

CAMINHOS PERCORRIDOS PARA A DIGITALIZAÇÃO DAS DISSERTAÇÕES DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO DA UNIVERSIDADE FEDERAL DA PARAÍBA UTILIZANDO A PLATAFORMA DSPACE COMO REPOSITÓRIO DIGITAL

PATHS FOLLOWED FOR THE DIGITIZATION OF THE DISSERTATIONS COMING FROM THE GRADUATE PROGRAM IN INFORMATION SCIENCE AT THE UNIVERSIDADE FEDERAL DA PARAÍBA USING THE DSPACE PLATFORM AS A DIGITAL REPOSITORY

Dulce Elizabeth Lima de Sousa*

Patrícia Silva**

Alba Lígia de Almeida Silva***

Guilherme Ataíde Dias****

RESUMO

As tecnologias digitais da informação e comunicação (TDIC) modificam o comportamento do indivíduo e alteram significativamente as relações sociais. Dentro dessa cadeia dinâmica da informação, percebemos que muitos trabalhos científicos que nasceram unicamente em meio impresso, estavam limitados geograficamente e o seu conteúdo não possuía visibilidade adequada. Esta pesquisa tem como objetivo principal relatar os caminhos percorridos para a digitalização das dissertações oriundas do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (PPGCI/UFPB), e disponibilizadas através da plataforma DSpace como repositório digital. O universo de pesquisa foi constituído por 174 dissertações. Os documentos foram digitalizados e processados através do Optical Character Recognition (OCR). Pode-se afirmar que 26% das dissertações existentes encontram-se disponíveis no repositório DSpace do PPGCI/UFPB. As estatísticas geradas mostram que essas dissertações estão sendo amplamente acessadas pela comunidade.

Palavras-chave: Digitalização. Repositório Digital. DSpace.

ABSTRACT

Digital information and communication technologies modify individual's behavior and change social relations significantly. Within this dynamic chain of information, we realize that

Biblionline, João Pessoa, v. 6, n. 2, p. 44-55, 2010.

many scientific studies deployed only in printed matter, were limited geographically and its contents did not have adequate visibility. This research aims to report the paths followed for the digitization of the dissertations coming from the Graduate Program in Information Science at the Universidade Federal da Paraíba (PPGCI/UFPB) and made available through the DSpace platform as a digital repository. The research universe was composed by 174 dissertations. The documents were scanned and processed using Optical Character Recognition (OCR). It can be stated that 26% of the existing dissertations are available at the PPGCI/UFPB repository. Statistics show that these dissertations are being widely accessed by the community.

Keywords: Digitizing. Digital Repository. DSpace.

1 INTRODUÇÃO

A explosão informacional gerada no Pós-Guerra desencadeou um sequência de ações inovadoras para época, a informação se tornou bússola norteadora das ações políticas, gerando novos parâmetros que ultrapassaram as fronteiras geográficas dos países e alcançaram uma amplitude mundial. Para Barreto (1994, p.3) “a informação, quando adequadamente assimilada, produz

conhecimento, modifica o estoque mental de informações do indivíduo e traz benefícios ao seu desenvolvimento e ao desenvolvimento da sociedade em que ele vive”.

Essa explosão informacional veio acompanhada da revolução tecnológica, por isso é possível observar que os processos de produção, acesso, recuperação e disseminação da informação, sofrem uma forte influência das Tecnologias Digitais da Informação e Comunicação (TDIC), e isso nos impulsiona a pensarmos no desenvolvimento de ferramentas, que auxiliem na operacionalização dessa gama de informações cada vez mais crescentes.

Os aparatos tecnológicos, tais como: base de dados, *sites*, *blogs*, repositórios digitais, entre outros, nos proporcionam uma estrutura mais acessível e variada aos conteúdos produzidos por diversos seguimentos, portanto, atualmente é comum encontramos instituições acadêmicas que depositam suas produções científicas em repositórios digitais. Grande parte da produção acadêmica está disponível publicamente, e por estarem amplamente acessíveis proporcionam maior visibilidade e impacto da produção científica da instituição. “A distribuição globalizada da informação científica promoveu novas parcerias e incentivou o desenvolvimento da ciência nas regiões mais pobres do planeta. (ARELLANO; VIANA; SHINTAKU, 2005, p. 4).

Existe mais de uma definição para repositórios, de acordo com essas mesmas, podemos dizer que repositórios digitais são sistemas de informação que servem para armazenar, preservar, organizar e disseminar amplamente os resultados de pesquisas. Este estudo é um relato de experiência, resultante de um projeto financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), sendo nossa abordagem principal, mostrar os caminhos percorridos na digitalização das dissertações oriundas do Programa de Pós-Graduação em Ciência da

Informação da Universidade Federal da Paraíba (PPGCI/UFPB).

2 NOVAS FERRAMENTAS E NOVOS CONTEXTOS

Se analisarmos o contexto histórico da criação da tipografia, perceberemos que o desenvolvimento da imprensa alterou significativamente o modo de pensar e agir da sociedade. Burke (2002) em seu artigo intitulado: “Problemas causados por Gutenberg: a explosão da informação nos primórdios da Europa moderna” aprofunda mais essa discussão.

A sociedade do XVI presenciou a quantidade de informações que estavam sendo geradas, devido à invenção da impressora manual. Atualmente estamos vivenciando algo semelhante, diversos seguimentos sociais estão usando recursos oferecidos pela Internet, no desenvolvimento de suas atividades.

Então, podemos dizer que todo processo de adaptação e mudança ocasionado por novas descobertas tecnológicas, não é algo incomum para a humanidade.

No início do século XX, surgiram outros dispositivos relacionados, de alguma maneira, a comunicação: o cinema falado, a televisão, o gravador, a caneta esferográfica, a fotocopiadora e os primeiros computadores (DIAS, 1999, p. 271).

Dessa forma, a transição de contextos é algo que esta ligada diretamente ao homem. O qual possui a capacidade de produzir, aceitar, rejeitar e ser indiferente com suas próprias criações. E por meio desses novos contextos ele acaba se modificando.

Setores como a indústria, comércio e as relações interpessoais, estão sendo reestruturadas. Houve uma mudança significativa na forma da sociedade pensar,

trabalhar e se relacionar. Considerando que as tecnologias digitais da informação e comunicação possuem uma vertente intelectual, podemos dizer que “não é a primeira vez que a aparição de novas tecnologias intelectuais é acompanhada por uma modificação das normas do saber” (LEVY, 1993, p. 19).

O uso cada vez mais frequente de ferramentas digitais, nos remete a pensar como ocorrerá o processo de busca da informação. Um dos pioneiros da Ciência da Informação, Mooers (1960) foi responsável por denominar essa busca de informação de Recuperação da Informação (*information retrieval*). Para ele “A recuperação de informação engloba os aspectos intelectuais da descrição da informação e de sua especificação para a busca, bem como qualquer sistema, técnica ou máquina que são utilizadas para realizar a operação”.

Com a chamada Web 2.0 as opções de serviços são incontáveis, a interação e colaboração exercidas entre os usuários contribuem significativamente para o aumento da circulação de informações e assim, ocupando nos últimos anos, uma posição fundamental dentro do Movimento de Acesso Livre. O contexto dinâmico da informação vai se moldando ao longo do tempo, encontrando caminhos permeáveis juntamente com as TDIC para sua disseminação. Nesse contexto, a informação científica vivencia um momento ímpar, onde o processo tradicional de divulgação das pesquisas está sendo substituído por novos meios, como os periódicos eletrônicos e repositórios digitais sendo cada vez mais utilizados.

O tratamento da informação, no sentido do termo, visa precisamente a criação de <<memórias>>, passíveis de serem utilizadas sempre que houver necessidade de recuperar dados (informação) nelas armazenados. Isto implica procedimentos de controle da informação, de **criação de**

meios de acesso às referidas memórias e de desenvolvimento de dispositivos susceptíveis de acionar os meios de acesso, com vista à recuperação da informação armazenada. (SILVA et. al, 2002, p. 27, grifo nosso).

Pensar em meios que proporcionem, dependendo da natureza do documento, a disseminação, o acesso e a preservação da informação, principalmente quando se trabalha com recursos tecnológicos, é de responsabilidade do profissional da informação, seja ele, arquivista, bibliotecário, museólogo ou outro.

3 REPOSITÓRIOS E ACESSO LIVRE: INOVAÇÃO NA COMUNICAÇÃO CIENTÍFICA

Para Arellano e Viana (2006) “Um repositório digital é uma forma de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar material por longos períodos de tempo e prover o acesso apropriado”. Complementando a idéia dos autores, podemos acrescentar que para garantir o acesso ao material armazenado nesses repositórios em qualquer época, é imprescindível o uso de ferramentas adequadas que não apenas resguardem o documento, mas, que possam garantir sua manutenção e acesso, protegendo a mídia e garantindo fidedignidade do conteúdo.

Ao pensar em repositório institucional universitário, Lynch (2003) conceitua como sendo um conjunto de serviços que uma universidade oferece para os membros de sua comunidade gerenciar e disseminar os materiais digitais criados pela universidade e membros de sua comunidade.

Os repositórios nasceram de uma iniciativa do Laboratório Nacional de Los Alamos nos EUA, voltado para as áreas de Ciência da Computação, Física e Matemática e outras. Criou-se um repositório digital denominado arXiv. Esta iniciativa foi vista com uma alternativa para o sistema de comunicação

científica que estava sofrendo uma crise, onde os preços dos periódicos estavam elevados para aquisição.

Os trabalhos depositados, não passavam pelo sistema *peer review*, como acontecia no modelo tradicional. Essa característica proporcionou uma maior flexibilidade, o leitor poderia tecer críticas e o autor poderia melhorar os seus trabalhos. Segundo Vidotti et al (2004) em paralelo ao arXiv existiu uma outra iniciativa através de Stevan Harned denominada Archiving Initiative, que tinha como objetivo se sobrepôr as barreiras comerciais das editoras universitárias.

Em julho de 1999, foi organizada uma Convenção em Santa Fé, no Novo México (EUA), resultando em um conjunto de operações que visavam a interoperabilidade entre os repositórios, nascendo o modelo do *Open Archives Initiative* (OAI). Foi proposto um modelo com provedores de dados (*data providers*) e provedores de serviços (*service providers*), os quais são necessários para o processo de busca. Essa convenção estabeleceu o protocolo o *Open Archives Initiative Protocol for Metadata Harvesting*.

Trata-se de um protocolo de comunicação que possibilita a coleta de metadados a partir de determinado provedor de dados. O provedor de serviços, para realizar a coleta de metadados, deve utilizar um programa chamado Harvester (mecanismo de colheita), o qual implementa esse protocolo, o OAI-PMH (KURAMOTO, 2006, p. 93).

Basicamente para existir um diálogo ente o provedor de dados, que expõe os metadados, e o provedor de serviço, responsável por recuperar a informação, é necessário que o *Harvester*, o mecanismo de colheita, faça essa intermediação tendo geralmente como padrão de metadados o *Dublin Core* (<http://dublincore.org/>).

Essa nova tecnologia e as dificuldades de publicação nos meios tradicionais, foi essencial para o escopo do Movimento do Acesso Livre. A Declaração de Berlim de 2003 que foi aprovada pela comunidade científica define acesso livre como uma fonte universal do conhecimento humano e do patrimônio cultural.

Na premissa da interoperabilidade diversos produtos de *software* foram criados, servindo como plataforma para o auto-arquivamento. O Arno, o EPrints, o MyCoRe, o Fedora e o mais utilizado no Brasil, o DSpace.

Estas plataformas são responsáveis pela implantação de repositórios institucionais e temáticos. Segundo Weitzel (2006, p.59) os termos “repositórios institucionais” ou “temáticos” são adotados para caracterizar os repositórios digitais que reúnem respectivamente a produção científica de uma instituição e de uma área.

4 DSPACE: ACESSO E PRESERVAÇÃO

O *software* que dá suporte ao repositório do PPGCI/UFPB é o DSpace. Este programa foi desenvolvido em conjunto pelo *Massachusetts Institute of Technology* (MIT) e o *Hewlett-Packard* (HP), o qual é uma aplicação com o código fonte aberto dentro dos termos da licença de distribuição BSD. O DSpace também preserva e torna fácil o acesso aberto a todos os tipos de conteúdo digital, incluindo textos, imagens estáticas, imagens em movimento, mpegs e conjuntos de dados (DSpace, 2010).

A plataforma DSpace tem como natureza preservar os objetos digitais. A recuperação dos trabalhos acontece inicialmente pelas “Comunidades e coleções”. Nas comunidades podemos encontrar sub-comunidades e realizar a recuperação de documentos através das opções de data, autor, título e assunto.

A primeira versão em português deste *software* foi da Universidade do Minho, atendendo assim uma iniciativa do Governo

Português em criar a universidade eletrônica. Segundo Baptista e outros (2004) “a UMinho foi a primeira universidade da comunidade mundial de língua portuguesa (de aproximadamente 200 milhões de pessoas) e uma das primeiras na Europa a traduzir e implementar o DSpace”.

Segundo Arellano e Viana (2006) o Instituto Brasileiro de Informação Ciência e Tecnologia (IBICT) depois da experiência do RepositorioUM(<http://repositorium.sdum.uminho.pt/>)– Portugal, resolveu experimentar o DSpace, posteriormente após um período o repositório foi adotado pelas Instituições de Ensino Superior (IES). No site do IBICT existe uma seção voltada para o DSpace, nela existe uma série de recomendações quanto ao uso dessa plataforma, diretrizes gerais para sua adoção, podendo ser encontrado alguns *links* para outros repositórios no Brasil e no exterior e há disponibilização da versão 1.4 para download.

Outra iniciativa portuguesa é um repositório que agrega outros através *links* de repositórios cadastrados, se tornando uma espécie de indexador geral. Através do mesmo é possível acessar qualquer repositório português, desde que esteja cadastrado. Esse projeto é uma iniciativa da UMIC – Agência para Sociedade do Conhecimento que foi concretizada pela FCCN – Fundação para a Computação Científica Nacional.

Autores como Bainbridge, Huang e outros (2007) acreditam que o DSpace pode preservar, organizar e armazenar o conteúdo disponibilizado nele. Qualquer organização pode ter essa plataforma sem nenhum custo, além disso, existem cerca de 80 desenvolvedores espalhados pelo mundo, pois cada aprimoramento que é feito nessa plataforma o conhecimento tem que ser transmitido. Segundo dados disponibilizados no site, existem cerca de 700 instituições que adotaram o DSpace.

Nessa plataforma se distingue dois níveis de preservação digital: Preservação em nível de *bit* e preservação funcional. Na preservação em nível de *bit* o arquivo permanece o mesmo, imutável com o passar do tempo, nenhum único *bit* é alterado, na preservação funcional o arquivo é alterado com o passar do tempo, mas a informação continua sendo acessível tal qual na sua primeira disponibilização enquanto os formatos digitais evoluem.

Alguns formatos de arquivos tais como imagens em *Tagged Image File Format* (TIFF) ou documentos em *eXtensible Markup Language* (XML) podem ser funcionalmente preservados diretamente através de migração de formato direta. Os demais formatos por razões diversas, ou por serem proprietários são bem mais difíceis de preservar a sua funcionalidade.

5 PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO DA UFPB

O PPGCI tem memória que remonta ao Curso de Mestrado em Biblioteconomia (CMB) datado de 1977 a 1996 e, posteriormente, ao Mestrado em Ciência da Informação (CMCI) que permaneceu ativo de 1997 a 2001 e posteriormente reativado em 2007. É oportuno ressaltar que o Curso de Graduação em Biblioteconomia começou a funcionar desde 1970, e o Mestrado somente foi oficializado em 1977, com a oferta de vinte vagas e o ingresso de candidatos de vários Estados brasileiro. Os argumentos para a implantação de um Curso desse nível na Região Nordeste são apresentados visto que o Brasil dispunha de apenas quatro cursos de pós-graduação na área, surgindo, então, a urgência de se implantar um “laboratório” para o desenvolvimento de pesquisas, visando ao aprofundamento de conhecimento técnico-científicos e a inserção de pesquisadores (as), discentes e docentes em comunidades socialmente vulneráveis.

A partir de então, o CMB contou com a colaboração de doutores de vários departamentos da UFPB, que ministraram disciplinas, participaram como membros do Colegiado do Curso, orientaram dissertações e participaram de bancas de qualificação e de defesas.

Em sua perspectiva de criação de um novo curso de Pós-Graduação em CI, o corpo docente do Departamento de Ciência da Informação (antigo Departamento de Biblioteconomia) entendeu que um programa de pós-graduação deve conter produção científica, Mestrado acadêmico, Especialização, Mestrado e Doutorado, com liberdade de fluxo entre os dois níveis. Assim, para incremento dos objetivos de implantação de uma nova pós-graduação em CI, foram credenciados mais três doutores, conforme procedimentos determinados pela legislação pertinente junto ao referido Programa.

Assim, o PPGCI iniciou suas atividades em março de 2007, com uma turma de vinte e um discentes, o que resultou em vinte e uma dissertações de Mestrado defendidas nas linhas de pesquisa “Memória, organização, acesso e uso da informação” e “Ética, gestão e políticas de informação”.

Nesse universo, o objeto de estudo da Ciência da Informação (CI) – a informação – preocupa-se com os princípios e as práticas de produção, organização, disseminação, representação, recuperação, acesso e uso da informação e seus fluxos, sua transmissão por meio de diversas formas e canais de comunicação. Barreto (2001) entende que as pesquisas em CI constituem um processo orientado à expansão das fronteiras do conhecimento, representando uma investigação ordenada, original e coerente com a orientação teórico-metodológica, tendo a intenção de mostrar evidências de um método racional de ação e experimentação, com o propósito de descobrir novas informações ou desenvolver novos processos

de transformação para viabilizar produtos ou serviços de informação.

É oportuno destacar que a produção de conhecimento ou produção científica é um tema conhecido para estudiosos da área de CI. Pensando nos trabalhos produzidos pelo PPGCI/UFPB, como por exemplo, as dissertações que foram defendidas e depositadas após a reabertura do Programa no ano de 2007, e não esquecendo os anteriores a esta data, achamos de fundamental importância torná-las acessíveis em meio digital e através de seu repositório eletrônico (DSpace) que tem como objetivo o armazenamento, a preservação, a disseminação e o acesso à produção intelectual dos docentes, discentes e dos pesquisadores vinculados ao PPGCI.

No repositório do PPGCI/UFPB são arquivados artigos, teses, dissertações e resultados dos grupos de pesquisas. A maior parte dessas pesquisas já nasceu em meio digital, contudo o acesso aos trabalhos que não foram produzidos em ambiente computacional é imprescindível, visto que estes foram precursores de muitos conceitos aplicados atualmente. Assim sendo,

os resultados alcançados por determinado pesquisador são frequentemente retomados por outros cientistas, teóricos ou aplicados, que dão continuidade ao estudo, fazendo avançar a ciência ou produzindo tecnologias ou produtos neles baseados (PELLO; CENDÓN; KREMER, 2000. p. 25).

Essa preocupação se atenua quando destacamos a importância do processo de digitalização como sendo a chave que possibilita o acesso (em meio digital) à informação mesmo tendo sido registrada em formato impresso.

O processo de digitalização vem sendo amplamente discutido na área da Arquivologia, geralmente essa discussão é voltada ao campo da gestão de documentos,

porém algumas premissas podem ser utilizadas para permitir o acesso a documentos/textos que não nasceram em formato digital e que possuem um acesso é limitado. Assim estas dissertações não possuem um caráter de documentos arquivísticos, mas não deixam de ter um valor científico e cultural relevantes no meio acadêmico.

A utilização destes avanços tecnológicos, representados pelo uso de computadores isolados ou integrados em redes, com transmissão de dados em alta velocidade, sistemas gerenciadores de banco de dados, armazenamento digital, etc., **possibilitaram o acesso instantâneo à informação e a documentação** (BARBIERI; INNARELLI; MARTINS. 2002, p. 54, grifo nosso).

Dessa forma, o Conselho Nacional de Arquivos - CONARQ é um órgão colegiado, vinculado ao Arquivo Nacional da Casa Civil da Presidência da República, este órgão tem a responsabilidade de definir a política nacional de arquivos públicos e privados, com orientações normativas para a gestão documental.

Em abril de 2010 o CONARQ lançou Recomendações para Digitalização de Documentos Arquivísticos Permanentes, nesse documento ele apresenta quatro pontos necessários para digitalização:

- Contribuir para o amplo acesso e disseminação dos documentos arquivísticos por meio da Tecnologia da Informação e Comunicação;
- Permitir o intercâmbio de acervos documentais e de seus instrumentos de pesquisa por meio de redes informatizadas;
- Promover a difusão e reprodução dos acervos arquivísticos não digitais, em formatos e apresentações diferenciados do formato original;

- Incrementar a preservação e segurança dos documentos arquivísticos originais que estão em outros suportes não digitais, por restringir seu manuseio.

O acervo do PPGCI/UFPB não se caracteriza como documento arquivístico, visto que é formado por dissertações de mestrado, contudo, estes trabalhos compõem a memória do Programa. É importante salientar que os assuntos abordados nessas dissertações contribuem para o resgate das representações históricas da região e do país. Nesse sentido, o processo de digitalização das dissertações vem atender a ampliação do acesso, permitindo a preservação dos originais e contribuindo para a divulgação científica.

Uma publicação nada mais é que o ato de tornar públicos a metodologia e os resultados de uma pesquisa. Se a pesquisa realizada enfocou um problema relevante para a sociedade, é de se esperar que essa mesma sociedade, por meio de seus grupos e representantes, venha a se interessar por tais resultados e pelas possibilidades de sua utilização para a resolução de seus problemas (PEREIRA JR, 2007, p. 308).

Concordando com o autor, ao disponibilizar as dissertações no DSpace do PPGCI/UFPB estaremos possibilitando o acesso na íntegra do conhecimento científico produzido pelos pesquisadores envolvidos com as temáticas da CI.

6 PROCESSO DE DIGITALIZAÇÃO E A ANÁLISE DOS RESULTADOS

O universo de pesquisa se constituiu de 174 dissertações, as quais 131 delas não nasceram em formato digital. O primeiro passo foi analisar o acervo para planejar as ações e o processo de digitalização, com olhar sempre atento, a fim de conservar, a fidedignidade dessas dissertações.

Na definição do processo metodológico, partimos do pré-suposto que, seria de fundamental importância destacar que o acesso aos registros digitalizados é o reconhecimento fundamental de três princípios: a) a fidedignidade dos registros digitalizados; b) a autenticidade dos registros digitalizados; e finalmente c) o acesso ao longo do tempo bem como, o desenvolvimento de ferramentas e estratégias capazes de lidar com o problema da obsolescência tecnológica. Os repositórios neste momento são “as chaves, para uma maior democratização do conhecimento e uma concretização da inclusão e do envolvimento de pesquisadores e da sociedade” (DOLLAR, 1994, p. 13).

A metodologia utilizada foi pautada nos parâmetros estabelecidos por Silva (2005) de acordo com o Manual de Digitalização de Acervos. Segundo este manual, alguns itens devem ser levados em consideração na digitalização de documentos, são eles: os estratégicos, organizacionais, estruturais, operacionais e técnicos. Como o objeto de digitalização já era definido por natureza material permanente, então apenas alguns itens foram considerados.

A digitalização ocorreu de acordo com o estado de conservação das dissertações. Os campos para a inserção dos metadados, necessários no processo de busca, já estavam definidos no repositório. Os dispositivos de *hardware* escolhidos foram um *scanner* de mesa simples e uma câmera Nikon DSLR. Para que os textos fossem navegáveis, optamos por um produto de *software* com algumas especificações, portanto, escolhemos trabalhar com uma das técnicas do reconhecimento de padrões de imagens, o *Optical Character Recognition* (OCR).

O OCR reconhece os caracteres impressos ou datilografados. A leitura desse tipo de programa é realizada com base nos

caracteres, ele não reconhece as palavras ou frases.

A precisão de um OCR é normalmente expressa em percentual (%) de caracteres que o programa acerta. Valores comuns são de ordem superior a 90%, no entanto o nível de precisão é por caractere, e não por palavras ou sequências de números (BALDAM et al, 2002. p. 119).

É importante dizer que o reconhecimento dos caracteres não é total, visto que foi realizada uma revisão minuciosa por toda dissertação digitalizada, fontes e parágrafos. Outros componentes como: gráficos, tabelas, questionários e imagens não são reconhecidos pelo OCR.

O objetivo de um sistema OCR é reconhecer os caracteres de uma imagem dada como entrada e extrair o texto para um arquivo texto editável. [...] Dessa forma, não é feita a separação de região textual de regiões não textuais como fotos, figuras geométricas e outras (MENA-CHALCO et. al, 2006, p. 2).

Depois de todo o processo de digitalização e formatação em editores de texto, inserimos nas partes escritas as imagens, tabelas e gráficos que não eram reconhecidas pelo software. Assim, buscamos manter a versão digital mais próxima possível da versão impressa.

Em seguida iniciamos a fase de conversão. Neste momento em que a dissertação está pronta, ela é convertida para o formato *Portable Document Format* (PDF), pois, este formato é o mais indicado até o momento para preservação digital, o qual independe de aplicativo, *hardware* e sistema operacional. Posteriormente enviamos a dissertação para a avaliação do autor com o pedido de autorização da disponibilização da mesma. Só depois deste procedimento a dissertação é arquivada no repositório do PPGCI/UFPB. A figura 1, mostra a interface do Repositório DSpace do PPGCI/UFPB.

Figura 1 – Interface do Repositório DSpace do PPGCI/UFPB



Fonte: <http://dci2.ccsa.ufpb.br>

No acervo do Programa de Pós-Graduação é possível encontrar dois tipos de dissertação: impressas e datilografadas. Existe uma diversidade na qualidade de impressão dessas dissertações. As datilografadas requerem maior dedicação para digitalização, porque o programa que já não é preciso em sua totalidade vai encontrar mais barreiras no reconhecimento desse tipo de texto.

Com essa peculiaridade há um aumento no tempo de revisão. Conseqüentemente o processo se torna moroso. Sabemos que a transição do impresso para o digital é uma ação que por natureza descaracteriza o próprio documento, então, todos os cuidados necessários foram tomados para que essa ação fosse minimizada.

É importante ressaltar que no processo de digitalização, não nos prendemos as regras da Associação Brasileira de Normas Técnicas (ABNT), na medida em que, entendemos que o contexto para as regras criadas na época era diferente. Os eventuais erros gramaticais também não foram corrigidos, pois assim como as regras da ABNT os padrões gramaticais também eram diferentes. Além disso, estaríamos modificando os textos originais, dessa forma, estaríamos contrariando o objetivo proposto, manter a fidedignidade.

De acordo com o gráfico 1 é possível observar o percentual das dissertações já acessíveis no repositório DSpace do PPGCI/UFPB (<http://dci2.ccsa.ufpb.br>), bem como ter uma noção do quanto ainda há pra fazer.

Gráfico 1 - Dissertações do PPGCI/UFPB



Fonte: Dados da pesquisa, 2010.

7 CONSIDERAÇÕES FINAIS

O reconhecimento do valor que a informação possui é primordial para a definição de novas diretrizes sociais, esse ação está ultrapassando barreiras geográficas. A denominação “Sociedade da Informação” retrata bem a situação que vem se desenvolvendo desde a criação da imprensa tipográfica, que começou a rascunhar o que denominaríamos posteriormente de *boom* informacional.

O desenvolvimento de sistemas automatizados está concomitantemente possibilitando uma recuperação da informação mais abrangente e modificando estruturas sociais do saber. A comunicação científica modificou os seus nortes e o modelo *Open Archives* forneceu subsídios, como por exemplo, os repositórios digitais.

A utilização do DSpace desperta investigações futuras no que diz respeito a aceitação tecnológica, bem como a realização de um levantamento memorialístico para a identificação das linhas de pesquisa mais trabalhadas.

Até o momento o que podemos dizer é que dentro do repositório há um campo que é possível saber qual a dissertação que está sendo mais acessada, e para a nossa

satisfação a dissertação com maior número de acesso, foi a primeira dissertação digitalizada, sendo sua defesa feito no Programa de Pós-Graduação em 1979.

É importante destacar que um repositório possui uma visibilidade muito extensa, os trabalhos depositados lá vão desde Teses e Dissertações até as produções dos grupos de pesquisa. Espera-se que esse estudo venha contribuir para a disseminação de informações no ambiente digital e que ela seja estimuladora para que outras Universidades repensem suas políticas de produção, acesso, recuperação e disseminação da informação de documentos que não nasceram no meio digital.

Nossa pesquisa foi veementemente norteadada com os princípios da recuperação da informação e os resultados parciais estão disponíveis no *site* do Departamento de Ciência da Informação da UFPB (<http://dci2.ccsa.ufpb.br>) onde podem ser facilmente acessados.

REFERÊNCIAS

- BARRETO, A. A. A questão da Informação. **São Paulo em Perspectiva**, São Paulo, v. 8, n.4, 1994
- ARELLANO, M. A. M.; VIANA, C. L. M. Repositórios Institucionais baseados em Dspace e Eprints e sua

viabilidade nas instituições acadêmico-científicas. In: Seminário Nacional de Bibliotecas Universitárias, 14., 2006, Salvador. **Anais...** Salvador: Universidade Federal da Bahia, 2006. p. 1-15.

BAINBRIDGE, D; DON; K. J; HUANG, C; TANSLEY, R; WITTEN, I. H. **A Bridge between Greenstone and Dspace.** Disponível em: <http://dspace.ibict.br/dmdocuments/StoneD_A_Bridge_between_Greenstone_and_DSpace.pdf>. Acesso em 22 set. 2010.

BALDAM, R; CAVALCANTI. M; VALLE, R. **GED:** gerenciamento eletrônico de documentos. São Paulo: Érica, 2002.

BAPTISTA, A. A. et al. **RepositóriUM – Implantação do Dspace em português:** Lições para o futuro e linhas de investigação. Disponível em: <http://biblioteca.universia.net/html_bura/ficha/pararams/id/49410187.html> Acesso em: 20 ago. 2010.

BARBIERI, C. C. D.; INNARELLI, H. C.; MARTINS, N. do R.. Gerenciamento eletrônico de documentos: criação de um banco de informações e imagens no Arquivo Permanente da UNICAMP. In: CONGRESSO INTERNACIONAL DE ARQUIVOS BIBLIOTECAS CENTROS DE DOCUMENTAÇÃO E MUSEUS, 1., 2002. **Textos...** São Paulo: Imprensa Oficial, 2002. p.53-66.

BURKE, P. Problemas causados por Gutenberg: a explosão da informação nos primórdios da Europa moderna. **Estudos Avançados**, São Paulo, v. 16, n. 44, p. . 2002.

DECLARAÇÃO DE BERLIM. Disponível em: <<https://repositorium.sdum.uminho.pt/about/DeclaraaoBerlim.htm>>. Acesso em: 03 ago. 2010

DIAS, C. A. Hipertexto: evolução histórica e efeitos sociais. **Ciência da Informação**, Brasília, v. 28, n. 3, p.269-277, set./dez. 1999.

DOLLAR, C. **Tecnologias da Informação digitalizada e pesquisa acadêmica nas ciências:** o papel crucial da arquivologia. **Estudos Históricos**, Rio de Janeiro, v. 7, n. 13, p. 65-79, 1994.

DUBLIN Core. Disponível em: <<http://dublincore.org/>>. Acesso em: 23 set. 2010

DSPACE. Disponível em: <<http://www.dspace.org/>>. Acesso em: 12 maio. 2010.

INSTITUTO BRASILEIRO DE INFORMAÇÃO CIÊNCIA E TECNOLOGIA. Disponível em: <<http://www.ibict.br/>>. Acesso em: 23 set 2010.

KURAMOTO, H. **Informação científica:** proposta de um novo modelo para o Brasil. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 91-102, maio/ago. 2006.

LÉVY, P. **As tecnologias da inteligência:** o futuro do pensamento na era da informática. Rio de Janeiro: Ed. 34, 1993.

LYNCH, C. A. Institutional Repositories: essential infrastructure for scholarship in the digital age. **ARL Bimonthly Report**, n. 226, Feb. 2003. Disponível em:

<<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>. Acesso em: 15 jul. 2010.

MÁRDERO ARELLANO, M. A; VIANA, C. L; SHINTAKU, M. **Repositórios institucionais em ciência e tecnologia:** uma experiência de customização do dspace. Disponível em: <http://dspace.ibict.br/index.php?option=com_content&task=view&id=32&Itemid=95>. Acesso em: 23 set 2010.

MENA-CHALCO, J. et al. Universidade de São Paulo - Instituto de Matemática e Estatística MAC5749 - **Análise e Reconhecimento de Formas.** Disponível em:<<http://www.vision.ime.usp.br/~fabriciolopes/mac5749-ep2.pdf>>. Acessado dia 22 set. 2010.

MOOERS, C. N. Mooers' law or, Why Some Retrieval Systems Are Used and Others Are Not. **American Documentation**, v.11, n.3, July, 1960. Disponível em:<http://findarticles.com/p/articles/mi_qa3633/is_199610/ai_n8749122/>. Acesso em: 17 jul. 2010.

PELLO, B. S; CENDÓN, B. V; KREMER (org.). **Fontes de informação para pesquisadores e profissionais.** Belo Horizonte, UFMG, 2000.

PEREIRA JR, Alfredo. A publicação científica na atualidade. **J Vasc Bras**, Porto Alegre, v. 6, n. 4, 2007. Editorial

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO.. Disponível em: <http://dci.ccsa.ufpb.br/ppgci/>. Acesso em: Acesso em: 23 set. 2010.

SILVA, A. M. et al. **Arquivística**: teoria e prática de uma Ciência da Informação. Porto: Afrontamento, 2002. 254p.

SILVA, R. R. G. **Manual de digitalização de acervo**: textos, mapas e imagens fixas. Salvador: edufba, 2005.

SOUSA, D. E. L; DIAS, G. A. **Repositórios eletrônicos acadêmicos**: disponibilizando a produção científica do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba. Encontro Nacional de Estudantes de Biblioteconomia, Documentação, Gestão, e Ciência da Informação, 23., 2010, Disponível em: <<http://dci.ccsa.ufpb.br/enebd/index.php/enebd/article/viewFile/84/163>>. Acesso em: 10 mar. 2011.

VIDOTTI, S. A. B. G; OLIVEIRA, G. P. de; SARMENTO E SOUZA, M. F. **A iniciativa dos arquivos abertos como alternativa a publicações científicas**. Disponível em: <<http://cutter.unicamp.br/document/?code=8301>>. Acesso em: 23 set. 2010.

WEITZEL, S. R. O papel dos repositórios institucionais e temáticos na estrutura da produção científica. **Em Questão**, Porto Alegre, v. 12, n. 1, p. 51-71, jan./jun. 2006.

Dados sobre Autoria

*Aluna de Graduação em Arquivologia. Bolsista de Iniciação Científica CNPq/UFPB. E-mail: dulcelizabeth@gmail.com

**Mestre em Ciência da Informação (UFPB); Especialista em Gestão Estratégica de Sistemas de Informação (UFRN). Professora do Departamento de Ciência da Informação da UFPB. Co-orientadora. E-mail: silva.131313@gmail.com

***Mestre em Ciência da Informação; Especialista em Gestão de Unidades de Informação ambos pela UFPB. Professora do Departamento de Ciência da Informação da UFPB. Professora Colaboradora. E-mail: aligiasilva@gmail.com

****Doutor em Ciências da Comunicação (Ciência da Informação) pela Universidade de São Paulo (USP). Mestre em Administração pela Central Connecticut State University. Professor Adjunto III do Departamento de Ciência da Informação da UFPB. Orientador. E-mail: guilherme@dci.ccsa.ufpb.br

Artigo enviado em outubro de 2010 e aceito em fevereiro de 2011.