descrever de qual maneira eles devem ser mostrados.

O XML é apontado como uma linguagem ideal para armazenamento de informações, pois possui a facilidade de não ser necessário o uso de aplicações específicas de software para interpretar seus dados e quaisquer alterações na estrutura dos dados XML ou em seu conteúdo é fácil de realizar e não necessita que sejam feitas alterações nas aplicações que as tratam.

Essas linguagens de marcação e suas respectivas características viabilizam a existência de interoperabilidade entre diferentes sistemas e ambientes, permitindo a troca de informações estruturadas.

Tratando-se dos repositórios digitais, duas maneiras distintas de se armazenar metadados podem ser utilizadas: ligar os metadados ao documento que eles descrevem ou armazenar os metadados dentro do arquivo do documento. Páginas em linguagem HTML são ideais para armazenarem os metadados internamente o documento, sendo que as mesmas possuem a *tag* META, marca que tem a função de armazenar os metadados. Outros formatos de arquivos também possuem campos para metadados, como os arquivos de áudio no formato

MP3 (*MPEG - 1/2 Audio Layer 3*) que possuem as *tags* ID3 (*IDentify an MP3*), os arquivos de imagem no formato JPEG (*Joint Photographic Experts Group*) que possuem informações EXIF (*Exchangeable Image File*), entre muitos outros. Porém existem vários tipos de arquivos que não permitem o armazenamento interno de metadados, ou suas possibilidades para tais descrições são muito limitadas, nestes casos é necessário organizar os metadados externamente aos documentos. Para isso é necessário ligar os documentos a seus respectivos metadados, proporcionando maior flexibilidade, visto que podem ser exportados ou importados apenas os bancos de metadados, e não os documentos em si.

2.4.1 – Padrão MARC (*Machine Readable Cataloging*)

O formato MARC foi criado na década de 60 nos Estados Unidos pela *Library of Congress* (Biblioteca do Congresso Norte Americano), que é a instituição mais confiável quando se trata de catalogação nos Estados Unidos, portanto é uma instituição com prestígio e reconhecimento internacional nesta área. Considerou-se que computadores não são capazes de fazer a leitura diretamente das fichas catalográficas, criadas seguindo as normas do AACR2 (*Anglo American Cataloguing Rules Edition 2*). Portanto era necessário a criação de um formato legível por computadores (LIBRARY OF CONGRESS HELP DESK, 2009), permitindo assim a troca de registros catalográficos e bibliográficos e entre bibliotecas, com o uso de computadores.

Segundo Library of Congress Help Desk (2009) a estrutura do padrão de metadados MARC é descrita da seguinte forma: os registros bibliográficos são formados por diversos campos como título, autor, assunto, entre outros. A estrutura do padrão MARC é resumidamente a seguinte: cada campo do registro está associado a um número com três dígitos, que é chamado de *tag* MARC. Desta maneira cada *tag* identifica um campo, portanto o computador pode identificar a partir destas *tags* a que campo pertence o valor que segue.

A *tag* MARC formada pelos três dígitos é seguida por mais dois dígitos decimais que são usados como indicadores e são escritos juntos, mas não são dependentes, ou seja, não são um número com dois dígitos, mas sim dois dígitos seguidos (LIBRARY OF CONGRESS HELP DESK, 2009). Um exemplo seria a *tag* de título (*tag* 254) acompanhada de um indicador que tivesse a função de fazer com que o computador ignore-se determinada parte do título no momento de fazer a ordenação dos registros, assim poderia ser indicado para que

ignora-se a primeira letra do título “O Alquimista”, título no qual a parte interessante para ordenação seria apenas “Alquimista”.

Alguns campos podem não possuir nenhum indicador, ou até mesmo possuir apenas um indicador definido. Para identificar indicadores indefinidos utiliza-se o caracter especial “#”, que significa espaço em branco.

Os campos deste formato podem ser divididos em subcampos, onde serão armazenadas informações pertinentes ao conteúdo do campo a qual estes subcampos são relativos, como por exemplo, um título, subtítulo e título alternativo, que são situações aceitas no AACR2. Outro exemplo seria o campo que descreve a responsabilidade intelectual de uma obra, campo que pode possuir várias recorrências, como autor, editor, tradutor, revisor, entre outros.

Os subcampos são referenciados por um delimitador, formado por uma seqüência de caracteres, como por exemplo \$a ou \$b.

Utilizando os três conceitos vistos sobre o padrão MARC (subcampos, campos e indicadores), um registro do campo título, ou como usado no AACR2 área de título, poderia ser representado em formato MARC da seguinte maneira:

```
245 14 $a A ascensão do excluído
      $b a história verídica de João Pedro Alcântara.
```

Neste exemplo o número 245 é a *tag* MARC que representa a *área de título*, campo que pode conter título principal, título alternativo, subtítulo, entre outros. O número 1 é o indicador de que deve existir uma ficha no catálogo de títulos convencional para a obra que esta sendo descrita e o número 4 é um indicador que faz com que o computador ignore os dois primeiros caracteres do campo no momento da ordenação, neste caso o caractere A mais um espaço em branco. E por último os subcampos \$a que é o título principal e \$b que é o subtítulo da obra.

Mesmo não tendo sido desenvolvido especificamente para tratamento de objetos digitais, o formato MARC, mesmo sendo um padrão antigo, é muito robusto e flexível o bastante para ser adaptados, além de já ser estabelecido (BRANTON, 2004), porém por possuírem um conjunto de elementos complexos e bastante rígidos, necessitam de especialistas para a descrição de elementos, o que dificulta seu uso como um padrão para descrição de recursos disponíveis na internet, pois para este fim são necessários padrões que possam ter as descrições feitas pelos próprios autores do recurso ou por administradores de sites.

2.4.2 – Padrão METS (*Metadata Encoding and Transmission*)

Em bibliotecas convencionais o fato de não se manter os metadados estruturais a respeito de uma obra não é propriamente um problema, visto que caso os metadados sejam perdidos, o conjunto da obra não será perdido. Porém, tratando-se de obras digitais, caso as mesmas não possuam metadados estruturais, corre-se o risco de as mesmas se “desintegrarem”, ou seja, de não possuírem ligações entre as partes que a compõem. Outro tipo de problema causado pela falta de metadados estruturais em obras digitais seria o caso de uma obra digitalizada que não possui-se os metadados técnicos sobre o processo de digitalização, então pode ocorrer a perda de informações importantes a respeito da obra original, como por exemplo um livro que tenha sido produzido em uma prensa de Guttenberg e tenha sido digitalizado, a omissão desta informação nos elementos dos metadados da obra digital fará com que ela perca muito de seu valor para os pesquisadores (LIBRARY OF CONGRESS, 2006).

Essas necessidades levaram a *Library of Congress* (Biblioteca do Congresso Norte Americano) a desenvolver o padrão de metadados METS (*Metadata Encoding and Transmission* – Padrão de Codificação e Transmissão de Metadados) que possuem segundo Library of Congress (2006) elementos “administrativos e estruturais para trabalhos textuais e baseados em imagens”.

Também segundo Library of Congress (2006) “o METS disponibiliza um formato XML capaz de codificar os metadados necessários tanto para a gestão de objetos de bibliotecas digitais num repositório quanto para a troca destes objetos entre repositórios, ou ainda entre repositórios e seus usuários”.

Para Gartner (2002, p.03) o padrão METS é um padrão emergente, desenvolvido para codificar os metadados necessários para a realização de descrições completas de objetos digitais em bibliotecas digitais.

Um documento no padrão METS é uma aplicação XML, não proprietária e independente de software específico (NISO, 2004, p.04).

O METS descreve os recursos digitais de uma biblioteca digital e é formado por sete seções principais, descritas resumidamente a seguir segundo Library of Congress (2006) e Niso (2004, p.05):

1. **Cabeçalho:** contém metadados descrevendo o documento METS em si, incluindo informação como o autor, editor, título, etc;
2. **Metadados Descritivos:** a seção de metadados descritivos pode apontar para metadados descritivos externos ao documento METS, inclusive em formatos diferentes do METS, como por exemplo um registro MARC, ou pode conter os metadados descritivos internamente, ou ambos. Múltiplas instâncias de metadados descritivos, tanto internas como externas, podem ser incluídas na seção de metadados descritivos;
3. **Metadados Administrativos:** a seção de metadados administrativos oferece informação sobre como os arquivos foram criados e armazenados, direitos de propriedade intelectual (*copyright*), metadados sobre o objeto original a partir do qual o objeto digital foi criado, e informações sobre a proveniência dos arquivos que compõem o objeto digital. Tal como os metadados descritivos, os metadados administrativos podem ser tanto externos ao documento METS, ou codificados internamente;
4. **Seção de Arquivos:** lista todos os arquivos que formam a versão eletrônica do objeto digital;
5. **Mapa Estrutural:** apresenta uma estrutura hierárquica entre os componentes do objeto digital e faz a ligação entre estes componentes e os arquivos e metadados dos mesmos;
6. **Ligações Estruturais:** armazena a existência de *hiperlinks* entre os nós da estrutura hierárquica do Mapa Estrutural e
7. **Comportamento:** pode ser utilizada para associar comportamentos executáveis ao conteúdo do objeto METS. Possibilita que se registrem informações a respeito de como os componentes do objeto digital serão renderizados para o usuário, incluindo os aplicativos que deverão ser utilizados ou os parâmetros necessários para a renderização do arquivo.

2.4.3 – Padrão Dublin Core

O padrão de metadados Dublin Core tem sua raiz na 2ª Conferência Internacional sobre WWW onde membros de instituições que tratavam de informática e internet

conduziram uma discussão sobre semântica e *web*. A partir desta discussão foi organizado um evento com o propósito de se discutir como um conjunto semântico para recursos digitais baseados na *web* poderia ser extremamente útil para busca e recuperação de recursos na internet. Este evento, que contou com a participação de profissionais das mais diversas áreas, como ciência da computação, bibliotecários, profissionais da informação on-line, catalogadores, e diversos outros que trabalhavam com tratamento de dados e informações nos mais diversos ambientes, tinha como principal objetivo definir um conjunto mínimo de elementos para a *web*.

Segundo Desai (1997) pretendia-se tratar o problema da catalogação de recursos na internet, com a adoção, a extensão ou a modificação de padrões existentes e dos protocolos capazes de facilitar sua recuperação e acesso, atendendo assim as várias comunidades que utilizassem os metadados.

Para se chegar a um consenso sobre esse conjunto de elementos, foram definidas algumas suposições a serem seguidas durante as discussões, que segundo Desai (1997) foram as seguintes:

- chegar a um conjunto de elementos comuns;
- todos elementos deveriam descrever o recurso por si próprio, com exceção do elemento *fonte*;
- os elementos precisariam descrever um documento como um objeto;
- os elementos precisariam suportar a recuperação dos recursos da rede;
- nenhuma suposição deveria ser feita para dizer se o recurso descrito era acessível pela rede ou se era especificamente um recurso eletrônico;
- todos os elementos de metadados poderiam ser repetitivos;
- todos elementos deveriam ser opcionais;
- os elementos deveriam descrever as características intrínsecas ao recurso e;
- elementos que não fossem incluídos não seriam obrigatoriamente excluídos.

Vários dos participantes do evento admitiram a dificuldade de se trabalhar com padrões que fossem constituídos por muitos elementos e então procuraram determinar um conjunto de elementos não exaustivo para descrever os recursos, tornando-se um consenso entre todos os participantes que o conjunto de elementos não poderia ser extenso.

Ao conjunto de elementos criados deram o nome de Padrão de Metadados Dublin Core, pois o evento ocorreu na cidade de Dublin, Ohio, nos Estados Unidos.

Além do conjunto de elementos criado no evento, também foi constituída a *Dublin Core Metadata Initiative* (DCMI) que é Iniciativa dos Metadados Dublin Core, uma organização que possui os objetivos de promover a adoção de padrões de metadados e o desenvolver vocabulários especializados de metadados com a função de descrever recursos que facilitem a recuperação dos mesmos na internet.

O DCMI é aberto para a participação de qualquer pessoa ou instituição interessada em participar da busca contínua de um consenso mínimo no desenvolvimento de vocabulários de metadados.

Segundo o **DCMI (2009A)** as atividades principais do DCMI são:

- Desenvolvimento de padrões de metadados para a recuperação de informação nos domínios da internet;
- Definição de estruturas para a interoperabilidade de conjuntos de metadados;
- Organização de eventos internacionais e grupos de trabalho para o desenvolvimento e manutenção das recomendações do DCMI;
- Disponibilização de ferramentas, serviços e infraestrutura em metadados e;
- Alcançar outras comunidades de desenvolvimento de metadados, desenvolvendo a interação entre o DCMI e elas.

O DCMI é estruturado para realizar todas estas atividades é composto segundo DCMI (2009B) da seguinte maneira:

- um Conselho de Administradores, que possui a função de aconselhar os membros da Diretoria nas questões estratégicas e de utilização de recursos financeiros, possuindo membros de vários países;
- uma Diretoria, que é formada por um Diretor Executivo e um Diretor Administrativo, que supervisionam o gerenciamento e a coordenação das atividades dos Grupos de Trabalho;
- um Conselho Consultivo, que é formado pelos presidentes dos Grupos de Trabalho, dando consultoria a Diretoria e;
- vários Grupos de Trabalho, que são criados de acordos com as necessidades percebidas e compostos por voluntários, que trabalham na resolução de problemas específicos.

O DCMI organiza várias listas de discussão e mantém informações a seus respeito em sua página *web*¹.

Segundo Hansen (1999) o formato Dublin Core foi criado com o objetivo de descrever um recurso eletrônico, tornar simples a sua descrição, possibilitar a indexação, tornar mais fácil a pesquisa de recursos e o acesso a estes recursos e garantir a qualidade do gerenciamento dos recursos.

Hansen (1999) cita que a principais razões para se adotar o formato Dublin Core são:

- facilidade de criação;
- simplicidade de indexação;
- indexação com mais precisão do que as buscas *full-text* e;
- capacidade de interoperabilidade.

O padrão de metadados Dublin Core é um conjunto formado por quinze elementos, que possuem como objetivo descrever um recurso eletrônico. Niso (2004, p.03) diz que a simplicidade do Dublin Core vem do fato de seus quinze elementos serem opcionais, passíveis de repetição e podem ser codificados em XML.

O padrão de metadados Dublin Core possui as seguintes características segundo DCMI (2008B) e DCMI (2005):

- **Simplicidade:** como a maioria dos seus elementos possui um entendimento semântico bem simples, o padrão Dublin Core pode ser criado facilmente pelo responsável pelo documento, sem a necessidade de treinamentos e especializações para tal objetivo;
- **Interoperabilidade Semântica:** modelos com diferentes descrições interferem na habilidade das busca entre áreas distintas, no entanto com a existência de um modelo comum de descrição ocorre o aumento da possibilidade de interoperabilidade entre estas áreas distintas;
- **Consenso Internacional:** como ocorre a participação de mais de vinte países no DCMI para a busca de um escopo internacional na internet e de uma infraestrutura adequada ocorre uma contribuição para que haja um consenso internacional;
- **Extensibilidade:** o padrão Dublin Core é um modelo de descrição simplificado, mas que possui extensibilidade e flexibilidade na elaboração de

¹ ver <http://www.dublincore.org>

modelos, ou seja, permite que elementos sejam acrescentados para atender necessidades específicas de descrição de recursos. Os elementos padrões do Dublin Core mais os novos elementos permitem que diversas comunidades de áreas distintas utilizem o padrão, troquem informações entre si e tenham acesso a elas;

- **Flexibilidade:** os elementos opcionais podem ser repetidos dependendo das necessidades da descrição e modificados fazendo-se uso de qualificadores, que serão mostrados neste capítulo.

2.4.3.1 – Especificações do Padrão Dublin Core

Para compreender-se o funcionamento do padrão Dublin Core, será feita a análise da versão 1.1, que é composta por quinze elementos.

Cada elemento do conjunto de elementos do padrão Dublin Core é definido usando um conjunto de dez atributos padrões para a descrição dos elementos de dados (DCMI, 2008B).

Os dez atributos padrões são os seguintes:

- **Nome:** o nome atribuído ao elemento;
- **Identificador:** o identificador único dado ao elemento;
- **Versão:** a versão atual do elemento;
- **Registro de Autoridade:** a entidade autorizada a registrar o elemento;
- **Língua:** o idioma no qual o elemento está especificado;
- **Definição:** a declaração que representa de maneira clara o conceito e a natureza essencial do elemento;
- **Obrigação:** indica se o elemento é obrigatório ou não;
- **Tipo de dado:** indica o tipo de dado que se pode representar no conteúdo do elemento;
- **Máxima Ocorrência:** indica o limite para a repetição dos elementos;
- **Comentário:** uma observação sobre a aplicação dos elementos.

Desses dez atributos, seis possuem valor único e comum para todos os elementos do Dublin Core. Segundo DCMI (2008B) são os seguintes:

- **Versão:** 1.1;
- **Registro de Autoridade:** Dublin Core Metadata Initiative (DCMI);

- **Língua:** Inglês;
- **Obrigaç o:** Opcional;
- **Tipo de dado:** *String* de caracteres;
- **M xima Ocorr ncia:** ilimitada.

O atributo Identificador permanece na l ngua original, a inglesa, pois   neste formato que o mesmo deve ser mostrado na descri o de um recurso.

A descri o detalhada de cada um dos quinze elementos padr es do Dublin Core segundo DCMI (2008B) segue abaixo. Em cada um ser o analisados os atributos que n o s o comuns para todos, que s o os atributos Nome, Identificador, Defini o e Coment rio. Para fins de exemplo ser o usados o site do Dublin Core e um artigo retirado da internet:

Elemento 1 – T tulo

Nome: T tulo

Identificador: *Title*

Defini o: um nome dado a um recurso.

Coment rio: ser  o nome pelo qual o recurso   formalmente conhecido.

Exemplos:

- Site do Dublin Core: Dublin Core Metadata Initiative
- Artigo: Understandig Metadata

Elemento 2 – Autor

Nome: Autor ou Criador

Identificador: *Creator*

Defini o: a entidade respons vel pela cria o do conte do do recurso;

Coment rio: exemplos de autores incluem pessoas, organiza es ou servi os.

Exemplos:

- Site do Dublin Core: Dublin Core Metadata Initiative
- Artigo: NISO – National Information Standards Organization

Elemento 3 – Assunto

Nome: Assunto ou Palavras-chave

Identificador: *Subject*

Defini o: o tema, objeto ou ponto principal do conte do do recurso, ou seja, do que se trata o recurso.

Comentário: normalmente o assunto será expresso por palavras-chaves, frases-chav ou códigos de classificação que descrevem o tema do recurso. A prática mais aconselhável é escolher os valores de vocabulários controlados ou de esquemas de classificação formais.

Exemplos:

- Site do Dublin Core: Dublin Core, Metadados, Informação, Interoperabilidade.
- Artigo: Metadados, Padrões de Metadados, Interoperabilidade.

Elemento 4 – Descrição

Nome: Descrição

Identificador: *Description*

Definição: uma descrição do conteúdo do recurso.

Comentário: descrições podem incluir, mas não são limitadas a, resumos, tabela de conteúdos, referências para representações gráficas do conteúdo ou textos livres de descrição do conteúdo.

Exemplos:

- Site do Dublin Core: “The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices. “
- Artigo: “*Understanding Metadata* is a revision and expansion of *Metadata Made Simpler: a guide for libraries* published by NISO Press in 2001. *NISO extends its thanks and appreciation to* Rebecca Guenther *and* Jacqueline Radebaugh, staff members in the Library of Congress Network Development and MARC Standards Office, *for sharing their expertise and contributing to this publication.*“

Elemento 5 – Editor

Nome: Editor

Identificador: *Publisher*

Definição: é a entidade responsável por tornar o recurso disponível.

Comentário: Exemplos de editores incluem pessoas, organizações ou serviços.

Exemplos:

- Site do Dublin Core: DCMI (Dublin Core Metadata Initiative)
- Artigo: NISO - National Information Standards Organization

Elemento 6 – Colaborador

Nome: Colaborador

Identificador: *Colaborator*

Definição: entidade responsável por fazer contribuições para o conteúdo do recurso;

Comentário: exemplos de colaboradores incluem pessoas, organizações ou serviços.

Exemplos:

- Site do Dublin Core: W3C
- Artigo: Association of Information and Dissemination Centers

Elemento 7 – Data

Nome: Data

Identificador: *Date*

Definição: uma data associada a um determinado evento no ciclo de vida do recurso.

Comentário: normalmente a data será associada a criação ou a disponibilização do recurso. Recomenda-se o uso do formato YYYY-MM-DD, onde YYYY é o ano, MM é o mês e DD o dia para representar datas no Dublin Core.

Exemplos:

- Site do Dublin Core: 1995
- Artigo: 2004

Elemento 8 – Tipo

Nome: Tipo do recurso

Identificador: *Type*

Definição: a natureza ou espécie do conteúdo do recurso.

Comentário: Tipo inclui termos descrevendo categorias gerais, funções, espécies ou níveis de agregação para o conteúdo. Recomenda-se o uso de valores retirados de vocabulários controlados. Para descrição do formato físico ou digital do recurso deve-se usar o elemento Formato.

Abaixo segue a lista sugerida pelo DCMI (2008C) para identificar o elemento tipo do recurso:

1. “**Collection**”: usada para definir uma agregação de itens;
2. “**Dataset**”: usado para definir dados codificados em estruturas definidas, Alguns exemplos são: listas, banco de dados e tabela;
3. “**Event**”: usado para definir a ocorrência baseada em tempo. Alguns exemplos são: Conferencias e Workshops;
4. “**Image**”: usado para definir recursos que são representações visuais. Alguns exemplos são: fotografias, desenhos, filmes, vídeos, mapas, entre diversos outros;
5. “**Interactive Resource**”: usado para definir recursos que requerem interação com o usuário. Um dos exemplos de recursos interativos são as páginas *web*;
6. “**Service**”: usado para definir sistemas que agregam valores para o usuário final, como por exemplo os serviços de autenticação, serviços de banco, e diversos outros;
7. “**Software**”: usado para definir um programa de computador compilado ou um arquivo fonte;
8. “**Sound**”: usado para definir recursos cujo conteúdo é composto de áudio. Alguns exemplos são os sons, músicas, CDs de áudio, gravações de voz, etc;
9. “**Text**”: usado para definir recursos compostos de caracteres para leitura, como livros, artigos e diversos outros;

Exemplos:

- Site do Dublin Core: *Interactive Resource*
- Artigo: *Text*

Elemento 9 – Formato

Nome: Formato

Identificador: *Format*

Definição: a formato digital ou físico que o recurso se encontra.

Comentário: normalmente o formato do recurso inclui o tipo de mídia ou dimensões do recurso. O formato pode ser usado para definir o software, hardware ou outro equipamento necessário para mostrar ou operar o recurso. Exemplos de dimensões incluem tamanhos e durações. Recomenda-se o uso de valores retirados de vocabulários controlados.

Exemplos:

- Site do Dublin Core: *text/html*
- Artigo: *text/pdf*

Elemento 10 – Identificador

Nome: Identificador de recurso

Identificador: *Identifier*

Definição: uma referência não ambígua para um recurso dentro de um dado contexto.

Comentário: recomenda-se a identificação do recurso pelo significado de uma *String* ou número conforme um sistema de identificação formal. Exemplos de sistemas de identificação formal incluem o Identificador de Recursos Uniforme (*Uniform Resource Identifier* – URI), o Localizador de Recursos Uniforme (*Uniform Resource Locator* – URL), o Identificador de Objetos Digitais (*Digital Object Identifier* – DOI) e o Número Internacional Normalizado para Livros (*International Standard Book Number* – ISBN).

Exemplos:

- Site do Dublin Core: <http://dublincore.org>
- Artigo: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

Elemento 11 – Fonte

Nome: Fonte

Identificador: *Source*

Definição: uma referência para um recurso do qual o presente recurso é derivado.

Comentário: o presente recurso pode ser derivado da fonte do recurso inteira ou apenas em parte.

Exemplos:

- Artigo: A Framework of Guidance for Building Good Digital Collections, disponível em <http://www.niso.org/framework/forumframework.html>

Elemento 12 – Língua

Nome: Língua

Identificador: *Language*

Definição: a língua do conteúdo intelectual do recurso.

Comentário: Recomenda-se o uso de um código com quatro caracteres dispostos da seguinte maneira LL-CC, onde os dois primeiros (LL) indicam a língua e os dois últimos (CC), que são opcionais, indicam o país.

Exemplos:

- Site do Dublin Core: en-us (Inglês- Estados Unidos)

- Artigo: en-us

Elemento 13 – Relação

Nome: Relação

Identificador: *Relation*

Definição: uma referência para um recurso relacionado, como tradução de um ou parte de um trabalho.

Comentário: recomenda-se o uso de *String* ou número conforme um sistema de identificação formal.

Exemplos:

- baseado na obra de Jorge Amado;
- tradução da canção de Jim Morrison.

Elemento 14 – Cobertura

Nome: Cobertura

Identificador: *Coverage*

Definição: o âmbito ou o escopo do conteúdo do recurso.

Comentário: Cobertura normalmente irá incluir localizações espaciais, como nome de locais ou suas coordenadas geográficas, também período temporais, como um rótulo de tempo, uma data ou um período de tempo ou também jurisdições, como o nome de uma entidade administrativa.

Exemplos:

- Site do Dublin Core:1995-2009
- Artigo: Estados Unidos

Elemento 15 – Direitos

Nome: Gerenciamento de Direitos

Identificador: *Rights*

Definição: informações a respeito dos direitos do recurso.

Comentário: normalmente um elemento de Direitos irá conter uma declaração dos direito do recurso ou referências para serviços que providenciam tais informações. Informações de direitos normalmente tratam de Direito de Propriedade Intelectual

(*Intellectual Property Rights* – IPR), *Copyright* e vários outros direitos de propriedade. Se o elemento *Rights* é vazio, nenhuma suposição sobre a situação deste recurso pode ser feita.

Exemplos:

- Site do Dublin Core: Copyright © 1995-2009 DCMI All Rights Reserved.
- Artigo: Copyright © 2004 National Information Standards Organization.

Segue na Tabela 1, um exemplo de descrição de um recurso utilizando o padrão Dublin Core. Para fins de ilustrar este exemplo irá ser feita a descrição da página *web* do DCMI como sendo o recurso.

Tabela 1 – Exemplo de descrição de um recurso usando o padrão Dublin Core.

Elemento do Padrão Dublin Core	Conteúdo
DC.Title	Dublin Core Metadata Initiative – DCMI
DC.Creator	DCMI
DC.Subject	Dublin Core
DC.Subject	Metadados
DC.Subject	Informação
DC.Description	The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices.
DC.Publisher	DCMI
DC.Data	1995
DC.Type	Text
DC.Fomat	text/html
DC.Identifier	http://dublincore.org
DC.Language	en-us
DC.Coverage	1995-2009
DC.Rights	Copyright © 1995-2009 DCMI All Rights Reserved.

Fonte: próprio autor.

Este esquema que apresenta apenas os quinze elementos padrões do Dublin Core é normalmente conhecido como Dublin Core não-qualificado (*Unqualified Dublin Core*). Com a necessidade de se representar de maneira mais detalhada os recursos descritos pelos elementos Dublin Core, o DCMI desenvolveu qualificadores para o conjunto de elementos do padrão Dublin Core.

Estes qualificadores desenvolvidos foram estudados e baseados sobre vocabulários controlados e sistemas de classificação formais existentes e não pretendem ser um conjunto rígido.

Estes qualificadores podem ser livremente utilizados ou não, dependendo apenas das necessidades específicas de cada usuário do padrão Dublin Core.

Duas classes de qualificadores foram definidas pelo DCMI:

1 – Elemento de refinamento: tais qualificadores aumentam a especificidade de um elemento, aumentando desta maneira seu detalhamento.

2 – Esquema de codificação: estes qualificadores identificam esquemas para o valor do elemento. Neles incluem-se vocabulários controlados e notações formais de representação, que são os sistemas de classificação formais.

Os qualificadores desenvolvidos são mostrados de forma resumida na Tabela 2, tabela na qual pode-se observar que os elementos *Creator*, *Publisher*, *Contributor* e *Rights* não possuem qualificadores de nenhuma das classes.

Cada qualificador Dublin Core possui cinco atributos, que são definidos abaixo:

- **Nome:** símbolo único atribuído ao qualificador e utilizado, por exemplo, em uma representação em HTML;
- **Rótulo:** um rótulo legível dado ao qualificador para que possa ser feita a leitura humana;
- **Definição:** uma declaração que representa a natureza e também o conceito do qualificador;
- **Comentário:** informações adicionais associadas ao qualificador, que são descritas caso existam;
- **Veja também:** link para mais informações a respeito do qualificador, que é inserido caso existir.

Todos os qualificadores do padrão de metadados Dublin Core são representados junto com os elementos padrões do Dublin Core, tanto separados dos recursos quanto dentro deles, podendo-se a través do uso dos mesmos tornar as descrições dos recursos digitais mais detalhadas e mais específicas, dependendo das necessidades e políticas das instituições que os implementem.

O padrão de metadados Dublin Core é utilizado pela plataforma de construção de repositório digital que mais adiante foi escolhida para desenvolvimento de repositório proposto no trabalho e permite que tal plataforma possua interoperabilidade com outros sistemas.

Tabela 2 – Qualificadores do padrão Dublin Core

Conjunto de elementos DC	Elementos de refinamento	Elementos do esquema de codificação
1. Title	Alternative	-
2. Creator	-	-
3. Subject	-	LCSH MeSH DDC LCC UDC
4. Description	Table Of Contents Abstract	-
5. Publisher	-	-
6. Contributor	-	-
7. Date	Created Valid Available Issued Modified	DCMI Period W3C-DTF
8. Type	-	DCMI Type Vocabulary
9. Format	Extent	-
	Medium	IMT
10. Identifier	-	URI
11. Source	-	URI
12. Language	-	ISO 639-2 RFC 1766
13. Relation	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format	URI
14. Coverage	Spatial	DCMI Point ISO 3166 DCMI Box TGN
	Temporal	DCMI Period W3C-DTF
15. Rights	-	-

Fonte: <http://dublincore.org/documents/dcmes-qualifiers/>

CAPÍTULO 3 – INTEROPERABILIDADE

Interoperável significa “Capaz de operar em conjunto”.

Segundo a International Standard Organization (ISO) a interoperabilidade é a “habilidade de dois ou mais sistemas (de computadores, meios de comunicação, redes, *softwares* e outros componentes de tecnologia da informação) de interagir e de intercambiar dados de acordo com um método definido, de forma a obter os resultados esperados”

O Concise Oxford Dictionary define interoperabilidade como “a capacidade de um sistema ou produto trabalhar com outro sistema ou produto sem requerer esforço especial por parte do cliente”

Já dentro da literatura sobre engenharia de software considera a interoperabilidade como "o esforço exigido para se acoplar um sistema a outro", um fator de garantia de qualidade de software, conjuntamente com manutenibilidade, portabilidade, integridade, confiabilidade, etc. (PRESSMAN, 1995).

Segundo Arms (2002), o objetivo da interoperabilidade é desenvolver serviços coerentes para os usuários, a partir de recursos informacionais que são tecnicamente diferentes e gerenciados por diferentes organizações. Isto requer acordos de cooperação em três níveis:

- **Nível Técnico:** O nível técnico nos proporciona a interoperabilidade tecnológica;
- **Nível de Conteúdo:** remete a interoperabilidade semântica, onde a representação e organização do conhecimento são áreas chaves a serem estudadas;
- **Nível Organizacional:** se refere a interoperabilidade política, quando organizações se reúnem com intuito de alcançar a interoperabilidade, implementando padrões e tecnologias que cooperem com este objetivo. A interoperabilidade política viabiliza a gestão articulada entre as organizações e os sistemas de informação.

Um dos aspectos importantes para que ocorra a interoperabilidade de informações é a organização do conhecimento e representação da informação. Segundo Oliveira (2005, p.34) para que duas pessoas, duas bases de dados ou duas instituições possam trocar informações de forma eficaz é necessário que ocorra o entendimento dos códigos utilizados por ambas as

partes e que eles tenham o mesmo entendimento quanto ao significado destes códigos. Para isto é necessária a implementação de padrões e normas que possibilitem o entendimento entre eles, desenvolvendo e reforçando trabalhos cooperativos.

A busca da interoperabilidade é um objetivo perseguido com afincos, especialmente após o desenvolvimento da internet como conhecemos hoje.

Para viabilização do compartilhamento de informações, diversos padrões e metodologias foram desenvolvidas, tanto no campo de metadados, como o padrão MARC e o Dublin Core, como no campo dos protocolos de comunicação, como o OAI-PMH e o Z39.50.

Para Marcondes e Sayão (2002, p.42):

[...] com o advento da explosão informacional na metade do século XX, “a saída encontrada pelas bibliotecas foi a cooperação (...) Desde a invenção do computador na década de 50 as tecnologias de informação passaram a ser usadas pelas bibliotecas para prover acesso não só a documentos dos seus próprios acervos, mas também aos armazenados em acervos de outras bibliotecas”.

Estas estratégias de cooperação possibilitavam que as bibliotecas fossem capazes de operar em conjunto, isto é, fossem interoperáveis, ampliando os serviços aos usuários (OLIVEIRA, 2005).

Segundo Alves (2005, p.81-82) a interoperabilidade na área da computação é definida da seguinte maneira:

Interoperabilidade é a capacidade de compartilhamento de informações em diferentes sistemas e que, por meio de algumas ferramentas como linguagem de marcação adequada, uso de metadados e arquiteturas de metadados, estas informações registradas e armazenadas em diferentes estruturas e em diferentes comunidades do conhecimento poderão ser intercambiadas e trocadas nestes sistemas, fazendo com que haja um trabalho conjunto entre sistemas.

Interoperabilidade é dividida em três níveis de aplicação, sendo eles os seguintes:

- **Interoperabilidade Semântica:** Segundo Alves (2005, p.82-83) a interoperabilidade semântica permite a compreensão do significado de cada elemento que descreve o recurso, incluindo as associações nele embutidas. O uso de padrões de metadados e vocabulários específicos são essenciais e sem eles não é possível se garantir este nível de interoperabilidade.
- **Interoperabilidade Estrutural:** também segundo Alves (2005, p.83-84) a interoperabilidade estrutural define cada um dos elementos que compõem um padrão de metadados, descrevendo seus tipos, as escalas de valores possíveis para tais elementos e os mecanismos utilizados para se agrupar e relacionar tais elementos a fim de que possam ser processados automaticamente, Quanto

maior o nível de complexidade da estrutura do padrão de metadados, mais complexo deve ser o modelo de dados empregado para sua descrição. Exemplos: Padrão Dublin Core e MARC 21.

- **Interoperabilidade Sintática:** a interoperabilidade sintática, segundo Alves (2005, p.84), determina como os metadados devem ser codificados para que ocorra a transferência de informações entre os sistemas.

Para que haja interoperabilidade semântica, é necessário que haja acordos quanto à utilização destes descritores, isto é, que um termo tenha o mesmo significado que o utilizado em outra base (OLIVEIRA, 2005).

O uso de padrões de metadados, apresentados no capítulo anterior, visa garantir a interoperabilidade entre sistemas nos três níveis existentes.

3.1 – Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)

No ano de 1991 o físico Paul Ginsparg, do laboratório de Los Alamos no Novo México criou o primeiro repositório de documentos eletrônicos, o ArXiv¹. A partir de então os “pesquisadores passaram a criar arquivos eletrônicos de *preprints* e *posprints* como alternativa para a publicação direta de seus trabalhos em texto completo” (OLIVEIRA, 2005, p.43 apud MARCONDES, 2003, p.43). Tais arquivos foram chamados de “*open archives*”.

No ano de 1999 durante a Convenção de Santa Fé², no Novo México surgiu a *Open Archives Initiative* (OAI), a iniciativa dos arquivos abertos, com objetivo de apoiar o desenvolvimento da publicação eletrônica e tornar os repositórios interoperáveis.

Durante a convenção foram definidos os princípios básicos de uma nova filosofia de publicação científica. Entre estes princípios, segundo Triska e Café (2001, p.92) pode-se citar como principais os seguinte: o auto-arquivamento, a revisão pela comunidade e a interoperabilidade.

O auto-arquivamento é o direito de o próprio autor de uma obra a enviar para publicação sem o intermédio de terceiros. O objetivo deste princípio é tornar as obras disponíveis o mais rápido possível e oferecer acesso gratuito e também democrático as obras

¹ Ver <http://arxiv.org/>

² Ver http://www.openarchives.org/sfc/sfc_entry.htm

publicadas eletronicamente. Esta necessidade surgiu pelo fato de as grandes editoras deterem anteriormente o monopólio sobre as publicações científicas, o que diminuía a visibilidade das obras e o interesse dos pesquisadores, pois muitas vezes quando se conseguia publicar uma pesquisa, tais informações já haviam se tornado menos importantes devido ao grande tempo que foi levado para serem publicadas.

A revisão pela comunidade é o princípio que possui como propósito a transparência das críticas e sugestões feitas as obras depositadas em repositórios (TRISKA E CAFÉ, 2001). Desta maneira, repositórios que permitem a revisão pela comunidade, permitem que toda ela tenha acesso aos processos de revisão e aceitação das obras.

Para Triska e Café (2001, p.93) “a interoperabilidade é a palavra-chave para a criação dos arquivos abertos”.

Segundo Triska e Café (2001, p.93) a Convenção de Santa Fé estabeleceu os seguintes aspectos como mecanismos para atingir a interoperabilidade:

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

Para tais objetivos foi criado na convenção de Santa Fé a Open Archives Initiative (OAI) que é a iniciativa dos arquivos abertos, “um conjunto de especificações técnicas e princípios organizacionais bastante simples, porém potencialmente poderosos e de grande alcance, no objetivo de integração desses arquivos” (OLIVEIRA, 2005, p.44-45 apud MARCONDES, 2003, p.46).

O “open” de *open archives* não significa que tais arquivos tenham acesso gratuito e ilimitado, mas refere-se à arquitetura da proposta, pois o protocolo dos *open archives* que é aberto (CUSIN; ELVIS, 2007).

A base da iniciativa é o OAI-PMH que é o protocolo para coleta de metadados da iniciativa dos arquivos abertos, que permite aos sistemas que adotem tal padrão realizarem a troca e compartilhamento de seus metadados. O protocolo OAI-PMH segue o padrão Dublin Core (Item 2.5.3).

Os sistemas participantes da iniciativa se dividem nos seguinte grupos (GARCIA; SUNYE, 2007, p.03):

- **Provedores de Dados (*Data Providers* ou **DP**):** mantêm repositórios de documentos digitais que implementam o OAI-PMH como forma de expor os metadados de seus documentos.
- **Provedores de Serviços (*Service Providers* ou **SP**):** oferecem busca a os metadados mantidos nos provedores de dados e outros serviços que visem agregar valor a iniciativa.

O OAI-PMH introduz o conceito de Metadata Harvesting (coleta de metadados), um processo considerado unilateral, no qual os provedores de serviços, a partir da lista de repositórios, os provedores de dados, registrados na OAI, realizam buscas a estes servidores de dados periodicamente, colendo os metadados para exibição sob a forma de consultas realizadas por usuários do serviço.

A coleta de metadados pode ser total ou seletiva. As coletas seletivas são baseadas em dois critérios de coleta do *Harvesting* do OAI-PMH que são os seguintes (GARCIA; SUNYE, 2007, p.04):

- **Date-based (baseado em data):** apenas os metadados incluídos ou alterados após um data especificada serão coletados;
- **Set-based (baseado em conjuntos):** apenas metadados referentes a conjuntos da hierarquia do repositório, especificados na requisição, serão coletados.

Um provedor de serviços pode utilizar os seis “verbos” previstos no OAI para coletar metadados de um documento do provedor de dados. Estes seis “verbos”, segundo Oliveira (2005) e Garcia e Sunye (2007) são os seguintes:

- **Identify:** este verbo é utilizado para recuperar as informações sobre o provedor de dados. A resposta mínima que se espera desta requisição é o nome do provedor de dados, o endereço do repositório, a versão do protocolo implementada e o endereço eletrônico (*e-mail*) do administrador do repositório;
- **ListMetadataFormats:** este verbo recupera os formatos de metadados disponíveis no provedor de dados. O formato padrão é o Dublin Core, representado no protocolo por “oai_dc”;

- **GetRecord:** recupera apenas um único registro do repositório. Deve ser especificados o formato dos metadados e o identificador do registro que se deseja recuperar na utilização deste verbo.
- **ListRecords:** este verbo “colhe” os metadados do repositório, coleta que pode ser feita de maneira seletiva, baseada em conjuntos ou em datas. Também é obrigatório especificar o formato de metadados na sua utilização;
- **ListIdentifiers:** é uma versão menor do verbo *ListRecords*, que retorna apenas os cabeçalhos dos registros do repositório. Também pode ser realizada a coleta de maneira seletiva, baseada em conjuntos ou em datas. Também é obrigatório especificar o formato de metadados na sua utilização;
- **ListSets:** este verbo lista a estrutura do conjunto do repositório, ou seja lista a hierarquia do repositório.

3.2 – Protocolo Z39.50

O protocolo Z39.50, conhecido formalmente *como information retrieval protocol* (protocolo de recuperação de informação) segundo Rosetto (1997):

[...] é um protocolo de comunicação entre computadores desenhados para permitir pesquisa e recuperação de informação – documentos com texto completos, dados bibliográficos, imagens, multimeios – em redes de computadores distribuídos.

A necessidade de mecanismos capazes de normalizar a comunicação entre sistemas de computadores levou a NISO (*National Information Standards Organization*) à estabelecer um comitê para elaboração de um protocolo para recuperação de informação. Os estudos deste comitê iniciaram-se a partir dos estudos realizados nos anos 70 pela *Library of Congress* (Biblioteca do Congresso Norte Americano) (ROSETTO, 1997).

A primeira versão do protocolo Z39.50 data de 1988.

O protocolo Z39.50 funciona da seguinte maneira: um usuário consultando através de um programa cliente tem a impressão que os diversos catálogos consultados, gerenciado por programas-servidores em diferentes bases de dados, sejam um único catálogo.

O protocolo executa buscas imediatas as diferentes bases de dados, “[...] a interface de busca distribui a consulta (*broadcast search*) a diferentes sites [...] identificados como capazes de fornecer respostas satisfatórias, e os resultados são consolidados e integrados [...]” (OLIVEIRA, 2005, p.42 apud MARCONDES, 2003, p.27).

O protocolo Z39.50 possibilita que seja feita consultas unificadas a catálogos que estão distribuídos, através de uma rede e gerenciados de maneira descentralizada (OLIVEIRA, 2005, p.41 apud MARCONDES, 2003).

Rosetto (1997) esclarece que o protocolo Z39.50,

[...] pode ser implementado em qualquer plataforma. Isso significa que permite a interoperacionalização de diferentes sistemas de computação com diferentes sistemas operacionais, equipamentos, formas de pesquisa, sistemas de gerenciamento de bases de dados. Uma implementação Z39.50 habilita uma interface única para conexão com múltiplos sistemas de informação, permitindo ao usuário final um acesso quase transparente para outro sistema. Novos comandos e técnicas de busca não têm de ser aprendidos, pois os resultados da pesquisa são apresentados no sistema local – novamente, em formatos e estilos com os quais os usuários estão acostumados. Esse sistema é extremamente vantajoso para bibliotecas que querem uma interface única para o usuário, a fim de realizar pesquisas no catálogo on-line local e em bases de dados referênciais e remotas.

O protocolo Z39.50 funciona seguindo a arquitetura de cliente-servidor, onde máquinas servidores prestam serviços a máquinas clientes. Para que haja a interoperabilidade entre os sistemas que utilizam o protocolo Z39.50 é necessário que todas as máquinas envolvidas, tanto as clientes quanto as servidoras, utilizem o protocolo.

As máquinas servidoras utilizam um programa servidor Z39.50 e máquinas cliente utilizam um programa cliente Z39.50. Esta utilização por ambas as partes é um pré-requisito para que o protocolo Z39.50, que é um protocolo distribuído, funcione corretamente.

Uma evolução existente do Z39.50 é o SRW (*Search Retrieval Webservice*¹), um protocolo criado para realizar consultas a várias bases de dados disponíveis via internet. Também conhecido como protocolo Z39.50 de nova geração, o SRW tem o objetivo de incorporar novas tecnologias não existentes no surgimento do Z39.50, que ocorreu no ano de 1988, e desde então muitas tecnologias eficientes e necessárias foram desenvolvidas e devem ser adotadas, como a linguagem de marcação XML.

O SRW propõe um protocolo menos complexo e mais fácil de implementar do que o Z39.50 e também baseia-se na arquitetura cliente-servidor. As suas solicitações de pesquisa e as resposta são codificadas usando XML (OLIVEIRA, 2005, p.43).

A consulta no SRW é formulada através de uma linguagem padronizada, chamada de CQL (*Common Query Language*), que utiliza o conjunto de elementos do padrão Dublin Core (Item 2.5.3) para especificação dos índices das bases de dados.

¹ Ver <http://www.loc.gov/z3950/agency/zingsrw>

CAPÍTULO 4 – ANÁLISE DAS FERRAMENTAS DE CONSTRUÇÃO DE REPOSITÓRIOS DIGITAIS

Este capítulo trata da análise das ferramentas *open source*, ou *softwares* livres, disponíveis para implementação de repositórios digitais.

Todas as ferramentas analisadas possuem interface *web* e são softwares livres, por isso tais características não serão tratadas repetidamente, apesar de ocorrerem comentários a respeito dos tipos de licenças destes *softwares* e também dos *layouts* de suas interfaces.

Alguns requisitos de software são indispensáveis para o repositório digital proposto e portanto serão analisados individualmente, enquanto que os requisitos apenas desejáveis serão analisados e comentados durante as descrições de cada ferramenta.

A análise e descrição dos aplicativos levarão em conta os seguintes aspectos:

- Histórico de desenvolvimento e uso do software;
- Estrutura do repositório;
- Formatos de arquivos suportados;
- Opções e métodos de busca;
- Processo de submissão de arquivos;
- Opção de navegação pelas coleções existentes;
- Interação com o usuário;
- Documentação existente;
- Processo de inclusão dos metadados dos arquivos;
- Protocolos de comunicação e interoperabilidade suportados;
- Métodos de importação e exportação de metadados.

Os seguintes *softwares* não tiveram uma análise mais aprofundada de suas funções devido ao fato de a documentação dos mesmos ter demonstrado de maneira clara que não teriam as funcionalidades necessárias para suprir as necessidades de um repositório digital para o objetivo que está sendo proposto, que seria um repositório digital para um grupo de pesquisas. Seguem listados e acompanhados dos motivos que os levarão a serem descartados:

- **GNU/E-Prints**¹: possui como objetivo principal a literatura cinzenta, que são apenas teses, monografias, relatórios, e outros tipos de textos. Permite a

¹ Ver <http://www.eprints.org>

inclusão de alguns outros tipos de objetos digitais, mas como poucas opções e funcionalidades, devido ao fato de ter o foco principal na literatura cinzenta.

- **Fedora**¹: por possuir um grau de detalhamento de metadados muito alto, seu uso torna-se altamente complexo para os usuários finais do repositório proposto. É um software flexível para suportar diversos formatos de objetos digitais, porém não possui uma interface própria com o usuário, podendo ser utilizado como suporte para outras aplicações.
- **Open Journal Systems**²: seu foco principal são as publicações periódicas, como revistas e jornais.

4.1 – Comparação de Aplicativos

A Tabela 3 demonstra a lista dos requisitos necessários para que o *software* atenda as necessidades do repositório digital proposto.

Estes requisitos serão cruciais na escolha do software para desenvolvimento do repositório digital proposto, pois o não cumprimento de algum destes requisitos por algum dos *softwares* analisados irá eliminá-lo automaticamente da lista de possíveis *softwares* a serem utilizados, pois sem o cumprimento integral destes requisitos obrigatórios, muito provavelmente o uso deles implicará problemas tanto na fase de desenvolvimento quanto para os usuários finais do repositório digital.

Tabela 3 – Atendimento dos requisitos obrigatórios para o repositório digital proposto nos aplicativos analisados.

Requisitos obrigatórios	Softwares			
	DSPACE	CDSWare	Nou-Rau	Greenstone
<i>Software</i> livre	A	A	A	A
Interface <i>web</i>	A	A	A	A
Permite inclusão de mais um formato de arquivo por obra	A	A	NA	A
Utilização simples e intuitiva	A	A	A	A
Possibilita criação de coleções distintas	A	A	A	A
Possibilita navegação pelos campos autor, título, assunto ou data.	A	A	A	A
Permite busca por autor, título, assunto e por palavras-chave	A	A	A	A
Permite submissão pelo próprio autor	A	A	A	NA
A inclusão de metadados segue algum	A	A	A	A

¹ Ver <http://www.fedora.info>

² Ver <http://pkp.sfu.ca/ojs>

padrão aceito internacionalmente				
Permite importação e exportação de obras e metadados	A	A	A	A
Possui interação com usuário, através de <i>e-mail</i> e informações no repositório	A	A	A	NA
Documentação existente é satisfatória	A	A	NA	A

A = atende requisito; NA = não atende requisito

Fonte: Próprio Autor

A Tabela 4 apresenta uma lista dos requisitos desejáveis, mas não obrigatórios dos softwares analisados. Estes requisitos não são obrigatórios, mas tem peso importante na escolha do *software* a ser utilizado no desenvolvimento do repositório digital.

Tabela 4 – Atendimento dos requisitos desejáveis para o repositório digital proposto nos aplicativos analisados.

Requisitos desejáveis	Softwares			
	DSpace	CDSWare	Nou-Rau	Greenstone
Não precisar de <i>plugins</i> instalados na máquina do usuário final	A	A	A	A
Possui comportamentos distintos para tratamento de coleções distintas	A	A	NA	A
Interface agradável ao usuário e personalizável segundo as necessidades	A	A	A	A
Possui hierarquia entre as coleções do repositório	A	A	A	A
Possibilidade de navegação por vários campos, além de autor, título, assunto ou data.	NA	A	NA	A
Permitir buscas avançadas fazendo a combinação de elementos de metadados e texto-integral (<i>full-text</i>).	A	A	NA	NA
Possui mecanismos de <i>workflow</i> básicos	A	A	A	NA
Permite a interoperabilidade de maneira automática com outros sistemas na mesma base ou em bases distintas	A	A	NA	A
Permite configuração de uso de vocabulários controlados	A	NA	NA	A
Permitir exportação de coleções ou parte delas para consulta local do usuário.	A	NA	NA	A
Permite utilização de protocolos automatizados para realização de troca de registros	A	A	A	NA
Permite uso de identificadores persistentes	A	NA	NA	NA
Permite gerar relatórios de estatísticas de utilização	A	A	A	NA
Permite interação no estilo de portais, onde usuários cadastrados possuem acesso	A	A	A	NA

 personalizado aos serviços

 A = atende requisito; NA = não atende requisito

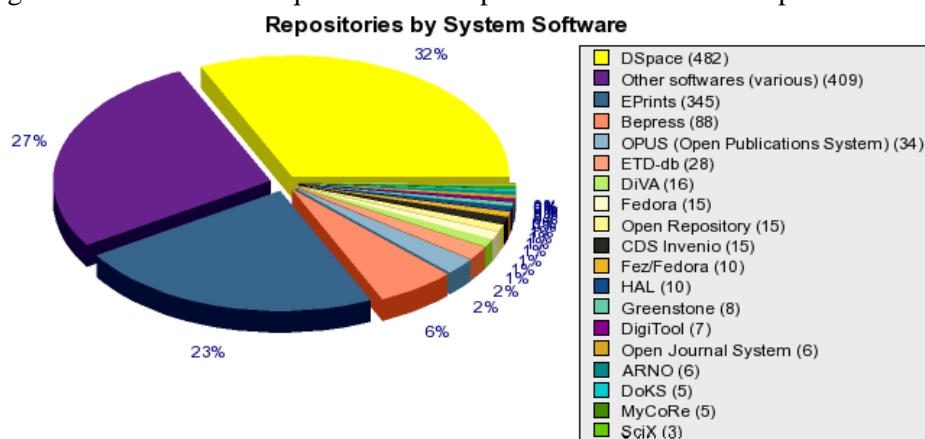
 Fonte: Próprio Autor

4.2 – Análise Detalhada da Ferramenta DSpace

DSpace é um sistema de repositórios digitais *Open Source*, baseado no modelo de licença BSD (*Berkeley Software Distribution*), desenvolvido pelo MIT¹ (*Massachusetts Institute of Technology*) em parceria com a HP² (*Hewlett-Packard*) e está disponível no próprio site do DSpace³. Seu nome completo é *Digital Space – Institucional Digital Repository System* (Espaço Digital – Sistema de Repositórios Digitais Institucionais).

Atualmente é a ferramenta para construção de repositórios digitais institucionais mais usada internacionalmente, segundo dados do OpenDOAR⁴ e do ROAR⁵, sites responsáveis por indexar os repositórios digitais existentes em todo o mundo e fornecer dados, números e gráficos a respeito do uso dos mesmos. A Figura 1 mostra o número de repositórios criados em várias ferramentas disponíveis diferentes disponíveis, e reforça a afirmação de que o DSpace é a mais usada entre elas.

Figura 1 – Gráfico com a quantidade de repositórios desenvolvidos por cada sistema



Fonte: http://roar.eprints.org/index.php?action=generate_chart&country=&version=&type=&chart_field=version&chart_type=pie&submit=Generate+Chart

Segundo os próprios desenvolvedores do sistema, o DSpace captura, distribui e também preserva objetos digitais (LEWIS; YATES, 2008).

¹ Ver <http://web.mit.edu>

² Ver <http://www.hp.com>

³ Ver <http://www.dspace.org>

⁴ Ver <http://www.opendoar.org/index.html>

⁵ Ver <http://roar.eprints.org/index.php>

Lewis e Yates (2008) relatam que o DSpace é uma plataforma que permite capturar itens em qualquer formato – em texto, vídeo, áudio e dados, os distribui através da *web*, indexa o trabalho, então usuários podem procurar e recuperar os itens, e ele preserva os trabalhos digitais a longo prazo.

Lewis e Yates (2008) também relacionam que o DSpace possui três objetivos principais, que são os seguintes:

- Facilitar a captura e inserção dos materiais, incluindo os metadados sobre os materiais;
- Facilitar o acesso fácil ao material, tanto por listagens quanto por buscas;
- Facilitar a preservação a longo prazo dos materiais.

Segundo a documentação do DSpace as idéias que motivaram seu desenvolvimento baseiam-se nas seguinte premissas:

- Muito dos materiais que nasceram digitais já estão perdidos;
- A maior parte do que ainda não foi perdido, corre risco de ser;
- É melhor realizar a preservação digitalmente, do que perder completamente;
- É necessário capturar tanta informação quanto for possível, para se suportar a preservação funcional;
- Relação custo/benefício favorável.

O DSpace é o sistema mais usado para desenvolvimento de repositório digitais em todo o mundo devido ao fato de ser desenvolvido em parceria por uma das instituições de ensino e um dos laboratórios de pesquisa mais respeitados em todo o mundo, formando assim o conceito de que o DSpace irá sempre se manter atualizado e com qualidade crescente, sempre atendendo as necessidades que forem surgindo e implementando as mais novas tecnologias existentes em suas funcionalidades. O DSpace encontra-se atualmente na versão 1.5.2, atualização lançada no ano de 2009. Sua primeira versão, a 1.0 foi lançada em novembro de 2002 e a partir daí a cada ano foi lançada uma nova versão, sendo a versão 1.1 em maio de 2003, a versão 1.2 em agosto de 2004, a versão 1.3 em agosto de 2005, a versão 1.4 em julho de 2006 e no único caso que passou do período de um ano em março de 2008 foi lançada a versão 1.5. Uma nova versão, a 1.6 já encontra-se em fase de estudos.

Outro fator importante para este sistema ser o mais adotado pelas instituições e universidades do mundo é este próprio fator, ou seja, como possui uma comunidade muito

grande de usuários e desenvolvedores e seu código ser aberto, tais desenvolvedores contribuem continuamente para melhorias e atualizações das funcionalidades do DSpace e também na correção dos possíveis problemas que possam vir a surgir como o seu uso.

Também pode se disser que uma das maiores vantagens do DSpace é o fato de ele suportar todos os tipos de formatos digitais. Alguns exemplos são:

- Documentos (artigos, *preprints*, relatórios técnicos, dissertações);
- Livros;
- Teses;
- Visualizações, simulações e outros modelos;
- Softwares;
- Publicações multimídia;
- Imagens;
- Arquivos de áudio;
- Arquivos de vídeo;
- Páginas *web*;
- E muitos outros.

A *DSpace Foundation* (Fundação DSpace) é uma fundação sem fins lucrativos fundada em 2007 para prover o crescimento da comunidade de instituições que adotam o DSpace. A missão da fundação é liderar o desenvolvimento colaborativo do *software open source* para possibilitar o permanente acesso as produções digitais.

Os principais objetivos da *DSpace Foundation* são:

- Desenvolver e gerenciar uma forte rede de provedores de serviços e recursos de treinamento;
- Promover o DSpace através de *newsletters* mensais, através da página *web* do DSpace, através de materiais de marketing e diversas outras ações;
- Construir e manter uma comunidade ativa de desenvolvedores e usuários;
- Garantir a integração do DSpace usando padrões abertos;
- Administrar e coordenar a plataforma DSpace e os lançamentos de softwares.

O DSpace é formado por diversos componentes, distribuídos por três camadas distintas:

- A *Storage Layer* que é responsável pelo armazenamento físico dos metadados e dos conteúdos;
- A *Business Logic Layer* que trata da gestão dos conteúdos do arquivo, dos utilizadores, das políticas de autorização e do *workflow* e;
- A *Application Layer* que contém os componentes que comunicam com o mundo exterior, como por exemplo a interface *web* do utilizador e o serviço de suporte ao protocolo de coleta de metadados da OAI.

Cada camada da arquitetura invoca apenas a camada imediatamente inferior, por exemplo, a *Application Layer* não pode invocar diretamente a *Storage Layer*. Cada componente da *Business Logic Layer* e da *Storage Layer* possuem uma API (*Application Programming Interface*).

Estas API's são classes, objetos e métodos em Java. O conjunto das API's públicas dos diversos componentes de cada camada, forma aquilo que se designa por *Layer API* e possui as designações de *Storage API* e *DSpace Public API*.

Um das grandes vantagens conhecidas em arquiteturas por camadas é a de permitir que o desenvolvimento de novas funcionalidades, o melhoramento da eficiência das existentes e a adaptação a outros ambientes (Sistemas Operacionais) possa ser feito apenas em uma camada, desde que se assegure que sua API mantenha o suporte a todas as funcionalidades anteriores ao desenvolvimento.

Esta preocupação em manter a arquitetura de três camadas do DSpace está presente na própria organização do código fonte, que está coerentemente repartido em três pacotes, que correspondem estritamente a cada uma das camadas. O pacote *org.dspace.app* que implementa a *Application Layer*, o pacote *org.dspace.Business* que implementa a *Logic Layer* e o pacote *org.dspace.storage* que implementa a *Storage Layer*.

Para seu correto funcionamento, o DSpace na versão mais recente, a 1.5.2, necessita dos seguinte pré-requisitos de softwares:

- **Sistema Operacional:** Windows ou UNIX-like.
- **Servidor Web:** Apache Tomcat 5.X ou mais novo.
- **Banco de Dados:** PostgreSQL 8.X ou mais novo ou Oracle 9 ou mais novo.
- **Adicionais:** Apache ANT 1.6.2 ou mais novo, Java SDK 1.5 ou mais novo e Apache Maven 2.0.8 ou mais novo.

4.2.1 – Metadados e Interoperabilidade no DSpace

O Dspace mantém três tipos de metadados a respeito dos seus recursos, os metadados descritivos, os administrativos e os estruturais.

Para descrição dos recursos, o DSpace utiliza o padrão de metadados Dublin Core, padrão aberto e amplamente aceito internacionalmente. Apenas três elementos do padrão Dublin Core são obrigatórios nas descrições de recursos no DSpace, o título (*title*), idioma (*language*) e data de depósito (*date*), mas todos os outros campos do Dublin Core podem ser preenchidos, dependendo apenas das necessidades.

Os metadados administrativos referem-se à preservação, proveniência e políticas de autorização. O Dspace contempla também outros elementos do Dublin Core, como o tamanho em *bytes* do *bitstreams*.

Os metadados estruturais incluem informação sobre como apresentar um item ou os seus *bitstreams*, ao usuário final bem como as relações entre as partes que formam o recurso.

O DSpace adota o protocolo de comunicação OAI-PMH para compartilhamento de recursos e capacidade de interoperabilidade, o que permite a coleta automática de metadados de documentos armazenados em arquivos eletrônicos.

O DSpace também possui módulos para exportação no formato METS, o que possibilita que as o que conteúdo digital das coleções armazenadas no DSpace juntamente com seus metadados possam ser exportados e compartilhadas tanto com repositórios DSpace quanto com repositórios que utilizem outras plataformas além do DSpace mas que sejam compatíveis com o OAI-PMH.

4.2.2 – Estrutura e Funcionamento do DSpace

A estrutura do DSpace é dividida da seguinte maneira:

- **Comunidades:** são o nível mais alto da hierarquia de conteúdos do DSpace. Podem ser departamentos, centros de pesquisa, laboratórios, etc. Fazem relação ao grupo de pessoas que submetem conteúdo à base. Possuem metadados descritivos a respeito delas mesmas e das coleções inseridas nelas.
- **Coleções:** são os conjuntos de itens pertencentes às comunidades. Coleções podem pertencer a uma comunidade ou a várias. Como as comunidades, as coleções possuem metadados descritivos a respeito delas mesmas e dos itens inseridos nelas.

- **Itens:** são os documentos, ou objetos digitais, em si. São a representação dos arquivos e dos metadados unidos para formar uma unidade atômica. Normalmente é o que os usuários estão buscando.
- **Bitstreams:** cada arquivo submetido ao DSpace ou criado por ele é considerado um *bitstream*. Um *bitstream* refere-se ao fato de cada arquivo ser um fluxo de ‘bits’ (0s e 1s) armazenados em um meio de armazenamento físico.

Resumindo a base de dados DSpace é composta por comunidades que representam as divisões da instituição. Cada comunidade possui suas coleções, que são compostas por itens. Cada item é formado por um ou mais arquivos em formatos diversos, chamados de *bitstreams*. Cada item também possui metadados do padrão Dublin Core, descrevendo as informações necessárias a respeito dos recursos.

Outra característica muito importante e que é encontrada entre as ferramentas analisadas apenas no DSpace é o uso de identificadores persistentes, que são endereços de internet baseados no sistema CNRI Handle System¹ que garantem que o recurso armazenado será sempre encontrado através do endereço eletrônico correspondente, o que permitirá a pesquisa e recuperação dos recursos em um futuro distante. A Figura 2 representa um exemplo da exibição dos metadados básicos de um recurso no DSpace. A informação <http://hdl.handle.net/2160/617> é o identificador persistente, portanto caso queira citar este item em uma referência bibliográfica, pode-se utilizá-lo com certeza de que este endereço será sempre referente ao documento “The DSpace Course - An Introduction to DSpace”.

O DSpace identifica dois níveis de preservação digital, que servem para garantir a preservação do material digital a longo prazo. Os dois tipos de preservação são os seguintes:

- **Preservação de bits:** assegura que um arquivo continua exatamente igual ao longo do tempo, não sendo modificado um único bit, enquanto que o meio físico que o rodeia e armazena evolui;
- **Preservação funcional:** vai além da preservação de bits. O arquivo não se modifica ao longo do tempo para que possa ser imediatamente usado na sua forma original, enquanto o formato digital e o meio físico evoluem ao longo do tempo.

¹ Ver <http://www.handle.net>

Figura 2 – Exemplo de representação dos metadados básicos no DSpace.

Please use this identifier to cite or link to this item: <http://hdl.handle.net/2160/617>

Title: The DSpace Course - An Introduction to DSpace

Authors: [Lewis, Stuart](#)
[Yates, Chris](#)

Issue Date: 13-Aug-2008

Abstract: This module will introduce DSpace, its history, and the DSpace Foundation. The module will introduce what DSpace is and what it can be used for. It will then describe the development of DSpace over time, and talk about the DSpace Foundation which now oversees the running of DSpace. The open source development model used by DSpace will also be described.

URI: <http://hdl.handle.net/2160/617>

Rights: This work is licensed by a Creative Commons 'by-attribution sharealike' licence

Sponsor: This course was created by the Repositories Support Project (<http://www.rsp.ac.uk/>) and was funded by JISC (<http://www.jisc.ac.uk/>)

Appears in Collections: [The DSpace Course](#)

Fonte: <http://cadair.aber.ac.uk/dspace/handle/2160/617>

Alguns formatos de arquivos podem ser preservados funcionalmente usando-se formatos de migração direta, como as imagens TIFF (*Tagged Image File Format*) e os documentos XML. Outros formatos de arquivos, por diversas razões, são mais complexos de serem funcionalmente preservados. A previsão de que tipo de formatos uma instituição irá usar para a criação de seus documentos não é sempre possível, então usam-se ferramentas de acordo com os fins pretendidos e os repositórios recebem todos eles. Por esta razão existem três níveis de preservação definidos para um determinado formato: suportado, conhecido ou não suportado. Abaixo segue detalhamento deste três níveis:

- Formatos suportados serão funcionalmente preservados através do uso do formato de migração ou técnicas de emulação conhecidas. Exemplos deste formato são os arquivos TIFF, SGML, XML, AIFF (*Audio Interchange File Format*) e PDF (*Portable Document Format*).
- Formatos conhecidos são aqueles que não se pode prometer a sua preservação, como os formatos proprietários ou binários, mas que são tão usuais que versões de migração poderão surgir e ajudar na migração destes formatos. Os exemplos

incluem arquivos do Microsoft Word, Excel e PowerPoint, Lótus 1-2-3 e WordPerfect.

- Formatos não suportados são aqueles que ainda não há conhecimento suficiente para ser feito algum tipo de preservação funcional. Incluem alguns formatos proprietários ou de *software* único.

Para os três níveis o DSpace realiza a preservação de bits.

As equipes de desenvolvimento do DSpace estão trabalhando continuamente em parceria com diversas instituições para desenvolverem novos procedimentos de *upload* para converter formatos não suportados ou conhecidos em suportados, e melhorar a capacidade do DSpace de preservar os metadados e proceder com as migrações periódicas de formatos.

4.2.3 – Processo de Depósito e Workflow no DSpace

O DSpace é a primeira ferramenta de repositório digital *open source* apresentar uma solução para resolver o complexo problema de como integrar e implementar os diferentes processos de submissão de itens necessários em sistemas multidisciplinares. As diversas comunidades, representando departamentos, centros de pesquisa, laboratórios, entre outros, possuem diferentes idéias de como, por quem, e com que restrições os materiais devem ser depositados no repositório. Questões como, quem deve ter permissão para submeter itens? Que tipo de documentos podem ser submetidos? Quem pode rever ou aprovar o depósito? Em que coleção um usuário pode submeter materiais? Quem pode ter acesso aos documentos no repositório? São respondidas pelos responsáveis de cada comunidade, e posteriormente definidas e configuradas no sistema.

Isto é possível devido a definição de papéis com determinadas permissões que são atribuídas aos usuários do sistema. Estas permissões podem ser definidas individualmente para comunidades, coleções e documentos. Podem ser definidos os papéis de *submitters*, *reviewers* e *metadata editors*.

O *workflow* é o processo de revisão a que um depósito é sujeito, pode ser composto por três passos. Cada coleção poderá ter grupos de usuários para executar cada um destes passos. Se um passo não estiver associado a nenhum grupo, esse passo é ignorado e passa-se ao passo seguinte, e se uma coleção não tiver grupos associados a nenhum dos passos de *workflow*, os itens a ela submetidos são inseridos diretamente no repositório.

Desta maneira o *workflow* de submissão de documentos é constituído pela seguinte seqüência: depois da coleção receber um pedido de submissão é examinado o passo 1 do *workflow* da coleção. Se algum grupo estiver associado ao passo 1, o grupo é notificado e o passo 1 é invocado. Senão o passo 1 é pulado, passando então a ser examinado o passo 2, que também só é invocado se algum grupo estiver associado a ele, e por fim o mesmo ocorrerá com o passo 3.

Quando um passo é invocado, as tarefas correspondentes a ele são colocadas no *pool* de tarefas do grupo respectivo. Cada membro deste grupo pode então aceitar essa tarefa, sendo então esta removida do *pool*, para evitar que outro elemento do grupo a aceite.

Conforme o passo do *workflow* invocado, o usuário que aceitou a tarefa tem as seguintes ações para executar:

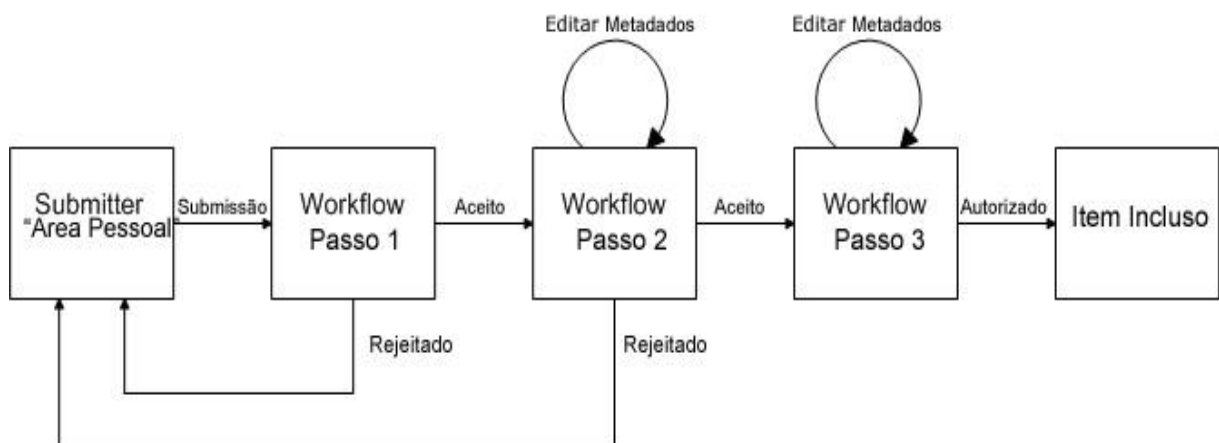
Passo 1: aceitar a submissão, e deixar o material ser incluso no repositório ou rejeitá-lo.

Passo 2: editar os elementos de metadados fornecidos pelo autor durante a realização do processo de submissão, aceitar a submissão ou rejeita-la.

Passo 3: não tem a capacidade de rejeitar a submissão, então é obrigado a aceita-la, mas antes pode editar os metadados fornecidos pelo autor.

Caso um revisor rejeite a submissão de um item, a razão para esta rejeição poderá ser enviada por e-mail à pessoa que realizou a submissão. O usuário que pretende depositar o item pode, a partir de sua página pessoal retomar o depósito do item, fazer as alterações necessárias e aconselhadas e voltar a submeter o mesmo, o que dará início a um novo *workflow*. A Figura 3 demonstra o processo de *workflow* no DSpace.

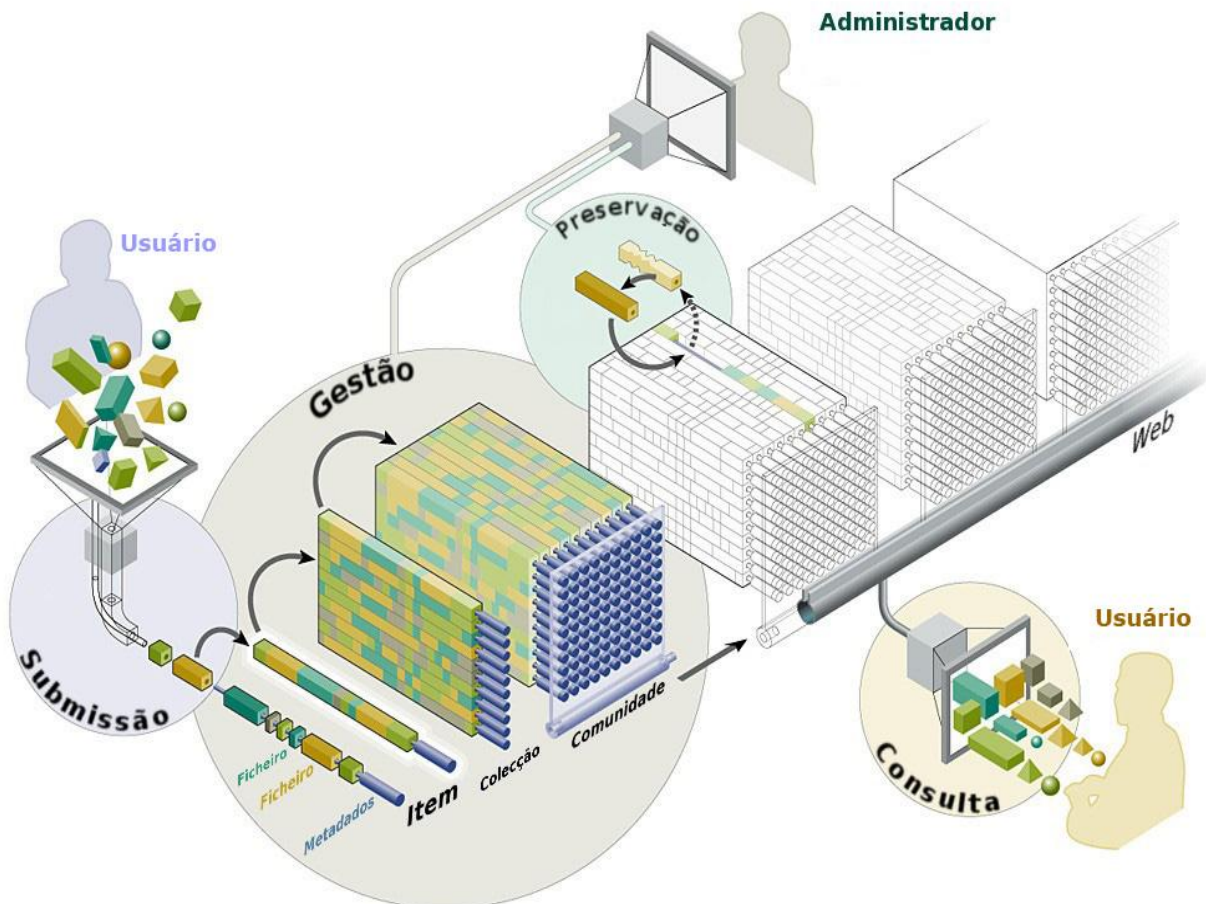
Figura 3 – Processo de *workflow* no DSpace



Fonte: Adaptado da imagem disponível em http://wiki.dspace.org/static_files/thumb/b/b8/Workflow-old.png/700px-Workflow-old.png

A Figura 4 resume os processos e o funcionamento do DSpace para cada um de seus usuários.

Figura 04 – Funcionamento do Dspace



Fonte: Adaptado da imagem disponível em <http://www.dspace.org/images/stories/dspace-diagram.pdf>

A Figura 4 mostra que um usuário pode realizar submissões, que são transformadas em itens, armazenadas dentro de coleções e relacionadas a comunidades. O administrador gere a aceitação e políticas de acesso aos documentos submetidos e novamente usuários podem realizar consultas a base de dados.

4.2.4 – Buscas e Navegação no DSpace

Pesquisas em repositórios DSpace podem ser realizadas de diversas maneiras diferentes.

O usuário pode pesquisar diretamente digitando os termos que deseja buscar ou pode optar por navegar pelas comunidades e coleções presentes no repositório. Pode ainda escolher navegar por título, autor e data, que são os elementos obrigatórios a serem preenchidos na descrição dos metadados de um recurso, e por isso é possível que seja feita a navegação por tais elementos, pois todo recurso no repositório obrigatoriamente possuirá estes três elementos preenchidos.

Um usuário possui a opção de se cadastrar no repositório, a fim de receber por *e-mail* informações sobre novos depósitos em cada coleção, e também quando recursos forem alterados.

O repositório possui dois tipos de áreas, as áreas livres que podem ser acessadas por qualquer usuário que acesse a página do repositório e as áreas de acesso controlado, onde somente usuários autorizados poderão ter acesso. A submissão de itens, como já descrito anteriormente, também só pode ser realizada por usuários autorizados.

Como citado acima, existe a opção de navegar pelas comunidades e coleções, por títulos, autores e datas. Também existe a opção de se digitar termos para uma busca geral e uma opção muito interessante, que é a de se digitar termos para buscar apenas dentro de coleções definidas de determinadas comunidades do repositório.

Nas buscas em que se digitam termos, sejam eles específicos por comunidades ou gerais, a pesquisa é realizada seguindo algumas regras, conforme definido a seguir:

- O asterisco é usado para fazer o truncamento, dessa maneira a busca *trab** irá retornar os resultados para *trabalho*, *trabalhando*, etc;
- Frases devem ser delimitadas por aspas duplas;
- Algumas palavras são ignoradas durante a busca. Na base em inglês algumas das palavras ignoradas são: “a”, “as”, “are”, “and”, “to”, etc;
- O símbolo “+” adicionado após uma palavra indica que obrigatoriamente ela deve constar nos resultados da busca. Dessa forma, a busca *+repositórios* digitais indica que “repositórios” deve aparecer nos resultados e “digitais” é opcional;
- O símbolo “-” é usado para negação;
- Os operadores booleanos AND, OR e NOT são suportados nas buscas, mas para isso devem ser escritos em caixa alta;

- Podem ser utilizados parênteses para combinar estratégia de busca, como por exemplo, (repositórios digitais OR repositórios institucionais) AND (metadados OR Dublin Core);
- As palavras têm seus finais expandidos para os finais que são mais comuns, com o objetivo de retornar mais resultados.

Com relação a interface com o usuário, pode se disser que é bastante simples, intuitiva e funcional, não acarretando problemas ou dificuldades maiores para os usuários finais do DSpace. O *layout* das páginas é extremamente agradável, embora razoavelmente simples, mas ainda assim muito agradável e conta muito para tornar a experiência no uso do DSpace para os usuários finais muito mais fácil.

4.3 – Análise Detalhada da Ferramenta Greenstone

O Greenstone é um sistema *Open Source*, baseado no modelo de licença GPL (*General Public Licence*), desenvolvido e distribuído pelo Projeto Biblioteca Digital da Nova Zelândia, na universidade de Waikato em parceria com a UNESCO (*United Nations Educational, Scientific and Cultural Organization*) e a Human Info NGO (*Humanitarian Information Non-governmental Organization*) da Bélgica e está disponível no próprio site do Greenstone¹.

Segundo Greenstone (2009), o Greenstone é uma ferramenta que visa construir e distribuir coleções em bibliotecas digitais.

O ambiente do Greenstone é dividido em duas interfaces distintas: uma disponível através do *browser* e que trata da interação com o usuário e uma interface gráfica baseada em Java, destinada ao administrador do sistema, onde o mesmo pode coletar itens para o acervo, adicionar metadados, projetar as funcionalidade de navegação e pesquisa que a coleção vai oferecer ao usuário final e construir e disponibilizar as coleções.

A construção de coleções também pode ser realizada na interface de administração, que é uma interface *web* que possui uma quantidade menor de funcionalidade que a interface Java.

O padrão de metadados Dublin Core é o nativo do Greenstone, mas novos padrões podem ser adotados através do uso de *plugins*.

¹ Ver <http://greenstone.sourceforge.net>

Os plugins também podem ser usados para a submissão de documentos no sistema, sendo que para documentos de texto existem plugins que dão suporte a diversos formatos de arquivos de texto. Também existem os plugins para os documentos multimídia, que dão suporte para diversos formatos de imagens, arquivos de áudio e também vídeo.

A documentação disponível sobre o Greenstone é ampla e conta até com exemplos completos de como construir uma coleção desde o início.

O processo de submissão é realizado a partir da interface Java, que inclui seções para coleta dos documentos, tanto a partir da internet quanto localmente. Também possui seções para definição de metadados para cada um dos documentos adicionados ao sistema, seleção das opções das coleções, como por exemplo quais documentos serão suportados, e a criação e disponibilização das coleções na interface *web* do usuário.

Para seu correto funcionamento, o Greenstone necessita dos seguinte pré-requisitos de softwares:

- **Sistema Operacional:** Windows ou UNIX-*lik* ou MAC-OSX.
- **Servidor Web:** Apache.
- **Banco de Dados:** GDBM (GNU Database Manager).
- **Adicionais:** PERL, Java SDK e plugins para padrões de metadados e formatos de arquivos adicionais.

4.4 – Análise Detalhada da Ferramenta CDSWare

O CDSWare é um sistema de repositórios digitais *Open Source*, baseado no modelo de licença GNU *General Public Licence*, desenvolvida pelo CERN (*European Organization for Nuclear Research*), que é a Organização Européia para Pesquisas Nucleares e está disponível no próprio site do CDSWare¹. Seu nome completo é CERN *Document Server Software* (CERN Software Servidor de Documentos).

Pepe et al (2005) afirma que o CDSWare é um conjunto de aplicativos que proporciona um ambiente e ferramentas para construir e gerenciar um servidor de biblioteca digital autônomo. Cada um de seus módulo é uma entidade independente que incorpora um aspecto específico de fluxo de trabalho de uma biblioteca digital.

A organização do CDSWare é feita de maneira a prover que seus vários módulos interajam entre si e com as camadas de interface e de armazenamento de dados. Os

¹ Ver <http://cdsware.cern.ch/cdsware/download.html>

documentos são organizados em coleções que são estruturadas pelo administrador do sistema em formato de árvores reais ou virtuais para facilitar a navegação dos usuários.

A submissão de recursos é realizada pelo próprio usuário e no momento da submissão o usuário deve informar a qual coleção o documento pertencerá. Após a realização da submissão de um documento, o mesmo passará por um processo de *workflow* que pode possuir ou não as etapas de revisão e aceitação.

Todos os formatos de arquivos aceitos podem ser submetidos ao repositório pelos usuários autorizados por e-mail ou através da interface *web* disponível. A maneira como tais arquivos serão exibidos pode ser personalizado, podendo assim existir comportamentos diferentes de acordo com o tipo de coleção. Também é permitido definir o tamanho máximo e mínimo de uma arquivo a ser adicionado à obra.

Os usuários podem criar cestas de documentos, nas quais irão constar os resultados das buscas realizadas por eles e também poderão compartilhar estas cestas entre grupos de usuários. O CDSWare também possui a funcionalidade de permitir que os usuários realizem comentários a respeito dos documentos existentes no repositório, funcionalidade que ajuda muito os usuários, pois os mesmos poderão contar com dicas e opiniões sobre a qualidade dos documentos inseridos no repositório.

O mecanismo de busca do CDSWare trabalha como uma sintaxe semelhante a usada pelo Google e permite que se combine busca nos metadados com buscas por texto-integral (*full-text*). O sistema possui uma opção de ordenação por frequência de palavras, que possibilita recuperar registros similares e também um método de *ranking* baseado nos valores específicos dos metadados.

O CDSWare também possui um módulo de indexação e *ranking* que possibilita classificar os resultados das buscas pela quantidade de *downloads* ou pelo número de citações dos registros. As buscas podem ser simples ou avançadas, sendo que os resultados das buscas podem ser agrupados também por coleção.

A interface de busca pode ser personalizada e encontra-se traduzida para diversas línguas, incluindo o português.

O CDSWare adota como padrão de metadados o MARC 21, e todos os metadados adicionados a ele são automaticamente convertidos para este padrão interno do software.

O padrão MARC 21 foi adotado por tratar de um formato já aceito e estabelecidos entre as bibliotecas do mundo, por possuir a característica de se integrar bem com as linguagens de marcação modernas, como o XML, é flexível o suficiente para garantir que

devido por muito tempo e pode ser estendida para se adaptar a qualquer estrutura de metadados existente (PEPE et al, 2005, p.04).

O CDSWare adotou o padrão MARCXML, padronizado pela *Library of Congress* (Biblioteca do Congresso Norte Americano), padrão que pode ser adotado por completo ou apenas em pequenos conjuntos de elementos específicos, dependendo apenas das necessidades de cada implementação. Normalmente quanto mais heterogênea for uma coleção, maior será o número de elementos de metadados que utilizará.

O padrão MARCXML usado no CERN possui atualmente mais de 150 elementos de metadados (PEPE et al, 2005, p.04)

O CDSWare também adota o protocolo OAI-PMH, portanto os metadados podem ser coletados de maneira automática e todos os metadados que são incluídos num repositório CDSWare são transformados para o formato nativo do *software*, para depois ficarem disponíveis. Um dos módulos existentes no CDSWare possui a função de traduzir os formatos recebidos, como Dublin Core, METS ou MARC 21 para o formato nativo do software no momento de realizar a exportação.

O CDSWare, dentro dos padrões estabelecidos pela OAI pode atuar tanto como sendo um provedor de dados quanto como sendo um servidor de serviços e possui a capacidade de fazer e também atender requisições a partir do protocolo OAI-PMH.

Segundo Pepe et al (2005) o CDSWare pode ser utilizado tanto como uma “solução genérica de gerenciamento de documentos” quanto como um “sistema de biblioteca digital ou um repositório digital” que pretende atender a acervos com médios e grandes portes.

A documentação disponível sobre o CDSWare é ampla e abrange características gerais e aprofundadas do sistema, porém algumas funcionalidades e diversos aspectos do funcionamento do CDSWare ainda não constam na documentação disponível.

Para seu correto funcionamento, o CDSWare necessita dos seguintes pré-requisitos de softwares:

- **Sistema Operacional:** UNIX-like.
- **Servidor Web:** Apache.
- **Banco de Dados:** MySQL.
- **Adicionais:** Python, PHP, WML, GNUplot, Implementação de Common LISP (CLISP, SBCL ou CMUCL) e conversores de formatos de arquivos para realização de indexação *full-text*.

4.5 – Análise Detalhada da Ferramenta Nou-rau

O Nou-rau é um sistema de repositórios digitais *Open Source* baseado no modelo de licença GPL, desenvolvido pelo Instituto Vale do Futuro em parceria com o Centro de Computação da Unicamp e está disponível no próprio site do Nou-rau¹.

O *software* tem como objetivo implementar um sistema *online* para arquivamento e indexação de documentos, provendo acesso controlado e mecanismos eficientes de busca (AKIMURA, 2005).

Os objetivos do sistema Nou-rau segundo Akimura (2005) são os seguintes:

- Armazenar qualquer tipo de documento;
- Manter sempre informações básicas e quando necessário informações específicas sobre cada documento;
- Permitir pesquisas tanto nos metadados quanto em texto-integral (*fulltext*);
- Permitir que haja um controle sobre as submissões aos acervos;
- Possibilitar a verificação externa de vírus nos documentos;

O Nou-rau é dividido dentro da seguinte organização funcional:

- **Documentos:** são os arquivos submetidos ao sistema e também seus metadados. Os documentos são armazenados, e quando necessário, são comprimidos internamente ao sistema, sendo retornados em sua forma original aos usuários;
- **Tópicos:** são os agrupamentos de documentos relacionados por assuntos específicos. Cada tópico do sistema possui um responsável por seu gerenciamento e podem ser organizados de forma hierárquica;
- **Categorias:** são os tipos de documentos aceitos, portanto podem-se definir quais tipos de documentos serão aceitos por cada categoria separadamente. O tamanho máximo que um documento pode ter também pode ser indicado por cada categoria separadamente;
- **Formatos:** Cada categoria do sistema aceita um ou mais formatos de arquivo. Podem ser definidos os formatos mais utilizados e que já são predefinidos pelo sistema e também podem ser criados grupos de formatos, contendo neles os formatos desejados e dando um nome que deseje e que torne fácil a identificação dos formatos aceitos.

¹ Ver <http://www.rau-tu.unicamp.br/nou-rau/>

Além do software básico necessário para o funcionamento do sistema, é preciso que sejam instalados conversores adicionais para indexar o conteúdo dos documentos para cada formato de arquivo, como PDF.

As buscas no Nou-rau são implementadas através da ferramenta Dig¹, que realiza a indexação dos documentos a partir das informações enviadas pelo sistema Nou-rau.

Atualmente são suportados pelo Nou-rau os seguintes formatos de arquivos: ASCII (*American Standard Code for Information Interchange*), HTML (*Hyper Text Markup Language*), RTF (*Rich Text Format*), SGML, WML, XML, documentos do Office Word, Excel e PowerPoint, PDF, PostScript, TeX, Látex e DVI.

Os usuários do Nou-rau podem assumir um ou mais papéis, dependendo dos seus objetivos no uso do repositório. Os papéis existentes são os seguintes:

- **Visitante:** acessa o repositório apenas com o objetivo de realizar consultas;
- **Colaborador:** possui a permissão para submeter documentos ao repositório, tornando-se então dono dos mesmos;
- **Responsável:** possui a função de administrar os tópicos do repositório e também de aprovar ou não as submissões realizadas a estes tópicos;
- **Administrador:** no nível hierárquico de usuário do Nou-rau é o usuário com papel de nível mais alto, sendo responsável pela manutenção do sistema, criação dos tópicos do repositório, manutenção das categorias e formatos de documentos e também a função de delegar funções aos outros usuários do sistema.

O processo de submissão em um sistema Nou-rau ocorre da seguinte forma: um *upload* é realizado por um colaborador, depois ocorre a avaliação pelo responsável do tópico ao qual o documento foi submetido e depois a possível verificação de vírus nestes documentos realizada pelo administrador. Caso um documento seja rejeitado em qualquer destas etapas, o mesmo é excluído da base de dados do sistema e um aviso é enviado ao colaborador que submeteu o documento.

As buscas podem ocorrer tanto nos metadados quanto no texto-integral (*full-text*) das obras que já foram indexadas. O sistema não permite realizar buscas apenas em um certo elemento dos metadados, como por exemplo realizar uma busca apenas por título ou autor.

¹ Ver <http://www.htdig.org/>

O Nou-rau possui uma interface semelhante a do DSpace, sendo também intuitiva e de fácil utilização, possuindo uma grande vantagem sobre as outras ferramentas: não é necessário o trabalho em cima de traduções, pois o sistema é desenvolvido por brasileiros.

A documentação disponível a respeito do sistema é muito pequena e percebe-se por pesquisas e acesso ao site do sistema que pouco tem sido realizado para melhorar as funcionalidades e muito menos no sentido de desenvolver novas versões que implementem as novas e importantes tecnologias existentes.

Para seu correto funcionamento, o Nou-rau necessita dos seguintes pré-requisitos de softwares:

- **Sistema Operacional:** UNIX-like.
- **Servidor Web:** Apache.
- **Banco de Dados:** PostgreSQL.
- **Adicionais:** PHP, FILE, HTDIG, PERL e conversores de formatos de arquivos para indexação *full-text*.

4.6 – Escolha do Sistema a ser usado para o Desenvolvimento do Repositório Digital Proposto

Após estudo dos softwares disponíveis, das teorias de metadados, interoperabilidade e dos repositórios digitais e também análise dos requisitos e necessidades do sistema proposto juntamente aos integrantes do Grupo de Pesquisas – Novas Tecnologias em Informação ficaram definidos os seguintes padrões e características necessárias no repositório proposto, e partindo destas definições ocorreu a definição de qual ferramenta utilizar como sistema para o repositório digital proposto.

O padrão de metadados adotado para descrição dos objetos digitais do repositório foi o Dublin Core e os motivos que levaram a esta escolha foram os seguintes:

- O nível de detalhamento dos metadados dos recursos submetidos ao repositório proposto não precisam ser excessivamente aprofundados;
- A utilização do padrão Dublin Core é simples e bastante intuitiva, o que torna fácil o entendimento dos profissionais das diversas áreas abrangidas pelo repositório, que estarão submetendo itens no repositório;
- O padrão Dublin Core é o padrão para as trocas de informações e capacidade de interoperabilidade previstas pela OAI (NSDL, 2005).

A adoção do padrão METS foi levado em consideração, mas apesar de o mesmo ser completo e considerado excelente pelos profissionais da informação, inclusive os profissionais pertencentes ao Grupo de Pesquisas – Novas Tecnologias em Informação, o mesmo é muito complexo para os usuários do repositório e portanto necessitariam de treinamento muito maiores do que para a utilização do Dublin Core.

Os *softwares* Nou-rau e o Greenstone foram eliminados devido ao fato de não atenderem a todos os requisitos obrigatórios (Tabela 3).

Além de o Greenstone não preencher tais requisitos, também fortaleceu a decisão de eliminá-los devido ao fato de a interface do administrador do sistema ser uma interface baseada em Java não podendo ser executada diretamente do *browser* e a submissão inicial dos documentos precisa ser realizada também pelo administrador do sistema, o que não seria possível de ser atendido no repositório proposto.

No Nou-rau existe também a dificuldade da escassez de documentação técnica e percebe-se que o desenvolvimento de atualizações e novas versões para melhoria do sistema não tem recebido a atenção necessária, pois sistemas *web* precisam estar em constante atualização e melhoria para que atendam as necessidades dos usuários. Outra dificuldade encontrada no Nou-rau é o fato de cada obra poder conter apenas um arquivo associado.

Portanto restaram apenas o DSpace e CDSWare como candidatos para adoção no desenvolvimento do repositório proposto, pois ambos atenderam todos os requisitos obrigatórios propostos. O CDSWare possui como grande vantagem seu mecanismo de busca muito poderoso e a existência de muitas opções de interação com o usuário do sistema.

O DSpace possui como grandes diferenciais o constante desenvolvimento e melhoria de seus sistemas, a utilização de identificadores persistentes, uma documentação extensa e uma comunidade de desenvolvedores e de instituições de ensino e empresas superior em quantidade a qualquer outro sistema analisado (Figura 1).

Uma desvantagem do DSpace sobre o CDSWare seria seu sistema de busca mais simples, mas que não é simples para os objetivos do repositório proposto e atenderia perfeitamente suas necessidades.

Os pontos fortes do DSpace, em principal o fato de ser o sistema de repositórios digital mais utilizado no mundo e de o mesmo possuir mais recursos que o CDSWare, pesaram muito na escolha e portanto o DSpace foi o *software* escolhido para a realização do desenvolvimento do repositório digital proposto.

CAPÍTULO 5 – IMPLANTAÇÃO E CUSTOMIZAÇÃO DO REPOSITÓRIO DIGITAL

Para desenvolvimento do repositório digital foi escolhida a plataforma DSpace (Capítulo 4). Foi utilizada a versão mais recente atualmente do *software*, a versão 1.5.2. A instalação e configuração do DSpace e o desenvolvimento do repositório digital foi realizada no sistema operacional Windows XP, com Service Pack 3.

Os componentes necessários para o correto funcionamento do repositório digital e sua respectivas versões usadas são citadas abaixo:

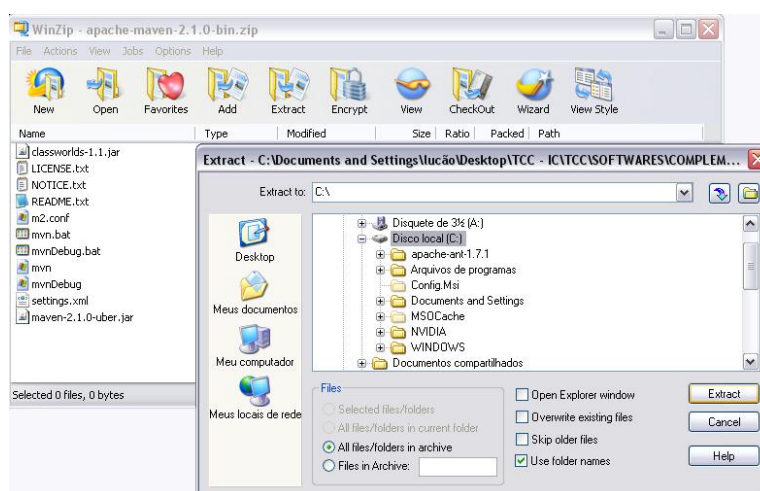
- **Servidor Web:** Apache Tomcat 6.0.18
- **Banco de Dados:** PostgreSQL 8.3.5.
- **Adicionais:** Apache ANT 1.7.1, Apache Maven 2.1.0 e Java JDK 6u12.

5.1 – Instalação dos Componentes e do DSpace

Primeiramente foi instalado o Java JDK 6u12 com todas as configurações padrões fornecidas pelo instalador.

Depois extraiu-se o conteúdo dos arquivos zipados “apache-ant-1.7.1-bin” e “apache-maven-2.1.0-bin” para o disco local (C:), conforme mostrado na Figura 5

Figura 5 – Extraíndo o arquivo “apache-maven-2.1.0-bin” para o disco local

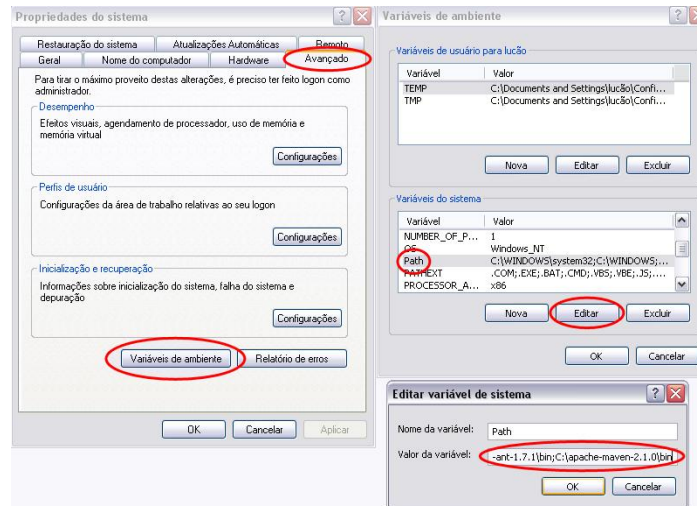


Fonte: Próprio autor.

Após a instalação do Java JDK e da extração do Apache Maven e do Apache Ant foi preciso configurar as variáveis de ambiente do computador, e para isto foi selecionada a opção “Propriedades” do item Meu computador. Na tela aberta após ter sido clicado sobre a

opção Propriedades clicou-se sobre a aba “Avançado” e então clicou-se sobre o botão “Variáveis de ambiente”. Já na caixa das variáveis de ambiente foi selecionada a variável de sistema chamada “Path” e então clicado na opção editar. Então adicionou-se ao campo “Valor da variável” os caminhos “C:\Arquivos de programas\Java\jdk1.6.0_12\bin” , “C:\apache-ant-1.7.1\bin” e “C:\apache-maven-2.1.0\bin”, todos seguidos por ponto e vírgula (;) e sendo o primeiro destes após o último já existente no campo, conforme mostrado na Figura 6.

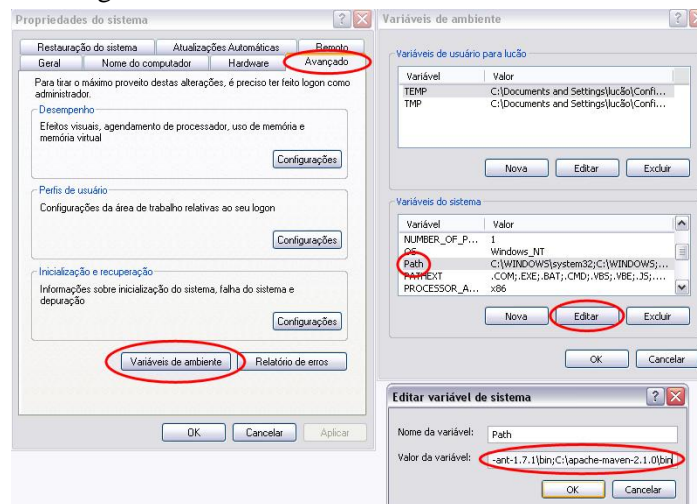
Figura 6 – Alterando as variáveis de ambiente.



Fonte: Próprio autor.

Após adicionados os três caminhos foi necessário criar duas variáveis do sistemas, a JAVA_HOME e a ANT_HOME. Para isso na mesma caixa das variáveis de ambientes, parte inferior, relativas as variáveis do sistema clicou-se em “Nova”, então foi preenchido o campo “Nome da variável” com JAVA_HOME e o campo “Valor da variável” com o caminho “C:\Arquivos de programas\Java\jdk1.6.0_12” e outra com o nome de ANT_HOME e o caminho “C:\apache-ant-1.7.1”, conforme mostrado na Figura 7.

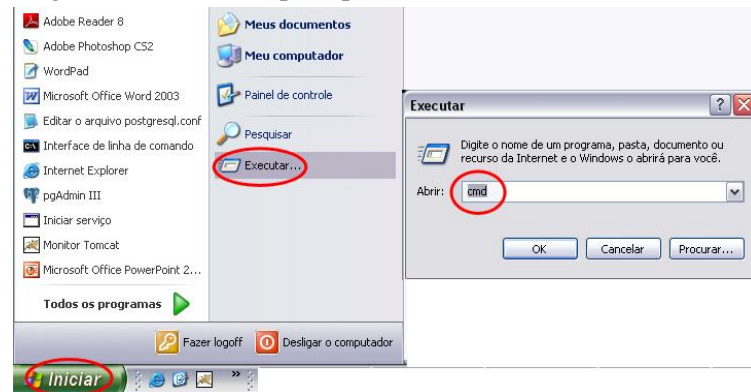
Figura 7 – Criando as novas variáveis do sistema



Fonte: Próprio autor.

Após criadas as novas variáveis de ambiente realizou-se a checagem das variáveis “Path”, “JAVA_HOME” e “ANT_HOME”. Para isso foi preciso abrir o prompt de comando do Windows, digitando “cmd” no executar, presente no “Iniciar”, conforme mostrado na Figura 8. Dentro do prompt de comando digitou-se os comandos “**mvn -version**” e também “**ant -version**”, e os resultados esperados e ocorridos são demonstrados na Figura 9.

Figura 8 – Abrindo o prompt de comando do Windows (cmd)



Fonte: Próprio autor.

Figura 9 – Resultados esperados na checagem das variáveis de ambiente.

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows [versão 5.1.2600.1
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\Lucão>cd..
C:\Documents and Settings>cd..
C:\>mvn -version
Apache Maven 2.1.0 (r755702; 2009-03-18 16:10:27-0300)
Java version: 1.6.0_12
Java home: C:\Arquivos de programas\Java\jdk1.6.0_12\jre
Default locale: pt_BR, platform encoding: Cp1252
OS name: "windows xp", version: "5.1" arch: "x86" Family: "windows"
C:\>ant -version
Apache Ant version 1.7.1 compiled on June 27 2008
C:\>

```

Fonte: Próprio autor.

Os resultados exibidos na Figura 9 significam que as variáveis de ambientes foram configuradas corretamente.

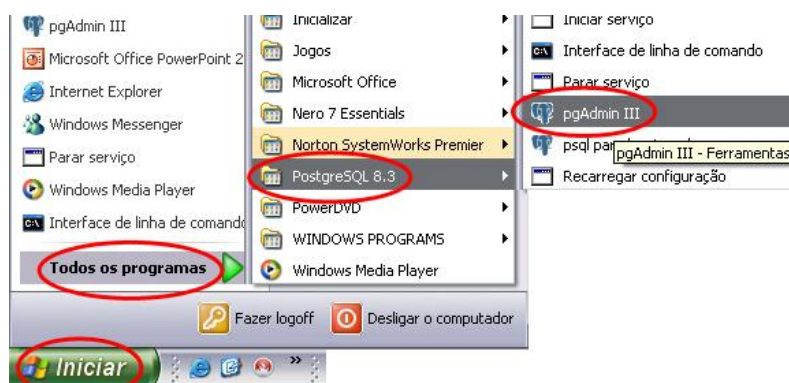
O próximo passo foi a instalação do sistema de banco de dados PostgreSQL, versão 8.3.5. Para isto foi executado o arquivo “postgresql-8.3” e então o instalador foi iniciado. Primeiramente foi selecionada a língua de preferência e clicado no botão “Start”. Na próxima tela do instalador foi avisado sobre a recomendação de se fechar todos os programas antes de continuar. Foram todos fechados e então clicou-se no botão “Próximo”. Na tela seguinte foi clicado novamente no botão “Próximo”. Na tela que veio a seguir foi configurado quais componentes desejava-se instalar. Foram selecionadas as opções que eram consideradas úteis na utilização do banco de dados, mas não foi retirado nenhum dos componentes padrões da

instalação. Na realização deste passo, foi selecionado a instalação dos componentes de suporte para idioma nacional e PL/Java.

Na tela seguinte foi preciso informar a senha do serviço do banco de dados, a qual foi definida como “admin” pelo autor. Ao clicar no botão “Próximo” uma caixa de dialogo surgiu dizendo que o usuário não foi encontrado e perguntando se desejava-se que o mesmo fosse criado. Foi escolhida a opção “Sim” e então uma nova caixa de diálogo surgiu dizendo que a senha informada (admin) parecia fraca e se desejava-se que fosse gerada uma senha aleatória. Foi selecionado a opção “Não”. Na tela que segue foi informada a senha do usuário, definida como “postgres” pelo autor e alterada a codificação para “UTF-8”. Após realizar as alterações clicou-se em “Próximo” e nas duas telas seguintes apenas clicou-se em “Próximo” também. Na tela seguinte foi mostrado diversos recursos que podem ser instalados, mas para o funcionamento do DSpace apenas é necessário deixar a opção default selecionada (Adminpack) e então foi clicado em “Próximo” e a instalação foi iniciada e se concluiu automaticamente, e após a mesma apenas se clicou em “Concluir” na última tela do instalador.

Após instalado o sistema de banco de dados foi aberto pgAdmin III do postgresQL, conforme mostrado na Figura 10, com o pgAdmin III aberto conectou-se ao banco de dados. Para isso foi clicado sobre “PostgreSQL Database Server 8.3 (localhost:5432)” e selecionada a opção conectar, conforme mostrado na Figura 11.

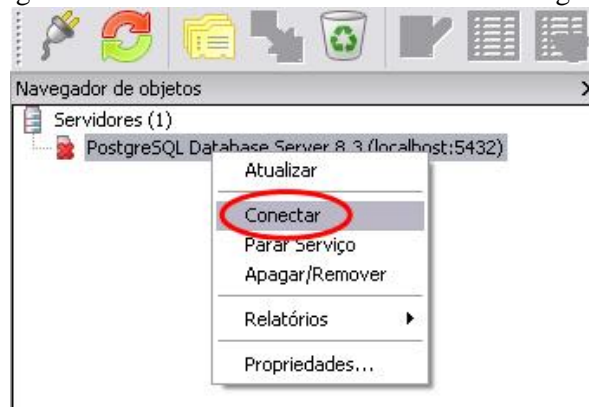
Figura 10 – Abrindo o pgAdmin III do postgresQL



Fonte: Próprio autor.

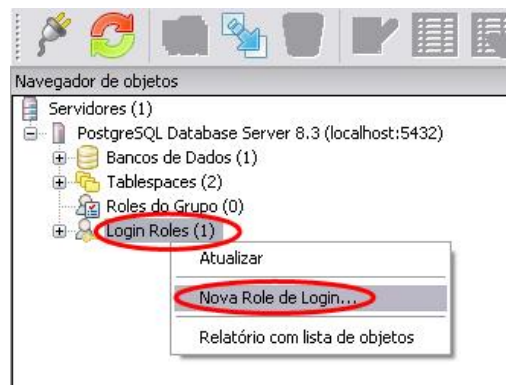
Após conectado ao banco de dados foi então criado um novo o “Login Role”, chamado de dspace e com a senha “dspace”. Para isso clicou-se com o botão direito sobre “Login Roles” e então selecionou-se a opção “Nova Role de Login...”, conforme mostrado na Figura 12 e então foi configurada a nova “Login Role”, conforme mostrado na Figura 13.

Figura 11 – Conectando ao banco de dados PostgreSQL



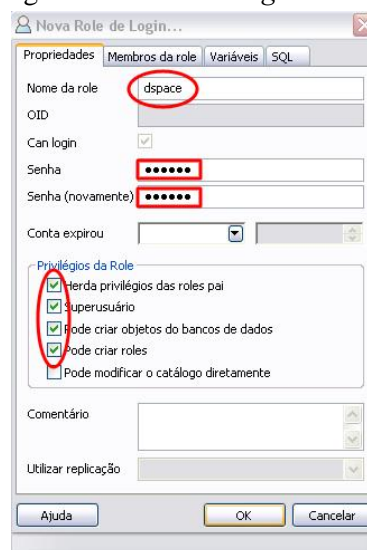
Fonte: Próprio autor.

Figura 12 – Criando um novo "Login Role" no PostgreSQL



Fonte: Próprio autor.

Figura 13 – Configurando o novo "Login Role" chamado dspace

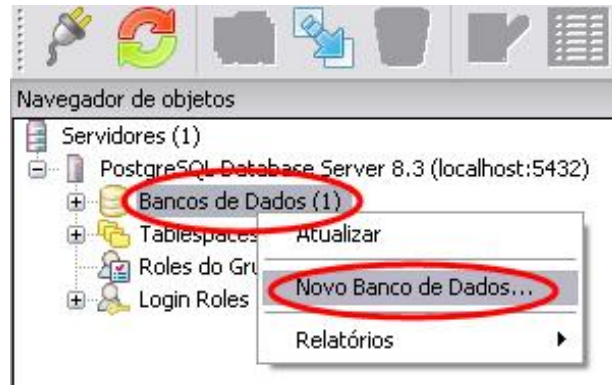


Fonte: Próprio autor.

Após criado o novo "Login Role" foi criado um novo banco de dados chamado dspace, que teve como dono o dspace e codificação UTF-8. Para criar o novo banco de dados

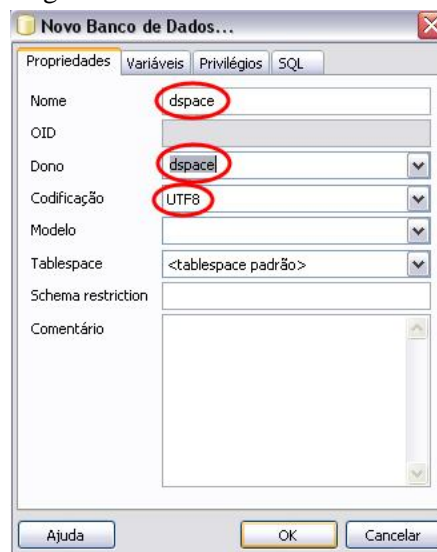
foi clicado com o botão direito sobre “Bancos de Dados” e selecionada a opção “Novo Banco de Dados...” conforme mostrado na Figura 14, e é configurado o mesmo conforme mostrado na Figura 15.

Figura 14 – Criando um novo Banco de Dados



Fonte: Próprio autor.

Figura 15 – Configurando o novo Banco de Dados chamado dspace



Fonte: Próprio autor.

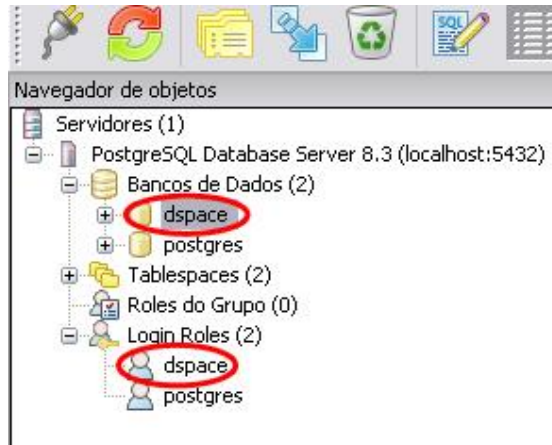
Após criado também o novo banco de dados, expandiu-se os itens “Banco de Dados” e “Login Roles” do pgAdmin III e então conferiu-se que o novo *login role* e o novo banco de dados foram criados corretamente, conforme mostrado na Figura 16.

O próximo passo foi a instalação do servidor *web* Apache Tomcat. Para isso foi executado o instalador “apache-tomcat-6.0.18” e seguidas todas as opções padrões da instalação, informando a senha “admin” na tela que foi pedido uma senha para o servidor *web*.

Após concluída a instalação do Apache Tomcat, o monitor do Tomcat passou a ser mostrado na área de notificação, na lateral inferior direita da tela, logo ao lado do relógio.

Neste ícone pode se parar o serviço do Tomcat e também iniciá-lo, conforme demonstrado na Figura 17.

Figura 16 – Conferindo o novo *login role* e o novo banco de dados criados



Fonte: Próprio autor.

Figura 17 - Ícone na área de notificação e opção de parar o serviço do servidor *web* Tomcat
(Stop Service)



Fonte: Próprio autor.

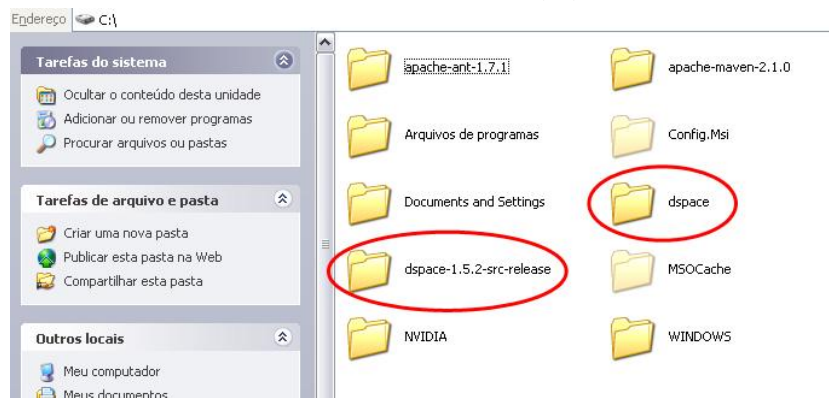
Após instalados todos os componentes foi extraído o conteúdo do arquivo zipado “dspace-1.5.2-src-release” para o disco local (C:), de modo que foi criada uma pasta no mesmo chamada de “dspace-1.5.2-src-release”.

Em seguida foi criada uma pasta chamada “dspace” na raiz do disco local (C:), então o disco local passou a mostrar as pastas do DSpace conforme mostrado na Figura 18.

Após criada a nova pasta foi iniciado o serviço do PostgreSQL, clicando sobre o menu “Iniciar”, depois sobre “Todos os programas”, depois “PostgreSQL 8.3” e finalmente sobre o atalho “Iniciar Serviço”, e então o serviço do banco de dados PostgreSQL se iniciou.

Antes de continuar foi necessário configurar o arquivo “dspace.cfg”, que encontra-se no diretório “C:\dspace-1.5.2-src-release\dspace\config”. Neste arquivo foi alterada primeiramente a linha “dspace.dir = /dspace” para “dspace.dir = C:/dspace”.

Figura 18 – Pasta “dspace-1.5.2-src-release” extraída para disco local e pasta “dspace” criada pelo usuário no disco local (C:)



Fonte: Próprio autor.

Aproveitando a configuração do “dspace.cfg” foi alterado o idioma padrão do repositório, que no repositório proposto foi definido como português do Brasil e também os idiomas que serão aceitos, que no caso do repositório proposto foram o português do Brasil e o inglês. Para isso foram realizadas as seguintes alterações no arquivo “dspace.cfg”:

- “default.language = en_US” para “default.language = pt_BR”;
- “default.locale = en” para “default.locale = pt_BR”;
- e adicionada a linha “webui.supported.locales = pt_BR” no arquivo.

Também foi alterado o nome do repositório digital, alterando a linha com o comando “dspace.name = DSpace at My University” para “dspace.name = Repositório Digital do Grupo de Pesquisas Novas Tecnologias em Informação”.

Após configurado o “dspace.cfg” foi gerado então o pacote de instalação do DSpace. Para isso devia-se estar conectado a internet e então abrir o prompt de comando do Windows (Figura 8), e executar o comando “**mvn package**” de dentro do diretório “C:\dspace-1.5.2-src-release\dspace”, conforme mostrado na Figura 19. Então muitos pacotes começaram a ser baixados e o pacote de instalação do DSpace foi criado dentro do caminho “C:\dspace-1.5.2-src-release\dspace\target\dspace-1.5.2-build.dir”. Depois de concluída a montagem do pacote de instalação do DSpace, irá mostrar no prompt de comando a mensagem “BUILD SUCCESSFUL”, conforme mostrado na Figura 20.

Esta mensagem significa que todos os pacotes necessários para a instalação do DSpace foram “baixadas” para o computador que executou o comando “mvn package” e então pode se dar início a instalação do DSpace.

Figura 19 – Criando do pacote de instalação do DSpace

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows XP [versão 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\lucão>cd C:\
C:\>cd dspace-1.5.2-src-release
C:\dspace-1.5.2-src-release>cd dspace
C:\dspace-1.5.2-src-release\dspace>mvn package
  
```

Fonte: Próprio autor.

Figura 20 – Criação do pacote de instalação concluída com sucesso

```

[INFO]
[INFO]
[INFO] BUILD SUCCESSFUL
[INFO]
[INFO] Total time: 59 minutes 19 seconds
[INFO] Finished at: Sat Nov 07 17:39:59 BRST 2009
[INFO] Final Memory: 32M/63M
[INFO]
C:\dspace-1.5.2-src-release\dspace>
  
```

Fonte: Próprio autor.

Antes de iniciar a instalação do DSpace foi preciso conectar-se ao banco de dados PostgreSQL, através do pgAdmin III, conforme demonstrado nas Figuras 10 e 11.

Após conectado ao banco de dados foi possível então instalar o DSpace para a pasta “dspace”, no diretório “C:\dspace”, executando a partir do prompt de comando do Windows (veja Figura 8) o comando “**ant fresh_install**” de dentro do diretório “C:\dspace-1.5.2-src-release\dspace\target\dspace-1.5.2-build.dir”, conforme mostrado na Figura 21, comando que fez com que a instalação do DSpace ocorresse para dentro da diretório “C:\dspace”. A mensagem “BUILD SUCCESSFUL” foi novamente mostrada, mas desta vez indicando a conclusão da instalação do DSpace, conforme demonstrado na Figura 22.

Figura 21 – Realizando a instalação do DSpace

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows XP [versão 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\lucão>cd C:\
C:\>cd dspace-1.5.2-src-release
C:\dspace-1.5.2-src-release>cd dspace
C:\dspace-1.5.2-src-release\dspace>cd target
C:\dspace-1.5.2-src-release\dspace\target>cd dspace-1.5.2-build.dir
C:\dspace-1.5.2-src-release\dspace\target\dspace-1.5.2-build.dir>ant fresh_install
  
```

Fonte: Próprio autor.

Figura 22 – Instalação do DSpace concluída com sucesso

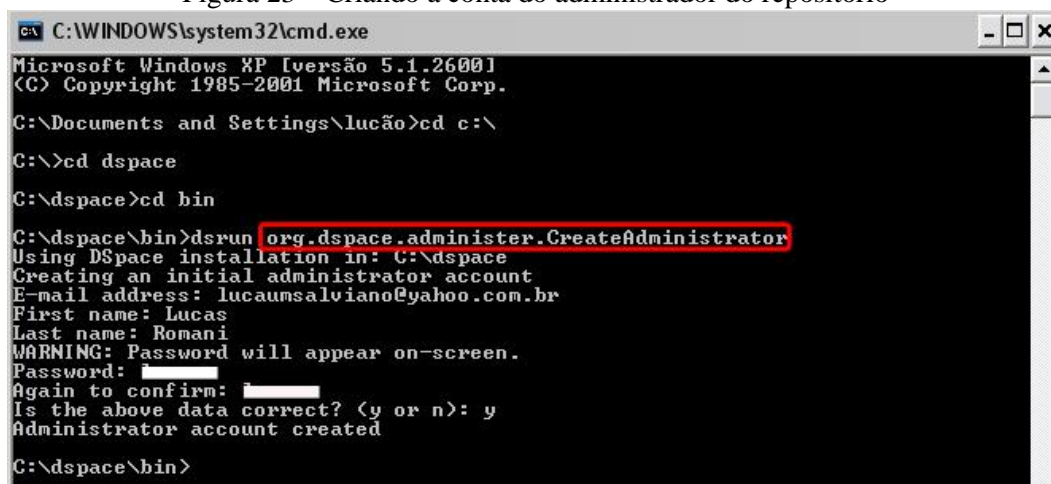
```

BUILD SUCCESSFUL
Total time: 38 seconds
C:\dspace-1.5.2-src-release\dspace\target\dspace-1.5.2-build.dir>
  
```

Fonte: Próprio autor.

Após instalado o DSpace foi criada a conta do administrador. Para isso foi novamente aberto o prompt de comando do Windows (Figura 8) e então aberto o diretório “bin” do DSpace (C:\dspace\bin). Então foi utilizado o comando “**dsrun org.dspace.administer.CreateAdministrator**” para dar início ao processo de criação da conta de administrador do DSpace. Neste processo o autor, que neste caso é o administrador, informou primeiramente o e-mail do administrador (*e-mail address*), depois o primeiro nome (*First name*), depois o último nome (*Last name*) e então informou a senha (*Password*), então foi novamente inserida a senha para confirmar e o processo perguntou se as informações estavam corretas, e como estavam foi digitado y (*yes - sim*) e apertado *Enter* e a conta do administrador foi criada. Caso algum dado digitado estivesse incorreto deveria ter sido digitado n (*no - não*) e recomeçado. Todo este processo encontra-se na Figura 23.

Figura 23 – Criando a conta do administrador do repositório



```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows XP [versão 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.
C:\Documents and Settings\lucão>cd c:\
C:\>cd dspace
C:\dspace>cd bin
C:\dspace\bin>dsrun org.dspace.administer.CreateAdministrator
Using DSpace installation in: C:\dspace
Creating an initial administrator account
E-mail address: lucaumsalviano@yahoo.com.br
First name: Lucas
Last name: Romani
WARNING: Password will appear on-screen.
Password: ██████████
Again to confirm: ██████████
Is the above data correct? (y or n): y
Administrator account created
C:\dspace\bin>

```

Fonte: Próprio autor.

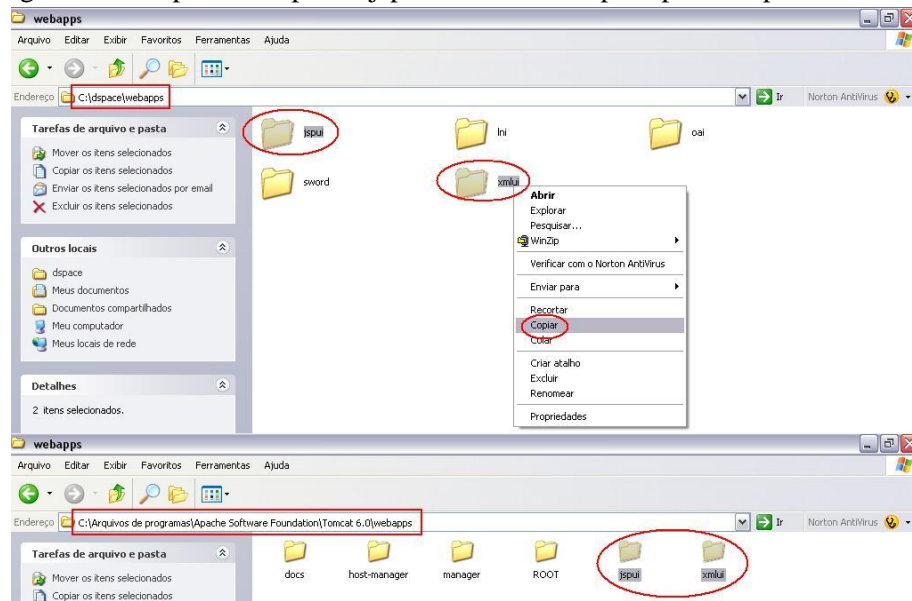
O próximo passo foi copiar as pastas jspui e xmlui, que se localizam no diretório “C:\dspace\webapps” para o diretório “C:\Arquivos de programas\Apache Software Foundation\Tomcat 6.0\webapps”, conforme mostrado na Figura 24.

Esta foi a última etapa para que então pudesse-se rodar o DSpace. Para isto foi reiniciado o Tomcat e abriu-se o browser do computador, que no caso foi usado o Internet Explorer, e então acessou-se o endereço <http://localhost:8080/jspui>, que abriu a página inicial do repositório digital. Também pode ser aberto o endereço <http://localhost:8080/xmlui>, que abriu o sistema Manakin.

Acessando o Dspace localmente no endereço <http://localhost:8080/jspui> obteve-se o primeiro contato como o repositório, que no caso das configurações descritas no processo de instalação do DSpace é o mostrado a Figura 25.

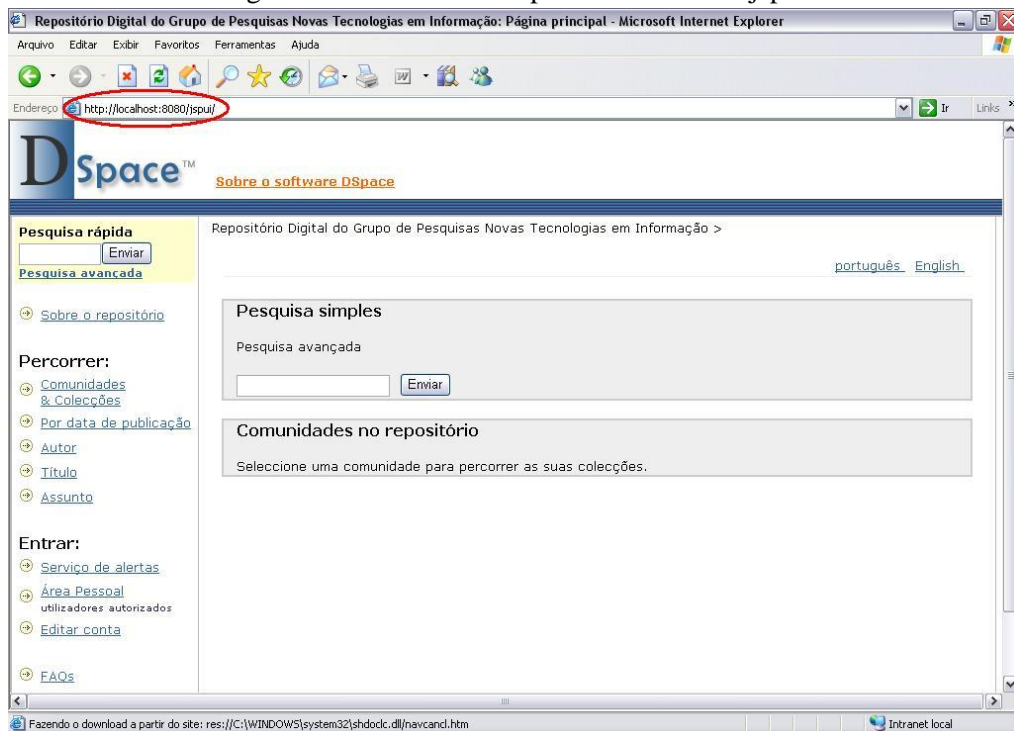
Acessando a partir do endereço <http://localhost:8080/xmlui> obteve-se o resultado demonstrado na Figura 26, que é a interface do Manakin dos repositórios digitais criados pelo DSpace, interface mais recente nas implementações do DSpace e usada para customizações visuais mais ricas e avançadas.

Figura 24 – Copiando as pastas jspui e xmlui do Dspace para o Apache Tomcat.



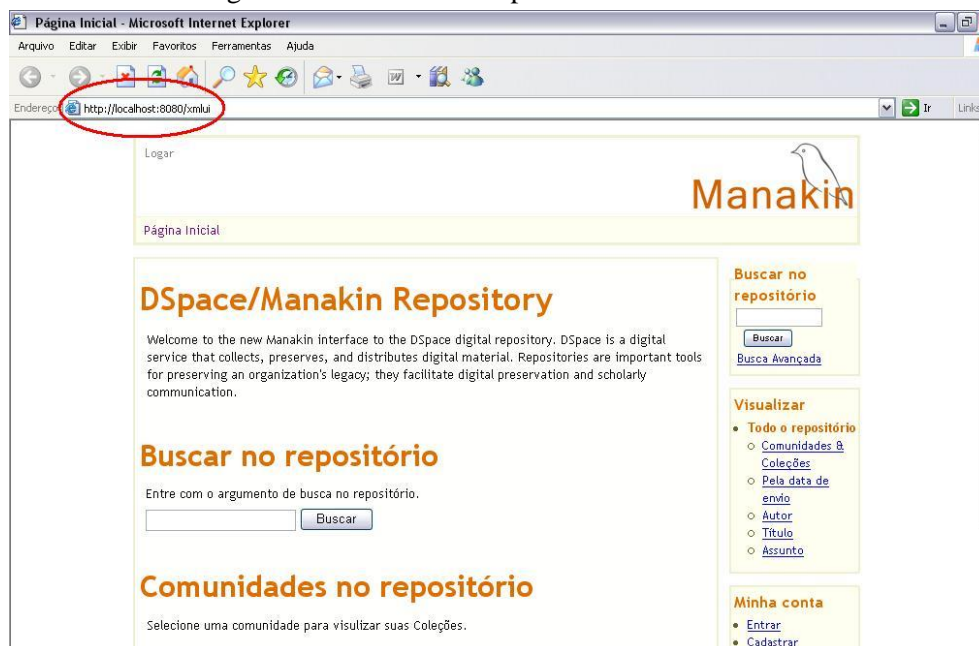
Fonte: Próprio autor.

Figura 25 – Acessando <http://localhost:8080/jspui>



Fonte: Próprio autor.

Figura 26 – Acessando http://localhost:8080/xmlui



Fonte: Próprio autor.

5.2– Customização do Repositório Digital

Neste tópico são apresentadas algumas das configurações realizadas no repositório a fim de customizá-lo para atender as necessidades do Grupo de Pesquisa – Novas Tecnologias em Informação, grupo para qual o repositório digital foi desenvolvido.

O primeiro procedimento foi a troca do banner original do DSpace pelo banner do repositório digital proposto. Para isto primeiramente foi criado pelo autor o referido banner e o mesmo foi copiado para o diretório “C:\Arquivos de programas\Apache Software Foundation\Tomcat 6.0\webapps\jspui\image” com o nome de “BannerGPNTI01.jpg”.

É válido saber que as páginas do DSpace são divididas em cinco partes, que são elas: Header, que é o cabeçalho da página; Footer, que é a parte inferior da página; Navigation Bar, que é a parte esquerda da página; Sidebar, que é a parte direita da página; e a parte central do repositório, que é chamada de Page Contents. A Figura 27 mostra estas divisões de maneira facilitada.

Após copiado o novo *banner* para o diretório adequado, o arquivo “header-default.jsp”, localizado no diretório “C:\Arquivos de programas\Apache Software Foundation\Tomcat 6.0\webapps\jspui\layout”, diretório no qual os arquivos que tratam dos *layouts* das páginas do DSpace se encontram, foi alterado. Dentro do arquivo o nome do arquivo que é referente ao *banner* do repositório foi alterado de “dspace-blue.gif”, que era o

arquivo do *banner* original do Dspace, para “BannerGPNTI01.jpg”, e a largura e altura também foram atualizadas com as configurações da nova imagem (400x80) e a partir de então o novo *banner* começou a ser mostrado no lugar do original.

Figura 27 – Divisão de uma página do DSpace.



Fonte: Próprio autor.

Foi decidido retirar o link “Sobre o software DSpace” do cabeçalho da página e para isso o mesmo arquivo “header-default.jsp” da alteração anterior foi modificado. Para que o *link* não fosse mais mostrado duas linhas tiveram de ser apagadas do arquivo. A Figura 28 mostra o novo *banner* e também qual link esta sendo tratado aqui e Figura 29 mostra quais as linhas que foram eliminadas do arquivo “header-default.jsp” para que o *link* não aparecesse mais.

Figura 28 – Novo *banner* e o link “Sobre o software DSpace”



Fonte: Próprio autor.

Figura 29 – Linhas a serem eliminadas do arquivo “header-default.jsp” para que o link “Sobre o software DSpace” deixar de ser mostrado

```
<!-- DSpace logo -->
<tr>
  <td>
    <a href="<%= request.getContextPath() %>/"> <!-- Make as wide as possible. cellpadding repeated for broken NS 4.x -->
    <a class="tagLineText" target="_blank" href="http://www.dspace.org/"><fmt:message key="jsp.layout.header-default.abov
  </td>
  <td nowrap="nowrap" valign="middle">
</td>
</tr>
```

Fonte: Próprio autor.

Para realizar alterações de cores de fontes, cores de fundo e propriedades referentes a estética da página foi alterado o arquivo “styles.css”, que fica localizado no diretório “C:\Arquivos de programas\Apache Software Foundation\Tomcat 6.0\webapps\jspui”.

Abaixo seguem alguns dos comandos alterados neste arquivo, mas não todos, devido ao fato de serem muitos. Apenas tais alterações já são suficientes para se entender o que foi realizado para que as alterações ocorressem.

Foi decidido alterar a cor da barra lateral esquerda (*Navigation Bar*) para o amarelo com código de cor #ffffcc. Para isto os seguintes comandos foram alterados.

No item “.navigationBar”, que contém as configurações da Navigation Bar em si, o background foi alterado para #ffffcc. No item “.navigationBarSublabel”, que possui as configurações dos títulos da Navigation Bar, que são “Per correr:” e “Entrar:” o background também foi alterado para #ffffcc e no item “.navigationBarItem” que altera as configurações dos itens mostrados na *Navigation Bar* também foi alterado o background para #ffffcc.

Para alterar a cor de fundo da parte central do DSpace deve-se alterar o background do item “.pageContents”, trocando assim a cor de fundo do Page Content, que como mencionado anteriormente é a parte central da página. Na customização do repositório digital proposto o plano de fundo central manteve-se o mesmo, ou seja, branco.

No item “.evenRowEvenCol” o background foi alterado para #ffff99, fazendo com que a cor de fundo dos quadros centrais da Page Content ficasse amarelo, só que numa tonalidade mais forte que a da Navigation Bar. A Figura 30 mostra esta alteração, mostrando para isso o trecho de código original do arquivo e depois alterado.

Figura 30 – Alterando a cor de fundo dos quadros centrais da Page Content

ORIGINAL

```
.oddRowEvenCol{ font-family: "verdana", "Arial", "Helvetica", sans-serif;
font-size: 12pt;
font-style: normal;
font-weight: normal;
color: #000000;
vertical-align: middle;
text-decoration: none;
background: #e0e0e0;
<%= padding %> }
```

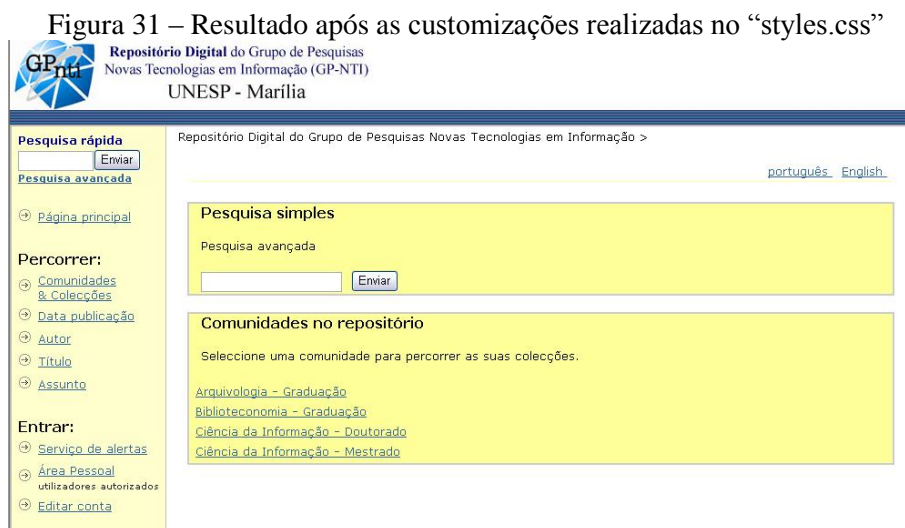
ALTERADO

```
.oddRowEvenCol{ font-family: "verdana", "Arial", "Helvetica", sans-serif;
font-size: 12pt;
font-style: normal;
font-weight: normal;
color: #000000;
vertical-align: middle;
text-decoration: none;
background: #ffff99;
<%= padding %> }
```

Fonte: Próprio autor.

Os resultados destas alterações podem ser visualizados na Figura 31.

A caixa que indica qual usuário esta logado no sistema também foi alterada, ficando com o fundo (background) do item “.loggedIn “ na cor laranja (código #ff9900), mantendo a fonte com cor preta e alterando o alinhamento do texto para centralizado, adicionando a linha “text-align: center;” dentro do item.



Fonte: Próprio autor.

Alterações nas cores e tamanhos das fontes apresentadas no repositório, cores de fundo, famílias de fontes aceitas, configurações de margens e bordas de tabelas, alinhamentos, estilos das fontes e demais configurações referentes a customização da estética do repositório devem ser realizadas neste arquivo, o “styles.css”, que como dito anteriormente encontra-se no diretório “C:\Arquivos de programas\Apache Software Foundation\Tomcat 6.0\webapps\jspui”.

Todas as modificações e customizações realizadas no repositório digital proposto foram testadas nos três navegadores mais usados atualmente, o Internet Explorer, na versão 8, o Mozilla Firefox, na versão 3.5.5 e o Google Chrome, na versão 3.0.195.33, e em todos os navegadores os resultados foram considerados satisfatórios, sem fornecer perca alguma das características e funcionalidades do repositório proposto.

5.3 – Customização do Repositório Digital a partir da interface de administrador.

Para realização de alterações no repositório digital a partir da interface de administrador, o usuário deve estar logado no sistema com a conta de administrador, configurada anteriormente (Figura 23). Logado como administrador foi acessada a área de Administrador, mostrada na Figura 32, e a partir desta interface podem ser realizadas ações como: criação das comunidades, sub-comunidades e coleções do repositório, cadastro de novos utilizadores do sistema, criação de novos grupos de usuários do sistema, dividindo os com relação as suas funções, alteração de registros (Handle), alteração dos metadados padrão do Dublin Core, alteração dos formatos de *bitstreams* conhecidos, acessar os processos de *workflow* pendentes, administrar as políticas de permissões do repositório, editar as notícias, tanto as laterais quanto as superiores, editar o arquivo de licença padrão do repositório, alterar as ordens de administração dos supervisores e visualizar as estatísticas, quando estas estiverem disponíveis no repositório.

Figura 32 – Link na *Navigation Bar* para acessar a interface de administrador e a própria interface de administrador



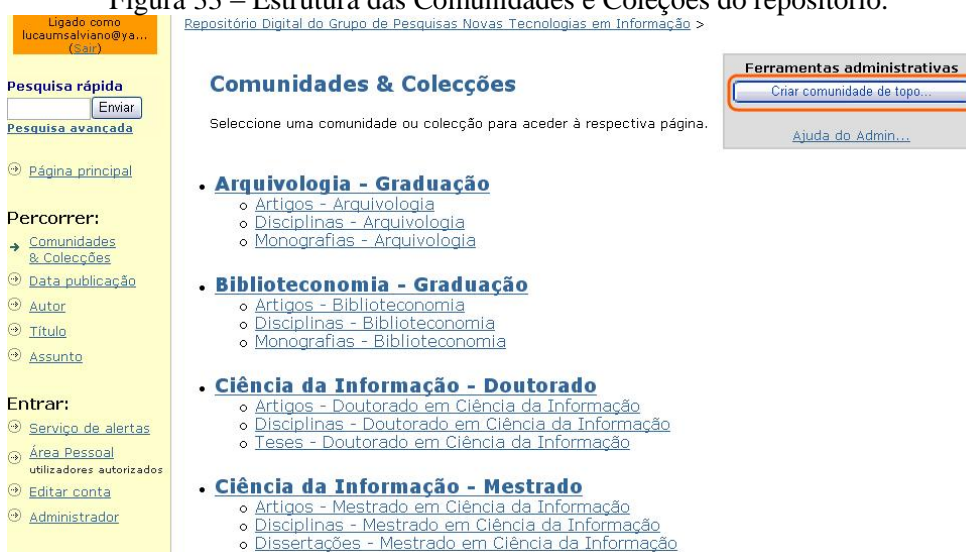
Fonte: Próprio autor.

A partir dos requisitos apresentados pelos membros do Grupo de Pesquisas – Novas Tecnologias em Informação foram criados quatro comunidades iniciais para o repositório proposto e para cada uma delas foram criadas algumas coleções ditas como necessárias pelos membros. A Figura 33 mostra o botão que leva a tela de criação de novas comunidades e também a hierarquia que se formou no repositório após a criação das comunidades e de suas respectivas coleções.

Também foram alteradas as notícias, através da interface do administrador do DSpace (Figura 32), na opção “Editar notícias”. Notícias não são nada mais do que os textos apresentados na parte lateral direita do repositório, que são as notícias da barra lateral e também na parte superior da *Page Contents*, que é a parte central da página do repositório, que são as notícias principais. As notícias laterais são normalmente usadas para apresentar novas informações a respeito do repositório e de seus usuários, mas pode-se apresentar qualquer texto desejado.

As notícias principais foram usadas para apresentar informações sobre o próprio grupo de pesquisa para qual o repositório foi criado.

Figura 33 – Estrutura das Comunidades e Coleções do repositório.



Fonte: Próprio autor.

5.4 – Tradução dos componentes do DSpace

Este último capítulo trata de uma configuração de extrema importância realizada no repositório, a tradução de seus componentes. Este passo é de extrema importância, visto que os textos padrões do DSpace são em língua inglesa.

Estes componentes são os seguintes: o arquivo “input-forms.xml”, localizado no diretório “C:\dspace-1.5.2-src-release\dspace\config”, que trata da apresentação dos metadados Dublin Core para o usuário do repositório, que são mostrados na Figura 34 já traduzidos e também os doze arquivos localizados no diretório “C:\dspace-1.5.2-src-release\dspace\config\emails”, que apresentam os textos enviados por e-mail aos usuários cadastrados no repositório.

Após a atualização destes arquivos no idioma traduzido, foram reconstruídos os pacotes do DSpace, utilizando o comando **“mvn package”** de dentro do diretório **“C:\dSPACE-1.5.2-src-release\dSPACE”** utilizando o prompt de comando do Windows (Figura 19) e então acessando o diretório **“C:\dSPACE-1.5.2-src-release\dSPACE\target\dSPACE-1.5.2-build.dir”**, também a partir do prompt de comando do Windows, e utilizado o comando **“ant update”**, que ao invés do comando **“ant fresh_install”** utilizado na instalação do DSpace, que criou os arquivos no diretórios **“C:\dSPACE”**, este comando apenas atualizou os necessários. Tal procedimento é demonstrado na Figura 35.

Figura 34 – Elementos do **“input-forms.xml”** traduzidos para o português.

Depositar: Descreva o seu registo

Por favor, introduza a informação necessária acerca do seu Depósito em baixo. Na maioria dos browsers pode usar a tecla TAB para navegar no formulário. [\(Mais ajuda...\)](#)

Entre com o(s) nome(s) do(s) autor(es) do Item, no campo abaixo.

Último nome ex. *Silva* Primeiro Nome ex. *Manuel*

Autor(es)

Entre com o título principal do Item.

Título

Entre com a série ou número designado ao Item pela comunidade.

Nome da série Relatório ou Paper N.

Séries/No. de relatório.

Caso o Item possua algum número de identificação ou código associado com ele, por favor, entre com esse identificador, abaixo.

Identificadores ISSN

Selecione o(s) tipo(s) de conteúdo do Item. Para seleccionar mais que um valor na lista, pode-se manter pressionado a tecla control - "CTRL" ou "Shift" e assim seleccionar ou marcar uma faixa.

Tipo

Fonte: Próprio autor.

Figura 35 – Atualizando os arquivos do diretório **“C:\dSPACE”**

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows XP [versão 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\lucão>cd..
C:\Documents and Settings>cd..
C:\>cd dSPACE-1.5.2-src-release
C:\dSPACE-1.5.2-src-release>cd dSPACE
C:\dSPACE-1.5.2-src-release\dSPACE>cd target
C:\dSPACE-1.5.2-src-release\dSPACE\target>cd dSPACE-1.5.2-build.dir
C:\dSPACE-1.5.2-src-release\dSPACE\target\dSPACE-1.5.2-build.dir>ant update
  
```

Fonte: Próprio autor.

Para finalizar foi alterado o número máximo de respostas a uma requisição ao protocolo OAI-PMH do repositório, com o objetivo de diminuir o tamanho máximo desta resposta, que com o valor padrão (100) seria muito grande, então o mesmo foi alterado para trinta (30). Para realizar tal alteração foi necessário alterar o conteúdo do arquivo Java **“DSpaceOAI Catalog”**, localizado no diretório **“C:\dSPACE-1.5.2-src-release\dSPACE-oai\dSPACE-oai-api\src\main\java\org\dSPACE\app\oai”**. O comando **“private final int**

MAX_RECORDS = 30;” foi adicionado ao arquivo para que esta alteração ocorresse. Novamente o DSpace precisou ter seus pacotes reconstruídos (Figura 19) e então atualizados (Figura 34).

Após finalizadas as customizações necessárias, que foram definidas a partir das necessidades e políticas do grupo ao qual o repositório é proposto, o repositório digital pode ser então hospedado em um servidor na internet para que possa então passar a ser utilizado pelos interessados e passar também a gerar o conhecimento explícito que o GP-NTI mostrou ter necessidade, pois a partir do desenvolvimento e hospedagem deste repositório, o grupo passará a ter controle de seus documentos desenvolvidos e poderá conforme o uso e disseminação deste novo sistema, passar a ter maior visibilidade de suas produções, podendo assim aumentar seu prestígio e renome perante outros grupos de pesquisadores, e também perante outras instituições.

CONCLUSÃO

Considerando-se a pesquisa realizada durante o desenvolvimento deste trabalho e o desenvolvimento do repositório digital proposto para atender as necessidades de armazenamento e disseminação das produções científicas geradas pelo Grupo de Pesquisa - Novas Tecnologias em Informação, pode-se dizer que todos os objetivos pré-definidos foram atendidos.

O estudo detalhado dos usos e modos de desenvolvimento de repositórios digitais é um assunto pouco abordado, principalmente tratando-se de texto em língua portuguesa. Mesmo tais sistemas sendo eficientes no objetivo de conservarem, gerirem e disseminarem as produções científicas e documentos gerados por instituições de fins diversos ocorre tal falta de literatura.

O fato de existir literatura um tanto escassa a respeito dos repositórios digitais enfatiza a importância e relevância deste trabalho, uma vez que o mesmo cumpre com os objetivos pré-definidos de estudar os repositórios digitais de forma a conhecer seus pontos positivos e negativos, levando em conta sempre o uso em qualquer tipo de instituição que produza documentos digitais. Também torna-se relevante devido ao fato de a documentação específica da ferramenta DSpace, que como citado no conteúdo deste trabalho é a mais usada em todo o mundo para desenvolvimento de repositório digitais, não aborda profundamente, novamente em especial em língua portuguesa, a instalação e customização do mesmo, sendo a instalação do sistema uma das maiores dificuldades encontradas pelo autor no desenvolvimento deste trabalho.

O estudo dos metadados e da interoperabilidade abriu ainda mais o leque de objetivos deste trabalho, uma vez que é abordado o uso dos metadados de vários padrões e também como se tornar possível que sistemas funcionando na internet tornem-se interoperáveis.

O repositório desenvolvido foi apresentado aos interessados no mesmo e tais pesquisadores demonstraram que realmente ele tem a capacidade de suprir as necessidades especificadas por eles antes do início deste trabalho, que são o armazenamento e distribuição eficiente e confiável das produções do GP-NTI. Portanto pode-se enfatizar que os repositórios digitais são sim capazes, através do seu correto uso e divulgação, de transformar os conhecimentos tácitos em conhecimentos explícitos, ou seja, fazer com que conhecimentos gerados passem a serem socializados para desenvolvimento de novos conhecimentos, fazendo

com que os documentos digitais gerados estejam sempre disponíveis aos interessados de uma maneira prática e eficiente e que também sejam preservados com o passar do tempo.

Também concluiu-se que os repositórios digitais podem ser muito eficazes dentro de instituições que não produzam produções científicas, mas que produzam documentos digitais referentes à administração e controle das instituições, pois com o uso eficaz destes sistemas, tais documentos podem ser preservados, o que permite o reuso dos mesmos, o que pode evitar muito trabalho extra, e também permite que tais documentos estejam centralizados, algo que muitas organizações procuram, devido ao fato de muito do que se é produzido por seus membros ficar armazenados em lugares únicos, ou seja, não sendo disponibilizados a outros membros da organização que poderiam fazer bom uso dos mesmos, tornando os processos da organização cada vez mais eficientes.

O uso dos metadados em sistemas computacionais é imprescindível quando deseja-se que documentos digitais armazenados em base de dados sejam encontrados facilmente, pois os elementos de metadados são capazes de criar descrições realmente precisas dos conteúdos de cada documento digital, podendo viabilizar a busca precisa e a fácil recuperação dos mesmos.

Também pode-se prover a interoperabilidade entre sistemas na internet, sejam eles os repositórios digitais, ou também outros sistemas capazes de implementar os protocolos de interoperabilidade existentes. A implementação de tais protocolos possibilita que bases de dados distintas possam trocar informações, ou seja, uma base de dados pode fazer requisições ao banco de dados de metadados de outra base de dados, possibilitando assim o maior aproveitamento dos documentos digitais descritos nestas bases de dados.

A partir deste trabalho pode-se entender o funcionamento real dos repositórios digitais e conhecer os benefícios que os mesmos podem oferecer as instituições que os implementem, então pode-se pensar a partir daqui no desenvolvimento de repositórios digitais para todas aquelas instituições que desejem desenvolver um sistema eficiente e que será capaz, a partir de um bom desenvolvimento e boa divulgação, de armazenar seus documentos digitais, preservando-os no tempo e podendo aumentar significativamente a visibilidade e também a relevância de tais documentos.

REFERÊNCIAS

AKIMURA, C.I. **Cadastro de Documentos para o Sistema Nou-Rau do DC-UEL**. Londrina. 2005. Disponível em: <<http://www2.dc.uel.br/nourau/document/?code=286>>. Acesso em 25 mai. 2009.

ALVARENGA, L. **Representação do conhecimento na perspectiva da Ciência da informação em tempo e espaços digitais**. Encontros Bibli, 1er. Semestre, n. 15, Universidade Federal de Santa Catarina, 2003. Disponível em: <redalyc.uaemex.mx/redalyc/pdf/147/14701503.pdf>. Acesso em 21 mai. 2009.

ALVES, R. C. V. **Web semântica: uma análise focada no uso de metadados**. 2005. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências - Universidade Estadual Paulista, Marília, 2005. Disponível em: <http://www.marilia.unesp.br/Home/PosGraduacao/CienciadaInformacao/Dissertacoes/alves_rcv_me_mar.pdf>. Acesso em 17 abr. 2009.

BARTON, M. R. **Creating an institutional repository: LEADIRS workbook**. Cambridge-MIT Institute, 2005. Disponível em: <www.ugr.es/~afporcel/construccion.pdf>. Acesso em 08 abr. 2009.

BAX, M. P. **Introdução às linguagens de marca**. Ci. Inf., Brasília, v. 30, n. 1, p. 32-38, 2001. Disponível em: <http://cuba.paradigma.com.br/paradigma/artigos/artigos_02.pdf>. Acesso em 18 mai. 2009.

BRANTON, A.; CHEN-GAFFEY, A. **MARC 21 Tutorial**. University of Southern Mississippi; Slippery Rock University, 2004. Disponível em: <http://www.lib.usm.edu/legacy/techserv/marc21_tutorial_ie/marcintroIE.htm>. Acesso em 28 mai. 2009.

CROW, R. **The case for institutional repositories: a SPARC position paper**. The Scholarly Publishing and Academic Resources Coalition, 2002. Disponível em: <http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf>. Acesso em 24 mar. 2009.

CUSIN, C. A.; FUSCO, E. **Seminário - Interoperabilidade**. Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista - UNESP, Marília, 2007.

DCMI. **Using Dublin Core**. 2005. Disponível em: <<http://www.dublincore.org/documents/usageguide/>>. Acesso em 15 mai. 2009.

DCMI. **DCMI Metadata Terms.** 2008A. Disponível em: <<http://dublincore.org/documents/dcmes-qualifiers/>>. Acesso em 15 mai. 2009.

DCMI. **Dublin Core Metadata Element Set, Version 1.1.** Dublin Core Metadata Initiative, 2008B. Disponível em: <<http://dublincore.org/documents/dces/>>. Acesso em 15 mai. 2009.

DCMI. **DCMI Type Vocabulary.** 2008C. Disponível em: <<http://dublincore.org/documents/dcmi-type-vocabulary/>>. Acesso em 15 mai. 2009.

DCMI. **About the Initiative.** 2009A. Disponível em: <www.dublincore.org/about/organization/>. Acesso em 08 mai. 2009.

DCMI. **Principles and Operation of the Dublin Core Metadata Initiative.** 2009B. Disponível em: <www.dublincore.org/about/organization/>. Acesso em 15 mai. 2009.

DESAI, B. C. **Supporting discovery in virtual libraries.** Journal of the American Society for Information Science, v. 48, n. 3, p. 190-204, 1997.

FUSCO, E. **Construção de Modelos Conceituais e o Processo da Catalogação:** Perspectiva de uso dos FRBR Orientado a Objetos. Qualificação (Doutorado em Ciência da Informação), Universidade Estadual Paulista – UNESP, Marília. 2009.

GARCIA, P. de A. B.; SUNYE M. S. **O protocolo OAI-PMH para Interoperabilidade em Bibliotecas Digitais.** 2007. Disponível em: <http://conged.deinfo.uepg.br/~iconged/Artigos/artigo_09.pdf>. Acesso em 22 mai. 2009.

GARTNER, R. **METS: Metadata Encoding and Transmission Standard.** Oxford University Library Services, 2002. Disponível em: <www.jisc.ac.uk/uploaded_documents/tsw_02-05.pdf>. Acesso em 30 mai. 2009.

GILL, T. **Los metadatos y la World Wide Web.** In: BACA, M. (Ed.). Introducción a los metadatos vías a la información digital, Los Angeles, 1998. p. 10-20.

GILLILAND-SWETLAND, A. J. **La definición de los metadatos.** In: BACA, M. (Ed.). Introducción a los metadatos vías a la información digital, 1998. p. 1-9.

GREENSTONE. **The Greenstone digital library software.** Disponível em: <<http://www.greenstone.org/>>. Acesso em 18 abr. 2009.

HANSEN, P. **User Guidelines for Dublin Core Creation**. Nordic Metadata Project, 1999. Disponível em: <http://www.sics.se/~preben/DC/DC_guide.html>. Acesso em 15 mai. 2009.

KNIGHT, G. **An introduction to metadata requirements for an eprint repository**. SHERPA Arts & Humanities Data Service, 2004. Disponível em: <http://www.sherpa.ac.uk/documents/D2-6_Report_on_Metadata_Issues.pdf>. Acesso em 21 mai. 2009.

LEITE, F. C. L.; COSTA, S. M. de S. **Repositórios institucionais sob a perspectiva da gestão do conhecimento científico**. 2005. Disponível em: <http://repositorio.bce.unb.br/bitstream/10482/1018/1/EVENTO_RepositorioInstitucionalSobPerspectiva.pdf>. Acesso em 28 mar. 2009.

LEWIS, S.; YATES, C. **The DSpace Course - Introduction to Dspace**. CADAIR, 2008. Disponível em: <<http://cadair.aber.ac.uk/dspace/handle/2160/617>>. Acesso em 22 abr. 2009.

LIBRARY OF CONGRESS. **METS: Introdução e tutorial**. The Library of Congress, Washington, 2006. Disponível em: <http://www.loc.gov/standards/mets/METSOverview.v2_port.html>. Acesso em 18 mai. 2009.

LIBRARY OF CONGRESS. **MARC Standards**. Library of Congress – Network Development and MARC Standards Office, 2009. Disponível em <<http://www.loc.gov/marc/>>. Acesso em 18 mai. 2009.

LIBRARY OF CONGRESS HELP DESK. **What is a Marc Record, and why is it Important?** The Library of Congress, Washington, 2009. Disponível em: <<http://www.loc.gov/marc/umb/um01to06.html>>. Acesso em 18 mai. 2009.

LYNCH, C. A. **Institutional repositories: essential infrastructure for scholarship in the digital age**. Association of Research Libraries, n. 226, 2003. Disponível em: <www.arl.org/bm~doc/br226ir.pdf>. Acesso em 23 abr. 2009.

MARCONDES, C. H. e SAYÃO, L. F. **Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T**. Ci. Inf., Brasília, v.31, n.3, p.42-54, 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n3/a05v31n3.pdf>>. Acesso em 12 mai. 2009.

NISO (National Information Standards Organization). **Understanding Metadata**. NISO Press, 2004. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em 15 mar. 2009.

NISO (National Information Standards Organization). **A Framework of Guidance for Building Good Digital Collections**. National Information Standards Organization - NISO, 3ª Edição, 2007 Disponível em: <<http://www.niso.org/publications/rp/framework3.pdf>>. Acesso em 08 abr. 2009.

NOERR, P. **The Digital Library Toolkit**. Sun Microsystems, 3ª edição, 2003. Disponível em: <http://www.ncsi.iisc.ernet.in/raja/is214/is214-2006-01-04/digital_library_toolkit-ed3.pdf>. Acesso em 08 abr. 2009.

NSDL. **Summary of OAI Metadata Best Practices**. Digital Library Federation, 2005. Disponível em: <<http://www.diglib.org/architectures/oai/imls2004/training/MetadataFinal.pdf>>. Acesso em 30 mai. 2009.

OLIVEIRA, V. S. **Buscando Interoperabilidade entre diferentes bases de dados: O caso da Biblioteca do Instituto Fernandes Figueira**. Dissertação (Mestrado em Gestão da Informação e Comunicação em saúde). Escola Nacional de Saúde Pública Sérgio Arouca – FIOCRUZ, Rio de Janeiro, 2005. Disponível em: <<http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=LILACS&lang=p&nextAction=lnk&exprSearch=494903&indexSearch=ID>>. Acesso em 08 mai. 2009.

PEPE, A. et al. **CERN Document Server Software: the integrated digital library**. CERN, Geneva, 2005. Disponível em <<http://cdsware.cern.ch/cdsware/doc/elpub2005.pdf>>. Acesso em 17 abr. 2009.

PRESSMAN, R. **Engenharia de Software**. São Paulo, Makron Books, 1995.

REPOSITÓRIO institucional. In: **GLOSSÁRIO**. Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasília, 2007. Disponível em: <http://dspace.ibict.br/index.php?option=com_content&task=view&id=43&Itemid=77>. Acesso em 07 abr 2009.

RODRIGUES, E. et al. **RepositóriUM: criação e desenvolvimento do Repositório Institucional da Universidade do Minho**. Serviços de Documentação da Universidade do Minho, Braga, 2004. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/422>>. Acesso em 02 abr. 2009.

ROSETTO, M. **Uso do Protocolo Z39.50 para recuperação de informação em redes eletrônicas**. Ciência da Informação. Print ISSN 0100-1965. Ci. Inf. vol.26, n.2, Brasília, 1997. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651997000200004>. Acesso em 23 mai. 2009.

SANTOS, J.; TEIXEIRA, C.; PINTO, J. S. **EABC**: um repositório institucional virtual. 2005. Disponível em: <http://www.iadis.net/dl/final_uploads/200508P001.pdf>. Acesso em 02 abr. 2009.

SOUZA, M. I. F.; VENDRUSCULO L. G.; MELO G. C. **Metadados para a descrição de recursos de informação eletrônica**: utilização do padrão Dublin Core. Ci. Inf., Brasília, v.29, n.1, p.93-102, 2000. Disponível em: <<http://www.scielo.br/pdf/ci/v29n1/v29n1a10.pdf>>. Acesso em 18 mai. 2009.

TRISKA, R.; CAFÉ, L. **Arquivos abertos**: subprojeto da Biblioteca Digital Brasileira. Ciência da Informação. Ci. Inf., Brasília, v.30, n.3, p.92-96, 2001. Disponível em: <http://www.scielo.br/scielo.php?pid=S0100-19652001000300012&script=sci_arttext&tlng=pt>. Acesso em 15 abr. 2009.