

# Arquivos abertos: subprojeto da Biblioteca Digital Brasileira

## **Ricardo Triska**

Coordenador-geral de administração do Ibict.

## **Lígia Café**

Coordenação de prospecção e desenvolvimento de tecnologias da informação. Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

E-mail: [cafe@ibict.br](mailto:cafe@ibict.br)

---

## **Resumo**

*Apresentação e descrição do subprojeto de implantação dos Arquivos Abertos coordenado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Após breve histórico sobre esta mecânica da comunicação científica, relatamos as experiências realizadas junto a comunidade científica da genética e os futuros empreendimentos pretendidos pelo IBICT.*

## **Palavras-chave**

*Arquivos abertos; Textos eletrônicos; Comunicação científica.*

## **Open Archives: the Brazilian Digital Library subproject**

## **Abstract**

*Describes the Open Archives subproject and its implementation, under the coordination of the Brazilian Institute for Information on Science and Technology (IBICT). Besides it includes a brief history on this scientific communication channel. The Genetic community was selected to evaluate the experiment and the results were given. Future achievements intended by IBICT are presented.*

## **Keywords**

*Open archives; Scientific communication.*

## **INTRODUÇÃO**

A implantação dos Arquivos Abertos, coordenada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), compõe o conjunto de ações que integram a Biblioteca Digital Brasileira.

Trata-se de uma proposta inovadora que possibilita não somente a publicação de textos eletrônicos diretamente na rede, mas também o envio de comentários e sugestões aos documentos eletrônicos disponíveis no repositório virtual. Esta iniciativa do Ibict tem a seguinte motivação:

- ampliar a visibilidade nacional e internacional da produção intelectual brasileira em C&T;
- melhorar o fluxo de comunicação científica e tecnológica;
- incrementar o ciclo de geração de novos conhecimentos.

Neste artigo, descreve-se um breve histórico desta nova mecânica da comunicação científica e relatam-se as experiências atuais e futuras realizadas pelo Ibict.

## **BREVE HISTÓRICO**

O surgimento da Internet e das novas tecnologias de comunicação disponíveis na rede têm provocado profundas mudanças no mundo da publicação científica. Neste contexto, surgem os arquivos abertos (*open archives*).

O conceito de arquivos abertos tem seu marco consolidado em outubro de 1999 durante a Convenção de Santa Fé ([http://www.openarchives.org/sfc/sfc\\_entry.htm](http://www.openarchives.org/sfc/sfc_entry.htm)) realizada no Novo México, sob os auspícios do Council on Library and Information Resources (CLIR), da Digital Library Federation (DLF), da Scholarly Publishing & Academic Resources Coalition (SPARC), da Association of Research Libraries (ARL) e do Los Alamos National Laboratory (LANL).

Durante esta Convenção foram definidos os princípios básicos de uma nova filosofia para a publicação científica que começava a se delinear. Destes princípios, citamos três que consideramos principais: o auto-arquivamento, a revisão pela comunidade e a interoperabilidade.

O auto-arquivamento refere-se ao direito de o próprio autor enviar o seu texto para publicação sem intermédio de terceiros. Trata-se de um conceito inovador cujos objetivos são tornar o texto disponível o mais rápido possível e favorecer o acesso democrático e gratuito das publicações eletrônicas, enfraquecendo o monopólio das grandes editoras científicas que até recentemente detinham em seu poder os direitos de publicação. Segundo Ginsparg (2000), os arquivos de textos eletrônicos são inteiramente dirigidos pelos cientistas e são flexíveis o bastante tanto para coexistir com os sistemas de publicação tradicional como para ajudar estes sistemas a se envolverem com algo mais próximo das necessidades dos pesquisadores.

No que diz respeito à revisão entre os pares, a filosofia dos arquivos abertos tem como propósito a transparência das críticas e sugestões feitas aos textos eletrônicos depositados no repositório. Desta forma, o ambiente possibilita que toda a comunidade tenha acesso ao processo de revisão e de versões do texto geradas com base nas sugestões. Assim, quebra-se o conceito de revisão sigilosa feita por um comitê científico, caracterizando mais uma vez o aspecto democrático desta iniciativa. Nesse sentido, Ginsparg (2000) menciona que a disseminação rápida oferecida pelos arquivos abertos não é incompatível com a tradicional revisão entre os pares. Para ele, a longo prazo, os arquivos abertos oferecem uma estrutura mais funcional de organização da literatura do que a organização atual existente no processo de revisão entre os pares.

A interoperabilidade é a palavra-chave para a criação dos arquivos abertos. Segundo Sompel & Lagoze (*apud* Senna 2000.74), na perspectiva deste grande movimento de publicação eletrônica, a interoperabilidade envolve uma série de aspectos, tais como conjunto mínimo de metadados, tipo de arquitetura subjacente do sistema, abertura para a criação de serviços de bibliotecas digitais de terceiros, integração com o mecanismo de comunicação já existente no meio científico, possibilidade de uso em contextos interdisciplinares e contribuição para criação de um sistema de medida de uso e de citação.

A Convenção de Santa Fé estabeleceu os seguintes aspectos como mecanismos para se atingir a interoperabilidade:

- definição de um conjunto mínimo de metadados;
- concordância no uso de uma sintaxe comum XML para representar e transportar tanto o OAMS (*Open Archives Metadata Set*) como os conjuntos de metadados específicos de cada repositório;

- definição de um protocolo comum – *Open Archives Dienst Subset* – para possibilitar a extração do OAMS e dos metadados específicos dos repositórios participantes.

No que se refere à criação de repositórios de documentos eletrônicos, a Convenção recomenda que sejam observados os seguintes aspectos:

- mecanismo de submissão;
- armazenamento a longo prazo;
- política de gerenciamento da submissão e da preservação dos documentos inseridos no repositório;
- interface aberta que permita que terceiros possam coletar dados do repositório.

Em agosto de 1991, antes da Convenção de Santa Fé, aparece o primeiro repositório de documentos eletrônicos baseado na filosofia dos arquivos abertos. Este repositório denominado ArXiv (<http://arxiv.org/>) foi criado pelo físico Paul Ginsparg do laboratório de Los Alamos no Novo México. Ele cobre as áreas de física e disciplinas correlatas, matemática, ciências não-lineares, linguística computacional e neurociências. Seus usuários podem recuperar ou submeter os artigos tanto por uma *interface on-line* na *world wide web* como via correio eletrônico. No ArXiv, é ainda permitida a atualização de artigos pelos autores, além de serem oferecidos serviços de alerta e mecanismos de busca adaptados a diferentes perfis de usuários.

Sompel & Lagoze (2000) citam outras iniciativas existentes no mundo:

CogPrints <http://cogprints.soton.ac.uk/> – localizado na Universidade de Southampton no Reino Unido. Segue o modelo do ArXiv e usa o *software e.print*. Abrange as áreas de psicologia, linguística, neurociências, ciência da computação, filosofia e biologia.

NCSTRL <http://www.ncstrl.org/> (Network Computer Science Technical Reference Library) – coleção internacional sobre relatórios de pesquisa em ciência da computação. É uma rede construída segundo um modelo descentralizado. Os documentos são armazenados em repositórios distribuídos e disponibilizados por serviços também distribuídos, via protocolo Dienst.

NDLTD <http://www.ndltd.org/> (Network Digital Library Thesis and Dissertations) – biblioteca eletrônica de teses e dissertações autorizadas por estudantes das instituições membros da rede. As pesquisas desenvolvidas para a criação desta rede envolvem assuntos como a criação de um fluxo (*workflow*) para submissão de ETD (Electronic Theses and Dissertations), o desenvolvimento de XML e de DTD (Document Type Definition) para ETDs e o suporte para a biblioteca digital de teses e dissertações eletrônicas.

RePEc <http://www.ndltd.org/> (Research Papers in Economics) atua na área de economia e foi desenvolvida seguindo um modelo distribuído. Segue o protocolo Guilford, que garante a interoperabilidade entre os arquivos da RePEc.

### **SERVIÇO DE BUSCA A REPOSITÓRIO DE ARQUIVOS ABERTOS**

Com objetivo de recuperar os documentos eletrônicos existentes nos mais diversos arquivos abertos criados no mundo, surgiu o primeiro serviço de busca denominado ARC (Cross Archive Searching Service - <http://arc.cs.odu.edu>). Este serviço utiliza arquitetura em Java e tem plataforma independente, podendo trabalhar com qualquer servidor web. Atualmente, a busca abrange os seguintes repositórios:

- A Celebration of Women Writers
- ArXiv
- California International and Area Studies Digital Repository
- Cogprints
- Dermatology Digital Repository
- Humbolt University of Berlin, Germany, Document Server
- Library of Congress Open Archive Initiative Repository 1
- LTRS
- M.I.T. Theses
- National Advisory Committee for Aeronautics
- NCSTRL

- NDLTD
- NSDL Open Archives Server at Cornell University
- OCLC Online Computer Library Center Theses and Dissertation Repository
- Open Video
- Perseus Digital Library
- PhsyNet, Oldenburg, Germany, Document Server
- Resource Discovery Network
- The Oxford Text Archive
- Tobacco Control Digital Repository
- University of Illinois at Urbana – Champaign, Digital Library Initiative

### **PROJETO DE ARQUIVOS ABERTOS COORDENADO PELO Ibict**

#### **Objetivos**

##### *Objetivo geral:*

- oferecer à comunidade científica um modelo que viabilize a publicação de textos na Internet, dinamizando o processo de comunicação entre os pares.

##### *Objetivos específicos:*

- subsidiar a formação de redes especializadas e dinâmicas de produção científica;
- divulgar as vantagens do uso do ambiente proposto pelos arquivos abertos

#### **Software escolhido**

Após a realização de estudos sobre o movimento dos arquivos abertos no mundo e do estabelecimento de contatos com especialistas envolvidos nesta iniciativa, foram testados alguns *softwares* para implementação desta alternativa de comunicação científica. Dentre os programas analisados, o *software e.print* (<http://www.eprints.org/>) caracterizou-se como o mais adequado.

O *e.print* é um *software* desenvolvido pela Universidade de Southampton no Reino Unido e é possível verificar o seu funcionamento nos seguintes repositórios:

### **Arquivos abertos: subprojeto da Biblioteca Digital Brasileira**

a) CogPrints Cognitive Sciences Eprint Archive

b) California Digital Library

Este *software* foi construído de forma bastante flexível, o que exige o mínimo de esforço de configuração e adaptação. A versão genérica do *e. print* é interoperacional com todos os OAI- *Compliant Open Archives*, o que significa que, independentemente de onde o artigo esteja arquivado, ele pode ser recuperado pelo protocolo OAI via um “arquivo virtual” global. Para tanto, deve ser usado um *Open Archives Service Providers* como o serviço ARC citado anteriormente.

Os serviços de recuperação de documentos oferecidos pelo *e. print* são:

- serviço de alerta (*subscription*);
- serviço de busca por área de conhecimento;
- serviço de busca (*search*) simples e avançada;
- serviço de busca pelo número do documento no repositório.

Com o *software* escolhido, entramos na etapa de adaptação às especificidades da publicação da produção científica nacional.

#### **Experimento com a Sociedade Brasileira de Genética**

A primeira experiência de arquivos abertos coordenada pelo Ibict foi feita com a Sociedade Brasileira de Genética (SBG). Escolhemos a área da genética devido à sua reconhecida projeção científica, em termos de produção e de participação em fóruns internacionais. Outro fator importante foi a decisão da SBG de investir em uma nova alternativa para visibilidade da produção intelectual de seus filiados. Segundo Jordan (1999), a área da biologia que trata dos genomas tem utilizado a web como uma forma de compartilhar novos conceitos, de visualizar grandes conjuntos de dados e de construir assim uma informação coletiva.

Além do objetivo de criar um ambiente web de publicação eletrônica baseado nos princípios dos arquivos abertos, a aliança com a SBG caracteriza-se por:

- desenvolver experimento que possibilite a geração de dados para análise do modelo;
- fornecer subsídios para a melhoria dos arquivos abertos criados pelo Ibict.

A estratégia de implantação foi dividida em três etapas. A primeira foi dedicada à adaptação do *software e. print* para o contexto brasileiro. Para esta etapa, foram efetuadas as seguintes atividades:

- definição dos tipos de documentos contemplados no programa;
- definição do conjunto de campos para cada documento, inclusive atribuindo obrigatoriedade quando fosse o caso;
- tradução e adaptação do manual;
- tradução e adaptação das telas de ajuda;
- elaboração da *home page*;
- teste do *software*;
- elaboração de instrumento de apoio à avaliação externa realizada junto à comunidade científica específica\*;
- implantação de um mecanismo de envio e resposta automática ao questionário de avaliação.

Na segunda etapa do projeto, foi realizada uma divulgação ampla para todos os membros da SBG por meio de uma carta que convidava os associados a visitar e utilizar o ambiente web disponível no *site* [www.sbg.ibict.br](http://www.sbg.ibict.br).

A terceira etapa tratou do lançamento oficial do serviço. Durante o 47º Congresso Brasileiro de Genética, foi proferida, pelo diretor do Ibict, palestra sobre o tema e foi montado um estande de demonstração. A presença dos especialistas na palestra e no estande nos leva a constatar o crescente interesse pela utilização deste serviço.

#### **Projetos futuros**

O Ibict tem por objetivo ampliar o uso do produto resultante deste experimento, oferecendo uma alternativa para implantação em ambientes externos ao instituto e para administração de acervos locais. Estes acervos serão integrados por processo de partilhamento de metadados reconhecidos por um procedimento comum de busca.

#### **Resultados esperados**

---

\* Esta etapa consistiu na elaboração de um questionário reduzido para ser enviado automaticamente a cada vez que o autor depositasse um documento nos arquivos abertos. O questionário tem como objetivo coletar sugestões de melhoria ao *software*.

O Ibict vem, por meio da implantação dos arquivos abertos, apresentar-se como parceiro no acompanhamento da criação de serviços desta natureza junto à comunidade científica brasileira, observando as questões concernentes à interoperabilidade, citadas anteriormente.

Dessa forma, o Ibict estará :

- promovendo a criação de serviços de informação inovadores que reflitam demandas da sociedade e que possam ser integrados à Biblioteca Digital Brasileira;
- promovendo e encorajando o estabelecimento de soluções conhecidas como auto-arquivo, por meio do desenvolvimento de mecanismos que assegurem alto grau de interoperabilidade entre essas soluções.

Com os arquivos abertos, o Ibict está colocando em prática as suas ações de:

- criação de uma metodologia baseada na iniciativa mundial dos *open archives*;
- operação de servidores de publicações eletrônicas;
- repasse dessa metodologia às instituições detentoras de arquivos digitais importantes para as áreas de C&T;
- oferta à comunidade científica e tecnológica de um ambiente para submissão, busca e recuperação de textos completos em meio digital que utilize a metodologia de arquivos abertos;
- oferta de um serviço de informação e comunicação inovador, baseado na moderna filosofia de auto-arquivamento pelo autor/pesquisador sem interferência de terceiros;
- ampliação da *performance* dos modelos de comunicação científica e tecnológica existentes atualmente no Brasil.

À medida que forem sendo criados e consolidados novos repositórios de arquivos abertos, o Ibict deverá conduzir uma análise das iniciativas já em operação ou em processos de implantação, tendo por referência a sua integração à Biblioteca Digital Brasileira.

#### REFERÊNCIAS BIBLIOGRÁFICAS

arXiv.org e-Print archive. Disponível em: < <http://arxiv.org/> > .

Cogprints. Disponível em: < <http://cogprints.soton.ac.uk/> > .

Cross Archive Searching Service (ARC). Disponível em: < <http://arc.cs.odu.edu/> > .

E. print.org. Disponível em: < <http://www.eprints.org/> > .

GINSPARG, Paul. Creating a global knowledge network. In: FREEDOM OF INFORMATION CONFERENCE: the impact os open access on biomedical research, 2000, New York Academy of Medicine, BioMed Central. New York, 2000. Disponível em: < <http://www.biomedcentral.com/info/ginsparg-ed.asp> > .

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. (Brasília, DF). *Projeto Biblioteca Digital Brasileira: descrição dos subprojetos*. Brasília, 2001.

JORDAN, Shardon M. Preprint servers : status, challenges, and opportunities of the new digital publishing paradigm. In: FORUM'99, 1999. [S. l. : s. n.], 1999. Disponível em: < <http://www.osti.gov/inforum99/papers/jordan.html> > .

LIU, Xiaoming, et al. ARC: an OAI service provider for cross-archive searching. Disponível: < <http://arc.cs.odu/help/jcdl.pdf> > .

OPEN archives initiative. Disponível em: < <http://www.openarchives.org/> > .

SENA, Nathália Kneipp. Open archives: caminho alternativo para a comunicação científica. *Ciência da Informação*, Brasília, v. 29, n. 3, set./dez. 2000.

THE SANTA FÉ CONVENTION FOR THE OPEN ARCHIVES INITIATIVE. Disponível em: < [http://www.openarchives.org/sfc/sfc\\_entry.htm](http://www.openarchives.org/sfc/sfc_entry.htm) > .

SOMPEL, Herbert van de; LAGOZE, Carl. The Santa Fé Convention of the Open Archives Initiative. *D-Lib Magazine*, v. 6, n. 2, Feb. 2000. Disponível em: < <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html> > .