

X CINFORM – Encontro Nacional de Ensino e Pesquisa em Informação
19 a 22 de setembro de 2011 • Porto Bello Hotels & Resorts • Salvador – Bahia
Responsabilidade social na representação, preservação e disseminação de conteúdos

Repositórios brasileiros implementados com DSpace **Brazilian institutional repositories implemented by DSpace**

Milton Shintaku

Doutorando em Ciência da Informação, IBICT/ UnB,
milton.shintaku@gmail.com

Sueli Mara Mara Soares Pinto Ferreira

Doutora em Ciência da Informação, USP, smferrei@usp.br

Jaime Robredo

Doutor em Ciência da Informação, UnB, Robredo@unb.br

Resumo: Há uma variedade de tipos de repositórios disponíveis na Web, desenvolvidos com vários tipos de softwares. Nesse contexto, o presente trabalho tem por finalidade analisar diferentes tipos de repositórios implementados com DSpace. Nesse caso, verificar a tipologia de repositórios passíveis de implementação com esse software. Para tanto, o presente estudo analisa questões relacionadas aos esquemas de metadados e interoperabilidade, com o propósito de, não apenas verificar a implementação, mas de analisar questões resolvidas por esses repositórios relacionados à catalogação e interoperabilidade. Visto que a im

Palavras-chave: Repositorios; DSpace; Esquemas de metadados; Interoperabilidade

Abstract: There are a lot of types of repositories available on the Web, developed by various types of software. In this context, the main of this paper is to analyze different types of DSpace repositories implemented by DSpace. In this case, check the implementation of repository developed by this software. Then, this study examines issues related to metadata schemas and interoperability, to verify the implementation and resolved problems with description and interoperability

Keywords: Repositories; DSpace; Metadata schema; Interoperability

Introdução

Repositório, no ambiente Web, pode ser definido de várias formas, dependendo da finalidade. Em todos os casos, apresentam-se como iniciativas para armazenamento e compartilhamento de objetos digitais. Entretanto, diferem-se quanto ao tipo de acesso, podendo ser restrito ou amplo. Outro diferencial é a questão financeira, uns repositórios podem ser de uso comercial, ou seja, ter acesso pago, e outros de uso gratuito.

A finalidade do repositório, por sua vez, transcende a questão da pura iniciativa que gerencia objetos digitais. Na maioria das vezes, os repositórios implementam diversas facilidades, que de forma particular, os contextualizam com o tipo de acervo e de usuários. Criando, dessa forma, tipos de repositórios baseado no acervo, público alvo e tipo de facilidades oferecidas.

Atualmente, um tipo de repositório muito comum na Web é o repositório de softwares, que pode assumir configurações diversas. Eshkevari, Arnaoudova, e Constantinides (2008) declaram que um repositório de software mantém um acervo composto por diferentes tipos de objetos, de programas fontes, documentação, até manuais. Assim, mantendo não apenas o software, em si, mas informações sobre configuração, modificações e a relação entre os componentes do software.

Segundo van Vugt (2008) um repositório de software pode ser considerado uma fonte para instalar programas. Tornando o repositório parte indispensável aos procedimentos de instalação e manutenção dos sistemas informatizados. Por esse motivo, os repositórios de distribuição de software, nos dias atuais, são essenciais para, criação, manutenção e atualização dos parques informatizados.

Nesse contexto, os repositórios de software possuem um acervo compostos de itens complexos. Em que cada item, desse acervo, possui um software como núcleo, em seus diferentes níveis de complexidade, juntamente com a documentação. Assim, facilitando a reutilização, atualização, distribuição, compartilhamento, preservação, controle, entre outros, desses softwares.

Outro tipo comum de repositório, disponíveis na web, são os repositórios científicos. Esses repositórios possuem acervo composto de documentos científicos. Linch (2003) considera os repositórios acadêmicos como um serviço de gerenciamento de disseminação de literatura científica. Viana, Márdero Arellano e Shintaku (2005), destacam o serviço de preservação dos documentos nos repositórios científicos.

Swan (2005) destaca o autoarquivamento como possibilidade de alimentação dos repositórios, contextualizando-o com os sistemas mais interativos, em que os depósitos são efetuados pelos próprios autores. Assim, a facilidade de autoarquivamento alia-se as outras facilidades disponíveis nos repositórios, como as de recuperação de documento e controle de acesso.

Os repositórios científicos podem ser classificados, de acordo com a origem dos documentos, como institucionais ou temáticos. Enquanto nos repositórios institucionais, todos os autores possuem vinculação com a instituição mantenedora do repositório, os repositórios temáticos aceitam documentos provenientes de várias instituições, desde que possuam um assunto compatível com os tópicos tratados pelo repositório.

Os repositórios temáticos têm como principal característica um acervo composto por documentos que abrangem uma área do conhecimento. Para Kuramoto (2006), os repositórios temáticos, mantidos por uma entidade, oferecem serviços para a gestão e disseminação de documentos técnicos e científicos vinculados a uma área ou subárea do conhecimento.

Os repositórios temáticos tem como marco a implantação do ArXiv, um repositório temático, especializado em ciências exatas, mais precisamente em física e disciplinas correlatas. Para Haque e Ginsparg (2009) o ArXiv faz a mediação entre a literatura e a comunidade científica, permitindo o acesso amplo aos artigos em formato de pré e pós-print.

Café et al. (2003) revela que os repositórios temáticos tem sido uma solução à disseminação de resultados de pesquisa. Dessa forma, apresentando-se como uma ferramenta que facilita o acesso à literatura científica, ampliando o alcance dessa literatura.

Os repositórios institucionais, por outro lado, agregam a produção intelectual de uma instituição, sem, necessariamente, ter um acervo vinculado a um tópico. Para Ware (2004), o repositório institucional, em geral, deve estar disponível em Web e ter acervo acumulativo e preservado, de acesso livre e interoperável.

Os repositórios institucionais tornam-se, assim, ferramentas para preservação e disseminação da produção intelectual de uma instituição. Bjorg (2005), em seu modelo para comunicação científica, coloca os repositórios como facilitador de acesso à informação científica, ampliando o processo de comunicação científica para além da publicação.

Leite e Costa (2006) revelam a possibilidade do uso de repositórios como ferramenta de gestão do conhecimento. Assim, coloca o repositório não apenas como disseminado, mas como parte do processo de compartilhamento da informação. Baseados na espiral de Nonaka e Takeuchi (1979) discutem o papel do repositório em processos como socialização, externalização, combinação e externalização do conhecimento.

Por fim, mas não menos importantes, os repositórios educacionais. Esses repositórios gerenciam um acervo composto por objetos educacionais, tradução para object learning, também definidos como objetos de aprendizagem, com algumas diferenças. No cadastro do ROAR (2011), quinze repositórios se declaram como repositórios educacionais e utilizam o DSpace como base, que revela a importância dessa iniciativa, pois já aparece como uma opção de classificação.

Os objetos educacionais são entidades em formato digital que apóiam o processo de aprendizado (IEEE, 2002). Esses objetos podem ser utilizados, reutilizados ou apenas referenciado durante o processo de aprendizado, intermediado pela tecnologia. Para o padrão de metadados LOM (Learning Object Metadata) os objetos educacionais podem ser classificados conforme a sua granularidade, ou nível de agregação, iniciando no nível um até o quatro.

Nesse contexto, os repositórios educacionais têm por finalidade o compartilhamento de objetos educacionais, com vista a sua reutilização. Essa reutilização, por sua vez, permite a adaptação do objeto educacional a outros contextos, ou melhoria, que pode resultar em uma mudança de nível. Por esse motivo, os repositórios educacionais apresentam-se com opção atual como apoio ao ensino.

Independentemente do tipo do repositório, nota-se que essas iniciativas têm por finalidade manter e recuperar objetos digitais, dos mais variados tipos. Para tanto, é necessário manter, não somente o objeto digital em si, mas, também, a sua descrição, nos vários aspectos. Os repositórios, para essa descrição, fazem uso de metadados..

ESQUEMAS DE METADADADOS

De forma mais simplista, diz-se que metadados é a informação estruturada sobre um recurso informacional, ou seja, dado sobre dado. NISO (2004) expande essa definição para informação estruturada que descreve, explica, localiza, facilita a localização, usa ou gerencia uma fonte de informação, ou seja, enumera os processos que os metadados podem operar.

Em muitos casos, o metadado é a única forma de recuperação de um objeto digital armazenado em sistemas de informação. Em caso de objetos não textuais, a recuperação pelos buscadores atuais atua somente na indexação dos metadados. Assim, os metadados são pontos de recuperação dos objetos mantidos na Web e sua qualidade é um fator de impacto à recuperação.

Nesse contexto, os metadados podem ser classificados como descritivos, estruturais e administrativos. Os metadados descritivos contêm elementos de descrevem o objeto, enquanto os estruturais indicam sua posição em um conjunto. Para os administrativos, podem conter informações sobre direitos de acesso e preservação.

Ikematu, por sua vez, divide os metadados em duas categorias, os metadados técnicos e os metadados de negócios. Os metadados técnicos englobam as informações necessárias para que ferramentas automatizadas possam manipular o objeto. Os metadados de negócio, por sua vez, descrevem o objeto conforme a finalidade do negócio.

Dessa forma, pode-se ver que os metadados podem ser classificados quanto ao seu conteúdo ou quanto a sua finalidade. Por exemplo, Márdero Arellano (2004) descreve que os metadados de preservação, mesmo sendo parte dos metadados descritivos, possuem informações técnicas que garantam a preservação.

Caplan (2003) ressalta que os metadados não podem ser simplesmente, acumulados ou representar um conjunto de elementos, mas devem ser gravados conforme um esquema de metadados documentado. Dessa forma, os metadados são organizados em esquemas que padronizam os seus elementos, permitindo uma compreensão melhor da utilização desses elementos.

Intner, Tseng e Larsgaard (2003) descrevem os componentes de um esquema de metadados como, os elementos, as regras e os atributos. Sendo que os elementos são as unidades informacionais, com semântica própria e independente. As regras definem o comportamento dos elementos, aplicando restrições ou indicando se o elemento é repetível, opcional, obrigatório ou recomendável. Os atributos qualificam os elementos, dando informações suplementares ou especificando o elemento.

Para Nelson (2008) os esquemas de metadados são úteis para detalhar o conteúdo marcado de forma econômica, pois os esquemas de metadados não precisam ser tão descritivos. Nesse sentido, deve apenas organizar a informação, mas não restringir o seu uso a especificações muito rígidas.

Para organizar e padronizar os metadados há vários esquemas de metadados disponíveis. Cada esquema com particularidades que adequam-se a tipos específicos de objetos digitais, entretanto, na maioria dos casos possuem um conjunto de metadados descritivos. Assim, os esquemas de metadados asseguram a descrição de um objeto digital, apoiando a recuperação informatizada desse objeto.

O quadro abaixo apresenta alguns dos esquemas de metadados mais comuns na descrição de objetos digitais e utilizados em interoperabilidade entre sistemas de informação. Esses esquemas, mesmo com suas peculiaridades, possuem características mais descritivas e podem ser representados em formato XML, por esse motivo, mais propícios à utilização em processos de intercâmbio de metadados.

Quadro explicativo dos esquemas de metadados		
Esquema	Mantenedor	Descrição
Dublin Core	Dublin Core Metadata Initiative - DCMII http://dublincore.org/	Simples e flexível, possui apenas 15 elementos e apenas o título é altamente recomendável. Os elementos podem ser qualificados, aumentando a especificação. Por exemplo, o elemento "contributor" que representa pessoa ou entidade que participou da elaboração do documento, pode ser qualificado com "advisor" para orientador, "illustrator" para ilustrador, entre outros. Se usado apenas os elementos pode ser denominado de DC ou OAI_DC, se for qualificado é denominado de QDC.
METS	Library of Congress http://www.loc.gov/standards/mets/	<i>Metadata encoding & Transmission Standard</i> - METS é um esquema de metadados multinível, que por um lado o torna mais complexo, por outro permite agrupar informações. Possui representação em XML, possuindo sete divisões: cabeçalho, metadados descritivos, metadados administrativos, seção de diretórios, mapa estrutural, links estruturais e comportamento.

			Possui uma finalidade maior para descrever livros digitalizados, mas pode ser utilizado para outros tipos de documentos digitais.
RDF	Consórcio W3C http://www.w3.org/RDF/	-	<i>Resource Description Framework</i> é um modelo padrão para intercâmbio de informações pela Web, podendo, inclusive, estar internamente em outros esquemas. Possui quatro divisões básicas: ICS – <i>Internet Content Selection</i> , os metadados simples da Web, mapas de sites e outras ferramentas e P3P – <i>Platform Privacy Preferences Project</i> . Nos metadados simples é utilizado o Dublin Core, pela facilidade de implementação. Nota-se a abrangência das informações centradas para os recursos da Web.
ETD-MS	NDLTD - Networked Library of Theses and Dissertations http://www.ndltd.org/	Digital ans	<i>Electronic Theses and Dissertations - Metadata Schema</i> é uma extensão do <i>Qualified Dublin Core - QDC</i> . Adiciona um novo elemento chamado de <i>theses.degree</i> (tendo os outros elementos mantido a forma padrão de <i>dc.elemento</i>), que descreve particularidades das teses e dissertações. Com quatro qualificadores, <i>name</i> (para o nome do grau obtido), <i>level</i> (para o nível do grau obtido), <i>discipline</i> (para a área de formação) e <i>grantor</i> (para a instituição vinculada com a tese ou dissertação). O elemento <i>degree</i> e todos esses qualificadores são opcionais e repetíveis.
MTD-BR	Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT		O Padrão Brasileiro de Metadados de Teses e Dissertações é extremamente complexo, multinível, que permite agregar muitas informações sobre teses e dissertações brasileiras. Desenvolvido no projeto da Biblioteca Digital de Teses e Dissertações - BDTD, com o propósito de oferecer produtos e serviços baseados em metadados. Possui informações que descrevem não apenas o documento, mas a biblioteca, autor, contribuidores e outros.

Electronic Theses and Dissertations - Metadata Schema é uma extensão do *Qualified Dublin Core - QDC*. Adiciona um novo elemento chamado de *theses.degree* (tendo os outros elementos mantido a forma padrão de *dc.elemento*), que descreve particularidades das teses e dissertações. Com quatro qualificadores, *name* (para o nome do grau obtido), *level* (para o nível do grau obtido), *discipline* (para a área de formação) e *grantor* (para a instituição vinculada com a tese ou dissertação). O elemento *degree* e todos esses qualificadores são opcionais e repetíveis.

MTD-BR Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT
O Padrão Brasileiro de Metadados de Teses e Dissertações é extremamente complexo, multinível, que permite agregar muitas informações sobre teses e dissertações brasileiras. Desenvolvido no projeto da Biblioteca Digital de Teses e Dissertações - BDTD,

com o propósito de oferecer produtos e serviços baseados em metadados. Possui informações que descrevem não apenas o documento, mas a biblioteca, autor, contribuidores e outros.

A padronização, com o uso de um esquema de metadados, permite, entre outros, a interoperabilidade entre sistemas de informação. Assim, implementando uma das características mais marcantes dos arquivos abertos. Por esse motivo, a implementação de um esquema de metadados conhecido possibilita troca de informação entre sistemas informatizados, que implementem o mesmo esquema.

ARQUIVOS ABERTOS E INTEROPERABILIDADE

Os arquivos abertos, muitas vezes, são confundidos com o movimento de acesso aberto, principalmente por terem as mesmas siglas, tanto em português (AA, Acesso Aberto, Arquivo Aberto) quanto em inglês (OA, Open Access e Open Archives). Apesar das semelhanças nota-se que o acesso aberto possui um viés mais político, enquanto o arquivo aberto é uma questão mais técnica.

Os arquivos abertos têm como marco o congresso de Santa Fé, nos Estados Unidos, em 1999. Esse evento definiu os conceitos básicos que orientou os planejamentos dos serviços voltados ao acesso aberto, baseados em depósito de documento digital. Esses serviços, por sua vez, basearam-se na interoperabilidade de iniciativas que mantêm acervos de acesso aberto (SANTA FE CONVENTION, 1999).

Sobre os documentos digitais, o evento ressaltou a necessidade do autoarquivamento e da preservação. Nesse caso, apresenta duas etapas necessárias as iniciativas que implementam os arquivos abertos. Primeiro o autoarquivamento, que é a possibilidade do autor executar o depósito, permitindo a facilidade de interação com o sistema. Segundo, a manutenção dos documentos digitais, preservando-os por longos períodos, mas sempre com a possibilidade de acesso.

Para orientar as iniciativas, em relação ao processo de submissão e preservação, políticas devem ser implementadas. Assim, o gerenciamento das iniciativas devem ser baseadas em regras claras que orientam o funcionamento da iniciativa que implementa o acesso aberto, ou seja, políticas. Essas orientações devem reger todos os processos disponíveis nas iniciativas.

Em relação à interoperabilidade entre iniciativas que implementam os arquivos abertos é necessário dois elementos, um que fornece os dados e outro que coleta os dados. Assim, foram conceituados os provedores de dados e provedores de serviços. Os provedores de serviço coletam, automaticamente, os metadados, enquanto os provedores de dados permitem a coleta.

Por fim, um conjunto mínimo de metadados capaz de descrever o documento digital. Com o depósito do documento digital em seu formato completo é preciso descrevê-lo por metadados, de forma completa, com o propósito de apoiar a recuperação. Nesse contexto, a união do documento digital e o conjunto de metadados que o descreve é denominado de registro.

Atualmente, os conceitos do congresso de Santa Fé foram atualizados, mas continuam focados na interoperabilidade. Para Triska e Café (2001) “A interoperabilidade é a palavra-chave para a criação dos arquivos abertos”. No entanto, para implementar a interoperabilidade acordou-se na utilização do protocolo Open Archives Initiative - Protocol Harvesting Metadata - OAI-PMH.

A troca de informação entre iniciativas, denominadas de repositórios, dá-se através do uso do protocolo OAI-PMH, com as informações em formato XML. Essa troca de informação, feita por processo de coleta automática de metadados ou harvesting, funciona como em um processo de requisição e resposta. Assim, a resposta a uma coleta será um registro em formato XML, contendo as informações do documento digital.

A interoperabilidade se concretiza através de um esquema de metadados codificado em formato XML. Assim, torna-se independente a forma que o repositório (iniciativa que

implementa os arquivos abertos) opera, desde que responda a coleta com os registros em formato XML com um esquema de metadados conhecido e validável.

DSPACE

Repositórios científicos, temáticos, educacionais e outros podem ser implementáveis de várias formas. Atualmente, os softwares livres tem colaborado para essas implementações, substituindo o processo de desenvolvimento pelo de adaptação. Assim, tornando a implementações mais rápidas e funcionais, visto que o software cria uma comunidade de usuários, permitindo o compartilhamento de conhecimento sobre o software.

Dos diversos softwares disponíveis para implementação de repositórios, segundo o Registry of Open Access Repository - ROAR, o DSpace é o mais utilizado com 42% das implementações mundiais, pouco mais de 900 repositórios. Mesmo que o cadastramento no ROAR seja opcional, o percentual revela a grande penetração desse software. Como o ROAR é dinâmico, com resultados variando conforme os cadastramentos, as informações apresentadas são de julho de 2011.

No Brasil, o software DSpace foi a plataforma escolhida pelo projeto de implementação de repositórios institucionais, coordenado pelo IBICT e financiado pela Financiadora de Estudos e Projetos - FINEP. Esse projeto distribuiu, entre outros, equipamentos com o DSpace pré-instalado, facilitando a implementação de repositórios, abrangendo mais de 40 instituições de ensino superior e instituições de pesquisa.

Para Baudoin e & Branschovsky (2004) o Dspace é um software para implementar repositórios, um projeto e um serviço de informação. É um software livre de código aberto, que significa que é gratuito e os programas fontes são distribuídos, permitindo alterações mais profundas. Um projeto mantido pelo Duraspace e pela comunidade usuária, em muitos casos, inovações e correções são efetuadas pela comunidade.

O DSpace como serviço, no entanto, merece um destaque, pois apresenta-se como um sistema de informação, que oferece serviços de compartilhamento de documentos digitais, facilidades de recuperação de documentos, ferramentas de submissão e interoperabilidade, entre outros. Esse conjunto de serviços está centrado na comunicação científica e o papel dos repositórios científicos, entretanto, há indícios que são úteis para outros tipos de repositórios.

Nota-se que todas as facilidades estão intimamente integradas. Como exemplo pode-se ver a questão da organização do repositório, que possível no DSpace. Organizada hierarquicamente em comunidades, coleções e itens. As comunidades podem representar estrutura organizacional ou grandes temas. As coleções, por sua vez, representam tipologia documental. Os itens são as unidades informacionais, composta pelos metadados e os objetos digitais.

As permissões e restrições são implementadas nos elementos da organização. Dessa forma, pode-se restringir todo o repositório, uma comunidade, uma coleção, ou somente um Item. Nota-se que as facilidades de segurança foram desenvolvidas para atuarem na organização. As restrições e permissões são implementadas tanto na submissão quanto na recuperação de documentos digitais.

Nesse sentido, o serviço de submissão só é executado em coleções. Assim a forma de organização afeta esse serviço. Da mesma forma, podem-se listar todos os documentos de uma coleção ou comunidade, portanto, a organização pode facilitar a recuperação. Outras facilidades de recuperação como listagens e ferramentas de buscas, também, podem ser aplicadas na organização do repositório.

A interoperabilidade, por sua vez, no Dspace é implementado tanto como provedor de serviço, quanto provedor de dados. Em outras palavras, repositórios implementados com DSpace podem, tanto coletar metadados de outras iniciativas de acesso aberto, quanto permitir que coletem metadados do acervo.

No que concerne ao processo de coletar metadado de outras iniciativas de acesso aberto, no DSpace só é possível fazê-lo em uma coleção, ou seja, ao criar, ou alterar uma

coleção, informa-se que a origem dos documentos dessa coleção será proveniente de um processo de harvesting.

A coleta de metadados, por sua vez, pode ser feita em todo o acervo, ou seja, coletar todos os metadados do acervo ou apenas os metadados dos documentos de coleção. Assim, cada coleção, para o processo de harvesting, se torna um set. Assim, a organização também afeta o processo de disponibilizar os metadados para a coleta.

Nesse contexto, muito provavelmente pelas facilidades e serviços, o DSpace é tão utilizado mundialmente. Dentre os repositórios implementados com DSpace destaca-se o Repositório Institucional da Universidade de Cambridge, pelo quantitativo de seu acervo. Atualmente com mais de 190 mil documentos em julho de 2011.

Para os países lusófonos, o Repositório, da Universidade do Minho é o repositório de língua portuguesa com maior visibilidade. Um conjunto de fatores torna esse repositório um destaque mundial. Não apenas o acervo, que passa muito os onze mil documentos, mas a posição de vanguarda faz desse repositório um modelo para os repositórios brasileiros;

REPOSITÓRIOS IMPLEMENTADOS COM DSPACE NO BRASIL

O presente estudo apresenta a análise de alguns repositórios brasileiros implementados com DSpace. Sem juízo de valor, os repositórios foram selecionados por serem representativos dentre as suas categorias. Assim revelando um panorama dos repositórios brasileiros implementados com o Dspace.

O presente estudo verificou elementos como estrutura informacional, tipologia de documentos, metadados e implementação do protocolo OAI-PMH para interoperabilidade. Assim, levantando informações que permitem criar um panorama sobre os repositórios no Brasil, que foram implementados com DSpace.

A análise de iniciativas distintas que utilizam o mesmo software, também, permite verificar a flexibilidade desse software para uso em diferentes contextos. A adaptabilidade do software embasa-se na tendência atual de sistemas flexíveis e customizáveis, de forma a se moldar as necessidades de cada instituição, sem a necessidade de desenvolvimento de sistemas proprietários.

Os repositórios institucionais científicos são a maioria, daqueles implementados com DSpace. Um levantamento no portal do Dspace no Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT (<http://dspace.ibict.br>) tem-se cadastrado 22 repositórios de universidades e institutos de pesquisa. É importante citar que esse cadastro não contempla todos os repositórios desenvolvidos com base no DSpace, mas representa uma boa amostra.

O repositório ARCA da Fundação Oswaldo Cruz - FIOCRUZ (<http://www.arca.fiocruz.br/>) possui características temáticas e institucionais. Temática por possuir certa unidade sobre os tópicos nos documentos do acervo, institucional por agregar a produção dessa instituição. Essa característica é comum em repositórios institucionais de unidade de pesquisa, que diferentes de repositórios institucionais de universidade não são multitemáticos.

Seu acervo representa a memória digital da produção técnico-científico da FIOCRUZ. Assim, sua estrutura reflete essa preocupação, representando as diversas entidades que compõe a FIOCRUZ, com seus institutos, centros, editora e revistas. A tipologia de documentos, por sua vez, abrange desde artigos publicados em periódicos científicos até relatórios técnicos, cobrindo uma vasta tipologia.

Pela amplitude de tipos de documentos, os metadados são contextualizados, ou seja, diferenciados para cada tipo de documentos. A utilização dos metadados relacionados ao ETD-MS, para as coleções de teses e dissertações, facilita a interoperabilidade com iniciativas internacionais de teses e dissertações, além de descrever melhor o documento.

Cabe relatar, também o uso de vocabulários controlados baseado nos descritores em Ciências da Saúde - DECS para a catalogação dos documentos digitais (MARTINS, 2009). Assim, normalizando campos de metadados e facilitando a organização da informação, além de facilitar a recuperação.

Quanto à interoperabilidade, o sistema ARCA, disponibiliza os metadados para coleta em vários esquemas de metadados, tais como o OAI-DC, RDF e METS. Cabe salientar o vanguardismo desse sistema em implementar o Open Archives Initiative - Object Reuse and Exchange - OAI-ORE, permitindo a disseminação do objeto digital, não apenas de seu metadado, salientando que o documento é mais importante que o sistema de informação.

O repositório da Universidade de Brasília - UnB (<http://repositorio.bce.unb.br/>) é um repositório multitemático que agrega a produção científica de estudiosos vinculados com a UnB. Mantido pela Biblioteca Central da UnB - BCE - UnB, tem como propósito disseminar a produção acadêmica dessa universidade e, assim, preservar esses documentos, além de melhorar a visibilidade dos seus estudiosos (UnB).

Com acervo tipicamente científico, contempla os documentos avaliados pelos pares, como os artigos de periódicos, teses e dissertações, entre outros. Sua estrutura reflete a organização da UnB, com comunidades representando as unidades, como faculdades e institutos, e as coleções a tipologia de documentos.

Utilizam os mesmos metadados para todos os tipos de documentos. Nesse caso, simplificando em um conjunto mínimo, mas significativo na descrição dos documentos. A não particularização dos metadados pode restringir na descrição, mas, com certeza facilita o processo de catalogação e simplifica o processo.

O Repositório da UnB - RiUnB responde ao harvester com os esquemas de metadados mais tradicionais. Respondendo com os esquemas OAI-DC, RDF e METS atende com os principais esquemas de metadados. Nesse caso, cumpre os principais requisitos para a interoperabilidade.

O repositório da Universidade Federal do Rio Grande do Sul, LUME (<http://www.lume.ufrgs.br/>), conta com acervo superior a 28 mil documentos. Como todo repositório universitário é multitemático. Assim, representando bem a produção intelectual da instituição.

A estrutura informacional do repositório difere um pouco dos outros repositórios. Enquanto a maioria dos repositórios reflete a estrutura da universidade no repositório, o LUME cria uma estrutura baseada na tipologia dos documentos que compõe o seu acervo. Essa estrutura, por sua vez, facilita, em alguns casos, a navegação.

O acervo do LUME é bem variado, com teses e dissertações, artigos e fotografias históricas, entre outros. Assim, os metadados são particularizados conforme a coleção. Nota-se na coleção de teses e dissertação o uso do elemento do ETD-MS, que descrevem melhor esses documentos. Da mesma forma, nas fotografias históricas a utilização de metadados que contextualizam a fotografia em seu período histórico.

A interoperabilidade, por sua vez, é possibilitada por vários esquemas de metadados. O LUME responde ao harvester com esquemas de metadados mais comuns como o OAI-DC, RDF, QDC e METS. Cabe salientar que o LUME responde, também, com o esquema MTD-BR. Essa opção permite que o LUME responda ao sistema da Biblioteca Digital de Teses e Dissertação - BDTD.

Os repositórios não acadêmicos, por sua vez, mesmo que ainda em minoria, possui representantes de destaque. Por abranger várias áreas e possui uma amplitude maior de tipos de documentos, esses repositórios tem se tornado comuns, principalmente em instituições públicas que desejam disponibilizar documentos digitais.

A BDJur foi primeiro repositório não acadêmico brasileiro implementado com o Dspace. Disponibilizado em 2004, esse repositório possui características temáticas e institucionais, semelhantes àqueles mantidos por institutos de pesquisa. Nesse caso é um repositório institucional do STJ, com acervo temático relacionado a documentos de caráter administrativo e jurídico.

A organização da BDJur compreende três partes, administração, doutrina (jurídico) e memória. Com tipologia própria dos documentos administrativo e jurídico, possui parte do acervo restrito, pois tem como propósito o compartilhamento, tanto interno, quanto externo.

Posteriormente foi criado um consórcio BDJur (<http://www.consorciobjur.gov.br/>), que agrega informações de outros repositórios. Atualmente conta com documentos

provenientes de seis repositórios, que possuem acervo composto por documentos de características jurídicas. Esse consórcio funciona baseado em harvesting, ou seja, como um provedor de serviço que oferece facilidades de recuperação de informação dos repositórios coletados.

O Acervo Tom Jobim (<http://www.jobim.org/jobim>) é um repositório, que como o nome revela, compõe-se de documentos digitais relacionados ao compositor brasileiro “Tom Jobim”. Mantido pelo Instituto Antonio Carlos Jobim, esse repositório se apresenta com características que podem ser associadas a acervos de museus, com o diferencial de ser digital e multimídia.

Os repositórios musicais são comuns na Web, na maioria dos casos, são iniciativas comerciais e disponibilizam apenas as músicas. O Acervo Tom Jobim, por sua vez, possui uma proposta diferenciada. Disponibiliza informações sobre o autor e possui várias diferenças, se comparado com os repositórios acadêmicos.

Como é um acervo relacionado a uma pessoa, a organização do repositório centra-se em organizar documentos relacionados ao Tom Jobim. Assim, a organização baseia-se em tipografia de documentos, no nível principal, com refinamentos nos outros níveis, semelhante às coleções. Por exemplo, no nível principal “Partituras” o segundo nível é “Publicados” e “Manuscrito”. Essa estrutura organizacional foi descrito pelo repositório como “Conjuntos e Coleções” no lugar de “Comunidades e Coleções”.

Com um acervo de 3021 itens, o Acervo Jobim agrega uma variedade de tipos de documentos digitais, é um repositório tipicamente multimídia. Por centrar-se na questão de acervo, possibilita o depósito de uma vasta gama de tipo de objetos digitais e tipos de documentos, desde que possua relacionamento com o Tom Jobim.

Quanto à interoperabilidade, não foi possível verificar a sua implementação. Possivelmente, por não ser a finalidade desse repositório, é bem possível que esse repositório não seja interoperável. Nota-se que a possibilidade de disponibilizar a interoperabilidade é opcional e depende das políticas de informação da instituição mantenedora do repositório.

O Instituto Antonio Carlos Jobim, atualmente, mantém outros acervos, como o Acervo Caymmi, Acervo Lúcio Costa e Acervo Chico Buarque. Possivelmente pela facilidade técnica de implementação desses acervos, essa iniciativa tem expandido para abranger obras e documentos digitais relacionados a outros grandes músicos. Assim, as dificuldades situam-se apenas nas questões de curadoria do acervo.

Dentre as implementações com DSpace, vale a pena destacar a Brasileira (<http://www.brasiliana.usp.br/>), que também possui características de acervo. Entretanto, Ao contrário do Acervo Tom Jobim, a Brasileira representa o acervo da Biblioteca Brasileira Guita e José Mindlin, do Instituto de Estudos Brasileiros - IEB, com o acervo digitalizado diretamente dos originais em papel, pela equipe da Brasileira.

Mantida pela Universidade de São Paulo - USP, tem início em 2007 e abriga obras raras, permitindo o acesso de livros raros a todos via Web gratuitamente (GOMI, 2009). Para os seus coordenadores, a Brasileira é um instrumento de democratização da informação, podendo ser instrumento educativo, visto que, permite o acesso de obras de valor inestimável de forma fácil via Web.

Quanto às questões técnicas, pode-se observar várias implementações que incrementam as funcionalidades padrão do DSpace. A agregação das facilidades do Software de gerência de conteúdo DRUPAL, aumentou o dinamismo na apresentação de informações como notícias e outros. Assim, disponibilizando uma página inicial mais amigável.

A Brasileira possui uma organização baseada na tipologia de documentos que compõe o acervo. Essa organização é mais condizente para a apresentação de museus, visto que, é composto por obras raras. Com 2586 documentos, compõem-se de livros, manuscritos, mapas, periódicos e imagens. Cabe salientar a utilização de softwares visualizadores de imagens, que fornece facilidades de navegar pelas imagens, aproximando, afastando ou indo a um ponto específico da imagem.

Quanto à interoperabilidade não foi possível detectar a sua implementação. É muito provável que essa facilidade não esteja disponível, visto que, por ser um sistema de informação muito específico é muito difícil encontrar outros sistemas que possa interoperar.

O Banco Internacional de Objetos Educacionais - BIOE é mantido pelo Ministério da Educação - MEC. Mesmo sendo mantido por uma instituição reconhecida, possui características mais temáticas, pois seu acervo possui objetos digitais de proveniências diversas, ou seja, não representa-se como produção intelectual do MEC. Assim, como o acervo possui um eixo temático, torna-se um repositório temático de educação.

A organização do BIOE espelha o sistema educacional brasileiro, abrangendo todos os níveis, educação infantil, ensino fundamental, médio, superior e profissional. Assim, facilita a navegação do usuário classificando os objetos educacionais conforme o nível educacional de aplicação. As coleções, por sua vez, representam a vasta tipologia que compõe o acervo, que no repositório é denominado de tipo de recurso.

O acervo do BIOE compõe-se de objetos educacionais, que nesse caso, é uma tradução do inglês "learning object". Assim possui objetos educacionais nos quatro níveis de complexidade apresentado por esse tipo de objeto, de simples figuras e mapas até programas de simulação. Com um quantitativo vasto, com mais de quatorze mil objetos educacionais, em várias línguas.

Por abranger uma tipologia diversa e descrever objetos educacionais, os metadados do BIOE foram baseados no esquema de metadados Learning Object Metadata - LOM, mesmo que utilize o esquema de metadados Dublin Core. Destaque para o campo de metadado relacionado aos objetivos educacionais e ao campo da origem, que contextualiza o objeto educacional e indica a proveniência do objeto.

Dentre os repositórios não acadêmicos educacionais, o BIOE, é um dos poucos que está aberto à coleta automática de metadados. Respondendo ao harvester com os principais esquemas de metadados, OAI-DCC, RDF e METS, o BIOE também fornece um webservice que permite a conexão automática com o BIOE.

Outro repositório educacional mantido pelo MEC é o Acervo Digital Acessível - ADA. Esse repositório, com aspectos temáticos, possui semelhanças de metadados e estrutura com o BIOE, mas seu acervo é restrito aos usuários, que estão vinculados aos centros de educação para os portadores de deficiência visual, possuindo apenas uma área pública com artigos, legislação e outros de acesso livre.

O grande diferencial desse repositório trata-se das implementações para acessibilidade para deficientes visuais. Essas facilidades agregadas ao DSpace contemplam ferramentas de ajuda para portadores de baixa visão e cegueira. Assim implementa aumento de fonte, autocontraste e teclas de atalho. Entretanto, o maior diferencial está na limpeza do código, facilitando a leitura por programas.

O ADA, por ter acervo restrito não implementa a interoperabilidade, mesmo utilizando o DSpace. Entretanto, revela a capacidade desse software para uso de repositório educacional acessível aos portadores de deficiência visual. Destaca-se, também, a questão da utilização para acesso restrito, mesmo que os usuários se espalhem por todo o Brasil.

CONSIDERAÇÕES FINAIS

O presente estudo revelou a diversidade de tipo de repositórios que podem ser implementados com DSpace. Nesse caso, apresentando a flexibilidade e adaptabilidade do software em ajustar-se para finalidades tão distintas. Dos repositórios para disseminação e preservação de documentos científicos até documentação técnica ou administrativa, o DSpace revela-se adaptável a várias outras finalidades, desde que esteja contextualizando no propósito de compartilhamento de documentos digitais.

A variedade de metadados presentes nos repositórios revelou a necessidade de particularizar as descrições para os diferentes tipos de documentos digitais. Por um lado revela uma riqueza na descrição, com detalhes mais específico dos documentos, que facilitam a recuperação, por outro lado, podem ocorrer problemas de normalização, criando variações de conteúdos que podem afetar futuras implementações.

Quanto a interoperabilidade, as variedades de esquemas de metadados possíveis ao harvesting, revela que a questão de conjunto mínimo de metadados apresentando nos arquivos abertos apresenta-se resolvido. Tradutores de esquemas de metadados denominados de “crosswalk” estão sendo utilizados para fazer a interface entre o esquema de metadados interno ao repositório e a interoperabilidade. Assim o esquema de metadados interno ao repositório não afeta a troca de informação.

A diversidade de nomenclatura das iniciativas implementadas com DSpace, por outro lado, apresenta uma questão mais relacionada à finalidade. Mesmo que apresente alguns problemas conceituais, mostra que as iniciativas acadêmicas preferem o termo repositório, ao passo que as iniciativas não científicas podem assumir nomenclaturas diversas. Podem-se destacar os “acervos”, que se centram na própria característica da formação das suas coleções.

Nesse contexto, o DSpace revela-se como opção viável à implementação de repositórios de finalidades diversas. A flexibilidade desse software aberto de código livre apresenta-se como opção para implementação de sistemas de informação de forma rápida, mudando a questão de desenvolvimento de sistemas para a adaptação de ferramentas existentes.

REFERÊNCIAS

BDJUR, Manual técnico.

Baudoin, Patsy and Margret Branschofsky, "[Implementing an Institutional Repository: The DSpace Experience at MIT](#)" (pdf, available at MIT only), Science & Technology Libraries, Volume 24, Issue 1/2, June 2004, p.31-45.

CAFÉ, L. et. al. Repositórios institucionais: nova estratégia para publicação científica na Rede. In: ENDOCOM – Encontro Nacional de Informação em Ciências da Comunicação, 13., 2003.

CAPLAN, P. Metadata Fundamentals for all Librarians. Chicago: American Library Association, 2003

ESHKEVARI, L. M. ARNAOUDOVA, V. CONSTANTINIDES, C. Comprehension and Dependency Analysis of Aspect-Oriented Programs through Declarative Reasoning. In Proceedings of PADL. 2008, 35-52.

GOMI, E. BRASILIANA DIGITAL: Livros Raros ao Alcance de Todos.

A. Haque and P. Ginsparg, "[Positional effects on citation and readership in arXiv](#)," J. Am. Soc. Inf. Sci. Technol., vol. 60, pp. 2203-2218, 2009.

IEEE. Learning Technology Standardization Committee (LTSC). The Learning Object Metadata Standard. [On-Line]. Disponível em: <http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf>.

[Sheila S. Intner](#), [Sally C. Tseng](#), [Mary Lynette Larsgaard](#), Electronic cataloging: AACR2 and metadata for serials and monographs. 2003 by CRC Press

KURAMOTO, Hélio. Informação científica: proposta de um novo modelo para o Brasil. Ci. Inf., Brasília, v. 35, n. 2, p. 91-102, maio/ago 2006.

NELSON, G. Metadata 100 success secrets 100 most asked questions on metadata how to management, repositories, softwares, standards, tools and databases. Emereo Publishing, 2008.

NISO (National Information Standards Organization). Understanding Metadata, Niso Press. 2004

VAN VUGT, S. Beginning Umbutu server administration: from novice to professional. Apress, 2008.

YANG, H. Advances in UML and XML-based software evolution. Idea Group Publishing, 2005