



30^o CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO



25 a 29 de novembro 2024

**Bibliotecas Fortes:
Sociedade Democrática Recife, PE**

Eixo 6 – O mundo digital: apropriação e desafios

Modalidade: trabalho completo

Reconhecimento de elementos para a implementação de um repositório digital para o Observatório Magnético de Tatuoca

Recognition of elements for the implementation of a digital repository for the Tatuoca Magnetic Observatory

Afonso Santos Serejo – Universidade Federal do Pará (UFPA)

Jacquelin Teresa Camperos Reyes – Universidade Federal do Pará (UFPA)

Resumo: Dada a necessidade de intervenção com fins de preservação de acervos técnico-científicos, tornam-se imperativas estratégias para acessibilidade, curadoria e aproveitamento dos recursos informacionais produzidos. Esta pesquisa busca identificar aspectos estruturais para implantação de repositórios digitais que divulguem produções técnico-científicas, como as contidas no acervo do Observatório Magnético de Tatuoca. A metodologia usada foi a revisão de literatura, recuperando subsídios teóricos voltados à curadoria digital, para organização e disponibilização de recursos informacionais, integrando áreas como Biblioteconomia, Arquivística e Ciência da Informação. Conclui-se dissertando sobre o valimento dos repositórios digitais para acesso e uso de produções dos diversos campos científicos.

Palavras-chave: Repositórios Digitais. Observatório Magnético de Tatuoca. Preservação Digital. Ciência Aberta.

Abstract: Given the need for intervention with purpose of preserving technical-scientific collections, strategies for accessibility, curatorship and use of the informational resources produced become imperative. This research seeks to identify structural aspects for implementation of digital repositories that disseminate technical-scientific productions, such as those contained in the Tatuoca Magnetic Observatory collection. The methodology used was the literature review recovering theoretical subsidies headed to digital curation, organization and availability of informational resources, integrating the areas Library Science, Archival Science and Information Science. It concludes by discussing the value of digital repositories for access and use of productions from the various scientific fields.

Keywords: Digital Repositories. Digital Preservation. Tatuoca Magnetic Observatory. Open Science.





1 INTRODUÇÃO

Ao longo das últimas décadas, houve grande aumento na produção técnico-científica das instituições pesquisadoras no mundo. A quantidade cada vez mais crescente de informação, acabou por sujeitar os integrantes do meio científico a adequar-se às condições que possibilitem que tais informações sejam adquiridas, armazenadas, preservadas e redistribuídas para o público em geral. É importante destacar que o acesso à informação é, inicialmente, um direito que o ser humano possui; direito esse que não está limitado apenas à produção informacional atual, mas também toda aquela ao longo da história. Esse preceito leva a entender que as instituições acadêmico-científicas têm por dever, encontrar as mais variadas maneiras de lidar com a informação, para que a humanidade venha a ter acesso ao conhecimento.

Os repositórios digitais revolucionaram a forma como a pesquisa acadêmico-científica é compartilhada, armazenada e acessada — principalmente ao se considerar como o célere avanço tecnológico vem facilitando a transmissão de informação. Atuam como plataformas que permitem à geração atual e mesmo às futuras, guardar e preservar seu patrimônio informacional, contribuindo para o “conhecimento coletivo” de uma comunidade. Ou seja, são ferramentas essenciais para a preservação e divulgação da informação produzida. Tão logo, este artigo tem por objetivo identificar aspectos estruturais para implantação de repositórios digitais que divulguem produções técnico-científicas, como as contidas no acervo do Observatório Magnético de Tatuoca. Para além disso, fazer uma sucinta descrição do papel das instituições enquanto gestoras dos repositórios e informações neles contidas.

2 MÉTODOS

A revisão de literatura se direcionou na busca do marco teórico que envolve o uso de repositórios digitais e a identificação de atributos de sistemas de disponibilização de recursos informacionais da mesma natureza que o observatório de Tatuoca.

Foram utilizados estudos de pesquisadores que contribuíram em amplo subsídio no contexto de repositórios digitais no Brasil. A fim de demarcar as mais apropriadas opções de referenciais teóricos, fez-se uso das plataformas Google Scholar, Scielo e do

Instituto Brasileiro de Informação em Ciência e Tecnologia. Na primeira plataforma, utilizou-se a ferramenta de busca para procurar o termo “repositórios digitais”. Para a delimitação dos resultados obtidos, averiguou-se a lista dos 100 primeiros na ordem de relevância, constatando-se os nomes dos autores marcados na listagem como os mais proeminentes e com maior número de citações, com o intuito de definir o nicho dos pesquisadores mais significativos para a pesquisa. A quantidade de citações foi um dos critérios escolhidos pelo fato de que, quanto mais citado o nome de um autor, maior é sua notoriedade em sua área de pesquisa; isto é, em tese, número de citações reflete o grau de importância que um autor agrega a determinado nicho de estudo.

Após a definição dos autores principais, passou-se para a utilização das segunda e terceira plataformas — a fim de encontrar-se as produções científicas relacionadas com o assunto “repositórios digitais”. Utilizando-se suas respectivas ferramentas de busca para encontrar produções científicas desses pesquisadores, fez-se nova delimitação, considerando-se quais dos mesmos possuíam o maior número de citações e produções científicas — o que auxiliou a defini-los como os maiores especialistas no assunto “repositórios digitais”. Deste modo, escolheu-se quais estudiosos comporiam o referencial teórico desta pesquisa, ao propósito de obter maior nível de qualidade de conteúdo para ela.

Como critério secundário, buscou-se nas três plataformas, também estudos de caso, pois, para explicar o desenvolvimento de um repositório digital, a análise de casos práticos deve obrigatoriamente compor o referencial. Os estudos de caso escolhidos foram os mais similares ao contexto de Tatuoca: repositórios acadêmico-científicos — incluindo *softwares* utilizados — cercando a área da geociência.

A metodologia se norteia essencialmente em manifestar a importância dos repositórios digitais, ferramentas de *software* optados por instituições acadêmico-científicas para comportar suas estruturas, além das medidas que devem ser consideradas para a preservação da informação e patrimônio arquivístico como as contidas no Observatório Magnético de Tatuoca.



3 A IMPORTÂNCIA DE REPOSITÓRIOS DIGITAIS E A PRESERVAÇÃO DA INFORMAÇÃO

A partir da Segunda Guerra Mundial, o mundo prosseguiu com uma crescente percepção do papel central da Ciência para o desenvolvimento, e, conjuntamente, o papel estratégico da chamada Informação em Ciência e Tecnologia (ICT) como insumo. A partir daí, gerenciar e otimizar os fluxos da ICT tornaram-se atividades **estratégicas** de governos e órgãos internacionais (Marcondes; Sayão, 2009a, grifo nosso).

Repositórios digitais são ambientes informacionais destacados no cenário científico e tecnológico atual, permitindo o armazenamento, disseminação e preservação de produções intelectuais, científicas e/ou artísticas de uma instituição (repositórios institucionais); de uma área do conhecimento (repositórios temáticos); de uma comunidade não necessariamente ligada a uma instituição de ensino e pesquisa, em um contexto de “acesso aberto”; ou de dados advindos de pesquisas científicas (repositórios de dados de pesquisa), que podem ser compartilhados por pesquisadores de uma comunidade científica, a visar sua (re)utilização no contexto da Ciência Aberta (Vechiato *et al.*, 2017). O Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT (2005, p. 1), por exemplo, reforça em seu *Manifesto Brasileiro de Apoio do Acesso Livre à Informação* que:

A informação científica é o insumo básico para o desenvolvimento científico e tecnológico de uma nação. Trata-se de um processo contínuo em que a informação científica contribui para o desenvolvimento científico, e este, por sua vez, gera novos conteúdos realimentando todo o processo. No entanto, a comunidade científica enfrenta dificuldades no acesso à informação científica, se considerado o modelo tradicional de publicação científica. Tradicionalmente, os artigos ou trabalhos científicos são publicados em revistas especializadas, e a forma de acesso [...] dá-se mediante assinatura das publicações pelas bibliotecas ou pelo pesquisador (IBICT, 2005, p. 1).

A lógica que preside no surgimento dos repositórios no cenário internacional da ICT é a de que os resultados da atividade científica, na forma de diferentes publicações, devem necessariamente também ser públicos, especialmente em razão de muitas vezes serem obtidos às custas de pesados investimentos públicos; devem poder ser utilizados amplamente, sem qualquer forma de apropriação privada. Desse modo, as instituições científicas ou acadêmicas e suas comunidades, devem manter cópias de sua produção científica em seu repositório institucional de livre acesso (Marcondes; Sayão, 2009a).

O *Glossário de Documentos Arquivísticos Digitais* do Conselho Nacional de Arquivos – CONARQ (2020, p. 39) define Preservação Digital como o “conjunto de ações

gerenciais e técnicas exigidas para superar as mudanças tecnológicas e a fragilidade dos suportes, garantindo o acesso e a interpretação de documentos digitais pelo tempo que for necessário”. Como destaca Márdero Arellano (2004, p. 15):

Na era da informação digital se está dando muita ênfase à geração e/ou aquisição de material digital, em vez de manter a preservação e o acesso a longo prazo aos acervos eletrônicos existentes. [...] A preservação dos documentos continua a ser determinada pela capacidade de o objeto informacional servir às utilizações que lhe são imputadas, às suas atribuições que garantem que ele continue a ser satisfatório às utilizações posteriores (Márdero Arellano, 2004, p. 15).

Isso significa que a informação precisa circular em um intervalo de tempo aceitável, com o devido registro e com um armazenamento adequado. A informação digital, nesse quesito, pode ser definida como um tipo de informação onde se mantém os objetivos de uso, mas com uma especificidade referida a como é produzida, organizada, administrada, distribuída, acessada e preservada. Pode estar armazenada em diversos suportes acessíveis por múltiplas maneiras, e essa diversidade de meios de suporte e acesso é um dos principais atributos que diferenciam uma informação disponível em meio digital de uma armazenada em meio analógico (Grácio; Fadel, 2010).

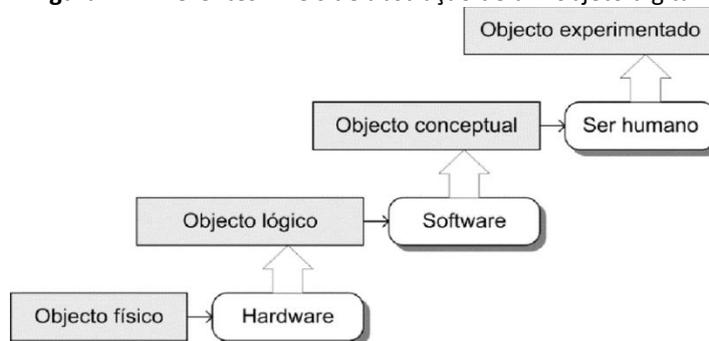
A fragilidade documental deve ser uma preocupação em ambos os ambientes real e virtual. Ferreira (2006, p. 19) comenta que:

A obsolescência tecnológica não se aplica somente ao nível dos suportes físicos. No domínio digital, todo o tipo de material tem obrigatoriamente de respeitar as regras de um determinado formato. Isto permite que as aplicações de *software* sejam capazes de abrir e interpretar adequadamente a informação armazenada. À medida que o *software* vai evoluindo, também os formatos por ele produzidos vão sofrendo alterações (Ferreira, 2006, p. 19).

É pressuposto que a preservação digital consiste justamente na capacidade de garantir que a informação permaneça acessível e com qualidade para ser usada no futuro, através de uma plataforma tecnológica diferente da utilizada no instante de sua criação. Um objeto digital pode ser definido como todo e qualquer objeto de informação representável por uma sequência de dígitos binários; definição que acomoda tanto a informação nascida em contexto tecnológico digital — objetos nato-digitais — como a informação digital obtida via suportes analógicos — objetos digitalizados. Para um objeto digital ser decifrado, é preciso que ocorra, essencialmente uma série de transformações partindo dos atributos iniciais, considerando sua origem como objeto

físico, o domínio junto aos símbolos, *hardware*, e *software* de manipulação (Ferreira, 2006). A figura 1 representa os níveis de abstração do objeto digital.

Figura 1 – Diferentes níveis de abstração de um objeto digital.



Fonte: Ferreira (2006, p. 23).

Descrição: Figura explicando os níveis de abstração de um objeto digital. O objeto físico segue o *hardware*; o *hardware* segue o objeto lógico; o objeto lógico segue o *software*; o *software* segue o objeto conceptual; o objeto conceptual segue o ser humano; e o ser humano segue o objeto experimentado.

Os epicentros da problemática da preservação de documentos em meio digital são a “atualização de suporte” e a “normalização de formato”. O CONARQ (2020, p. 11 e 36) define a primeira como a “técnica de migração que consiste em copiar os dados de um suporte para outro, sem mudar sua codificação”; e a segunda como a “conversão de formatos de arquivo para um elenco gerenciável de formatos apropriados para preservação e acesso”. Repositórios digitais são concebidos como ferramentas para contribuir na busca de soluções para a preservação, acesso e disponibilização de objetos informacionais produzidos no interior das organizações acadêmicas e científicas.

4 RESULTADOS E DISCUSSÕES

Para se criar um repositório institucional é preciso, primeiramente, definir a necessidade de criá-lo. Somente após as políticas de gestão e de uso estarem fixadas, e o elenco de serviços que serão ofertados a uma comunidade definida de usuários estarem estabelecidos, será o momento de avaliar as questões operacionais necessárias (Marcondes; Sayão, 2009b). Portanto, os mais preocupados são os grupos responsáveis pelas bibliotecas e arquivos nas instituições; grupos esses que exigem estratégias metodológicas bem definidas para o desenvolvimento de padrões e mecanismos legais para lidar com arquivos eletrônicos (Márdero Arellano, 2004).

Porém, é importante ter em mente que a opção da instituição não é o desenvolvimento de uma solução própria. Para tal, é preciso compor um conjunto de requisitos que refletirão as diretrizes e políticas gerais adotadas pelo repositório. São esses requisitos que estabelecerão uma métrica para avaliação dos pacotes de *software* disponíveis e, para elaborar esse conjunto e avaliar as alternativas, normalmente as instituições formam comitês integrados por: membros da administração da biblioteca, profissionais de tecnologia da informação e representantes dos segmentos de potenciais usuários do repositório (Marcondes; Sayão, 2009b).

Sendo o repositório digital um guardador abissal de conteúdo, a preocupação com o armazenamento deve ser tomada com extrema cautela, pois refletirá tanto na garantia de não perda de informação, bem como no volume final de circulação de dados no repositório. É preciso estabelecer como a interface adequar-se-á às condições de uso, tanto dos indexadores quanto dos usuários. Facilidade de manipulação, interligação e reutilização, são essencialmente características da informação digital que se traduzem em um alargamento do campo da preservação digital (Ferreira, 2006). Para que esta ocorra, são necessárias abordagens tecnológica e sistêmica, pois as instituições têm imensa responsabilidade na manutenção, garantia de acesso e utilização do patrimônio que está sendo digitalmente preservado. E os repositórios digitais consistem, justamente de maneira sistêmica, em uma integração de indivíduos, sistemas e tecnologias responsáveis por proteger e fornecer acesso à informação para a sua comunidade de interesse (Corujo; Ferreira, 2015).

Shintaku e Meirelles (2010, p. 20) destacam que “uma das decisões do administrador ou da equipe responsável pelo repositório é exatamente decidir qual das interfaces o repositório irá utilizar”. Ao se pensar na construção de um repositório, é importante antecipar a gama e tipos de documentos que serão nele armazenados, e a forma mais apropriada de recuperá-los. O projeto da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), por exemplo, atende a mais de 140 instituições e agrega mais de 900 mil documentos (BDTD, c2024); e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), entidade vinculada ao Ministério da Educação do Brasil (MEC), estabeleceu, por meio da Portaria nº13/2006, a obrigatoriedade de disponibilizar ao público em geral, as teses e dissertações produzidas pelos programas de mestrado e doutorado em formato digital — medida que visa fomentar a disseminação científica e

também evidenciar a postura favorável do Brasil em relação ao acesso livre e integral às teses e dissertações via ambientes digitais, tais como os repositórios das instituições de ensino superior e de pesquisa (Shintaku; Vidotti, 2016).

4.1 Softwares de código aberto para repositórios digitais

Existem inúmeras opções de *softwares* agregadores de *websites*. Repositórios também estão inclusos, mas com a diferença de que contêm uma armazenagem imensa de documentos e em inúmeros formatos. Por todos os motivos anteriormente mencionados, é importante analisar as opções a se utilizar no desenvolvimento de um repositório digital, pois cada repositório, com suas características e necessidades próprias, precisa ser criado a partir de um *software* capaz de conter toda a sua documentação e permitir o rápido e denso acesso de usuários. Assim, a seguir faz-se uma análise e posterior comparação entre três populares *softwares* utilizados mundialmente na construção de repositórios digitais.

4.1.1 GREENSTONE DIGITAL LIBRARY (GSDL)

Desenvolvido e distribuído em cooperação com a UNESCO e a ONG *Human Info*, possui como objetivo a capacitação de usuários — especialmente em universidades, bibliotecas e outras instituições de serviço público — para que sejam capazes de construir as suas próprias bibliotecas digitais (Martins; Silva; Siqueira, 2018). Este *software* provê uma maneira de organizar a informação e publicá-la tanto na rede como em qualquer outra plataforma digital, como DVDs e *flash drives USB*; isto é, permite a navegação em um meio ambiente não conectado à rede. As bibliotecas digitais construídas pelo GSDL possuem recursos digitais completamente buscáveis e dirigidos por metadados. Porém, o GSDL não deve ser considerado uma biblioteca digital em si, mas uma ferramenta que permite a sua construção. Além disso, com ele há uma incapacidade para atualizar e gerenciar conteúdo interativo, identificar arquivos duplicados, um processo de construção de coleções deveras dificultoso (Verma; Kumar, 2018, tradução nossa).



4.1.2 EPRINTS

Apesar de compartilhar muitas das características comumente observadas em sistemas de gestão de documentos, é usado principalmente para repositórios institucionais e periódicos científicos. Foi desenvolvido pela *School of Electronics and Computer Science* na *University of Southampton* e disponibilizado sob a licença GPL (*General Public License*). Por objetivo, busca criar um repositório *web* altamente configurável, sendo aplicado não apenas para pesquisa de artigos, mas também para outros tipos de arquivo, tais como imagens, dados, áudio, ou qualquer item armazenável digitalmente (Martins; Silva; Siqueira, 2018).

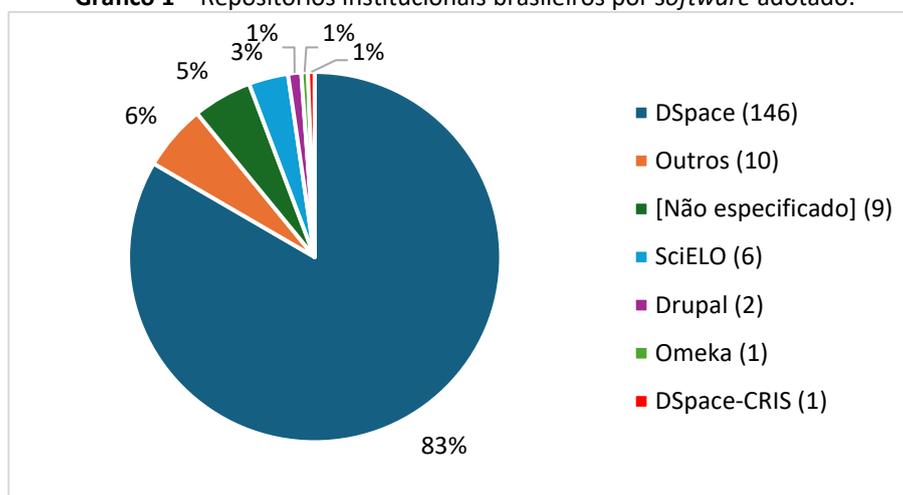
Inicialmente, requeria plataformas de *software* como Linux, Apache, MySQL e Perl, mas atualmente também é aplicável ao sistema operacional Windows. Permite aos usuários inserir documentos em formulários *web* — caso haja metadados suficientes para tal; se conecta à base de dados SHERPA/RoMEO, que ajuda os autores a verificar seus direitos a respeito de suas submissões nos repositórios e a remover qualquer publicação não autorizada; possibilita que a informação dos metadados seja provida com cópias eletrônicas do documento; e sua fácil customização de metadados permite aos administradores disponibilizar somente campos realmente necessários para o contexto de cada repositório. Todavia, ainda que a categoria de navegação e customização de busca — que facilita a recuperação de documentos — sejam pontos positivos, não é possível inserir múltiplos registros de uma vez — um empecilho para colocar grandes volumes de documentos no repositório. Múltiplos arquivos é totalmente possível, mas apenas se pertencerem a um mesmo registro; a migração para a biblioteca do EPrints também pode falhar caso não existam as mesmas coleções no novo repositório, isto é, cada registro tem de ser criado individualmente, além de não ser possível criar registros em comum; e a falta de operadores booleanos limita a funcionalidade de busca (Verma; Kumar, 2018, tradução nossa).

4.1.3 DSPACE

Criado em 2002, pelo Instituto de Tecnologia de Massachusetts e pelos Laboratórios Hewlett-Packard, tem o propósito de prover um mecanismo de armazenamento para documentos digitais oriundos de pesquisas ou destinados à educação, e distribuído via licença de código aberto (Ferreira; Saraiva; Rodrigues, 2012).

Hoje, o DSpace é o sistema agregador da plataforma Capes e o mais utilizado no Brasil. Da lista de lista de 175 repositórios brasileiros registrados no OpenDOAR (c2024), 146, isto é, 83% do total, utilizam o DSpace (vide gráfico 1). Dentre eles, encontra-se o *site* do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), cujo apoio reflete diretamente no número de instituições que usam a ferramenta, dado o fato de prover segurança de ter uma instituição tradicional de pesquisa dando sufrágio a um *software* (Shintaku; Vechiato, 2018).

Gráfico 1 – Repositórios institucionais brasileiros por *software* adotado.



Fonte: Baseado em OpenDOAR (c2024).

Descrição: Gráfico em pizza comparando o percentual de *softwares* utilizados em um total de 175 repositórios brasileiros. De cima para baixo, lê-se: “DSpace (146)”, representando 83%; “Outros (10)”, representando 6%; “Não especificado (9)”, representando 5%; “SciELO (6)”, representando 3%; e “Drupal (2)”, “Omeka (1)” e “DSpace-CRIS (1)”, representando 1% cada.

O DSpace é um sistema gerenciador e protetor de objetos digitais que funciona como um serviço de informação que entrega arquivos digitais aos usuários de maneira facilitada, criando assim uma extensa rede de serviços informacionais. Sua criação baseou-se na comunicação científica, tendo por objetivo principal a divulgação da literatura científica em formato de artigos, nomeadamente aqueles que foram previamente publicados em revistas acadêmicas. Possui uma estrutura hierárquica, que melhora a organização das coleções e facilita a recuperação dos objetos digitais depositados, podendo ser dividida em (Shintaku; Meirelles, 2010):

- a) **comunidades:** possuem subcomunidades ilimitadas, que não contêm diretamente os objetos digitais, mas permitem a organização por temas, onde as comunidades representam grandes temas e as subcomunidades os refinam em subtemas, tornando a organização mais precisa e detalhada;



- b) **coleções:** podem ter administradores que controlem aspectos e itens nelas contidos, via permissões especiais para distribuir responsabilidades em diferentes níveis;
- c) **itens:** compostos por descrições e objetos digitais, são formados por campos descritivos associados a esses objetos, que juntos formam uma unidade. Contêm fluxo de submissão — influenciado pelo princípio científico de avaliação dos registros antes da publicação —, processo pelo qual o objeto digital é depositado e ajustado conforme as necessidades da coleção, até a disponibilização para acesso.

No entanto, o DSpace também apresenta desvantagens que podem atrapalhar o uso do repositório, tais como: a estrutura de arquivos *Flat* e metadados complexa; falta de escalabilidade e extensibilidade; API, capacidades de relatório e atributos de metadados limitados — em contraposto às subcomunidades ilimitadas —; e falta de suporte para dados entrelaçados (Verma; Kumar, 2018, tradução nossa).

4.1.3.1 DSpace aplicado a um repositório geocientífico

Existem muitas instituições da área da Geociência imersas em realidades similares à de Tatuoca e que conferem sobre as vantagens do DSpace. Uma delas é a IUGONET – *Inter-University Upper Atmosphere Global Observation Network* (Rede Interuniversitária de Observação Global da Atmosfera Superior) do Japão, um projeto iniciado em 2009, com o intuito específico de criar uma rede de metadados — isto é, dados que descrevem certos aspectos de informação dos dados — de observações atmosféricas de vários observatórios magnéticos e instituições afins.

À época, o Instituto Nacional de Investigação Polar do Japão e as universidades de Nagoya (Laboratório do Ambiente Solar-Terrestre), Quioto (Instituto de Investigação para a Humanosfera Sustentável, Centro de Análise de Dados para Geomagnetismo e Magnetismo Espacial, Observatórios de Kwasan e Hida), Kyushu (Centro de Investigação do Ambiente Espacial) e Tohoku (Centro de Investigação do Plasma Planetário e da Atmosfera) enfrentavam problemas com uso de dados, devido à necessidade de uma investigação multidisciplinar que combinasse variados tipos de observações terrestres. Os dados ou bases de dados das observações eram comumente mantidos e disponibilizados à comunidade por cada organização/grupo de investigação que as



efetuava; todavia, eram dificilmente utilizados por pesquisadores em outras áreas de investigação, em resultado da falta de informação sobre os dados. Para além disso, estes vinham sendo utilizados somente por um número muito reduzido de profissionais que participaram das pesquisas, e a disponibilidade para outros pesquisadores era pouco conhecida (Hayashi *et al.*, 2013, tradução nossa).

Dois principais pontos levaram os idealizadores a escolher o DSpace para desenvolvimento do repositório: a necessidade de evitar problemas como propriedade dos dados, autenticação e autorização; e construir um ambiente virtual de base de dados integrada para partilhar os metadados, incluindo o Localizador Uniforme de Recursos (URL) dos ficheiros de dados. Com sua alta interoperabilidade, o DSpace facilitaria o compartilhamento técnico de informação, customização e o gerenciamento na instalação de sistemas; além de ser muito adaptável à escalabilidade. Numa situação em que registros de metadados tornassem-se muito grandes, seria difícil gerenciá-los em um único servidor; assim, uma busca cruzada que utilizasse uma interface externa, possibilitaria conexões entre as múltiplas bases de dados da rede (Abe *et al.*, 2014, tradução nossa). E, como o *software* contém funções fundamentais para registrar, recuperar, fornecer, e recolher dados digitais escritos em vários formatos — além de permitir o gerenciamento dos mais comuns formatos de metadados —, o DSpace era o melhor candidato para propiciar um sistema estável de base de dados de metadados no curto período de 6 anos em que a IUGONET foi desenvolvida (Hayashi *et al.*, 2013, tradução nossa).

4.2 Comparativo entre os *softwares*

Não há nenhum *software* totalmente inexecutável; todos apresentam características favoráveis e desfavoráveis, sendo preciso compará-los para chegar a uma escolha. Como mostra o quadro 1, o GSDL apresenta maior gama de opções para formato de metadados e funciona mesmo em versões suplantadas de sistemas operacionais, tendo também sua própria base de dados. Já o EPrints não tem base de dados própria, mas pode ser integrado com um número maior de bases. E, dentre os três, o DSpace é o que consegue interagir com muito mais *softwares* e operar em sistemas e com linguagens de programação tão modernos e populares quanto os demais. Os três *softwares* compartilham de um mesmo formato de metadados, o popular Dublin Core;



todavia, o DSpace apresenta dois quesitos importantes que o diferem dos demais: o sufrágio que recebe do IBICT, e sua popularidade no Brasil, país no qual será implantado o repositório digital do Observatório Magnético de Tatuoca.

Quadro 1 – Comparativo entre os *softwares* GSDL, DSpace e EPrints.

Atributos	GSDL	DSpace	EPrints
Sistema operacional	Windows 95/98/Me/NT/2000/XP/10, Unix/Linux, e MAC OS-X	Windows (NT/2000/XP/10), Linux, BSD, UNIX-símiles, OSX	Linux, Unix, Windows
Linguagem de programação	C++, Perl, Java	Java e JSP	Perl
Base de dados	Sua própria	Oracle, PostgreSQL	MySQL, Oracle, PostgreSQL, Cloud
Formato de metadados	Dublin Core (qualificado e desqualificado), METS, RFC1807' NZGLS (<i>New Zealand Government Locator Service</i>), AGLS (<i>Australian Government Locator Service</i>)	Dublin Core (qualificado), METS	Dublin Core, METS
<i>Softwares</i> associados	Apache <i>web server</i> , Java 1.4.0 ou superior, Image Magick, Ghost <i>script</i> e navegador <i>web</i>	Java JDK5 ou superior, Apache Ant 1.6.2 ou superior, Apache Maven 2.0.8 ou superior, Java 1.4 ou superior, PostgreSQL 7.3 ou superior, Apache Tomcat 4.x/5.x e navegador <i>web</i>	Linux ou Unix, Apache, Perl
Custo de licença/custo de atualização	Gratuito	Gratuito	Gratuito

Fonte: Adaptado de Verma e Kumar (2018, p. 365-366, tradução nossa).

Descrição: Quadro de 7 linhas e 4 colunas fazendo um comparativo entre os *softwares* GSDL, DSpace e EPrints, respectivamente. A primeira coluna contém os atributos dos respectivos *softwares*. Como atributo “sistema operacional”, tem-se “Windows95/98/Me/NT/2000/XP/10, Unix/Linux, e MAC OS-X” para o GSDL; “Windows (NT/2000/XP/10), Linux, BSD, UNIX-símiles, OSX” para o DSpace; e “Linux, Unix, Windows” para o EPrints. Como atributo “linguagem de programação”, tem-se “C++, Perl, Java” para o GSDL; “Java e JSP” para o DSpace; e “Perl” para o EPrints. Como atributo “base de dados”, tem-se “sua própria” para o GSDL; “Oracle, PostgreSQL” para o DSpace; e “MySQL, Oracle, PostgreSQL, Cloud” para o EPrints. Como atributo “formato de metadados”, tem-se “Dublin Core (qualificado e desqualificado), METS, RFC1807' NZGLS (*New Zealand Government Locator Service*), AGLS (*Australian Government Locator Service*)” para o GSDL; “Dublin Core (qualificado), METS” para o DSpace; e “Dublin Core, METS” para o EPrints. Como atributo “*softwares* associados”, tem-se “Apache *web server*, Java 1.4.0 ou superior, Image Magick, Ghost *script* e navegador *web*” para o GSDL; “Java JDK5 ou superior, Apache Ant 1.6.2 ou superior, Apache Maven 2.0.8 ou superior, Java 1.4 ou superior, PostgreSQL 7.3 ou superior, Apache Tomcat 4.x/5.x e navegador *web*” para o DSpace; e “Linux ou Unix, Apache, Perl” para o EPrints. Como atributo “custo de licença/custo de atualização”, tem-se “gratuito” para todos os três *softwares*.

4.3 Política institucional voltada à preservação

Como previamente mencionado, as questões operacionais necessárias procedem o processo de fixação das políticas de gestão e de uso de repositórios institucionais. À medida que cresce a dependência das pessoas e organizações por dados digitais, torna-se cada vez mais crucial estabelecer métodos e políticas cooperativas que garantam a durabilidade e disponibilidade da informação (Ferreira; Saraiva; Rodrigues, 2012). A alta direção da organização ou entidade deverá obrigatoriamente dar sua aprovação à proposta de elaboração de uma política que garanta a preservação de documentos. Assim, é imprescindível criar uma equipe multidisciplinar — que deverá seguir à risca, a metodologia delineada no processo de planejamento —, idealmente liderada por profissional arquivista ou com experiência na área, para realizar todos os meios de coleta de informações necessários para o desenvolvimento dessas políticas (Holanda, 2019).

As instituições de ensino e pesquisa apontam uma crescente procura de atualização de seus serviços de informação, o que exige a digitalização e/ou reprodução dos seus acervos. A produção e disseminação de informação científica estão sendo cada vez mais impactadas pela velocidade das redes de comunicação, pelos ambientes virtuais de aprendizagem e pelos esforços colaborativos. Para acompanhar esse crescimento, há uma necessidade crescente de sistemas de informação que possam armazenar e recuperar dados eficientemente. Logo, a preservação digital far-se-á necessária na conjunção de vários fatores, tais como tecnologias emergentes, comunicação científica, depósito legal e direitos autorais (Márdero Arellano, 2008).

É importante saber que repositórios digitais operam num sistema de confiança de três níveis: a forma como estabelecem confiança nas suas respectivas comunidades de interesse; a forma como confiam nos fornecedores externos; e a forma como os consumidores confiam nas informações fornecidas. Assim, para atingir os seus objetivos, um repositório digital deve aderir a um conjunto predefinido de expectativas, que exigem que ele exista dentro de uma estrutura organizacional que facilite a preservação dos dados e a longevidade do próprio repositório. Além disso, esta estrutura deve assumir a responsabilidade de manter os ativos digitais ao longo do tempo, alinhando-se com os interesses daqueles que nela depositam informações, e dos atuais e futuros usuários (Corujo; Ferreira, 2015).



A relevância das estratégias atuais de preservação é posta à prova pela rápida obsolescência tecnológica. A capacidade de implementá-las depende da consistência dos avanços tecnológicos e das restrições orçamentais enfrentadas pelas instituições. Como resultado, é crucial criar planos econômicos para a preservação digital (Márdero Arellano, 2008). Ademais, a organização também deve realizar análise minuciosa de seu contexto institucional; análise que pode ser bifurcada em duas partes. A primeira está associada ao contexto das exigências de fontes externas à instituição, que impactam não só as suas operações, mas também a criação, manutenção e disponibilidade de documentos. A segunda se refere ao contexto interno — conjunturas de proveniência, procedimentos e tecnologia —, cujo levantamento inclui normas, regulamentos e práticas relacionadas com a produção, manutenção e acesso aos documentos arquivísticos digitais (Holanda, 2019).

Para ser considerado credível, um repositório deve demonstrar a sua responsabilidade pela longevidade financeira; cumprir as recomendações e regulamentos globais relativos à gestão, acesso e proteção de recursos digitais; estabelecer padrões para avaliar a fiabilidade dos seus sistemas, de acordo com as expectativas da comunidade; e manter políticas, procedimentos e avaliações de desempenho transparentes, sujeitos à revisão por entidades independentes. Igualmente significativas são as preocupações relacionadas com a infraestrutura tecnológica do repositório, incluindo *software*, *hardware* e medidas de segurança, sendo que todos esses fatores culminam em um processo de auditoria e certificação que segue parâmetros normativos (Corujo; Ferreira, 2015).

5 CONSIDERAÇÕES FINAIS

O estudo realizado almejou destacar a importância e impacto dos repositórios institucionais/digitais para os ambientes de informação técnico-científico. Na era digital em que a sociedade hoje vive, os métodos mais tradicionais de compartilhamento de pesquisas — tais como publicações em livros e/ou periódicos em meio físico, a exemplo — podem ser limitados por barreiras geográficas e inúmeros problemas de acesso. A comunidade acadêmica-científica mundial, investe muito tempo e grande esforço nos mais variados campos, e, portanto, sem uma plataforma que permita que as produções



sejam facilmente acessadas, seu impacto será reduzido. Ademais, as formatações nos tipos de arquivos digitais que, a todo momento são atualizadas, fazem com que os repositórios digitais também precisem atualizar-se com elas.

Entretanto, não se deve tratar repositórios como uma “rede social” para troca de informação e produção científica. Sua criação tem, por essência, a meta de preservar essa produção para uso posterior. Qualquer espécie de documento produzido em meio físico corre grande risco de perda. Esses objetos em si já possuem sua própria importância, mas as informações neles contidas são “sem preço”. Assim, compreender a função dos repositórios digitais é um dever essencial no cenário acadêmico atual.

Tão logo objetivou-se com este artigo, destrinchar os primeiros escopos de construção de repositórios digitais/institucionais, e, assim, em estudos futuros, identificar atributos para o projeto do repositório digital do Observatório Magnético de Tatuoca. Conforme foi-se desenvolvendo o estudo, viu-se que será necessário mais aprofundamento no assunto, analisar repositórios já existentes mais detalhadamente, almejando implementar os mais adequados atributos para o tipo de documentação que será nele armazenada. Destarte, constata-se que a preservação digital traz situações complexas a serem consideradas e resolvidas, visando proporcionar facilidade de acesso a recursos informacionais.

As funcionalidades e benefícios dos repositórios tornam-se desafios para os administradores das instituições, uma vez que estes trabalham em função de seus repositórios, e não o contrário. A conversão de um documento físico para o digital, por exemplo, é uma tarefa demorada, que deve ser feita criteriosa e cautelosamente; mas, mesmo assim, guardar a informação para sua futura recuperação é de maior importância, pois o ser humano tem o direito irrevogável de poder acessá-la.

Conforme mencionado, dentre os *softwares* estudados, o que mostrou maior vantagem para Tatuoca foi o DSpace. Sua estrutura hierárquica dada pela divisão ilimitada em subcomunidades, poderia permitir organizar os documentos (itens) — como magnetogramas, livros do ano, folhas do absoluto, a exemplo — por coleções, considerando subdivisões por períodos, formatos do arquivo físico original, e outras características de contexto arquivístico. Apesar de estar alguns passos atrás em certos quesitos comparado a outros *softwares*, o DSpace vem sendo atualizado constantemente, possui código aberto — uma das propostas buscadas para o repositório

de Tatuoca —, e foi criado especificamente para comportar conteúdo acadêmico-científico. Para além disso, sua escolha para unificar os diversos observatórios da rede IUGONET prova que é uma excelente opção para repositórios geocientíficos; e sua popularidade no Brasil mostra a confiança que as instituições brasileiras têm nele.

As áreas de Ciência da Informação, especialmente a Biblioteconomia e Arquivística, têm uma importância fundamental nesse processo de construção de repositórios digitais. Ao fato de que estes dois campos tratam diretamente com os objetos de preservação nos repositórios e com o público usuário, é importante também valorizá-los, ainda que não sejam os seus profissionais aqueles que programam o repositório diretamente. Todavia, sem suas orientações, os especialistas das áreas de informática e tecnologia da informação não conheceriam a melhor forma de lapidar as plataformas; do contrário, haveria um vácuo na armazenagem e preservação da informação. Assim, pode afirmar-se que a construção de repositórios digitais implica em processos de colaboração e de pesquisa interdisciplinar.

6 AGRADECIMENTOS

Esta pesquisa foi financiada pela Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA), pela concessão da bolsa de pesquisa de iniciação científica.

REFERÊNCIAS

ABE, S. *et al.* Progress of the IUGONET system – metadata database for upper atmosphere ground-based observation data. **Earth, Planets and Space**, [s. l.], v. 66, n. 133, p. 1-7, out. 2014. DOI: <https://doi.org/10.1186/1880-5981-66-133>. Disponível em: <https://link.springer.com/article/10.1186/1880-5981-66-133>. Acesso em: 11 maio 2024.

BDTD. Biblioteca Digital Brasileira de Teses e Dissertações. **Biblioteca Digital Brasileira de Teses e Dissertações**: acesso e visibilidade às teses e dissertações brasileiras. c2024. Disponível em: <https://bdtd.ibict.br/vufind/>. Acesso em: 19 mar. 2024.

CONARQ. Conselho Nacional de Arquivos. Câmara Técnica de Documentos Eletrônicos. **Glossário de documentos arquivísticos digitais**. 8ª versão. [Brasília, DF]: CONARQ/CTDE, 2020. 50 p. Disponível em: https://www.gov.br/conarq/pt-br/assuntos/camaras-tecnicas-setoriais-inativas/camara-tecnica-de-documentos-eletronicos-ctde/glosctde_2020_08_07.pdf. Acesso em: 03 nov. 2023.



CORUJO, L.; FERREIRA, M. Estudo comparativo de referenciais normativos para avaliação da confiabilidade e certificação em repositórios digitais. *In*: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS DE PORTUGAL, 12., 2015, Évora. **Anais** [...]. Évora, Portugal: Associação Portuguesa de Bibliotecários, Arquivistas, Profissionais da Informação e Documentação – BAD, 2015. p. 1-13.

Disponível em:

<https://publicacoes.bad.pt/revistas/index.php/congressosbad/article/view/1372>.

Acesso em: 22 nov. 2023.

FERREIRA, M. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006. 88 p. Disponível em: <https://hdl.handle.net/1822/5820>. Acesso em: 10 nov. 2023.

FERREIRA, M.; SARAIVA, R.; RODRIGUES, E. **Estado da arte em preservação digital**. [S. l.]: Repositório Científico de Acesso Aberto de Portugal – RCAAP, 2012. 60 p. Disponível em: <https://hdl.handle.net/1822/17049>. Acesso em: 22 nov. 2023.

GRÁCIO, J. C. A.; FADEL, B. Estratégias de preservação digital. *In*: VALENTIM, M. (org.). **Gestão, mediação e uso da informação** [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2010. p. 59-83. Disponível em: <https://books.scielo.org/id/j4gkh/pdf/valentim-9788579831171-04.pdf>. Acesso em: 08 nov. 2023.

HAYASHI, H. *et al.* Inter-university Upper Atmosphere Global Observation NETWORK (IUGONET). **Data Science Journal**, [s. l.], v. 12, p. 179-184, abr. 2013. DOI: <https://doi.org/10.2481/dsj.WDS-030>. Disponível em: <https://datascience.codata.org/articles/10.2481/dsj.WDS-030>. Acesso em: 12 maio 2024.

HOLANDA, A. P. de. **Recomendações para elaboração de política de preservação digital**. Rio de Janeiro: Arquivo Nacional, 2019. 24 p. (Série recomendações para gestão de documentos nos órgãos e entidades do Poder Executivo Federal, 2). Disponível em: https://www.gov.br/arquivonacional/pt-br/servicos/gestao-de-documentos/orientacao-tecnica-1/recomendacoes-tecnicas-1/politica_presercacao_digital.pdf. Acesso em: 28 nov. 2023.

IBICT. Instituto Brasileiro de Informação em Ciência e Tecnologia. **Manifesto brasileiro de apoio ao acesso livre à informação**. 2005. Disponível em: <https://livroaberto.ibict.br/Manifesto.pdf>. Acesso em: 26 out. 2023.

MARCONDES, C. H.; SAYÃO, L. F. À guisa de introdução: repositórios institucionais e livre acesso. *In*: SAYÃO, L.; TOUTAIN, L. B.; ROSA, F. G.; MARCONDES, C. H. (org.). **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 9-22. Disponível em: <https://repositorio.ufba.br/handle/ufba/473>. Acesso em: 09 nov. 2023.

MARCONDES, C. H.; SAYÃO, L. F. *Software* livres para repositórios institucionais: alguns subsídios para a seleção. *In*: SAYÃO, L.; TOUTAIN, L. B.; ROSA, F. G.; MARCONDES, C. H.



(org.). **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 23-54. Disponível em: <https://repositorio.ufba.br/handle/ufba/473>. Acesso em: 09 nov. 2023.

MÁRDERO ARELLANO, M. Á. **Critérios para a preservação digital da informação científica**. 2008. 354 f. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, 2008. Disponível em: <https://repositorio.unb.br/handle/10482/1518>. Acesso em: 27 nov. 2023.

MÁRDERO ARELLANO, M. Á. Preservação de documentos digitais. **Ciência da Informação**, [s. l.], v. 33, n. 2, p. 15-27, maio/ago. 2004. DOI: <https://doi.org/10.1590/S0100-19652004000200002>. Disponível em: <https://revista.ibict.br/ciinf/article/view/1043>. Acesso em: 08 nov. 2023.

MARTINS, D. L.; SILVA, M. F.; SIQUEIRA, J. Comparação entre sistemas para criação de acervos digitais: análise dos *softwares* livres DSpace, EPrints, Fedora, Greenstone e Islandora a partir de novas dimensões analíticas. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 9, n. 1, p. 52–71, mar./ago. 2018. DOI: 10.11606/issn.2178-2075.v9i1p52-71. Disponível em: <https://www.revistas.usp.br/incid/article/view/134333>. Acesso em: 7 set. 2024.

OPENDOAR. Browse by country and region. **Sherpa Services**, c2024. Disponível em: https://v2.sherpa.ac.uk/view/repository_by_country/Brazil.html. Acesso em: 7 set. 2024.

SHINTAKU, M.; MEIRELLES, R. F. **Manual do DSpace**: administração de repositórios. Salvador: EDUFBA, 2010. 84 p. Disponível em: <https://repositorio.ufba.br/handle/ri/769>. Acesso em: 25 nov. 2023.

SHINTAKU, M.; VECHIATO, F. L. Histórico do uso do DSpace no Brasil com foco na tecnologia. **Revista Informação na Sociedade Contemporânea**, [s. l.], v. 2, n. 1, p. 1-16, jan./ jun. 2018. DOI: <https://doi.org/10.21680/2447-0198.2018v2n0ID13097>. Disponível em: <https://periodicos.ufrn.br/informacao/article/view/13097>. Acesso em: 22 nov. 2023.

SHINTAKU, M.; VIDOTTI, S. A. B. G. Bibliotecas e repositórios no processo de publicação digital. **BIBLOS – Revista do Instituto de Ciências Humanas e da Informação**, [s. l.], v. 30, n. 1, p. 61-80, 2016. Disponível em: <https://periodicos.furg.br/biblos/article/view/5762>. Acesso em: 20 nov. 2023.

VECHIATO, F. L. *et al* (org.). **Repositórios digitais**: teoria e prática. Curitiba: EDUTFPR, 2017. 272 p. Disponível em: <http://repositorio.utfpr.edu.br/jspui/handle/1/2495>. Acesso em: 10 nov. 2023.

VERMA, L.; KUMAR, N. Comparative analysis of open source digital library softwares: a case study. **DESIDOC Journal of Library & Information Technology**, v. 38, n. 5, p. 361-368, 5 set. 2018. DOI: <https://doi.org/10.14429/djlit.38.5.12425>. Disponível em: <https://publications.drdo.gov.in/ojs/index.php/djlit/article/view/12425>. Acesso em: 7 set. 2024.