

DOI: 10.5748/9788599693131-14CONTECSI/PS-4665

**DIGITAL INSTITUTIONAL REPOSITORIES OF SCIENTIFIC INFORMATION:  
IMPLEMENTATION WITH DSPACE SOFTWARE AS A TECHNICAL SOLUTION**

Renato Machado Sobral (Universidade de São Paulo, São Paulo, Brasil) -  
renato.sobral@usp.br

Cibele Araújo Camargo Marques dos Santos (Universidade de São Paulo, São Paulo,  
Brasil) - cibelear@usp.br

Teaching and research institutions have as a fundamental part of their existence the production of scientific information by teachers, students and researchers, who participate in a scientific communication system for the dissemination of this knowledge. The pressure for access to information presents the need to scale their own intellectual production, as a means of grouping, organizing and disseminating scientific works, often the result of public research funding. In order to do so, they need a technical alternative that can account for an increasingly digital communication context and the speed it provides. The answer lies in the creation of repositories. This work aims to conceptualize the institutional repositories, as digital tools, presenting the main characteristics that define them. The second purpose is to present DSpace software as a widely used instrument for the implementation of institutional repositories.

Keywords: Scientific Communication, Institutional Repositories, Information Services, Electronic Document Management, DSpace

**REPOSITÓRIOS INSTITUCIONAIS DIGITAIS DE INFORMAÇÃO CIENTÍFICA:  
IMPLEMENTAÇÃO COM O SOFTWARE DSPACE COMO SOLUÇÃO TÉCNICA**

Instituições de ensino e pesquisa têm como parte fundamental de sua existência a produção de informação científica pelos professores, alunos e pesquisadores dessas entidades, que participam de um sistema de comunicação científica para a disseminação desses conhecimentos. A pressão pelo acesso à informação apresenta a estas instituições a necessidade de dimensionar sua própria produção intelectual, como meio de agrupar, organizar e divulgar os trabalhos científicos, muitas vezes fruto de financiamentos públicos de pesquisa. Para tanto precisam de uma alternativa técnica que possa dar conta de um contexto de comunicação cada vez mais voltado ao digital e a velocidade por ele proporcionado; a resposta reside na criação de repositórios. Este trabalho busca conceituar os repositórios institucionais, enquanto ferramentas digitais, apresentando as principais características que os definem. O segundo propósito é apresentar o software DSpace como instrumento largamente utilizado para implementação de repositórios institucionais.

Palavras-chave: Comunicação Científica, Repositórios Institucionais, Serviços de Informação, Gerenciamento Eletrônico de Documentos, Dspace

## Introdução

Nos meios de disseminação da informação, cada vez mais, faz-se necessário o desenvolvimento de aparatos que permitam que tal processo transcorra de forma rápida e segura, ao mesmo tempo, que preserve a integridade documental e forneça a possibilidade do acesso e recuperação dos dados de maneira eficiente.

Em relação à informação produzida no meio acadêmico é fundamental divulgar e comunicá-la para o maior número possível de pessoas. Essa necessidade de “publicidade” advém, principalmente, do alto valor social que a maior parte dos estudos acadêmicos representa, onde recursos financeiros destinados às pesquisas têm nas formas de fomentos públicos o seu maior aporte.

Diante de tal questão, as instituições de ensino e pesquisa responsáveis por construir a informação científica de maior complexidade encontram-se diante da premência de disponibilizar todo o *corpus* produzido sob seus auspícios de modo que possam devolver à sociedade a devida dimensão dos estudos elaborados por seus colaboradores.

O ciclo social da informação preconizado por Le Coadic (2004) no qual a informação percorre caminho contínuo através dos processos de construção, comunicação e uso é um bom parâmetro ilustrativo da pertinência buscada por instituições de ensino e pesquisa que necessitam estabelecer tal circulação informacional para continuar sendo capazes de construir novos saberes que impactem de forma positiva a sociedade.

Se nos ativermos, por exemplo, apenas à forma de comunicar a informação, uma instituição engajada na instauração desse ciclo precisará rever seu processo de comunicação para além do sistema de publicação científica, fortemente estabelecido, mas que traz em si grandes barreiras para livre circulação.

Definida a pertinência, o próximo questionamento imediatamente nos remete ao “como fazê-lo?”, ou seja, quais recursos podem ser empregados para coletar, referendar, preservar e comunicar a informação científica construída pelos membros da comunidade institucional? No esteio do crescente movimento internacional pelo livre acesso à informação os repositórios institucionais surgem como opção capaz de instrumentalizar as necessidades acima mencionadas.

Pelo entendimento demonstrado por Leite & Amaro (2012: 7) podemos definir repositório institucional (RI) como um repositório de acesso aberto que constitui um serviço de informação científica, em ambiente digital e interoperável, dedicado ao gerenciamento da produção científica e/ou acadêmica de uma instituição (universidades ou institutos de pesquisa). Nesta definição, contempla a reunião, armazenamento, organização, preservação, recuperação e, sobretudo, a ampla disseminação da informação científica produzida na instituição.

Um dos aspectos fundamentais na definição acima estipula o repositório institucional enquanto recurso de ambiente digital, portanto sujeito às necessidades técnicas características da tecnologia da informação e comunicação, paradigma esse tomado como o ponto de partida nesta pesquisa, que tem por objetivos apresentar o conceito de repositórios e repositório institucional e sua relevância como instrumento de preservação e disseminação da informação científica, bem como, apresentar o software DSpace enquanto solução técnica para a implantação, desenvolvimento e gerenciamento de um repositório institucional.

## Metodologia

O percurso adotado para elaboração do trabalho consistiu em eleger de início a abordagem conceitual sobre repositórios desdobrando-se até o específico dos repositórios institucionais com especial ênfase aos postulados desejáveis para implantação e desenvolvimento de um RI.

Dentro desse arcabouço destacou-se o aspecto do software adotado face sua importância crucial para o sucesso da empreitada. O software DSpace foi escolhido para, novamente, através de uma abordagem conceitual, destacar suas principais virtudes e aplicabilidades para gestão de repositórios.

## Repositórios Institucionais: conceitos e características

Repositórios institucionais são definidos por Leite, Amaro, Batista, & Costa (2012) como um serviço de informação científica, operando em ambiente digital e interoperável, e dedicados ao gerenciamento da produção científica e/ou acadêmica de universidades e institutos de pesquisa; contemplando a reunião, armazenamento, organização, preservação, recuperação e a ampla disseminação da informação científica produzida na instituição.

A palavra repositório tem seu significado diretamente ligado ao aspecto de “lugar” e “coleção”, ou seja, temos um local destinado à preservação organizada de algo cuja pertinência demande tal iniciativa.

Dentre as diversas atividades capitaneadas por uma biblioteca, especialmente quando as tomamos sob a luz das especificidades das suas diferentes categorias, os tipos de serviços por elas prestados, bem como as necessidades informacionais pelas quais pretendem responder, a constituição e o desenvolvimento de repositórios podem ser compreendidos como atividade pertinente às bibliotecas especialmente àquelas que lidam com informação destinada a públicos especializados.

Assim, bibliotecas ligadas a instituições de ensino, pesquisa e desenvolvimento são aquelas mais propícias ao estabelecimento de repositórios que surgem como uma coleção, com diferentes graus de extensão e pertinência, que provavelmente estará sob a égide da biblioteca, arquivo ou algum outro serviço de informação que detém a expertise necessária para não apenas salvaguardar o seu conteúdo, mas também disponibilizá-lo de maneira inteligível através dos sistemas de classificação, indexação e recuperação de dados.

Portanto, o repositório tem para a biblioteca um caráter de acervo, detentor de características que pressupõem o tratamento informacional mais elaborado e poderá ser uma coleção digital.

Aliás, cabe ressaltar que muitos autores e serviços, que lidam de maneira teórica ou pragmática com o tema, consideram repositório (no domínio das Ciências da Informação) estritamente enquanto ferramenta digital. Uma dessas definições nos é dada pelo Repositório Científico de Acervo Aberto de Portugal (RCAAP)<sup>1</sup>:

Um repositório é uma coleção de documentos digitais. Repositórios OAI-compliant partilham os mesmos metadados, tornando assim os seus conteúdos interoperáveis com outros. Os seus metadados podem ser recolhidos para arquivos “virtuais” globais que são de fácil navegação por qualquer utilizador (tal como um índice comercial ou uma base de

<sup>1</sup> Disponível em: <<http://projeto.rcaap.pt/index.php/lang-pt/consultar-recursos-de-apoio/faqs/3-repositorios-cientificos/8-o-que-e-um-repositorio>>. Acesso em: novembro de 2016

dados abstrata é navegável, mas com acesso ao texto completo).

Destaca-se, no conceito apresentado, a assertiva pelo digital como definição de repositório, a capacidade de compartilhar metadados e acesso ao texto completo; o que ressalta os seus aspectos de preservação e disseminação.

Se considerarmos, por extensão, que ao se falar em repositórios estamos tratando de um objeto virtual, sem materialidade física, se faz necessário, também, elucidar conceitualmente o repositório digital.

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), uma das principais entidades responsáveis pelo fomento da comunicação científica nacional, assim o define (IBICT, 2016):

Os repositórios digitais (RDs) são bases de dados online que reúnem de maneira organizada a produção científica de uma instituição ou área temática. Os RDs armazenam arquivos de diversos formatos. Ainda, resultam em uma série de benefícios tanto para os pesquisadores quanto às instituições ou sociedades científicas, proporcionam maior visibilidade aos resultados de pesquisas e possibilitam a preservação da memória científica de sua instituição. Os RDs podem ser institucionais ou temáticos. Os repositórios institucionais lidam com a produção científica de uma determinada instituição. Os repositórios temáticos com a produção científica de uma determinada área, sem limites institucionais.

Leite, Amaro, Batista & Costa (2012) corrobora com tal entendimento e amplia o conceito referendando, através de exemplos, as diferenças entre repositórios digitais do tipo temático e institucional além de acrescentar um terceiro tipo à equação, os repositórios de teses e dissertações, que lidam exclusivamente com essa tipologia documentária.

Outra questão colocada pelos autores diz respeito ao entendimento de que todo repositório digital pode ser considerado um tipo de biblioteca digital, sendo que o repositório, na essência, teria seu conteúdo limitado pelos horizontes temáticos ou institucionais, enquanto bibliotecas digitais seria resultado de políticas de desenvolvimento de coleção deliberadas, além de incluírem um importante aspecto de serviço (referência e assistência à pesquisa, conteúdo interpretativo ou quaisquer outros recursos especiais).

Silva & Tomaél (2008) examinam a questão com minúcia propondo uma nova categorização de repositórios que pretende ampliar, de forma não excludente, as características pelas quais podem ser definidos.

Segundo esse entendimento a complexidade por trás dos propósitos de um repositório permite analisá-los pelos seguintes atributos quanto aos seus objetivos são: *acadêmicos/científicos*, que tornam visível a produção de instituições de ensino e pesquisa; ou *artísticos, culturais e sociais*, que trabalham com informação voltada para o desenvolvimento cultural e social.

Em relação ao conteúdo os repositórios são *temáticos*, que tratam de um ramo do conhecimento específico; ou *institucionais* que tratam do total da produção informacional de uma determinada instituição. Quanto ao tipo de documento podem ser *monodocumentais*, lidam apenas com uma tipologia documental, por exemplo, artigos científicos, relatórios etc; ou *multidocumentais* e não fazem distinção de tipologia. Considerando à administração os repositórios são *consorciados*, fruto da cooperação entre instituições variadas; *centralizados* com administração do repositório concentrada em um único local; *descentralizados* com administração distribuída entre várias instâncias da organização; e *governamentais* com repositórios mantidos por organizações ligadas ao Estado.

No que se refere ao tipo de objeto são *textuais*, a informação arquivada está em documentos com extensões de formato texto como, por exemplo, DOC, PDF, TXT etc; ou

*multimídias* cujos objetos do repositório estão em formatos variados de textos, sons, vídeos como, por exemplo, MP4, JPG, WAV etc.

Existe também a possibilidade de análise quanto à natureza da informação: *primários* que contemplam fontes primárias diretamente no servidor; ou **agregadores** que reúnem metadados de outros repositórios ou, também, documentos alocados em outros sistemas, que representam as fontes secundárias. Por fim, quando analisado pelos sistemas de avaliação podem ser *peer-reviewed*, armazenam documentação científica formal avaliada por pares; ou *não avaliados*, contemplam pre-prints e outros documentos não avaliados.

O estabelecimento desses atributos serve para demonstrar como um repositório digital pode ser um sistema informacional multifacetado, abrigando objetos de tipologias variadas, com vários propósitos em consonância aos objetivos dos seus mantenedores e público alvo.

### Critérios para implantação de repositórios institucionais

As demandas que levaram à concepção e implantação dos repositórios institucionais estão estreitamente ligadas às necessidades de preservação, organização e disseminação da informação científica com as quais instituições de ensino e pesquisa; sobretudo aquelas de nível superior sempre estiveram envolvidas.

A maior parte da informação científica produzida no seio dessas instituições encontra-se indexada nas mais distintas bases de dados referenciais, mas quase nunca as próprias entidades produtoras são capazes de dimensionar o quão relevante possa ser tal informação, já que têm dificuldades de controle sobre o fruto dos seus esforços. Assim, mesmo a atividade fundamental de dar visibilidade, de modo coeso e inteligível, à sua própria produção científica se mostra uma barreira de transposição deveras árdua colocando-as no centro do seguinte paradoxo: a grande capacidade de produzir informações científicas é inversamente proporcional ao controle sobre as mesmas.

Como bem observa Marra, em seu trabalho de revisão sobre o papel das bibliotecas universitárias na comunicação científica, as instituições de ensino superior são a maior fonte de conhecimento científico de um país, porém, há imensa dificuldade de tornar o conhecimento produzido visível e acessível, seja devido aos paradigmas sobre o livre acesso à informação ou, até mesmo, devido à dificuldade de encontrar ferramentas adequadas e de fácil manuseio pelo público interessado (Marra, 2012). Uma das soluções viáveis seria, justamente, a implantação dos repositórios institucionais.

Essa mesma visão também nos é transmitida por Lynch (2003) quando afirma que os repositórios institucionais surgem da necessidade de modernizar e aperfeiçoar a comunicação acadêmica, como uma alternativa e/ou um complemento aos padrões mais clássicos de publicação e, em consonância com o desenvolvimento tecnológico. Constituem-se em uma ferramenta importante para compartilhamento do conhecimento produzido por uma instituição entre seus membros ou ao restante da comunidade científica, e incentivam a adoção de novas formas de comunicação científica.

A urgência na transmissão dos resultados alcançados pelas pesquisas dentro do sistema de publicações, fortemente hierarquizado e pautado por critérios de qualidade, visibilidade e impacto do veículo de divulgação, demonstra que os periódicos ainda são a forma mais utilizada para disseminação da informação na comunidade científica, mesmo constituindo uma forma de transmissão de conhecimento não muito eficaz, pois, os processos para aceitação e publicação de um artigo são demorados. Alternativas como publicações eletrônicas, bibliotecas virtuais e repositórios institucionais têm surgido numa tentativa de aumentar a velocidade de transferência de conhecimento e estes são uma alternativa para socializar o conhecimento e dar visibilidade às pesquisas das universidades.

Além disso, os repositórios institucionais servem para prestar contas, à sociedade, dos estudos que estão sendo desenvolvidos no país, uma vez que, o financiamento das pesquisas é em grande parte proveniente de verbas públicas.

Dentro desse contexto, que opõe editoras e seus altos custos para disseminação da informação, e a busca pelo livre acesso, os repositórios institucionais aparecem como instrumentos para assegurar o implemento da “via verde” idealizada por Stevan Harnad. O sistema de “vias” de Harnad<sup>2</sup>, basicamente, estabelece caminhos para o acesso aberto à informação científica: a via dourada versa sobre produção e disseminação de periódicos eletrônicos de acesso aberto na rede; a via verde preconiza a criação de repositórios institucionais para a ordenação e divulgação da produção científica das instituições de ensino e pesquisa (Leite F. C., 2009).

Para Crow (2002) os repositórios institucionais apresentam respostas convincentes para duas questões estratégicas, pois fornecem um componente crítico na reforma do sistema de comunicação acadêmica; um componente que expande o acesso à pesquisa, reafirma o controle sobre os estudos pela academia, e aumenta a concorrência ao reduzir o poder de monopólio das revistas trazendo alívio econômico e maior relevância para as instituições e bibliotecas que as suportam. Além disso, possuem o potencial de servir como indicadores tangíveis da qualidade de uma universidade e demonstrar a relevância científica, societária e econômica de suas atividades de pesquisa, aumentando a visibilidade, status e valor público da instituição.

O RI deve proporcionar ao profissional acadêmico um meio para compartilhamento de seus trabalhos com livre acesso e permitir um meio de disseminação rápido, já que utiliza a Internet. (Barton & Waters, 2004).

Segundo a organização SPARC (*Scholarly Publishing and Academic Resources Coalition*)<sup>3</sup>, um repositório institucional (RI) é definido por: ser institucional, conter material acadêmico, ter caráter acumulativo e incessante, ter acesso livre e interoperabilidade.

A criação de um Repositório Institucional deve partir de um planejamento básico que contemple questões importantes para o sucesso da empreitada. Inicia-se o processo a partir do conhecimento do que é um repositório e através da análise de modelos já existentes, segue-se à produção de um cronograma que englobe custos, tempo e regras para implantação e distribuição (marketing) e, por fim, é necessário o estabelecimento de software e hardware e, sobretudo, de recursos humanos.

O tipo de abordagem que contemplará o repositório institucional também é critério fundamental na etapa de planejamento, podendo sê-la rígida ou flexível, entende-se por abordagem rígida aquela que vai dar prioridade para conteúdos publicados nas fontes formais: periódicos com avaliação por pares visa, assim, dotar de credibilidade o RI e contribuir para maximização dos resultados dos trabalhos científicos aumentando sua disseminação para além dos canais oficiais (Leite F. C., 2009). Essa abordagem é consonante com os argumentos do movimento de Acesso Aberto à Informação Científica. Já a abordagem flexível, aumenta seu alcance, pois contempla, além da informação avaliada por pares, outros conteúdos produzidos pela comunidade institucional: dados brutos de pesquisa, objetos de aprendizagem, relatórios de pesquisa, imagens, vídeos e etc. O autor comenta que é mais recomendável iniciar o funcionamento do Repositório

<sup>2</sup> Os sistemas de vias, caminhos percorridos para publicação da informação científica bem como os impactos causados podem ser melhor compreendidos na apresentação de Tim Brody e Stevan Harnad: *The research impact cycle*. 2004. Disponível em: <<http://opcit.eprints.org/feb19oa/harnad-cycle.ppt>>. Acesso em: 2016.

<sup>3</sup> A SPARC é uma coalização que preconiza o “livre” como padrão para educação, acesso e dados. Disponível em: <<http://sparcopen.org/>>. Acesso em: novembro de 2016.

Institucional adotando-se uma política rígida e que poderá ser flexibilizada ao longo do tempo (se assim for intuito da instituição), pois o contrário traria estranhamento à comunidade envolvida no projeto.

A Figura 1 elenca as etapas desejáveis para construção de repositórios institucionais, leva-se em consideração desde questões pragmáticas como a disponibilidade de recursos humanos adequados e escolha de softwares até o marketing responsável por disseminar entre o corpus intelectual da instituição as qualidades que tornariam imprescindível a submissão de suas obras no veículo ora preparado.



Figura 1: Fases da construção de repositórios institucionais.

Fonte: (Leite F. C., 2009)

Para ficar circunspecto aos parâmetros aludidos nos objetivos previstos, doravante, a análise será direcionada aos aspectos relacionados ao software necessário para o funcionamento de um repositório.

#### Softwares para implantação de repositórios institucionais

O primeiro modelo com o conceito de RI surgiu na década de 90 nos EUA, quando os pesquisadores do Los Alamos National Laboratory criaram o ArXiv que permitia o armazenamento de artigos não publicados ou não avaliados da maneira convencional, pelos pares. Daí em diante, outros repositórios surgiram no restante do mundo, o que trouxe a necessidade de ferramentas para que esses repositórios tivessem interação e operassem entre si (Marra, 2012).

Com base no entendimento de que cada instituição tem características individuais não há um modelo único de plataforma para todos os tipos de repositórios institucionais. Deste modo é importante o conhecimento de alguns softwares disponíveis, bem como das características de navegação por plataformas já existentes, para auxiliar na seleção de um modelo que atenda às necessidades da instituição.

Uma equipe formada por bibliotecários e profissionais capacitados em programação e tecnologia é de fundamental importância no estabelecimento de um repositório, criando mecanismos acessíveis e de fácil manuseio para os usuários, além de garantir a interoperabilidade com outras bases e a alocação de materiais nos mais diversos formatos

(áudio, vídeo, documentos de texto, imagens). Não deve ser negligenciado o fato de que a instituição deve possuir um conjunto de softwares básicos (Windows, Linux ou similar, base de dados, etc.) para a implementação de um repositório.

Para os softwares que oferecem uma plataforma para a criação de um repositório há as seguintes opções: softwares pagos, para os quais uma licença de uso é adquirida e possíveis atualizações são cobradas; softwares gratuitos, cujo download é feito sem custos e suas atualizações disponibilizadas pela instituição desenvolvedora (por exemplo, o DSpace); e, ainda, há opção da contratação de profissionais específicos para o desenvolvimento e gerenciamento de uma plataforma que atenda aos requisitos e objetivos traçados pela instituição.

A escolha de uma plataforma e o traçado das características deve levar em consideração o que é almejado pela instituição para o repositório ao longo dos anos. Assim, a escolha deve ser direcionada a atender metas futuras no que diz respeito aos materiais que compõem o acervo e sua forma de compartilhamento. Além disso, deve haver zelo pela adaptação das mídias ao avanço tecnológico, garantindo, desta maneira, a preservação das informações ao longo do tempo.

Sayão & Marcondes (2009) apoiados nos critérios apontados por Marill & Luczak (2009), Barton & Waters (2004) e Powel (2005) sistematizaram uma série de requisitos recomendados para avaliação do software adequado ao perfil do RI desejado. Ao contrário das gerações de programas predecessores, especialmente voltados para automatização de bibliotecas e que na sua maioria são alternativas pagas, os softwares para criação de repositórios institucionais são, de modo geral, programas com distribuição livre de código aberto, refletindo sua origem majoritariamente ligada à projetos que unem universidades, agências governamentais, Organizações Não Governamentais (ONGs) e, em menor escala, algumas empresas privadas. Nos quadros seguintes (1 a 6) os parâmetros avaliadores aparecem como indicativos para a escolha pautada em fundamentos sólidos sendo recomendável a instauração de um comitê multidisciplinar (bibliotecários, profissionais de tecnologia da informação, potenciais usuários e etc) para melhor julgar as propriedades buscadas.

No quadro 1 temos características gerais para avaliação de softwares que servem perfeitamente para avaliação de repositórios institucionais.

Critérios	Descrição
Escalabilidade	capacidade de crescimento do sistema
Extensibilidade	capacidade do programa de integrar ferramentas externas
Facilidade de implantação	grau de facilidade para integração com outros programas necessários ao funcionamento do repositório
Plataforma computacional	componentes necessários ao sistema: sistema operacional, servidores, gerenciador de banco de dados e etc.
Implantações de sucesso	usuário e organizações que já usam o sistema com sucesso
Suporte do sistema	capacidade da rede de desenvolvedores e usuários em resolver problemas técnicos
Base de conhecimento das comunidades envolvidas	quão capacitadas são as comunidades envolvidas no desenvolvimento e uso do programa (informações atualizadas, fóruns, eventos e etc.)
Estabilidade da organização de desenvolvimento	grau de confiança estabelecido pela organização responsável pelo desenvolvimento
Perspectivas para o futuro	capacidade de evolução e incorporação de inovações
Limites do sistema	limites do programa: volume de dados, coleções, registros e etc.
Documentação disponível/cursos/publicações	informações sobre o programa provenientes de fontes distintas: livros, tutoriais, cursos etc.

Quadro 1: Critérios para escolha de software: características gerais.

Fonte: elaborado pelo autor, de acordo com descrito por Sayão & Marcondes (2009).

No Quadro 2, são levados em consideração os requisitos de software, hardware, mão de obra especializada e aqueles que costumam gerar impacto financeiro imediato caso não haja disponibilidade dos recursos dentro da instituição mantenedora.

Critérios	Descrição
Hardware	disponibilidade dos requisitos mínimos de hardware na instituição
Software	disponibilidade dos requisitos mínimos de software na instituição
Expertise técnica gerencial e metodológica	disponibilidade mão de obra qualificada para instalação, operação e gestão do sistema
Sustentabilidade	disponibilidade de fundos para planejar, implantar, operar e manter o repositório em funcionamento

Quadro 2: Critérios para escolha de software: infraestrutura técnica, gerencial, metodológica.

Fonte: elaborado pelo autor, de acordo com o descrito por Sayão & Marcondes (2009).

O Quadro 3 versa sobre a interface do sistema e a incorporação de objetos no RI, descrevendo o quão personalizável podem ser as possibilidades de uso que garantam fluxo de dados constante no sistema.

<b>Cr�terios</b>	<b>Descri�o</b>
<b>Diversidade de servi�os</b>	estima a diversidade de produtos e servi�os oferecidos pelo programa nativamente ou implementado por outros programas
<b>Interatividade</b>	interatividade em tempo real entre usu�rio e sistema
<b>Tempo de resposta</b>	espa�o de tempo entre uma requis�o e a resposta pelo sistema
<b>Interface default</b>	se o sistema oferece interfaces prontas
<b>Padr�es de interface</b>	se as interfaces est�o em conformidade com padr�es apropriados
<b>Customiza�o</b>	se a interface pode ser customizada segundo cr�terios da institui�o
<b>Ergonomia e usabilidade</b>	se a interface � ergon�mica facilitando usabilidade e navega�o
<b>Acessibilidade</b>	se implementa padr�es b�sicos de usabilidade
<b>Multil�ngue</b>	com quais idiomas � poss�vel interagir com o sistema, possibilidade e adicionar outros idiomas
<b>Software cliente</b>	necessidade e a disponibilidade de software para entrada de dados que n�o sejam os navegadores padr�o
<b>Formatos aceitos</b>	capacidade de trabalhar com documentos em diversos formatos (PDF, MP3, HTML etc.
<b>Importa�o de dados de outras fontes</b>	capacidade de importatr dados de outras fontes: bases de dados, outros reposit�rios
<b>Modelo de objetos complexos</b>	se o software implementa um modelo de objeto que permita incorpora�es de vers�es, inst�ncias e etc. e padr�es de empacotamento de metadados mais comuns como METS e MPEG-21
<b>Autoarquivamento/autosubmiss�o/revis�o/apr ova�o</b>	disponibilidade de autoarquivamento e m�dulo para reten�o do trabalho para revis�o, cr�tica e edi�o dos metadados at� libera�o pelo administrador do sistema
<b>Objetos de dados distribu�dos</b>	se o conte�do pode ser referenciado n�o apenas no ambiente local mas em qualquer lugar na web
<b>Fluxo de trabalho</b>	disponibilidade de fu�oes no controle do processo de entrada de dados
<b>Armazenamento</b>	as formas de armazenamento dos conte�dos e metadados submetidos ao reposit�rio

Quadro 3: Cr terios para escolha de software: interface do sistema e a incorpora o de objetos.

Fonte: elaborado pelo autor, de acordo com o descrito por Say o & Marcondes (2009).

O Quadro 4 analisa como o software d  suporte para a organiza o da informa o e sua recupera o, ajudando a verificar se   poss vel distribuir os objetos de maneira org nica refletindo a estrutura institucional e quais m todos de recupera o est o dispon veis.

<b>Critérios</b>	<b>Descrição</b>
<b>Métodos de indexação</b>	qual é o método de indexação utilizado pelo pacote; se é nativo ou implementado por outro software
<b>Texto completo</b>	se indexa o texto completo
<b>Adequação da indexação</b>	se o método de indexação para as exigências de recuperação do repositório
<b>Criação de índices</b>	se o programa cria índices (autor, assunto, etc)
<b>Vocabulários controlados</b>	se tem algum módulo ou possibilidade de integrar vocabulários controlados, lista de autoridades e etc.
<b>Modalidade de buscas disponíveis</b>	modalidades de busca disponíveis: simples, avançada etc; possibilidade de consulta aos índices para formulação de buscas
<b>Browsing</b>	se possibilita navegar pelo índices
<b>Disseminação de informação</b>	se tem nativamente ferramentas para disseminação da informação: rss, alertas etc
<b>Apresentação dos resultados</b>	opções de exibição dos resultados da busca
<b>Acesso aos conteúdos</b>	opções de acesso aos conteúdos: downloading, streaming etc
<b>Interface de consulta</b>	como é a interface de consulta segundo parâmetros de customização, usabilidade, acessibilidade etc

Quadro 4: Critérios para escolha de software: organização e recuperação da informação.

Fonte: elaborado pelo autor, de acordo com o descrito por Sayão & Marcondes (2009).

O Quadro 5 trata dos padrões de metadados e interoperabilidade que dotarão o repositório de capacidade para correta descrição dos conteúdos arquivados e troca de dados com outros sistemas.

<b>Critérios</b>	<b>Descrição</b>
<b>Esquemas de metadados aceitos</b>	esquemas de metadados aceitos pelo pacote
<b>Dublin Core</b>	aceitação do Dublin Core; se permite extensão do esquema com uso de qualificadores
<b>Criação/edição de metadados</b>	se permite criação e edição de esquemas de metadados
<b>Norma ISO/OAIS</b>	grau de adesão à norma OAIS
<b>Outras metodologias</b>	se o software aplica metodologias como a conversão para formatos padronizados (XML, texto puro etc)
<b>Migração</b>	ferramentas de apoio à migração de dados
<b>Outras estratégias de preservação digital</b>	qualquer outra estratégia de preservação
<b>Protocolo Open Archives</b>	se implementa o protocolo OAI/PMH
<b>Protocolo Z39.50</b>	se implementa o protocolo Z39.50
<b>Nome padronizado</b>	se implementa algum sistema de identificação permanente: URN, DOI, Handle System
<b>Esquema específico</b>	se implementa alguma estratégia específica de identificação dos seus conteúdos

Quadro 5: Critérios para escolha de software: metadados e interoperabilidade.

Fonte: elaborado pelo autor, de acordo com o descrito por Sayão & Marcondes (2009).

No Quadro 6 são relacionados tópicos sobre administração e segurança: avalia a existência de ferramentas que auxiliem a gestão das informações incorporadas no repositório e assegurem a sua integridade e proteção contra ingerências e acessos descabidos.

Critérios	Descrição
Gestão de coleções	possibilita gestão diferenciada de coleções no mesmo repositório
Estatísticas	possui módulo para geração de estatísticas de uso
Controle de direitos	possui módulo de apoio a gestão de direitos
Interface administrativa	interface de apoio a gestão segundo parâmetros de ergonomia, usabilidade e etc
Preservação digital	se possui dispositivos de apoio à preservação digital
Autenticação e direitos de acesso	se há funcionalidades para gestão da segurança da informação: controle de acesso, autenticação de usuários, níveis de permissão
Backup e restauração	suporte ao processo de backup e restauração
Criptografia	suporte a cifragem de dados durante transmissão de conteúdos

Quadro 6: Critérios para escolha de software: administração e segurança.

Fonte: elaborado pelo autor, de acordo com o descrito por Sayão & Marcondes (2009).

Dentre os principais softwares livres disponíveis para desenvolvimento de repositórios podemos apontar como os de maior propagação: DSpace, Eprints, Greenstone, Fedora (*Flexible Extensible Digital Object and Digital Repository Architecture*) e OPUS (*Open Publications System*).

Há dois diretórios virtuais dedicados a listar repositórios de acesso livre em todo mundo: o OpenDOAR (Directory of Open Access Repositories)<sup>4</sup> e o ROAR (Registry of Open Access Repositories)<sup>5</sup>.

No OpenDOAR estão listados 3272 repositórios e no ROAR esse número chega aos 4358; em ambos os casos é possível realizar pesquisa nos diretórios utilizando como critério de busca o tipo de software empregado pelo repositório.

No Quadros 7 e Gráficos 1 e 2 são demonstradas a distribuição de repositórios em relação aos principais softwares já destacados.

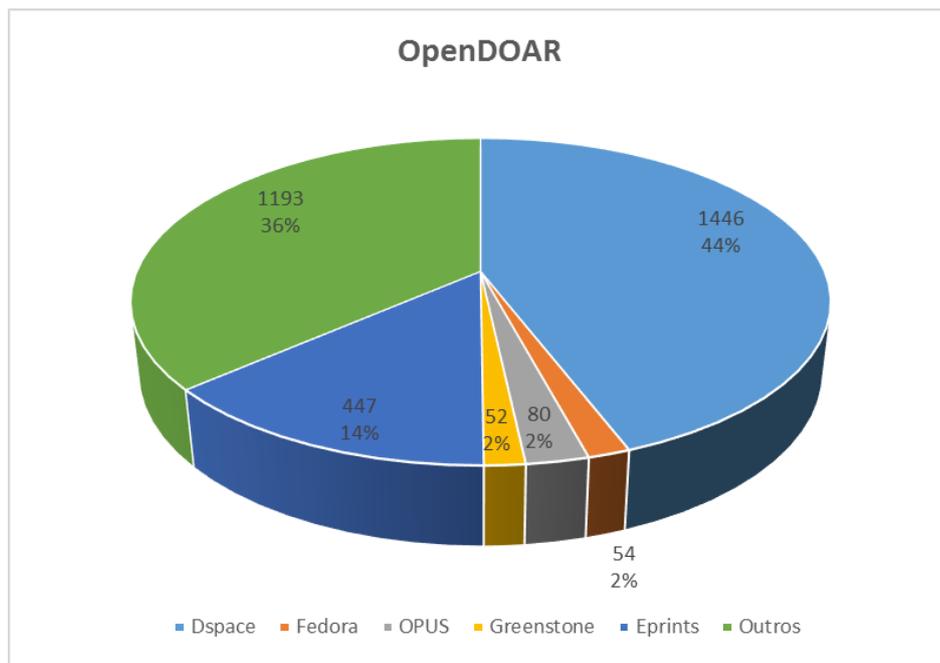
Software	ROAR	OpenDOAR
DSpace	1709	1446
Fedora	54	54
OPUS	87	80
Greenstone	21	52
Eprints	608	447
Outros	1879	1193
<b>Total</b>	<b>4358</b>	<b>3272</b>

Quadro 7: Distribuição de softwares de repositórios.

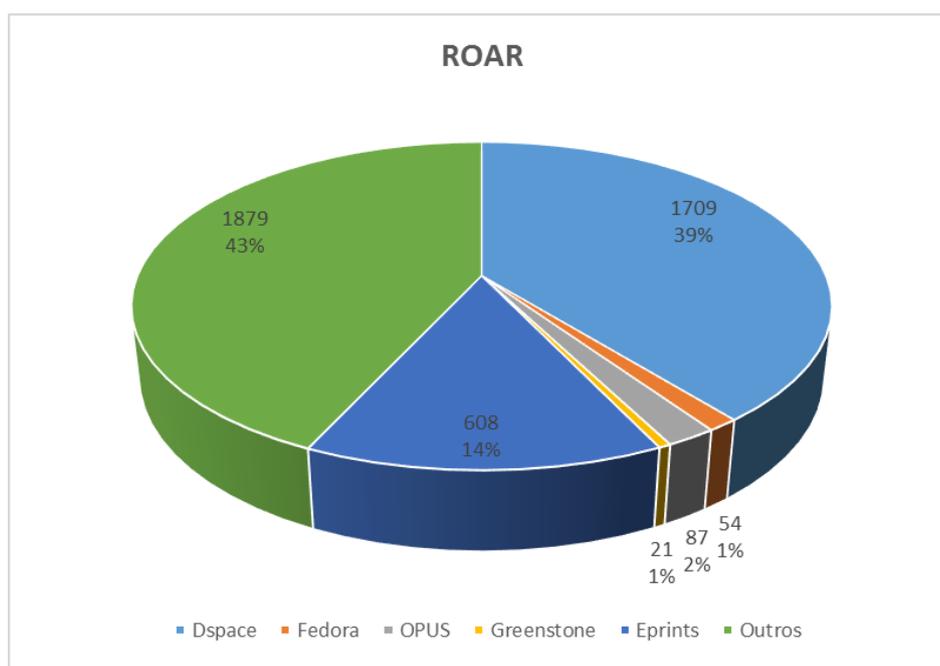
Fonte: sites do ROAR e OpenDOAR (Nov/2016).

<sup>4</sup> Disponível em: <<http://www.opendoar.org/>> Acesso em: novembro de 2016

<sup>5</sup> Disponível em: <<http://roar.eprints.org/>>. Acesso em: novembro de 2016



**Gráfico 1:** Porcentagem participativa dos principais softwares  
 Fonte: OpenDoar <http://www.opendoar.org/>



**Gráfico 2:** Porcentagem participativa dos principais softwares para repositórios  
 Fonte: Registry of Open Access Repositories <http://roar.eprints.org/>

### Sobre o DSpace

O DSpace pode ser definido com um software livre projetado para criação de repositórios digitais atendendo funções de armazenamento, gerenciamento, preservação digital e divulgação da produção científica de instituições de ensino e pesquisa, permitindo o auto

arquivamento de documentos e a sua marcação com metadados.

Cabe ressaltar que a concepção de livre acesso, norteadora do DSpace, está intimamente ligada ao surgimento dos repositórios. Esse conceito vem desde o início da década de 80 e, atualmente, encontramos uma série de licenças de uso com essa proposta de distribuição livre, dentre as quais podemos citar: GNU *General Public License* (GPL), BSD (*Berkeley Software Distribution*) *License*, GNU *Lesser General Public License*, MIT (*Massachusetts Institute of Technology*) *License*, *Mozilla Public License* e Apache, cada qual contendo particularidades e termos de uso distintos. Dentre as inúmeras razões para a adoção de softwares de acesso livre pelas instituições, uma das mais determinantes é a ausência de custos, o que possibilita a adesão de instituições que não possuam muitos recursos financeiros. (Biswas & Paul, 2010)

O DSpace é capaz de promover o gerenciamento de uma grande diversidade de materiais digitais com foco na proteção e disponibilidade dos conteúdos arquivados ao longo do tempo, o que, por si só, é uma característica deveras importante tendo em vistas que a velocidade com a qual recursos eletrônicos são implementados está em paridade com a possibilidade real de perdermos capacidade de recuperar a informação contida nos mais diversos suportes em que se encontra depositada. O software do DSpace pode lidar com documentos, livros, teses, imagens, arquivos de áudio e vídeo, páginas web, coleções de bibliotecas digitais, memoriais, bases de dados variadas e etc; praticamente não há limites para sua abrangência.

Podemos resumir seus principais atributos: software livre de código aberto, está coberto pela licença BSD (*Berkeley Software Distribution*); foco no uso dos repositórios institucionais; completamente personalizável para atender às necessidades do usuário; Gerencia e preserva todo o formato de conteúdo digital (PDF, Word, JPEG, MPEG, TIFF etc.); armazena num único local a produção científica; identifica quais documentos são publicados, em quais versões e quantos downloads foram realizados para cada um; tem um identificador único para cada documento; permite pesquisa personalizada; tem instalação Unicode; possui um conjunto básico de funcionalidades que podem ser estendidas ou integradas com serviços e ferramentas complementares no ecossistema acadêmico maior; roda nos sistemas operacionais: Linux, Unix, Ubuntu, Windows, Mac OS.

O DSpace é bastante usado por bibliotecas acadêmicas e de pesquisa como um repositório de acesso aberto para o gerenciamento dos trabalhos científicos do seu corpo institucional. Há também muitas organizações usando o software para hospedar e gerenciar informações em mídias diversas ou conjuntos de dados. Apesar de o foco maior ocorrer na criação de repositórios institucionais, a página oficial do DSpace provê acesso à uma lista com o registro de mais de 2 mil repositórios onde temos uma descrição básica do seu escopo<sup>6</sup> e finalidades distintas para as quais ele pode ser empregado. A seguir estão relacionados alguns exemplos de repositórios gerenciados pelo DSpace com aplicações variadas<sup>7</sup>:

**Repositório Institucional**: Repositório para funcionários, docentes e / ou documentos de estudantes, pesquisa, artigos e publicações em uma instituição. Exemplo apresentado na Figura 2.

<sup>6</sup> O DSpace User Registry está acessível em <<http://registry.duraspace.org/registry/dspace>>, o Brasil é o quinto país com mais repositórios cadastrados. Acesso em: novembro de 2016

<sup>7</sup> Exemplos disponibilizados em: <<http://www.dspace.org/use-cases>>. Acesso em: novembro de 2016

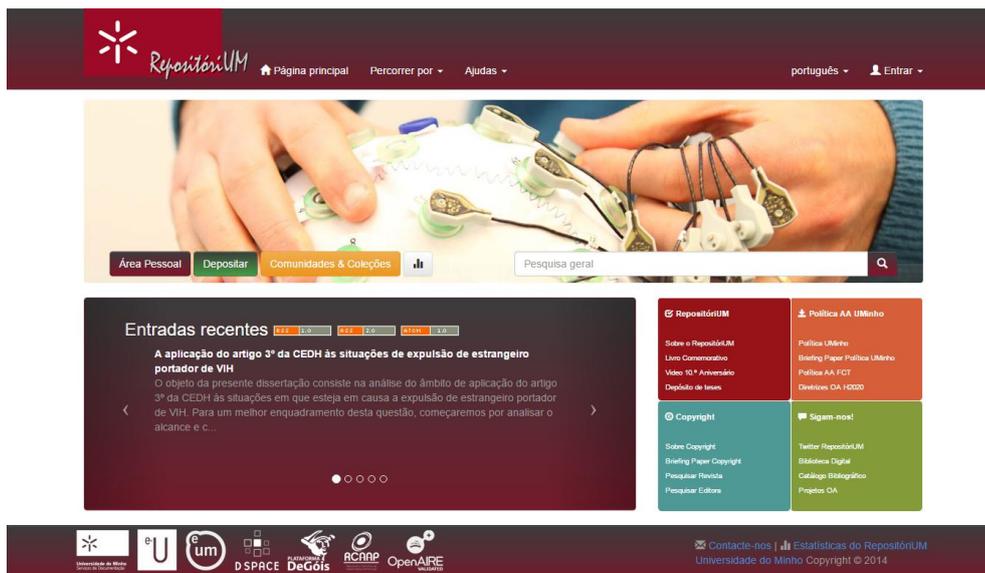


Figura 2: Página inicial do RepositoriUM, Universidade do Minho, Portugal.  
 Fonte: <https://repositorium.sdum.uminho.pt/>

**Repositório de Imagens:** Repositório para imagens digitalizadas de arte ou itens históricos para facilitar o uso dos materiais e preservar um registro dos itens originais, conforme exemplificado na Figura 3.

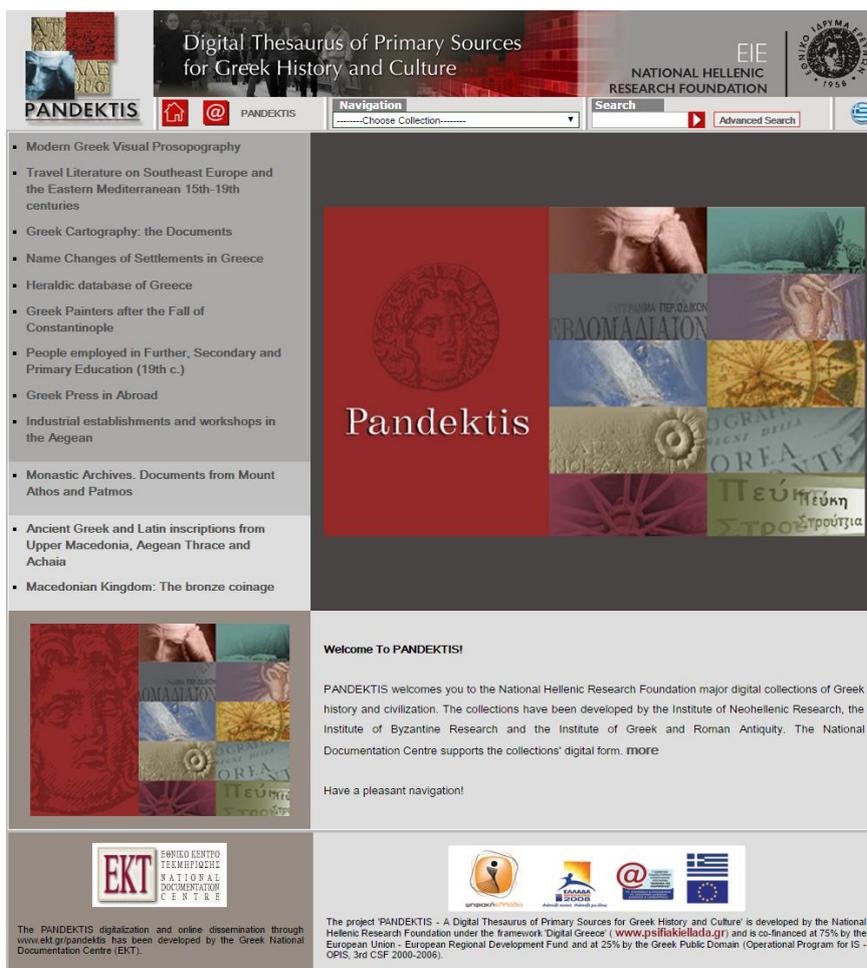


Figura 3: Página inicial do Pandektis, National Hellenic Research Foundation, Grécia.  
 Fonte: <http://pandektis.ekt.gr/pandektis/?locale=en>

**Repositório de Áudio e Vídeo:** Repositório para armazenar arquivos de áudio e vídeo, conforme mostrado na Figura 4.

The screenshot shows the SUNY Digital Repository interface for 'The Writers Forum'. The header includes the SUNY logo and 'Digital Repository' text, with an 'Entrar' button in the top right. Below the header, there's a breadcrumb trail: 'Página Inicial / SUNY College at Brockport / The Writers Forum'. The main content area is titled 'The Writers Forum' and includes a search bar, navigation options (e.g., 'Por data do documento', 'Autores', 'Títulos', 'Assuntos'), and a search input field. The page also contains introductory text about the forum's history and a list of recent video submissions, each with a filmstrip icon and a QuickTime player icon.

**Submissões recentes**

- The Writers Forum: Leslie Ullman**  
Ullman, Leslie (Guest) (SUNY Brockport, 2000-10-26)
- The Writers Forum: Deborah Tall**  
Tall, Deborah (Guest) (SUNY Brockport, 2002-09-04)
- The Writers Forum: Joanna Scott**  
Scott, Joanna (Guest) (SUNY Brockport, 2003-12-02)
- The Writers Forum: David Rothenberg**  
Rothenberg, David (Guest) (SUNY Brockport, 2005-11-02)
- The Writers Forum: John Poch**  
Poch, John (Guest) (SUNY Brockport, 2005-03-09)
- The Writers Forum: Carl Phillips**  
Phillips, Carl (Guest) (SUNY Brockport, 2001-11-14)
- The Writers Forum: Edith Pearlman**  
Pearlman, Edith (Guest) (SUNY Brockport, 2005-04-06)
- The Writers Forum: Alan Michael Parker**  
Parker, Alan Michael (Guest) (SUNY Brockport, 2004-10-27)

**Figura 4:** Página inicial da SUNY Digital Repository The Writers Forum Videotape Library, University of New York, EUA.

Fonte: <https://dspace.sunyconnect.suny.edu/handle/1951/29064>

**Repositório de Museus e Patrimônios Culturais:** Repositório para preservar documentos, publicações, imagens e outros itens de museus e patrimônios culturais bem como fornecer acesso dos conteúdos à comunidade, como pode ser exemplificado pela Figura 5.

Página inicial

**AMNH Library Digital Repository**

The AMNH Library Digital Repository is an archive maintained by the Research Library for AMNH Scientific Publications, AMNH scholarly output and other original and published materials digitized by the Library. All information in the repository is freely accessible to scholars around the world to support their research.

The Library's image database, Digital Special Collections can be accessed at:  
<http://images.library.amnh.org/digital/index.php>.

**Comunidades no DSpace**

Selecione uma comunidade para navegar nas coleções.

- **Center for Biodiversity and Conservation**
  - [Supplemental Data](#)
- **Earth And Planetary Sciences**
  - [Supplemental Data](#)
- **Herpetology**
  - [Supplemental Data](#)
- **Ornithology**
  - [Manuscripts](#)
- **Research Library**
  - [American Museum Journal](#)
  - [Annual Reports](#)
  - [Catalogs](#)
  - [Hall Guides](#)
  - [James Arthur Lecture on the Evolution of the Human Brain](#)
  - [Manuscripts](#)
  - [Museum Publications](#)
  - [Natural History Magazine](#)
- **Richard Gilder Graduate School**
  - [Dissertations](#)
- **Scientific Publications**
  - [American Museum Novitates](#)
  - [Anthropological Papers of the American Museum of Natural History](#)
  - [Bulletin of the American Museum of Natural History](#)
  - [Memoirs of the American Museum of Natural History](#)
  - [Other Scientific Publications](#)
  - [Supplemental Material](#)

**Submissões recentes**

[Rollo Beck's collections of birds in northeast New Guinea. \(American Museum novitates, no. 3873\)](#)  
 Lecroy, Mary.; Diamond, Jared M. (American Museum of Natural History., 2017-02-15)  
 In 1928-1929 Rollo Beck discovered in New Guinea a spectacular new species of bowerbird, *Sericulus bakeri*, which according to his specimen labels he collected near the previously well-studied lowland town and former colonial ...

**Buscar DSpace**

Advanced Search

**Discover**

**Autor**

- American Museum of Natural History (243)
- Allen, J. A. (Joel Asaph). 1838-1921 (178)
- Platnick, Norman I. (145)
- Central Asiatic Expeditions (1921-1935) (140)
- Simpson, George Gaylord. 1902- (121)
- Curran, Charles Howard. 1894-1972 (117)
- Nichols, John T. (John Treadwell). 1883-1958 (114)
- Brown, Barnum (106)
- Rozen, Jerome George. 1928- (105)
- American Museum Congo Expedition (1908-1915) (104)
- ... Ver mais

**Assunto**

- Natural history museums -- New York (State) -- New York -- Periodicals. (134)
- American Museum of Natural History -- Periodicals. (133)
- Skull. (116)
- Natural history -- Periodicals. (109)
- Insects -- North America (97)
- Birds -- South America (93)
- American Museum of Natural History (85)
- Paleontology -- West (U.S.) (82)
- Paleontology -- Mongolia. (76)
- Spiders. (66)
- ... Ver mais

**Data de publicação**

- 2000 - 2017 (855)
- 1900 - 1999 (5233)
- 1851 - 1899 (254)

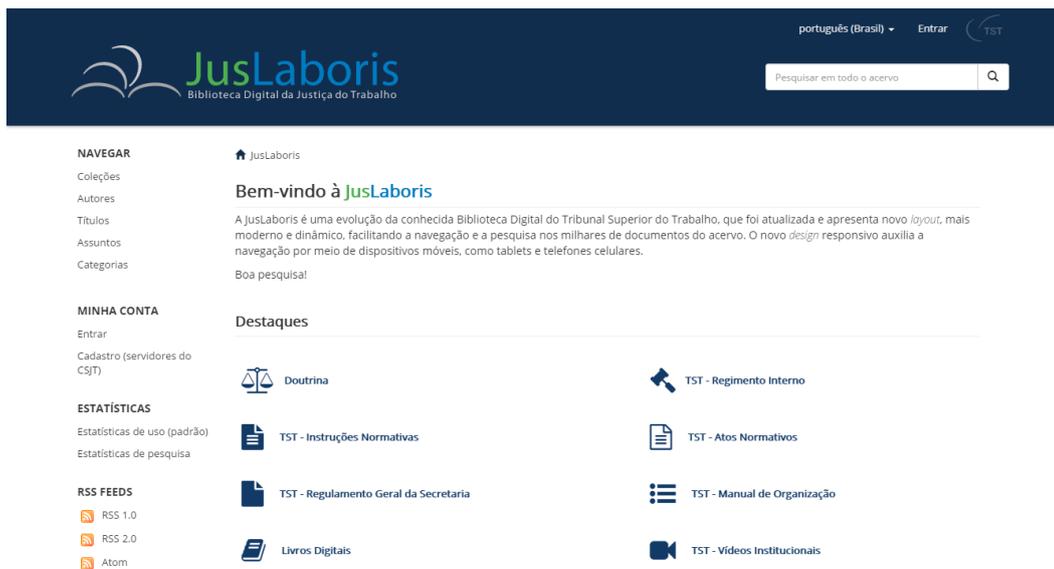
**Navegar**

- Todo o repositório
- Comunidades e Coleções
- Titulos

**Minha conta**

**Figura 5:** Página inicial do AMNH Library Digital Repository, American Museum of Natural History, EUA.  
 Fonte: <http://digitallibrary.amnh.org/>

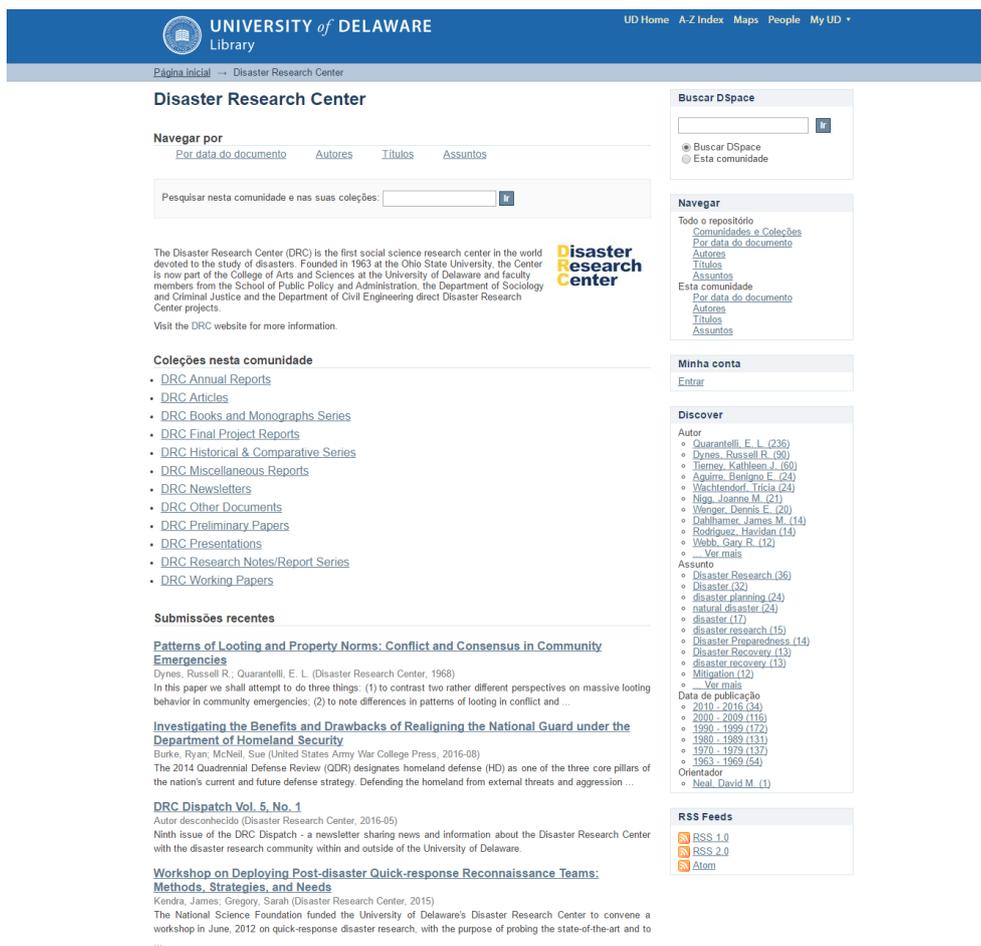
**Repositório de Registros e Relatórios Governamentais:** Repositório para preservação e distribuição de documentos e registros legais, como apresentado na Figura 6.



**Figura 6:** Página do JusLaboris: repositórios institucionais da Biblioteca Digital da Justiça do Trabalho, Tribunal Superior do Trabalho, Brasil.

Fonte: <https://juslaboris.tst.jus.br/>

**Repositório Temático:** Repositório focado em um assunto ou área específica como demonstrado na Figura 7.



**Figura 7:** Página inicial do Disaster Research Center, University of Delaware, EUA.

Fonte: <http://udspace.udel.edu/handle/19716/35>

No Brasil o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) é o responsável pela tradução e distribuição do DSpace desde 2004<sup>8</sup>. Atualmente a última versão estável do Dspace é a 6.x (com suporte garantido no mínimo até 2018)<sup>9</sup>.

O DSpace surgiu como um esforço de desenvolvimento conjunto da *Hewlett Packard* (HP) e do *Massachusetts Institute of Technology* (MIT), iniciado em 2002, foi desenvolvido como software de código aberto para gerenciar pesquisa, trabalhos acadêmicos e outros conteúdos publicados em um repositório digital, com foco no armazenamento, acesso e preservação de longo prazo. À medida que os usuários do DSpace aumentaram, um grupo de instituições formou a *DSpace Federation* em 2004, que determinou a governança de desenvolvimento futuro para o DSpace. Em julho de 2007, o crescimento do software levou a HP e MIT a formar a *DSpace Foundation*.

Paralelamente ao avanço do DSpace, outro software criado em meados da década de 1990 pela Cornell University chamado Fedora (*Flexible, Extensible Digital Object Repository Architecture*) também alcançou destaque como uma solução de acesso aberto na criação e gerenciamento de repositórios. De 2001 a 2007, Cornell e a University of Virginia combinaram esforços para criar o *Fedora Project*. Ainda em 2007 foi criado o *Fedora Commons*, uma organização sem fins lucrativos com objetivo de construir uma comunidade de usuários acadêmicos, científicos e educacionais para continuar o desenvolvendo do software.

Em maio de 2009, a colaboração em projetos relacionados e sinergias crescentes entre a *DSpace Foundation* e *Fedora Commons* levou à união das duas organizações para prosseguir com a sua missão em comum, foi, então, fundada uma nova instituição sem fins lucrativos chamada *DuraSpace*. Atualmente, a comunidade de softwares e usuários do DSpace recebe liderança e orientação da *DuraSpace* (DuraSpace, 2016).

### Funcionamento do Dspace

A seguir vamos fornecer uma visão sintética do funcionamento do DSpace como recurso para desenvolvimento de repositórios institucionais. Os autores Rodriguez-Gairín & Duesa (2008) e o Repositório Científico de Acesso Aberto de Portugal – RCAAP (RCAAP, 2016) são fontes elucidativas nesse intuito.

O DSpace é um sistema com interface web que permite o auto arquivamento de documentos e a sua marcação com metadados. Foi desenhado para suportar qualquer tipo de formato desde os documentos de texto simples (formatos PDF, TXT, DOC e etc) até arquivos multimídia, e ainda, é capaz de registrar formatos não reconhecidos para que possam ser identificados no futuro.

No sistema os arquivos (também denominados "*bitstreams*") estão organizados em conjuntos. Cada um dos *bitstreams* tem um determinado formato e informação técnica específica. Esta informação técnica é armazenada juntamente com os *bitstreams* para ajudar na sua preservação ao longo do tempo.

Um item é a menor unidade do repositório e consiste no agrupamento de informação relacionada com o arquivo que são os seus metadados. Os metadados dos itens são expostos para indexação e pesquisa, no próprio repositório e externamente, como, por

<sup>8</sup> O Dspace em língua portuguesa pode ser baixado a partir do endereço eletrônico:

<<http://www.ibict.br/pesquisa-desenvolvimento-tecnologico-e-inovacao/Sistema-para-Construcao-de-Repositorys-Institucionais-Digitais/download>>, onde também há informações pertinentes sobre o trabalho do IBICT com o programa. Acesso em: 2016

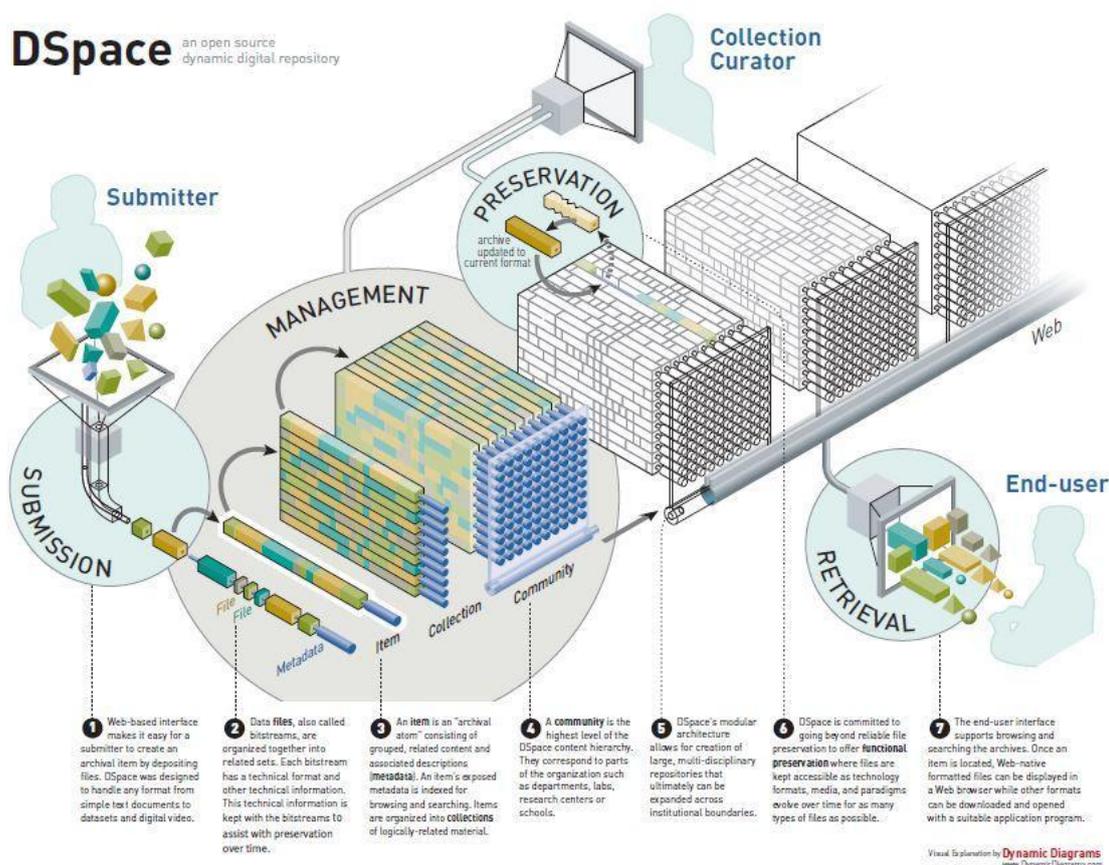
<sup>9</sup> A Fundação DuraSpace mantém uma página wiki com toda documentação referente ao desenvolvimento das versões. Disponível em: <<https://wiki.duraspace.org/display/DSPACE/Documentation>>. Acesso em: novembro de 2016

exemplo, no *Google Scholar*. Estes itens podem ser alocados em coleções, comunidades e subcomunidades conforme a necessidade de organização proposta pela instituição mantenedora. A comunidade é o nível mais alto na hierarquia do DSpace e pode corresponder ao organograma da sua instituição relacionando faculdades, institutos, departamentos, laboratórios e afins, ou, se assim for necessário, a organização das comunidades poderá ser baseada em áreas temáticas. (RCAAP, 2016)

A modularidade do sistema permite adicionar e criar novas funcionalidades incorporando outros softwares ou personalizando o DSpace conforme os recursos desejáveis. Rodriguez-Gairín (2008) salientam a necessidade de um profissional de informática e programação para dar suporte à implantação do repositório adaptando-o às necessidades tanto dos profissionais que vão alimentar a base como do usuário final.

A interface do DSpace permite ao utilizador final a procura e pesquisa de documentos que poderão visualizar diretamente pelo navegador ou descarregar para o seu dispositivo de acesso, uma das funcionalidades mais recentes do DSpace, possibilita a adaptação do repositório para uma versão mobile adaptando-o ao uso cada vez mais disseminado de smartphones, *tablets* e outros dispositivos de acesso móvel.

Na Figura 8 temos um esquema que retrata sumariamente o funcionamento do repositório em DSpace.



**Figura 8:** Ilustração do funcionamento do DSpace.

Fonte: (DSpace, 2016)

Numa tradução livre dos itens descritos na imagem da Figura 8 temos os passos elencados no diagrama funcional como sendo:

1) Interface baseada na web torna mais fácil para o usuário criar um arquivamento depositando documentos. O DSpace foi projetado para lidar com qualquer formato de

documentos de texto simples, conjuntos de dados e vídeo digital.

2) Arquivos de dados, também chamados de *bitstreams*, são organizados juntos em conjuntos relacionados. Cada *bitstream* tem um formato técnico e outras informações técnicas. Esta informação técnica é mantida com *bitstreams* para auxiliar na preservação ao longo do tempo.

3) Um item é um "átomo arquivístico" que consiste em conteúdo agrupado, relacionado e descrições associadas (metadados). Os metadados expostos de um item são indexados para navegação e pesquisa. Os itens são organizados em coleções de material logicamente relacionado.

4) Uma comunidade é o nível mais alto da hierarquia de conteúdo do DSpace. Eles correspondem a partes da organização, tais como departamentos, laboratórios, centros de pesquisa ou escolas.

5) A arquitetura modular do DSpace permite a criação de grandes repositórios multidisciplinares que, em última instância, podem ser expandidos através das fronteiras institucionais.

6) O DSpace está comprometido em ir além da preservação confiável de arquivos para oferecer preservação funcional onde os arquivos são mantidos acessíveis à medida que os formatos de tecnologia, mídia e paradigmas evoluem ao longo do tempo para o maior número possível de arquivos.

7) A interface do usuário final oferece suporte à navegação e pesquisa nos arquivos. Uma vez localizado um item, os arquivos formatados na web podem ser exibidos em um navegador, enquanto outros formatos podem ser baixados e abertos com um programa de aplicação adequado.

Além desses dados é importante destacar outros pormenores técnicos do programa: os metadados, incluindo informações de acesso e configuração, são armazenados em um banco de dados relacional e suportam o uso do PostgreSQL e do banco de dados Oracle. Atualmente, o DSpace suporta duas interfaces web primárias: JSPUI que usa JSP e Java Servlet API e XMLUI (também conhecido como Manakin) baseado no Apache Cocoon, usando XML e XSLT.

Quando falamos sobre metadados cumpre ressaltar que é um requisito extremamente sensível para o sucesso do repositório. A *National Information Standards Organization* (NISO)<sup>10</sup> os define como informações estruturadas que descrevem, explicam, localizam ou tornam mais fácil recuperar, usar ou gerenciar um recurso de informações, são muitas vezes chamados de dados sobre dados ou informações sobre informações, denotando-se, portanto, a razão pela qual são uma prioridade. O *Dublin Core* é o formato de metadados padrão dentro da aplicação DSpace, entretanto é possível adicionar ou alterar qualquer campo para personalizá-lo para o tipo de aplicação necessária. Atualmente, o DSpace suporta qualquer esquema de nomes simples não hierárquicos, embora seja possível inserir outros esquemas hierárquicos de metadados como MARC e MODS.

No Quadro 8 temos a descrição dos 15 elementos fundamentais do padrão de metadados Dublin Core, como pode ser observado a seguir:

<sup>10</sup> O documento que traz essa e outras definições "Understanding Metadata" está disponível em: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>. Acesso em: Novembro de 2016

Elementos Dublin Core	Descrição
<b>Title: Título</b>	Um título será o nome pelo qual o recurso é formalmente conhecido, podendo ser o próprio título.
<b>Creator: Autor</b>	Pode uma pessoa, uma organização ou um serviço.
<b>Subject: Assunto/ palavras-chave</b>	O assunto será expresso com palavras-chave, descritores ou códigos de classificação que descrevem o tema do recurso (indica o conteúdo informativo).
<b>Description: Descrição</b>	descrição pode incluir tabelas de conteúdo, referências para uma representação de conteúdo ou um texto livre de relato do conteúdo.
<b>Publisher: Editor</b>	Inclui uma pessoa, uma organização ou serviço.( o nome do editor deve ser usado para indicar uma entidade).
<b>Contributor: Contribuidor/ colaborador</b>	Inclui uma pessoa, uma organização ou serviço.( o nome do editor deve ser usado para indicar uma entidade).
<b>Date: Data</b>	Data será associada a criação ou disponibilização do recurso. recomenda-se o uso da norma ISO 8601 e segue o formato AAAA/MM/DD.
<b>Type: Tipo do recurso</b>	Descrição de categorias gerais, funções, espécies ou níveis de agregação para o conteúdo, recomenda-se utilizar vocabulário controlado. ( para descrever manifestações física ou digital do recurso deve-se usar o elemento Formato).
<b>Format: Formato</b>	Pode incluir o tipo da mídia ou as dimensões do recurso, pode ser usado para determinar o software, hardware ou outro equipamento necessário para mostrar ou operar o recurso.
<b>Identifier: Identificador do recurso</b>	recomenda-se utilizar o string ou número conforme um sistema de identificação formal. Exemplo: (Uniform Resource Identificador URI) e outros.
<b>Source: Fonte</b>	O presente recurso pode ser derivado de uma fonte de recurso inteira ou em parte, recomenda-se utilizar o string ou número conforme um sistema de identificação formal.
<b>Language: Idioma</b>	A recomendação para o melhor uso dos valores do elemento língua é definida pela RFC 1766 que inclui um código de língua em 2 letras( do padrão ISO 639), seguido opcionalmente pelo código do país em 2 letras também (do padrão ISO 3166).
<b>Relation: Relação</b>	Recomenda-se utilizar o string ou número conforme um sistema de identificação formal.
<b>Coverage: Abrangência/ Cobertura</b>	Inclui localização espacial, período temporal ou jurisdição, recomenda-se utilizar vocabulário controlado.
<b>Rights: Gerenciamento de Direitos autorais</b>	Conterá uma declaração de gerenciamento de direitos para o recurso. Informações de Direitos frequentemente abrangem Direito de Propriedade Intelectual (Intellectual Property Rights IPR), Copyright, e várias propriedades de Direitos.

**Quadro 8:** Elementos fundamentais do padrão Dublin Core de metadados.

Fonte: elaborado pelo autor.

No Quadro 9 temos um esboço de compatibilidade entre os campos MARC (*Machine Readable Cataloging*), formato utilizado em catálogos de biblioteca, e os principais elementos do *Dublin Core*. Tomamos como exemplo esses padrões para demonstrar a viabilidade de interação entre um repositório institucional operando com metadados *Dublin Core* e um sistema desenvolvido para um catálogo geral de obras distribuídas pelos acervos das bibliotecas que opere com o formato MARC.

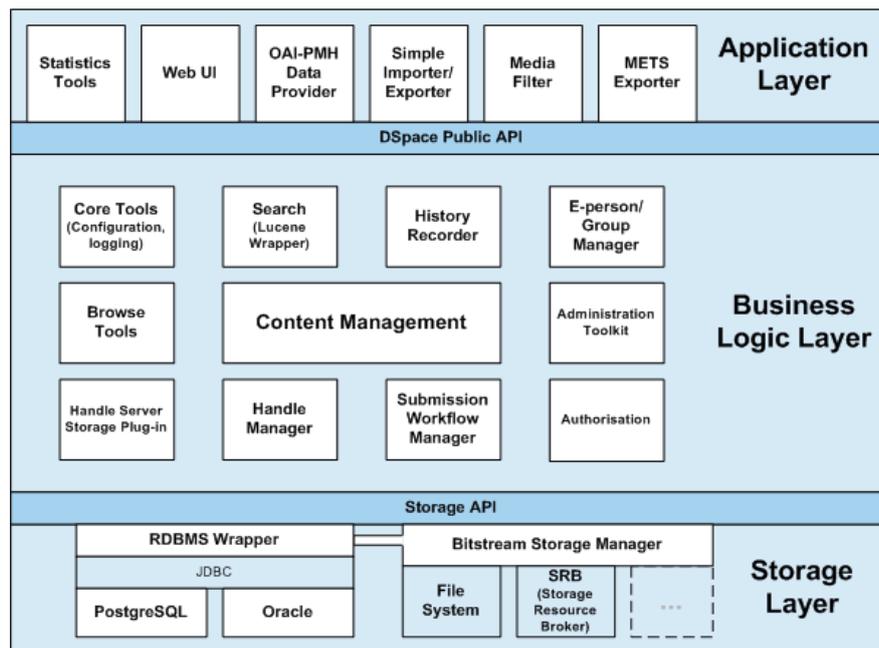
MARC fields	DC Element	Implementation Notes
100, 110, 111, 700, 710, 720	Contributor	
651, 662, 751, 752	Coverage	
008/07-10, 260\$c\$q	Creator	Creator element not used.
008/07-10, 260\$c\$q	Date	
500-599, except 506, 340, 856\$q	Description	
340, 856\$q	Format	
020\$a, 022\$a, 024\$a, 856\$u	Identifier	
008/35-37, 041\$a\$b\$d\$e\$f\$q\$h\$i, 546	Language	
260\$a\$b, 530, 760-787\$o\$t	Publisher	
530, 760-787\$o\$t	Relation	
506, 540, 534\$t	Rights	
534\$t, 786\$o\$t	Source	
050, 060, 080, 082, 600, 610, 611, 630, 650	Subject	
245, 246, Leader06, Leader07, 655	Title	Repeat dc:title for each.
Leader06, Leader07, 655	Type	

**Quadro 9:** Compatibilidade MARC Dublin Core

Fonte: <https://www.loc.gov/marc/marc2dc.html>

As sessões no DSpace são disponibilizadas principalmente através de uma interface web, mas também suporta o OAI-PMH v2.0 e é capaz de exportar pacotes METS (Metadata Encoding and Transmission Standard). O DSpace suporta os padrões comuns de interoperabilidade usados no domínio dos repositórios institucionais, como o *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE), SWORD, WebDAV, OpenSearch, OpenURL, RSS, ATOM. Versões mais recentes do DSpace também suportam a pesquisa facetada e a funcionalidade de navegação usando o Apache Solr.

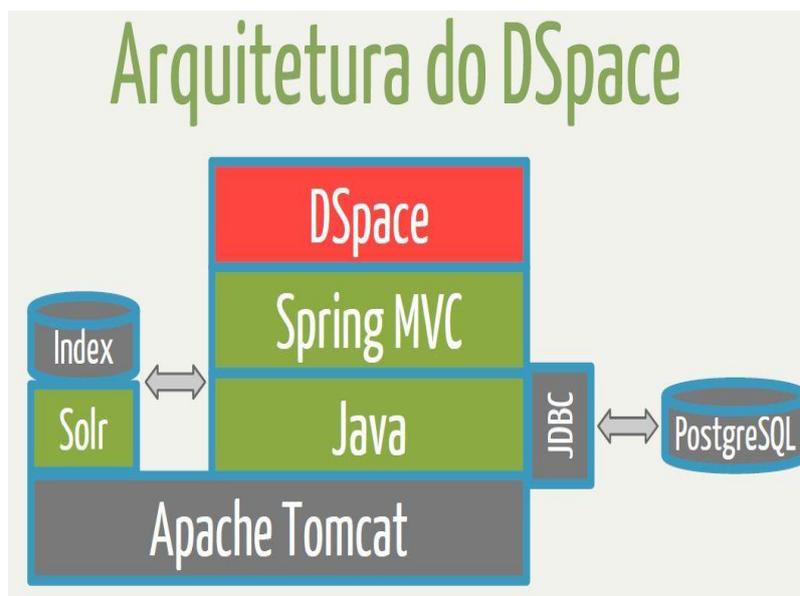
O DSpace está organizado em três camadas (Figura 9), cada uma das quais consiste em vários componentes. A camada de armazenamento (*storage layer*) é responsável pelo armazenamento físico de metadados e conteúdo. A camada de lógica de negócios (*business logic layer*) trata do gerenciamento do conteúdo do arquivo, dos usuários do arquivo (*e-people*), da autorização e do fluxo de trabalho. A camada de aplicativo (*application layer*) contém componentes que se comunicam com o mundo fora da instalação individual do DSpace, por exemplo, a interface do usuário da internet e o protocolo OAI para o serviço de coleta de metadados (DuraSpace, 2016).



**Figura 9:** Esquema da divisão em camadas do DSpace.  
 Fonte: (DuraSpace, 2016)

O Dspace não é particularmente exigente com os requisitos de hardware, com apenas 8 Giga Bytes de memória RAM e Hard Disk de 1 Tera Byte é possível gerenciar mais de 500 mil registros. Os requisitos de software fluem no mesmo caminho da simplicidade, a versão 5.x precisaria para uma instalação típica dos seguintes programas: Java 7; Apache Maven >= 3.0.5; Apache Ant >= 1.8; PostgreSQL >= 9.0; Apache Tomcat >= 7; Apache Solr (já vem embutido no pacote do DSpace).

Conceitualmente essa disposição de softwares resultará na estrutura demonstrada na Figura 10:



**Figura 10:** Arquitetura do DSpace.  
 Fonte: elaborada por Diego Ucha

Dentro dessa arquitetura os componentes em destaque desempenham as seguintes funções para o funcionamento do Dspace:

- 1) Solr - é o indexador, utilizado para mostrar os resultados de pesquisa no DSpace. Ele é composto de um índice (index no diagrama), que armazena o índice invertido (esse índice é construído com base em uma lista de documentos, tem esse nome por inverter a hierarquia da informação; ao invés de uma lista de documentos contendo termos, é obtida uma lista de termos que faz referência aos documentos).
- 2) Apache Tomcat - é a aplicação de servidor web, que serve as páginas do Dspace.
- 3) PostgreSQL - é o sistema gerenciador de banco de dados relacional, que se conecta ao Java via um plugin chamado JDBC.
- 4) Java - é a linguagem de programação na qual o DSpace está escrito.
- 5) Spring MVC - é um framework web que abstrai acesso ao banco de dados através de uma camada Java de entidades.

Expostas todas as principais características do programa as vantagens preconizadas pelo DSpace residem no fato de ser uma aplicação com código de acesso aberto, altamente customizável, utilizado por uma parcela majoritária de organizações que criaram repositórios institucionais; a grande dispersão do programa traz uma vantagem adicional que é a de formar uma vasta rede de colaboradores que auxiliam no desenvolvimento do aplicativo<sup>11</sup>.

## Conclusões

A tarefa de construir um repositório institucional será uma imposição cada vez mais recorrente para as instituições que lidam, principalmente, com informação científica e precisam de alguma forma mensurar a quantidade e impacto daquilo que são capazes de produzir. Boa parte da comunicação científica está alicerçada no contexto da publicação de livros e periódicos científicos avaliados por pares e indexados em bases de dados, o que promove uma disseminação incompleta da informação, pois apenas um número limitado de organizações e pessoas poderá ter pleno acesso aos conteúdos (muitas vezes limitado até aos autores da própria obra).

Os repositórios institucionais são, sem dúvida, alternativa recomendada para que as instituições sejam protagonistas e tenham relevância no ciclo da comunicação, não se trata de rejeitar o sistema estabelecido, mas operar em consonância com a missão que lhes é atribuída: preservação e divulgação. A opção pelo repositório institucional perpassa por uma série de decisões importantes, dentre as quais destacamos a escolha de um software adequado.

## Referências Bibliográficas

- Barton, M. R., & Waters, M. M. (2004). *Creating an Institutional Repository: LEADIRS Workbook*. Cambridge-MIT Institute.
- Biswas, G., & Paul, D. (Fevereiro de 2010). An evaluative study on the open source digital library softwares for institutional repository: Special reference to Dspace and greenstone digital library. *International Journal of Library and Information Science*, 2(1), 1-10.
- Crow, R. (2002). The Case for Institutional Repositories: A SPARC Position. *ARL Bimonthly Report*(223), 1-4.
- Directory of Open Access Repositories. (s.d.). *OpenDOAR*. Acesso em Novembro de 2016, disponível em <http://www.opendoar.org/>

<sup>11</sup> Podemos encontrar material sobre o DSpace em várias redes de colaboração, destaco especialmente a página wiki mantida pela DuraSpace. Disponível em: <<https://wiki.duraspace.org>>. No Brasil além do canal oficial oferecido pelo Ibtic destaco o grupo de discussões Dspace Brasil. Disponível em: <<https://groups.google.com/forum/#!forum/dspace-brasil>>. Acessos em: novembro de 2016

- DSpace. (Novembro de 2016). <http://www.dspace.org/introducing>. Acesso em 2016, disponível em DSPACE.ORG: [http://www.dspace.org/sites/dspace.org/files/media/DSpace%20Diagram\\_0.pdf](http://www.dspace.org/sites/dspace.org/files/media/DSpace%20Diagram_0.pdf)
- DuraSpace. (2016). *DURASPACE*. Acesso em novembro de 2016, disponível em <https://wiki.duraspace.org>
- DuraSpace. (2016). *DuraSpace History*. Acesso em 2016, disponível em DURASPACE.ORG: <http://www.duraspace.org/history>
- Fonseca, E. N. (2007). *Introdução à Biblioteconomia* (2 ed.). Brasília: Briquet de Lemos.
- IBICT. (2016). *Sobre Repositórios Digitais*. Acesso em 14 de Novembro de 2016, disponível em IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia: <http://www.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais>
- Leite, F. C. (2009). *Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto*. Brasília: Ibict.
- Leite, F., Amaro, B., Batista, T., & Costa, M. (2012). *Boas práticas para a construção de repositórios institucionais da produção científica*. Brasília: Ibict.
- Lynch, C. A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report*(226), 1-7.
- Marill, J. L., & Luczak, E. C. (Maio/Junho de 2009). Evaluation of Digital Repository Software at the National Library of Medicine. *D-Lib Magazine*, 15(5/6). doi:10.1045/may2009-marill
- Marra, P. d. (2012). O PAPEL DAS BIBLIOTECAS UNIVERSITÁRIAS NA COMUNICAÇÃO CIENTÍFICA: UM ESTUDO SOBRE OS REPOSITÓRIOS INSTITUCIONAIS. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 17(esp. 2 - III SBCC), 174-194. doi:10.5007/1518-2924.2012v17nesp2p174
- Powell, A. (Dezembro de 2005). Notes about possible technical criteria for evaluating institutional repository (IR) software. *UKOLN*. Acesso em 2016 de Novembro, disponível em JISC Information Environment (JISC IE): <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/ir-software.pdf>
- RCAAP. (2016). *Módulo 1 - Introdução ao Open Access e ao DSPACE*. Acesso em Novembro de 2016, disponível em RCAAP: [http://projeto.rcaap.pt/formar/mod1/contents/o\\_sistema\\_dspace.html](http://projeto.rcaap.pt/formar/mod1/contents/o_sistema_dspace.html)
- Registry of Open Access Repositories. (s.d.). *ROAR*. Acesso em Novembro de 2016, disponível em <http://roar.eprints.org/>
- Rodríguez-Gairín, J. M., & Duesa, A. S. (2008). DSpace: un manual específico para gestores de la información y la documentación. *BiD: textos universitaris de biblioteconomia i documentació*(20). Acesso em novembro de 2016, disponível em <http://bid.ub.edu/20rodri2.htm>
- Sayão, L. F., & Marcondes, C. H. (2009). Software livres para repositórios institucionais: alguns subsídios para a seleção. In: L. Sayão, L. B. Toutain, F. G. Rosa, & C. H. Marcondes (Eds.), *Implantação e gestão de repositórios institucionais : políticas, memória, livre acesso e preservação* (pp. 23-54). Salvador: EDUFBA.
- Silva, T. E., & Tomaél, M. I. (2008). Repositórios Institucionais e o Modelo Open. In: M. I. Tomaél, *Fontes de Informação na Internet* (pp. 123-149). Londrina: EDUEL.