

# REPOSITÓRIOS INSTITUCIONAIS EM CIÊNCIA E TECNOLOGIA: UMA EXPERIÊNCIA DE CUSTOMIZAÇÃO DO DSpace

VIANA, C. L. M.  
MÁRDERO ARELLANO, M. A.  
SHINTAKU, M.

Instituto Brasileiro de Informação em Ciência e Tecnologia  
E-mail: cassandra@ibict.br

## RESUMO

Foi planejada a construção de repositórios para as Unidades de Pesquisa do Ministério de Ciência e Tecnologia no Brasil utilizando o *software* DSpace. A implementação demandou a instalação de repositórios de caráter experimental e a adoção de metodologia envolvendo abordagens teórica e empírica. A sistemática de pesquisa abrangeu procedimentos de: simulação, adaptação, avaliação e testes de *software*. As avaliações operacionais e funcionais demonstraram que a *Web User Interface* (WUI) desta ferramenta torna-se mais complexa ao adentrar-se ao nível das funcionalidades administrativas, de acesso restrito aos administradores do sistema. A simulação dos fluxos de submissão/ depósito e revisão/ *workflow* permitiu confirmar a adequação do novo *design* do *template* de submissão (seu formato e conteúdo) quanto aos campos a serem preenchidos (formato) e ao seu conteúdo. A adaptação do procedimento de *upload* para objetos externos possibilitou contornar a impossibilidade de enviar *links* para o servidor (*upload*), garantindo ao repositório a capacidade de manutenção dos metadados desses itens. Os resultados dos testes de customização de metadados demonstraram que a consistência dos elementos e de seus qualificadores foi mantida e exibida nos registros Dublin Core dos itens tanto em formato simples como em formato completo. Concluiu ser viável a customização de metadados, garantindo a interoperabilidade, através de dois procedimentos diferentes: codificação e

configuração. Recomenda que a seleção e adaptação dos elementos e qualificadores tome por base pelo menos os três critérios testados, que se referem aos atributos dos campos a serem customizados.

**Palavras-Chave:** Repositórios digitais; Repositórios institucionais; Dspace; Metadado, Customização.

## **INTRODUÇÃO**

Nos últimos anos, a Iniciativa de Arquivos Abertos e o Movimento de Acesso Aberto à Informação Científica<sup>1</sup> vêm propondo que a informação científica seja disponibilizada gratuitamente, apesar dos esforços contrários por parte da indústria editorial e de órgãos públicos que guardam acervos valiosos para a pesquisa. As bibliotecas digitais, os repositórios institucionais e os periódicos de acesso aberto estão provocando uma mudança nas práticas de trabalho dos membros da comunidade científica e requerendo sua participação no esforço global para definir e classificar seus documentos segundo um novo modelo de comunicação científica. Desde 1998 algumas instituições já vinham criando seus próprios repositórios digitais, mas a maioria não possuía normas para o seu desenvolvimento nem para o estabelecimento de parcerias com outras instituições (HEDSTROM, MONTGOMERY, 1998).

Entre os sistemas que estão sendo utilizados para criação de grandes repositórios de informação digital estão aqueles que comportam diferentes tipos de conteúdos e formatos de arquivos digitais e são voltados para a captura, distribuição e preservação da produção intelectual de uma instituição.

O *DSpace Institutional Digital Repository System* (projeto colaborativo da MIT Libraries e a Hewlett-Packard Co.) é um destes vários projetos, atualmente em andamento. Por ser um *software* livre, o DSpace transfere para as organizações a responsabilidade pela gestão da informação técnico-científica, e pela garantia de que o acesso à essa informação será permanente (MÁRDERO ARELLANO, 2004).

Apresentamos a seguir os atributos dos repositórios institucionais e, mais especificamente os da ferramenta DSpace, configurado para ser utilizado como repositório institucional para as unidades de pesquisa do Ministério de Ciência e Tecnologia. Com esta iniciativa o IBICT pretende contribuir para que esse tipo de solução tecnológica venha a ser um instrumento valioso para a gestão e publicação científica no país, em complemento às formas tradicionais já consagradas.

## REPOSITÓRIOS DIGITAIS

Um repositório digital é uma forma de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar material por longos períodos de tempo e prover o acesso apropriado. Essa estratégia foi possibilitada pela queda nos preços no armazenamento, pelo uso de padrões como o protocolo de coleta de metadados da Iniciativa dos Arquivos Abertos (OAI-PMH), e pelos avanços no desenvolvimento dos padrões de metadados que dão suporte ao modelo de comunicação dos arquivos abertos.

Com o surgimento da Iniciativa dos Arquivos Abertos<sup>2</sup> (inicialmente conhecida como UPS: *Universal Preprint Service*), na última década do século XX, diminuíram as dificuldades que os autores enfrentavam para disponibilizar o

---

<sup>1</sup> O objetivo da Iniciativa, segundo a declaração chamada de “Budapest Open Access Initiative” é acelerar os avanços, no esforço internacional, para que os artigos científicos de todas as áreas estejam disponíveis abertamente na Internet (<http://bibliotecna.upc.es/rebiun/BOAI.pdf>).

<sup>2</sup> URL: [www.openarchives.org](http://www.openarchives.org).

resultado das suas pesquisas internacionalmente, aumentando a sua visibilidade e a estimativa de impacto<sup>3</sup>. O Protocolo de Coleta de Metadados da Iniciativa dos Arquivos Abertos (OAI-MHP) está sendo reconhecido pela comunidade de editores de periódicos científicos como uma forma de garantir que suas informações na Internet sejam visíveis de acordo com padrões de interoperabilidade. A distribuição globalizada da informação científica promoveu novas parcerias e incentivou o desenvolvimento da ciência nas regiões mais pobres do planeta. Através do compartilhamento de metadados, facilitado pelo modelo OAI, os autores e os editores interessados em disponibilizar suas publicações abertamente não precisam acrescentar novos serviços aos seus produtos de informação; eles são compartilhados com outras instituições.<sup>4</sup> Esta é uma das formas utilizadas pelas universidades para apoiar a divulgação dos resultados de pesquisas científicas, criando mecanismos para legitimar e estimular a publicação dos trabalhos produzidos (LYNCH, 2003).

A capacidade de contar com um espaço, onde possa ser depositada a produção científica de uma instituição, propicia o ambiente necessário para a discussão entre os pares, o intercâmbio de idéias dentro de uma comunidade de pesquisadores e a revisão constante das versões de um documento. O debate crítico é alimentado de forma aberta pelos participantes dos repositórios de arquivos abertos, onde não se tem limite para o tamanho dos documentos ou extensão temporal da discussão de qualquer tema. São resolvidos os problemas de distribuição dos canais formais e o acesso pode ser gratuito. Por serem um veículo de auto-publicação por natureza, os arquivos/repositórios abertos solicitam, apenas, uma promoção institucional para seu uso. As instituições reconhecem os materiais arquivados e os valorizam, estabelecendo as ligações entre a informação que precisam e aquela que desejam disponibilizar abertamente.

---

<sup>3</sup> O impacto provocado na comunidade é medido pelo grau de usabilidade alcançado (Harnad, 1999).

<sup>4</sup> Mais de 400 arquivos de acesso aberto (revistas e repositórios) registrados na OAI URL: <http://archives.eprints.org/>

Um repositório digital pode ser operado em qualquer organização do setor público, atualmente existem instituições encarregadas de armazenar e gerenciar material digital, mas são poucas as que afirmam ter a capacidade de armazenamento a longo prazo. Muitas empresas privadas afirmam que provem acesso e armazenamento, mas o acesso público é restringido e o proprietário original que gerencia o conteúdo desde uma página comercial passa o controle direto do material a mãos de terceiros. Nesses casos, segundo Simpson (2004), para manter os direitos e o controle sobre o material digital deve-se consultar a legislação, assim como especificar se as pesquisas foram financiadas, estipulando o material que continuará sob o controle e propriedade da instituição onde foi gerado.

Estes repositórios<sup>5</sup> incentivam a publicação na Rede gerenciada pelo pesquisador (*self-archiving*), utilizam novas tecnologias abertas (*open source*), e as informações ficam disponíveis para serem acessadas permanentemente por diversos provedores de serviços a nível nacional e internacional<sup>6</sup>. Esta mudança na publicação científica iniciou-se com a constituição dos chamados repositórios digitais temáticos.<sup>7</sup> A idéia evoluiu para o agrupamento destes repositórios sob a responsabilidade de instituições centrados na divulgação da produção científica local (CAFÉ et al, 2003).

---

<sup>5</sup> Ainda na sua fase inicial, poucos repositórios têm conseguido o tamanho e importância considerável, entre os principais *softwares* usados estão: FEDORA (Flexible Extensible Digital Object and Repository Architecture), CDSWare, Eprints, Archimene, ARNO, iTor, MyCoRe.

<sup>6</sup> Alguns provedores de serviço baseados no Protocolo OAI que permitem a busca em mais de 250 arquivos abertos são: Arc, Citebase, MyOAI, Perseus, Scirus, OAIIECSP, Callima, CYCLADES, OAIster, SAIL-eprints, TORII e PKP.

<sup>7</sup> Um dos primeiros projetos de repositórios temáticos foi o JSTOR (The Scholarly Journal Archive – URL: <http://www.jstor.org/>) criado para servir apenas como repositório de preservação digital de periódicos científicos de algumas áreas.

## REPOSITÓRIOS INSTITUCIONAIS

Desde suas primeiras implementações, os repositórios institucionais têm sido voltados para a informação científica, contendo os trabalhos de pesquisa e ensino de comunidades acadêmicas, assim como a documentação de uma ampla variedade de atividades das próprias universidades.

Os repositórios institucionais podem ser pensados como parte de uma ação dentro da gestão documental, que inclui sua criação, tratamento, transmissão e acesso. Crow (2002) define os repositórios institucionais como “um arquivo digital de produtos intelectuais criados por uma comunidade de pesquisadores, estudantes e professores de uma instituição”. Para Lynch (2003) eles são “um conjunto de serviços que a instituição oferece aos seus membros para o gerenciamento e disseminação de materiais digitais criados na instituição”.

A arquitetura da Iniciativa dos Arquivos Abertos (OAI-PMH) define o que seriam os provedores de dados e de serviços (LAGOZE, et al, 2002), onde os repositórios digitais estariam encarregados com o desenvolvimento das sua coleção digital, definindo a abrangência do seu serviço e níveis de qualidade (WOLPERT, 2002). Como provedores de dados, os repositórios institucionais basicamente expõem seus metadados para serem coletados por os provedores de serviços.

Existem, atualmente, mais de 200 repositórios institucionais em todo o mundo.<sup>8</sup> 100 deles estão localizados em universidades. A seguir na figura 3 é demonstrado o crescimento dos arquivos /repositórios institucionais e dos seus conteúdos.

---

<sup>8</sup> URL: <http://archives.eprints.org/index.php?action=browse#type>

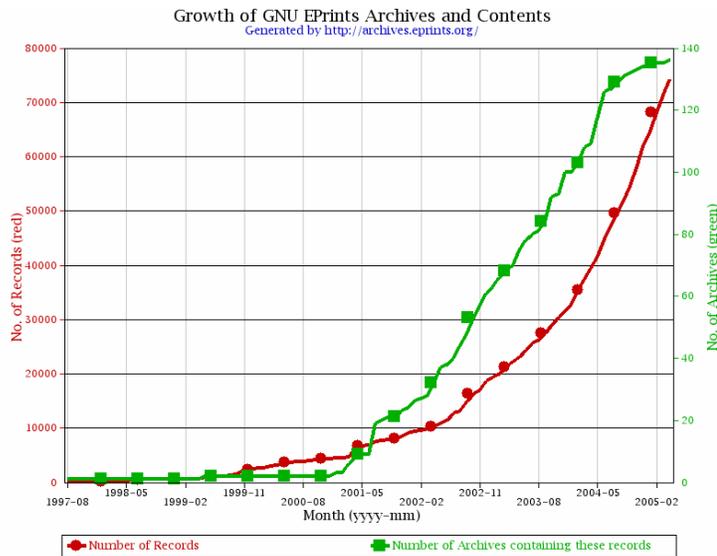


Figura 3 - Crescimento dos Arquivos/Repositórios Institucionais.

Um dos objetivos da publicação científica é salvaguardar a propriedade intelectual (OKUBO, 1997). Os repositórios institucionais lidam com essas tarefas e com o problema da cópia apropriada. Existem, também, alguns tipos de licença sobre a utilização de material intelectual (Common License/Creative License), que seguem as normas de direitos autorais. O único obstáculo atual para o Acesso Aberto através do uso de repositórios institucionais é a ausência de políticas institucionais que requeiram o auto-arquivamento da sua própria produção científica.<sup>9</sup>

A opção pelo auto-arquivamento temático e/ou institucional certamente é mais confiável. Além disso, na medida em que o repositório institucional se consolida, a coleta automática aos repositórios temáticos poderá auxiliar na segurança dos dados fornecendo *backups*, redundâncias e espelho dos metadados e documentos contidos nos repositórios temáticos.

Com os repositórios institucionais são colocados em funcionamento serviços para a os membros das comunidades que visam o gerenciamento e a disseminação de todo

<sup>9</sup> Berlin 3 URL: <http://www.eprints.org/berlin3/outcomes.html>

o material criado dentro de uma instituição e pelos seus membros fora dela. Através deles é garantido o gerenciamento das mudanças tecnológicas e a migração do conteúdo digital para novas versões de mídia eletrônica.

Assim como as novas tecnologias de informação estão sendo um desafio para as organizações, existe também um grande número de desafios relacionados com a habilidade dessas organizações para integrar o gerenciamento de materiais digitais na sua estrutura organizacional.

Pesquisas internacionais propõem a criação de repositórios que permitam o acesso e recuperação dos dados. Para Crow (2002) os repositórios institucionais deveriam estar acessíveis para qualquer usuário dentro ou fora da instituição, com poucas restrições. A solução ideal é a aliança entre uma organização habilitada no gerenciamento dos dados digitais e os criadores, de tal forma que possam criar conhecimento do material mantendo o controle das decisões sobre o conteúdo a serem preservados e os intervalos para essa atividade. Por esses motivos, os repositórios institucionais devem estar respaldados por política de gerenciamento que os reconheçam como vitais para a instituição.

Estabelecer um repositório institucional indica que a biblioteca esta mudando seu papel de custódia para contribuir ativamente na mudança do modelo de comunicação científica. As bibliotecas mantêm a responsabilidade de gerenciar e arquivar material impresso. Mas, a medida que o volume de material para pesquisa de acesso aberto em formato digital cresce, o papel e o valor das coleções impressas declinam proporcionalmente. Os orçamentos das bibliotecas começam a mudar no sentido de apoiara publicação de acesso aberto da sua comunidade e continuar tendo um papel relevante. Para as bibliotecas organizacionalmente os repositórios institucionais são uma resposta apropriada ao novo contexto da informação digital. As bibliotecas por estarem mais bem preparadas para prover a normalização na preparação dos documentos digitais, facilitam o trabalho dos autores a contribuir com suas pesquisas nos repositórios institucionais (PINFIELD, et al, 2002).

## DSPACE

Duas instituições pioneiras no desenvolvimento de *software* livre para repositórios digitais foram a *University of Southampton* na Inglaterra e o *Massachusetts Institute of Technology* (MIT). A primeira criou o *software* Eprints (<http://www.eprints.org/>) destinado para servir de repositório institucional ou temático (de uma área do conhecimento) de documentos digitais. O MIT, em colaboração com a *Hewlett Packard Corporation* criaram o modelo de repositório institucional chamado DSpace (<http://www.dspace.org/>), que atualmente está sendo o *software* para construção de repositórios institucionais mais usado internacionalmente. A rede internacional de repositórios institucionais que usam o DSpace prove amplo conteúdo e serviços entre instituições de ensino e pesquisa, além de possuir uma federação já constituída.

O DSpace adotou o protocolo para coleta de metadados da Iniciativa dos Arquivos Abertos (OAI-PMH v2.0) e é considerado um provedor de dados. O Protocolo foi implementado usando o *software* aberto OAICat, para fazer os registros dos itens disponíveis para coleta.

Entre as características do DSpace estão: a) ser um *software* livre, b) sua arquitetura de *software* é simples e eficiente, c) uso de tecnologia de ponta, d) direcionado para o acesso aberto, e e) intencionalmente implementado para servir de repositório institucional. No Dspace os dados estão organizados de forma a refletir a estrutura da instituição e se organizam em coleções.

A meta do projeto DSpace é estudar os repositórios institucionais, incluindo aspectos tais como: controle do acesso, direitos autorais, versões digitais de documentos, recuperação, receptividade por parte da comunidade acadêmica, e suas funcionalidades para a publicação (BARTON, WALKER, 2003).

## REPOSITÓRIOS INSTITUCIONAIS DSPACE NO BRASIL

Após a primeira implementação do DSpace em novembro de 2002, como repositório institucional em língua portuguesa feita pela Universidade do Minho, em Portugal (RODRIGUES et al, 2004), o IBICT iniciou seus trabalhos de prospecção com o *software*. Em 2004 a Biblioteca do Supremo Tribunal de Justiça teve assessoramento do IBICT para a criação do repositório institucional (BDJur), adotando o DSpace e instalando-o em uma das suas unidades administrativas (BASEVI, 2004). Também, o IBICT estabeleceu uma parceria com a Universidade de São Paulo (USP) para realizar a tradução da versão mais atual do *software* para o português do Brasil; essa experiência resultou na criação do Reposcom – Repositórios Institucionais em Ciências da Comunicação (FERREIRA, SOUTO, 2005). Paralelamente, a Universidade Federal do Paraná (<http://dspace.c3sl.ufpr.br/dspace/index.jsp>) estabeleceu sua Biblioteca Digital de Teses e Dissertações, usando o *software* DSpace.

O Apesar de nenhuma das experiências brasileiras poderem ser consideradas implementações de repositórios institucionais, já que elas não possuem políticas de depósito e acesso, nem representam toda a produção das suas instituições, seu surgimento manifesta o interesse dos responsáveis pelos acervos digitais no Brasil, em trabalhar com sistemas que permitam documentar e acessar a produção intelectual das suas instituições.

Desde 2003, no IBICT foram realizados estudos sobre ferramentas para repositórios institucionais (CAFÉ et al, 2003). O interesse do IBICT era maximizar a visibilidade e o uso da produção técnico-científica dos funcionários e bolsistas das Unidades de Pesquisa do Ministério de Ciência e Tecnologia por meio de repositórios institucionais. O objetivo final era integrar as iniciativas do Ministério de Ciência e Tecnologia referentes ao registro da produção técnico-científica institucional. Esse objetivo seria alcançado pela institucionalização de um

repositório no IBICT e a construção de repositórios institucionais para cada unidade de pesquisa do Ministério.

A concretização iniciou-se com uma pesquisa sobre as ferramentas mais adequadas para essa finalidade. Algumas das ferramentas analisadas foram: *Archimede* da Loyal University Library *Software*, o *CDSware* do CERN, o *FEDORA* da Universidades de Virginia e Cornell e o *DSpace* do MIT. O *software* que apresentou as características mais adequadas às necessidades do IBICT foi o *DSpace*.

O primeiro teste do sistema foi a criação do repositório piloto, o RIDI (Repositório Institucional Digital do IBICT). Na versão piloto visava-se recuperar a documentação de todas as atividades de pesquisa, publicando os conteúdos que se encontravam dispersos em diversos departamentos. Mas já na fase inicial de testes do sistema, percebeu-se a necessidade da elaboração de políticas e normas para depósito e publicação de conteúdos no repositório. Por meio de ordem interna (OI) da diretoria do IBICT, foi nomeada uma comissão de técnicos do instituto para que trabalhassem na elaboração do documento contendo essas políticas, no primeiro semestre de 2005. A disponibilização do RIDI no *Site* do IBICT e os testes finais dependem agora da aprovação do documento contendo as normas e da assinatura do termo de compromisso de repasse de documentos eletrônicos para o RIDI por parte dos responsáveis pelos projetos e programas.

## METODOLOGIA

Em virtude dos atributos do *DSpace* como repositório digital institucional que captura, armazena, indexa, preserva e redistribui os recursos da biblioteca existentes em formato digital (LITTLE, 2005), visualizou-se a perspectiva de usá-lo como o sistema base para a implementação de repositórios institucionais para as Unidades de Pesquisa do MCT.

A metodologia do trabalho consistiu em duas abordagens: uma pesquisa de caráter exploratório e descritivo visando a fundamentação teórica do uso da ferramenta; e a outra, pesquisa empírica consistindo em procedimentos de testagem operacional-funcional e a “resposta” adaptativa do *software*.

Estas abordagens foram aplicadas alternada ou simultaneamente, nas fases de: a) instalação e configuração da ferramenta; b) customização de metadados; c) manutenção de objetos digitais externos; d) operação do provedor de dados; e e) estudo do modelo de licença de distribuição.

## INSTALAÇÃO E CONFIGURAÇÃO DA FERRAMENTA

Por se tratar de um sistema executável em plataforma *WEB*, e desenvolvido em *open source*, o DSpace necessita de uma infra-estrutura (*hardware* e *software*) própria deste tipo de *software*.

Para sua instalação, os sistemas operacionais recomendados são aqueles baseados em UNIX/Linux. Optou-se por utilizar o sistema operacional Kurumim ([www.kurumin.org](http://www.kurumin.org)) e ainda o servidor de aplicações Tomcat ([jakarta.apache.org/tomcat/](http://jakarta.apache.org/tomcat/)) por recomendação da documentação técnica.

Dentre os bancos de dados apropriados: Oracle ([www.oracle.com](http://www.oracle.com)); e PostgreSQL ([www.postgresql.org](http://www.postgresql.org)); foi escolhido o segundo, por ser gratuito, não acarretando ônus adicional para instituição.

A preparação do ambiente computacional seguiu os padrões de instalação recomendados. O *software*, originalmente compactado em arquivo do tipo *Web Application Archive* – war, é descompactado pelo Tomcat, criando o sistema. O pacote do *software* inclui dois arquivos .war, um para a instalação em *Web* e outro para suporte ao protocolo *Open Archives Initiative – Protocol for Metadata Harvesting* (OAI-PMH).

Por ser uma aplicação desenvolvida por um consórcio, sua estrutura é bem organizada. O sistema tem três camadas: Aplicação, Negócio e Armazenamento, organizadas em três pacotes separados que contém o código-fonte. Dentro de cada camada estão distribuídos os diversos componentes, tais como: Busca; Interface *Web* ; Provedor de Dados do OAI-PMH; Bancos de Dados; Gerenciador de *workflow* de Submissão etc.

Possui um conjunto de funcionalidades especificamente relacionadas à recuperação e disseminação de metadados. No caso da recuperação dos itens depositados o DSpace utiliza o mecanismo de busca Lucene ([//lucene.apache.org/](http://lucene.apache.org/)), que integra seu conjunto de programas. Quanto à disseminação, o DSpace se utiliza do OAI-PMH para manipulação dos metadados.

Para esta pesquisa foram adotadas versões-padrão da ferramenta, isto é, que implementam apenas a camada de Provedor de Dados, permitindo a exposição dos metadados para acesso por outros repositórios, através do mesmo protocolo. Essa implementação foi ativada através do *software* servidor de aplicações *Web* denominado OAICat, da OCLC.

A instalação foi realizada em 2 etapas, envolvendo duas versões diferentes do DSpace, a saber:

- Instalação da Versão 1.2.1, que passaremos a denominar de 1º Repositório;
- Instalação da Versão 1.2.2, que passaremos a denominar de 2º Repositório.

Também a configuração original do sistema foi alterada, visando a construção de repositórios apropriados para as Unidades de Pesquisa do Ministério de Ciência e Tecnologia. Isto ocorreu no caso de ambas as versões (1º e 2º Repositórios), resultando na:

- Customização da interface gráfica ao padrão do IBICT: cores, logotipos, etc.;

- Adequação da estrutura de disponibilização de metadados. O sistema DSpace utiliza o padrão Dublin Core ([//dublincore.org/](http://dublincore.org/)) com todos os seus elementos e qualificadores, porém sua configuração completa é opcional;
- Utilização de interface padrão traduzida para o português. O DSpace fornece orientações sobre a possibilidade de utilização de interface internacionalizada, porém não a sua configuração automática.

Embora a internacionalização seja uma propriedade muito valorizada atualmente, tornando possível dispor de opções múltiplas de língua e permitindo ao usuário selecionar aquela na qual deseja que lhe seja exibida a interface *Web*, por envolver procedimentos que requereriam maior tempo de dedicação, optou-se por não implementar esta configuração.

#### Processo de testagem

O processo de testagem envolveu 4 tipos de procedimentos: a) adaptação; b) simulação; c) checagem; e d) testes funcionais.

Estes procedimentos foram aplicados a todas as funções de competência do Administrador do Sistema, a saber: gestão de comunidades; administração de coleções; administração de E-pessoas e Grupos; gerenciamento de políticas (de comunidades, coleções e itens); fluxos de submissão e revisão; modificação de licença de distribuição.

Foram observadas características operacionais da ferramenta, tais como: navegabilidade, “amigabilidade” etc.

A navegação, para os usuários “anonymous” e “depositantes” é bem simples e de fácil visualização. Entretanto, as “ferramentas administrativas” não estão acessíveis, no mesmo nível que os demais usuários, sendo necessário passar para um outro nível mais interno (hierarquicamente inferior) da interface gráfica, para a

definição de parâmetros como: notícias, estatísticas, permissões e alterações de E-pessoas, políticas de coleções, registro Dublin Core entre outros.

No caso do cadastramento de usuários e do *workflow* editorial, foi feita uma simulação do desempenho das funções relacionadas às E-pessoas e Grupos, através da criação de contas para usuários fictícios, com as permissões previstas nos fluxos de submissão e de revisão.

Quanto às características editoriais da ferramenta, foram checados: o processo de depósito de documentos; os *workflows* de revisão; o acompanhamento dos processos de depósito de cada coleção; e os níveis de permissões. Foi avaliado o fluxo editorial do DSpace, que inclui 5 papéis: Autor; Revisor de Conteúdo ou Editor de nível 1; Revisor (de Conteúdo e de Metadados) ou Editor de nível 2; Revisor de Metadados ou Editor de nível 3; e Administrador da coleção.

## CUSTOMIZAÇÃO DE METADADOS

Os benefícios da interoperabilidade bem sucedida estão amplamente documentados na literatura. Metadados dão significado semântico a dados aparentemente isolados, provendo um contexto ao processo de busca.

O potencial dos metadados para aumentar a relevância dos resultados de buscas em repositórios já havia sido ressaltado ainda na década passada, pela afirmação de Weibel e Lagoze, quanto à correlação entre a padronização na descrição de conteúdo e as capacidades de recuperação de recursos em ambientes distribuídos. Seus comentários constam, ainda hoje, na documentação da *Dublin Core Metadata Initiative* (DCMI) (DUBLIN CORE..., 2005).

A DCMI é apenas um dos esforços sistematizados de prover interoperabilidade de dados entre criadores e usuários. Mesmo assim, dentro deste padrão cada elemento é opcional visando dar ao criador de conteúdo a capacidade de “customizar” o processo de descrição.

Os esforços de padronização de metadados são valiosos quando a interoperabilidade dos componentes a serem descritos torna-se necessária. Porém esses padrões entram em desacordo quando aplicados à descrição de singularidades que não podem ser ignoradas. (NORTON, 2005)

Atualmente o DSpace aceita unicamente o padrão Dublin Core para a descrição de conteúdo dos itens, embora já existam pesquisas no sentido de desenvolver componentes que permitam o uso de esquemas de metadados aleatórios, usando o Resource Description Framework (RDF).

Foram testadas as propriedades da ferramenta para a descrição dos itens (documentos do MCT), os quais requeriam um processo de customização e, conseqüentemente, uma diferenciação de metadados e seus qualificadores.

O DSpace disponibiliza os 15 elementos básicos do Dublin Core, adotados isoladamente ou em associação com 46 qualificadores, desenvolvidos com base no Dublin Core Libraries Working Group Application Profile (LAP). Há, portanto, um total há 66 atributos disponíveis para esta descrição.

No DSpace o uso dos elementos DC e de seus qualificadores é opcional, com exceção do título, idioma e data de depósito. Além disto, os elementos podem ser repetidos quando necessário.

Assim, foram selecionados, dentre os elementos e qualificadores, os considerados mais apropriados para representar as propriedades dos itens pertencentes às coleções a serem criadas para as Unidades de Pesquisa do MCT.

## CODIFICAÇÃO

Os testes no 1º Repositório (versão 1.2.1) envolveram alterações em nível de Codificação, isto é, alteração nos programas do pacote DSpace. Os arquivos .jsp tiveram de sofrer alterações de modo a tornar possível a captação de outras características dos itens.

Foram testados os metadados e qualificadores apresentados na Tabela 1.

Tabela 1. Metadados customizados

Elemento	Qualificador
contributor	author
date	issued
description	
description	abstract
identifier	citation
identifier	other
publisher	
relation	ispartofseries
subject	
title	

Foi alterado o template de captação dos metadados, de tal modo que os campos referentes aos metadados acima receberam novas denominações. Essas mudanças foram absorvidas pelo registro DC em formato simples, exibindo as alterações efetuadas na denominação dos elementos descritivos, do qualificadores (em alguns casos) e das notas de escopo (quando necessário).

Entretanto, no caso do formato completo do registro DC, como já era previsto, as alterações só causaram impacto nos elementos e qualificadores que sofreram alterações de conteúdo e não na forma de preenchimento dos campos.

## CONFIGURAÇÃO

Os testes no 2º Repositório (versão 1.2.2) envolveram alterações apenas em nível da Configuração, sem modificar os programas do DSpace. Foram testados os mesmos metadados e qualificadores que no caso do 1º Repositório.

Entretanto, neste caso, as alterações se deram através do uso de valores pré-determinados dentro dos campos básicos do *template* de submissão. Os valores-padrão foram incluídos previamente, isto é, antes do processo de depósito/submissão, no momento da criação de cada coleção. As alterações, portanto, foram efetuadas apenas na interface de submissão, aplicando os metadados já existentes às novas características a serem descritas.

Foi avaliado também, apenas no caso deste 2º Repositório, o desempenho da ferramenta frente a diferentes tipos de configurações de coleções. Foram testadas coleções com diferentes configurações. Foram usadas pelo menos três diferentes estruturas de metadados, ainda dentro do padrão DC, em coleções diferenciadas, porém integrantes deste mesmo repositório.

A intenção era verificar o efeito dessas variações no funcionamento das coleções e do repositório como um todo, especialmente no que se refere à ocorrência de erros internos de sistema, bem como quanto à probabilidade de comportamento da ferramenta diante das variedades de situações possíveis em ambientes reais de sistemas de informação.

## ADAPTAÇÕES

Os procedimentos de adaptação foram aplicados à testagem do provedor de dados, à manutenção de objetos digitais externos e ao termo de licença de distribuição.

O Provedor de Dados precisou ser testado após a configuração do sistema, para checar seu funcionamento e, especialmente, no caso das customizações efetuadas.

Entretanto, o OAI-PMH não pode ser testado pelo *Repository Explorer* (<http://re.cs.uct.ac.za/>) uma vez que os repositórios experimentais encontravam-se instalados em servidor seguro (https). Neste caso, foi definida uma porta do TCP/IP através da qual foi estabelecida a comunicação com o servidor e testados os verbos de requisição do protocolo.

Em vista da ocorrência de objetos digitais originados fora dos repositórios e cuja localização necessitava ser gerenciada para atender aos requisitos dos repositórios das Unidades de Pesquisa do MCT, foi constatado que o DSpace não dispõe de tecnologia apropriada para esta funcionalidade. Trata-se, de documentos em HTML, Páginas *Web* etc., os quais não tem como ser “enviados” para o servidor (*upload*), mas precisam ser localizados pelo sistema, pois fazem parte de suas coleções. A adaptação adotada para esses casos demandou, além da descrição normal do item, por meio do *template* de submissão, a criação de documento HTML para “apontar”, através dos identificadores persistentes (*handles*) e redirecionar o sistema para a localização original do item na *Web*.

A última etapa do processo de submissão: a concessão da licença de distribuição, também requereu adaptação apropriada às características dos repositórios a serem implantados. Foi realizada tradução e feitos ajustes ao “Termo de Licença de Distribuição Não Exclusiva” para cada das versões (1.2.1 e 1.2.2) do 1º e 2º Repositórios experimentais

## RESULTADOS

As avaliações operacionais demonstraram que a *Web User Interface* (WUI) torna-se mais complexa ao adentrar-se ao nível das ferramentas administrativas, de acesso permitido apenas aos administradores do sistema. Somente neste nível é possível realizar: a administração de E-pessoas; a definição de funções dos membros de cada coleção (Grupos); a edição e remoção de itens; a edição do

registro DC completo; a edição de todos os registros contendo todos os formatos de bitstream cadastrados e devidamente descritos; o acompanhamento dos *workflows* correntemente ativos; as definições de políticas para comunidades, coleções, itens e políticas avançadas e a edição de notícias.

A simulação dos fluxos de submissão/ depósito e revisão/ *workflow* permitiu confirmar a adequação do novo *design* do *template* de submissão (seu formato e conteúdo) quanto aos campos a serem preenchidos (formato) e ao seu conteúdo.

Os testes de exposição automática de metadados, executados em relação a todos os verbos de requisição do protocolo OAI-PMH, retornaram registros em formato XML compatíveis com as requisições. Esta verificação foi feita após a consolidação das alterações, adaptações e testes efetuados e confirmou o funcionamento correto do Provedor de Dados.

No caso dos objetos digitais localizados externamente, a adaptação possibilitou contornar a impossibilidade de enviar o *link* para o servidor (*upload*), garantindo ao repositório a capacidade de manutenção também dos metadados desses tipos de itens.

O “Termo de Licença de Distribuição Não Exclusiva” recebeu uma tradução com terminologia mais adequada, adaptada ao contexto das Unidades de Pesquisa do MCT e do sistema jurídico brasileiro.

### **Quanto à configuração**

Os resultados obtidos demonstraram que:

- Só é possível a alteração em nível de configuração no momento da criação da coleção. Após criada a coleção não é mais possível selecionar outros metadados ou usar os valores-padrão;
- a interoperabilidade é mantida pela preservação da integridade do padrão DC;

- Os campos dos formulários, que foram acrescentados, aparecem no template de submissão, dentro dos campos já existentes;
- Os campos dos formulários, que foram acrescentados, aparecem na exibição dos elementos de metadados e de seus conteúdos no registro do item depositado (seja em formato simples ou completo);
- os novos elementos de metadados utilizados guardam estreita similaridade com os atributos de formato e conteúdo dos campos dos formulários do MCT.

A consistência dos metadados submetidos a alterações em nível da Configuração foi verificada através da forma de exibição dos elementos e qualificadores modificados, nos elementos DC, para cada coleção. Esta avaliação consistiu na comparação da forma e conteúdo dos elementos de metadados e de seus qualificadores no registro de itens depositados, antes e após as modificações efetuadas em nível de codificação, bem como de configuração.

### **Quanto à codificação**

Os resultados obtidos demonstraram que:

- Até a versão 1.2.1, o Dspace só aceita um único template padrão para todas as comunidades e coleções daquele repositório. Não foi possível fazer adaptações para comunidades ou coleções específicas como no caso das customizações baseadas em Configuração;
- Alguns dos novos metadados necessitaram de configuração específica em relação ao seu conteúdo (valor e formato dos dados a serem coletados), para serem corretamente exibidos no template de submissão;
- O processo de customização baseado em alteração no código-fonte demonstrou ser mais demorado e oneroso que a customização por configuração;

- O resultado final da alteração nos programas Java é mais satisfatório em função da maior amplitude de alterações possíveis.

A consistência dos metadados submetidos a alterações em nível da Codificação foi verificada da mesma forma que para os procedimentos relativos a Configuração, pela comparação da forma e conteúdo dos elementos de metadados e de seus qualificadores no registro de itens depositados, antes e após as modificações efetuadas em nível de codificação, bem como de configuração.

Em ambos os casos, isto é, tanto na configuração como na codificação, foram adotados três critérios para a seleção dos elementos e qualificadores mais adequados, a saber:

- a) a semelhança de escopo entre os elementos DC e o conteúdo dos campos dos formulários do MCT;
- b) o tipo de caracteres aceitos dentro dos elementos (valor do campo);
- c) o formato de apresentação dos dados (características dos itens) a serem descritos.

## **CONCLUSÕES**

Um repositório digital é um serviço que requer continuidade. Quando uma instituição cria um recurso automatizado de tal porte, ela está reconhecendo que está tomando para si um compromisso de longo prazo. Esta implementação é uma oportunidade para a instituição abrir-se à reflexão quanto a infra-estrutura necessária para dar suporte a suas atividades científicas no futuro. Para tanto, uma ampla discussão com todos os depositários de informação atribuiria um alto grau

de confiabilidade na definição dos elementos de metadados necessários para melhor descrever os objetos digitais a serem mantidos.

Desde o seu surgimento, os repositórios institucionais têm sido voltados para a comunicação científica, contemplando os trabalhos de pesquisa, ensino e extensão de diversas comunidades acadêmicas, assim como a documentação de uma ampla variedade de outras atividades das próprias instituições de ensino superior. Estas informações são de altíssimo valor como subsídio tanto para as pesquisas correntes como para a tomada de decisão quanto a futuras linhas de pesquisa.

Além das propriedades relativas à preservação, divulgação e recuperação dos objetos digitais, observou-se no DSpace uma representação bastante similar ao processo de editoração de um periódico científico. Este tipo de “administração” do sistema de publicação parece bastante adequado para a comunicação científica uma vez que reproduz o processo de avaliação pelos pares e ainda permite a distinção de duas funções editoriais importantes: uma referente à avaliação do conteúdo e da qualidade intelectual do trabalho, e a outra dedicada à descrição e à disseminação deste conteúdo, para a comunidade científica, e que é representada pelo *schema* de metadados adotado (neste caso, o Dublin Core – DC).

Como os procedimentos adotados ocorreram em repositórios de caráter apenas experimental, não foi possível garantir sua validade em situações reais de “sondagem” por Provedores de Serviços, para *harvesting*.

A pesquisa constatou a viabilidade da customização de metadados sem prejuízo do *schema* ou padrão adotado, ou seja, garantindo a interoperabilidade. Isto pode ser realizado através de dois procedimentos diferentes: codificação e configuração. Ambos foram testados e demonstraram ter pontos fortes e fracos.

Além disto, foi possível identificar pelo menos três parâmetros para a seleção dos elementos e qualificadores a serem manipulados. Os que foram adotados referiram-se aos atributos dos campos a serem customizados.

A possibilidade do acesso aos recursos de informação resultantes do capital intelectual disponível nas Unidades de Pesquisa forneceria uma ampla visão dos

“insumos” de informação pertencentes ao sistema de C&T nacional, possibilitando a geração de produtos, indicadores de acesso, uso etc., séries históricas e outros, que agregariam valor aos resultados produzidos em cada instituição, podendo ser utilizados para tomada de decisão, tanto no planejamento de projetos futuros, como para uma reorientação do papel da ciência e tecnologia visando o benefício de setores excluídos da sociedade brasileira.

Entretanto, o ponto de partida seria a definição do padrão de metadados mais apropriado. Para tanto, uma ampla discussão com todos os depositários de informação atribuiria um alto grau de confiabilidade na definição dos elementos e qualificadores necessários para melhor descrever essas informações.

É necessário valorizar e fazer uso da infra-estrutura tecnológica oferecida pelas ferramentas que trazem embutidos os componentes necessários para realizar a integração, dos repositórios institucionais, implementando a tecnologia de arquivos abertos, como é o caso do DSpace, que podem desempenhar um papel crucial nessa tarefa.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

BARTON, M. R.; WALKER, J. H. Building a business plan for DSpace... Journal of Digital Information, MIT Libraries' Digital Institutional Repository, v. 4, n. 2, maio 2003. Disponível em: < <http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Barton/>> Acesso em: jun. 2005.

BASEVI, T. BDJur Consortium – Juridical Digital Library: implementing DSpace in the Brazilian Judiciary. In: ICCO CONFERENCE ON ELECTRONIC PUBLISHING, 9. Proceedings. Leuven, Be: Katholieke UniverSiteit de Leuven, 2005.

CAFÉ, L.; MÁRDERO ARELLANO, M. A.; BARBOZA, E. M. F.; MELO, B. A.; NUNES, E. M. A. Repositórios Institucionais: nova estratégia de publicação científica na rede. In: ENDOCOM, 13, Belo Horizonte, MG, set. 2003. Anais. Belo Horizonte: 2003.

CROW, R. The case for institutional repositories: a SPARC position paper. [S.l.]: The Scholarly Publishing and Academic Resources Coalition, 2002. Disponível em: <<http://www.arl.org/sparc/IR/ir.html>>. Acesso em: junho 2005.

DSPACE system documentation: architecture. Cambridge, MA: [Massachusetts Institute of Technology](#); Palo Alto, CA: [Hewlett-Packard Company](#), 2005. Disponível em: <<http://www.dspace.org/technology/system-docs/architecture.html>>. Acesso em: 06 out. 2005.

DSPACE system documentation: contents. Cambridge, MA: [Massachusetts Institute of Technology](#); Palo Alto, CA: [Hewlett-Packard Company](#), 2005. Disponível em: <<http://dspace.org/technology/system-docs/>>. Acesso em: 06 out. 2005.

DSPACE system documentation: version history. Cambridge, MA: [Massachusetts Institute of Technology](#); Palo Alto, CA: [Hewlett-Packard Company](#), 2005. Disponível em: <<http://dspace.org/technology/system-docs/history.html>>. Acesso em: 06 out. 2005

DUBLIN Core Metadata Initiative: using Dublin Core. Disponível em: <<http://dublincore.org/documents/usageguide/>>. Acesso em: 10 out. 2005.

FERREIRA, S. M. S. P.; SOUTO, P. C. N. Federação de bibliotecas digitais lusófonas em ciências da comunicação. In. MARCONDES, C. H.; KURAMOTO,

H.; TOUTAIN, L. B.; SAYÃO, L. Bibliotecas digitais: saberes e práticas. Salvador: EDUFBA; Brasília: IBICT, 2005.

HARNAD, S. Free at last: the future of peer-reviewed journals. D-Lib Magazine, v. 5, n. 12, dez. 1999. Disponível em:

<<http://www.dlib.org/dlib/december99/12harnad.html>> Acesso em: maio 2005.

HEDSTROM, M.; MONTGOMERY, S. Digital preservation needs and requirements in RLG member institutions. [S.l.]: RLG, 1998. Disponível em:

<<http://www.rlg.org/preserv/digpres.html>>. Acesso em: junho 2005.

LAGOZE, C.; VAN DE SOMPEL, H.; NELSON, M.; WARNER, S. (eds.). The Open Archives Protocol for Metadata Harvesting: v. 2.0, 14 June 2002. Disponível em: <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>. Acesso em: out. 2005.

LITTLE, D. DSpace: SRB at UCSD libraries. Disponível em:

<[http://www.sdsc.edu/user\\_services/training/past/institute2004/PPT/3\\_7\\_DSpaceDL.ppt](http://www.sdsc.edu/user_services/training/past/institute2004/PPT/3_7_DSpaceDL.ppt)> Acesso em: 06 out. 2005.

LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. ARL Bimonthly Report, 26, 2003. Disponível em:

<<http://www.arl.org/newsltr/226/ir.html>>. Acesso em: maio 2005.

MÁRDERO ARELLANO, M. A. Repositórios DSpace e a preservação digital. In: ENDOCOM, 14, Porto Alegre, 2004. [Anais.] Porto Alegre: 2004.

MODESTO, F.; SEITI, R.; WEITZEL, S. R.; FERREIRA, S. M. S. P. Biblioteca digital lusófona em ciências da comunicação – BDCOM - DSpace versão 1.2 (beta 3): manual de utilização. São Paulo: INTERCOM, jul. 2004.

NORTON, R. J. Reuse of personal *software* assets: theories, practices, and tools. Tallahassee, FLA: Florida State University, 2005. Dissertação de Mestrado. Disponível em: <<http://etd.lib.fsu.edu/theses/available/etd-08282003-183319/>>. Acesso em: 10 out. 2005.

OKUBO, Y. Bibliometric indicators and analysis of research systems: methods and examples. Paris: OCDE/GD, 1997

PINFIELD, S.; GARDNER, M.; MACOLL, J. Setting up an institutional e-print archive. *Ariadne*, 31, 2002. Disponível em: <<http://www.ariadne.ac.uk/issue31/eprint-archives/intro.html>>. Acesso em: out. 2005.

RODRIGUES, E.; BAPTISTA, A. A.; RAMOS, I.; SARMENTO E SOUZA, M. F. Repositório – Implementing DSpace in Portuguese: lessons for the future and research pathways. In: ICCO CONFERENCE ON ELECTRONIC PUBLISHING, 8, Brasília – DF, Brasil, June 2004. **Proceedings**. Brasília: 2004.

WOLPERT, A. J. Institutional repositories: key policies. In: INSTITUTIONAL REPOSITORIES: A WORKSHOP ON CREATING AN INFRASTRUCTURE FOR FACULTY-LYBRARY PARTNERSHIPS, 18 October 2002, Washington, DC. [Papers]. Disponível em: <<http://www.arl.org/IR/wolpert/index.htm>>. Acesso em: jun. 2005.