



Ministério da Saúde
FIOCRUZ
Fundação Oswaldo Cruz



ICICT
Instituto de Comunicação e Informação
Científica e Tecnológica em Saúde

CURSO DE ESPECIALIZAÇÃO EM INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

OTIMIZANDO A RECUPERAÇÃO DE INFORMAÇÃO EM REPOSITÓRIOS INSTITUCIONAIS

por

Clea Mara Barradas dos Reis

Fundação Oswaldo Cruz/ Instituto de Comunicação e Informação Científica e Tecnológica em Saúde

Projeto apresentado ao Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, da Fundação Oswaldo Cruz, como requisito parcial para obtenção do título de Especialista em Informação Científica e Tecnológica em Saúde.

Orientadora: Maria de Fátima Moreira Martins
Mestre em Ciência da Informação

Rio de Janeiro, novembro de 2008

SUMÁRIO

1	INTRODUÇÃO	3
2	JUSTIFICATIVA	4
3	REFERENCIAL TEÓRICO	6
3.1	EVOLUÇÃO DA PUBLICAÇÃO CIENTÍFICA ELETRÔNICA	6
3.1.1	Cenário Mundial da Comunicação Científica	8
3.2	ACESSO ABERTO	9
3.2.1	Repositórios Institucionais	10
3.2.1.1	Recuperação da Informação em RI	11
3.3	ESTUDO DE CASO: ARCA	13
4	OBJETIVOS	16
4.1	OBJETIVO GERAL	16
4.2	OBJETIVOS ESPECÍFICOS	16
5	METODOLOGIA	17
6	RESULTADOS ESPERADOS	18
7	REFERÊNCIAS CONSULTADAS	19
8	CRONOGRAMA	21
9	ORÇAMENTO	22

1 INTRODUÇÃO

Os repositórios institucionais (RI) se constituem como um novo modelo de comunicação científica, capaz de agilizar, tanto o processo de comunicação científica, quanto alimentar a memória informacional da instituição. Nesta perspectiva, o ARCA se constitui no repositório do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT) da Fiocruz, que está sendo implantando com a finalidade reunir, preservar, dar acesso, disseminar e aumentar a visibilidade dos trabalhos acadêmicos gerados pela comunidade científica.

Inserir-se, também, no crescente movimento de constituição de repositórios e arquivos de acesso livre, no quadro da Iniciativa dos Arquivos Abertos. Projetado para apoiar a divulgação dos resultados de pesquisas científicas desenvolvidas, cria mecanismos para legitimar e estimular a publicação da produção científica na instituição, está sendo construído na plataforma *DSpace*, que é um *software* livre, intencionalmente implementado para servir de repositório.

Para descrever os itens existentes no repositório, o *DSpace* utiliza o padrão metadados *Dublin Core*, para coleta automática. Dessa forma, permite o auto-arquivamento, isto é, que o próprio autor submeta o trabalho, desde o preenchimento dos metadados, até o depósito do arquivo. Por outro lado, a falta de padronização dos metadados pode trazer como consequência, a ambigüidade em determinados campos, e provocar um grande desafio para a organização e recuperação da informação.

Neste sentido, este projeto tem como objetivo o uso de um vocabulário controlado para o preenchimento do metadado descritivo que represente o conteúdo temático dos documentos (assunto), possibilitando, assim, a maior eficácia na recuperação de documentos relevantes, e, também, dentro do domínio bibliográfico, resolver a questão por meio de construção de arquivos de autoridade e de nomes de instituições.

2 JUSTIFICATIVA

O movimento de criação de repositórios institucionais desenvolveu-se, de forma significativa, a partir de 2002, após o forte impulso das principais Declarações internacionais sobre o Acesso Livre (Declarações de Budapeste, de Bethesda e de Berlim).

E ainda, os repositórios institucionais diferenciam-se de outros repositórios, porque neles devem estar contidas todas as produções técnico-científicas da instituição. Os objetos devem ser inseridos pelo próprio autor (auto-arquivamento), acesso livre ao conteúdo, e requer a elaboração, por parte da instituição, de uma política para o repositório.

Neste contexto, o questionamento em torno da busca de mais eficiência e eficácia para a recuperação das informações armazenadas nos repositórios institucionais é ainda uma questão em aberto nas pesquisas sobre repositórios institucionais.

Percebe-se que a facilidade para criação e divulgação de repositórios institucionais levou o surgimento de diversos na rede, mas após seu surgimento sentiu-se a necessidade de ampliação de suas funcionalidades. Isto acontece porque os repositórios institucionais ampliam as características de conteúdo, a forma de apresentação, a interatividade e a capacidade de oferecer serviços.

No entanto, mesmo com o crescimento de repositórios institucionais na rede, a sua construção é um campo relativamente novo, e, por isso, pouca literatura científica é encontrada sobre o assunto. Faltam estudos mais específicos sobre o desenvolvimento de ferramentas para representação da informação, e que possibilitem a recuperação dos documentos neles contidos.

Segundo Vidotti e Sanches (2003), para a recuperação das informações na Internet se faz necessária uma forma de representação descritiva e temática adequada aos conteúdos, observando-se, antes da implementação desse sistema, a forma como os usuários potenciais desse ambiente realizam essas buscas, de maneira direta ou avançada, e seus diferentes tipos de necessidades informacionais, identificando, discriminando e visando suprir essas possíveis variantes.

Por outro lado, Segundo Tomaél et al. (2000, p.3), ainda não foram desenvolvidos meios para se organizar os documentos disponíveis na Internet, e o acúmulo de

informações sem relevância aponta para a necessidade de filtros que permitam a recuperação de informações de qualidade.

Possibilitar o acesso a toda e qualquer palavra contida no documento ou nos metadados, fazem com que certos repositórios institucionais não atinjam seus objetivos, além de provocar falsas recuperações, e levando os usuários a ter dificuldades para encontrar o que procuram, e localizar as informações relevantes.

Estas considerações levam ao seguinte problema de pesquisa: Que ferramentas podem ser utilizadas para aumentar a eficácia na recuperação de documentos relevantes depositados nos repositórios institucionais?

Nessa ótica, esse trabalho se justifica, ao contribuir para a utilização do vocabulário controlado DeCS para indexar (o metadado assunto) e recuperar efetivamente, os documentos depositados no ARCA (Memória Digital Institucional Informação Científica e Tecnológica em Saúde), objetivando melhorar o mecanismo de busca atual.

3. REFERENCIAL TEÓRICO

A revisão de literatura deste trabalho é composta de dois momentos distintos:

- evolução da publicação científica eletrônica;
- acesso aberto.

3.1 EVOLUÇÃO DA PUBLICAÇÃO CIENTÍFICA ELETRÔNICA

Já em 1946, Bernal formulou o conceito de comunicação científica como um “amplo processo de geração e transferência de informação científica”, constituindo, assim, a troca de informações gerada pelos pesquisadores. Nessa mesma linha de pensamento, Garvey (1979) afirma que a comunicação científica está relacionada ao conjunto de atividades de produção, disseminação e uso da informação.

Para Back (1972), é através dos canais de informação que ocorre a atualização dos profissionais e a disseminação de informações importantes ao desenvolvimento da ciência.

Neste contexto, Meadows (1999, p. 7) destaca que as diferenças entre os canais de informação utilizados pelo sistema de comunicação científica relacionam-se com as funções de disseminação e suporte: a formal “encontra-se disponível por longos períodos de tempo para um público amplo”, e está associada a um suporte físico, enquanto que, a informal, é “em geral, efêmera, sendo posta à disposição apenas de um público limitado”, e associada à comunicação oral.

Em linhas gerais, o termo comunicação científica refere-se à troca de informação e idéias entre cientistas. A comunicação científica é essencial para a ciência, permitindo não só a disseminação dos resultados das pesquisas, quanto a proteção da propriedade intelectual, aceitação dos resultados pelos pares e consolidação do conhecimento.

Boorstin (1989), narra que o periódico científico aparece em 1660, após a fundação das sociedades científicas na França e Inglaterra, e surgiu como uma evolução do papel das cartas pessoais trocadas entre cientistas e das atas de memórias de reuniões científicas.

A partir de então, o modelo de periódico científico começou ser difundido por toda a Europa, promovendo o crescente número de publicações e a disseminação das

mesmas, passando a desempenhar importante papel no processo de comunicação da ciência. Estes periódicos, em sua maioria, estavam diretamente vinculados a instituições científicas, e eram utilizados para disseminar as realizações de seus membros, no campo da ciência (STUMPF, 1996).

Em 1945, Bush reconhece o problema da explosão informacional, em função do surgimento dos periódicos especializados, que cobriam não só as grandes áreas do conhecimento, mas, também, a subárea desses campos.

Um outro aspecto a ser analisado, além do imenso crescimento do número de periódicos científicos, ocorridos no século XX, é o surgimento das novas tecnologias de informação, que permitiram a otimização da produção, acesso e disseminação da informação. Assim, a partir de 2000, as comunidades científicas foram as primeiras a usufruir das facilidades das tecnologias de informação e comunicação, o que, indiscutivelmente, contribuiu para mudanças significativas no cenário global da comunicação, como demonstrou Briquet de Lemos (2005).

É assim que, o acesso ao texto integral, via Internet, contribuiu para ampliar o alcance e a visibilidade das revistas, e o número crescente de artigos acadêmicos.

Até 2005, algo em torno de 11 mil periódicos científicos eletrônicos eram editados no mundo. Mas os números demonstram que o mercado editorial dedicado à comunicação científica tornou-se mais lucrativo e monopolista: em 2005, por exemplo, o grupo Reed Elsevier, uma das oito empresas que publicavam 70% da literatura médico-científica mundial, editava, sozinho, 2.600 títulos nessa área (BRIQUET DE LEMOS, 2005).

Por outro lado, estudos recentes de Tenopir (2005) mostram que as comunidades científicas passaram a ler mais artigos com o advento da Internet, onde dois terços do que lêem os pesquisadores provêm de fontes eletrônicas.

Nessa abordagem, a comunicação científica, seja formal ou informal, está passando por mudanças tão significativas, que cada vez mais se tornam indistintas suas barreiras (RUSSELL, 2001).

As publicações científicas eletrônicas se constituem, atualmente, um dos temas de maior repercussão dentro da comunicação científica, e, por conseguinte, da

própria ciência moderna, ao atuarem dentro de um campo-chave para o funcionamento do atual modelo de produção do conhecimento científico.

3.1.1. Cenário Mundial da Comunicação Científica

Como vimos, desde o seu surgimento, o periódico científico passou por inúmeras modificações, mas, foi a partir do século XX, que a publicação de revistas científicas cresceu acentuadamente. Weisman (1972) nos dá uma panorâmica detalhada desse crescimento quando ressalta que, no início do século XIX, existiam cerca de 100 periódicos; em 1830, este número aumentou para 500, e, em 1850, registravam-se 1.000 títulos. No ano de 1900, o número atingiu 10.000 títulos, e segundo uma avaliação feita pela “Library of Congress”, dos Estados Unidos, por volta de 1960, eram publicados no mundo, cerca de 30.000 títulos de periódicos técnicos científicos.

Um outro aspecto a ser analisado, além do imenso crescimento do número de periódicos científicos, ocorridos no século XX, é o aparecimento das tecnologias de informação e comunicação, que permitiram a otimização da produção, acesso e disseminação da informação, mudando o tradicional conceito de informação bibliográfica, baseada em documentos impressos. O acesso via Internet, a novos recursos informacionais, tem se tornado uma realidade cada vez mais presente no dia-a-dia dos pesquisadores e da sociedade em geral.

Atualmente, assiste-se a uma crise no modelo clássico de produção, distribuição e consumo dos periódicos científicos, crise esta que envolve a inexorável transição do suporte impresso para o eletrônico (WEITZEL, 2005).

Outro destaque é o alto custo das assinaturas dos principais periódicos científicos. No panorama mundial, e especialmente no brasileiro, a falta de recursos financeiros para aquisição e manutenção dos periódicos soma-se com o aumento do custo de mais de 01 (hum) mil por cento nas últimas duas décadas (SOARES, 2004). O lucro elevado das editoras tem, como maior exemplo, o editor Elsevier que detém atualmente 70% do mercado das revistas científicas e com isso, impõe duras condições de negociação para as assinaturas, com preços elevados para os consumidores, como as bibliotecas de universidades e instituições, fato conhecido como oligopolização (ORTELLADO e MACHADO, 2006).

É oportuno finalizar este item, destacando a importância dos arquivos abertos (*open archives*) para a comunicação científica. Assim, o movimento internacional em defesa do acesso livre (*open access*) à informação se tornou uma ferramenta de relevância, na tarefa de disseminação do conhecimento, promovendo maior agilidade na comunicação científica, além de democratizar o acesso à informação científica.

3.2 ACESSO ABERTO

Na última década, a questão do acesso livre (*open access*) tem sido ampla e calorosamente discutida, na literatura especializada de diversos campos do conhecimento, em função do modelo de publicação da pesquisa que domina o cenário mundial.

No final de 2001, ocorreu uma reunião em Budapeste, promovida pelo *Open Society Institute* (OSI), da qual resultou um dos mais relevantes documentos e iniciativas do movimento de Acesso Livre, conhecido como *Budapest Open Access Initiative* (BOIA)¹. Segundo Targino (2007), a declaração ressalta:

as iniciativas dos arquivos abertos e do acesso aberto como recurso para incrementar o avanço científico e tecnológico mundial, mediante a disseminação irrestrita de artigos científicos dos distintos campos do saber, resistindo às pressões da indústria editorial e dos órgãos públicos que mantêm acervos valiosos para pesquisas.

E outros movimentos se seguiram, como a Declaração de Berlim (2003), recomendação do Parlamento Britânico (2005), e, no Brasil, recentemente, o Acesso Livre tornou-se Projeto de Lei na Câmara, que se for aprovado, obrigará as instituições de ensino superior, de caráter público, assim como as unidades de pesquisa, a construir os seus repositórios institucionais nos quais deverão ser depositadas as produções técnico-científicas, resultado das pesquisas realizadas pelos seus pesquisadores e professores, financiadas com recursos públicos, para acesso livre na rede mundial de computadores – Internet.

Essas condições poderão propiciar um futuro menos dependente das revistas científicas comerciais

É essencial, portanto, considerar o que observa Alberts (2002), ao afirmar que a informação técnico-científica é, fundamentalmente, um bem público, que deve

¹ BRASIL. IBICT. Iniciativa de Acesso Livre à Informação Científica. Disponível em: <<http://www.ibict.br/openaccess>> Acesso em: 10 ago. 2007.

estar livremente disponível para o benefício de todos: “nesse contexto, o papel de modelos alternativos de comunicação científica, tais como repositórios institucionais, ao quebrar monopólios de editores e aumentar a *awareness* a respeito da produção intelectual”.

Resultados de uma pesquisa realizada por Lawrence (2001), por exemplo, estão de acordo com o que preconizam Brody e Harnard, pois mostraram um crescimento de 3.365, em média, nas citações a artigos disponíveis *online*, em relação a artigos publicados *offline*, na mesma fonte. Como observa Lawrence (2001), “para maximizar o impacto, minimizar a redundância e acelerar o progresso científico, autores e editores deveriam visar a tornar a pesquisa fácil de ser acessada”.

Um exemplo neste mesmo ano, é a SciELO (Scientific Electronic Library Online), que possui uma estrutura de editoração inspirada no modelo tradicional de periódico, onde os artigos avaliados pelos pares são disponibilizados para acesso livre junto com outros periódicos de uma ou mais áreas.

3.2.1 Repositórios Institucionais

Segundo Crow (2002), os repositórios institucionais são arquivos digitais que reúnem, preservam, divulgam e permitem o acesso à produção científica de comunidades universitárias. Podem ser criados de forma individualizada ou por grupos de instituições que trabalhem numa base cooperante. Nessa mesma linha de pensamento, Lynch (2003) ressalta que consiste em “um conjunto de serviços que a instituição oferece aos seus membros para o gerenciamento e disseminação de materiais digitais criados na instituição”.

Por sua vez, Crow (2002), referindo-se à memória, salienta que o repositório institucional representa, essencialmente, o compromisso da instituição de cuidar do material digital ali produzido, incluindo a preservação em longo prazo, a organização, o acesso e a disseminação.

Inserem-se no movimento conhecido por *Open Access Initiative* (OAI), que visa promover o Acesso Livre à literatura científica e acadêmica.

O repositório institucional permite aos pesquisadores, professores, alunos e pessoal da instituição, assim como toda a sociedade, facilitar o acesso à

informação científica produzida na instituição; maximizar a visibilidade da produção científica, acadêmica e institucional, potencializando o impacto das instituições na sociedade; preservar as produções científicas, acadêmicas e institucionais, garantindo a memória histórica da instituição; servir de plataforma de armazenamento e intercâmbio comum da informação produzida na instituição; mostrar a própria produção, avaliando o rendimento da instituição e facilitando a comunicação científica; dotar os pesquisadores, professores e alunos desta ferramenta de auto-arquivo, facilitando a comunicação com todos; e, ainda, promover a retroalimentação das pesquisas.

Por sua vez, caracterizam pela diversidade de tipologia de conteúdos e formatos que podem ser armazenados, como: artigos científicos, livros eletrônicos, capítulos de livros, *preprints*, *postprints*, relatórios técnicos, textos para discussão, teses, dissertações, trabalhos apresentados em eventos, material de ensino, arquivos, patentes e multimídia.

Atualmente, repositórios institucionais representam uma importante ferramenta no sistema político científico e acadêmico das instituições. Segundo Café et. al. (2003, p.6), “quanto mais instituições utilizarem repositórios institucionais, mais se beneficiarão com informações atualizadas. Este é mais um ponto positivo que o pesquisador deve considerar ao lutar pela implantação deste novo paradigma em sua instituição”.

3.2.1.1 Recuperação da Informação no RI

No que se refere à recuperação da informação e indexação em base de dados, Lancaster (1993, p. 229), afirma que o serviço de informação cria representações dos documentos, armazena-os em base de dados e proporciona algum tipo de dispositivo, que possibilita que se façam buscas nessas representações. Dessa forma, os conceitos são extraídos do documento, através de um processo de análise, e então traduzidos para os termos da linguagem documentária.

De acordo com Campos e Gomes (2004), existe uma insatisfação dos usuários com relação à recuperação da informação disponibilizada em rede e na Internet, devido à falta de ferramentas de acesso adequadas para o controle terminológico.

Segundo Shintaku e Brascher (2006), os metadados, além de fornecer informações sobre o objeto digital, servem de ponto de recuperação a esses

objetos. Para documentos textuais, porém, na maioria dos casos, há uma coincidência entre o conteúdo do documento e o metadado. Título, autores e resumo, por exemplo, são os mesmos no documento e nos metadados.

Para repositórios baseados no *DSpace*, que fazem uso da indexação de texto completo, essa repetição não ajuda a criar formas de recuperação do documento mais otimizados. O metadado palavra-chave, possui certa independência entre o metadado e o documento. Esse metadado cumpre a função de organizar, classificar e hierarquizar os documentos no repositório, e facilitar a recuperação por assunto, agrupando os documentos que possuem relação de assunto.

Problemas simples como o de grafia, que provocam a recuperação de documentos diferentes, revelam a necessidade de padronização e de um estudo mais profundo sobre vocabulários controlados, que auxiliem na qualidade dos metadados para a organização e recuperação da informação.

Segundo Borbinha (2000), o metadado palavra-chave, entre outros pertencentes ao esquema de metadados *Dublin Core*, é um caso que se encaixa na categoria dos metadados para os quais é recomendado o uso de vocabulário controlado.

Para evitar grandes variações e padronizar entradas, a utilização de vocabulário controlado para metadados permite restringir as possibilidades de preenchimento do assunto.

A falta de padronização dos metadados pode trazer como consequência a ambigüidade em determinados campos e provocar um grande desafio para a organização e recuperação da informação.

A literatura dá indicações de que um vocabulário estruturado melhora a experiência do usuário. Segundo Muramaki (2005), o uso de vocabulários controlados melhora a qualidade e disponibilidade como suporte para buscas em várias bases de dados.

Neste sentido, os vocabulários controlados são instrumentos de controle terminológico utilizados em sistemas de informação para traduzir a linguagem dos documentos, dos indexadores e dos pesquisadores numa linguagem controlada, usada na indexação e recuperação de informações. Além disso, possuem uma série de finalidades, como garantir o entendimento da terminologia utilizada pelo usuário, produtores da informação e indexadores, procurando reduzir o ruído no

processo de comunicação; controlar a terminologia, através do estabelecimento de relações semânticas, procurando, assim, alcançar a melhor representação dos conceitos; e, por fim, possibilitar a recuperação exaustiva da navegação da estrutura do tesouro, identificando termos mais precisos dentro da cadeia hierárquica.

3.3 ESTUDO DE CASO: ARCA

Ao longo dos últimos vinte e um anos, o atual Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT), construiu uma sólida trajetória no campo da Informação e Comunicação, que o credenciou como unidade de pesquisa e ensino na Fundação Oswaldo Cruz (Fiocruz). Somando a essa trajetória intelectual, desenhou-se também o caminho de uma prática que expressa, de forma singular, um projeto político de disseminação e democratização da informação científica e tecnológica em saúde, e da inclusão social.

Uma relevante análise mostra que a produção científica e acadêmica do ICICT encontra-se dispersa na Fiocruz, e em diversas publicações (periódicos científicos, anais de congressos, etc.).

Buscando reunir, registrar e preservar essa memória, foi criado o ARCA, com o objetivo de armazenar, gerir, preservar e divulgar, a nível internacional, via Internet, contribuindo, deste modo, para a proteção e valorização da memória intelectual do ICICT e para o aumento da visibilidade, acessibilidade e impacto do seu valor científico e cultural.

O ARCA vem sendo construído na plataforma *DSpace*, que é um *software* livre desenvolvido pelo Massachusetts Institute of Technology (MIT) e pelos Laboratórios Hewlett-Packard Company, intencionalmente implementado para servir de repositório.

Para descrever os itens existentes no repositório, o *DSpace* utiliza o padrão metadados *Dublin Core Resource Description* (DC), que é um formato menos estruturado e mais flexível, que adota a sintaxe do *Resource Description Framework* (RDF), que propicia um conjunto de 15 elementos padrão utilizados para descrever uma variedade de recursos, permitindo, também, a inclusão de

elementos adicionais para atender às particularidades de cada usuário, com isso, possibilitando a coleta automática e a interoperabilidade entre os sistemas.

Por outro lado, a falta de padronização dos metadados pode trazer como consequência a ambigüidade em determinados campos, perda da especificidade e trazer resultados pouco confiáveis, provocando, assim, um grande desafio para a organização e recuperação da informação.

O acesso aos ítems bibliográficos depositados no repositório institucional requer, desse modo, uma linguagem comum de representação temática, que contemple, às demandas informacionais e da produção de conhecimento pelas diferentes linhas de pesquisa.

Para enfrentar esse desafio, optou-se pela utilização de um vocabulário controlado da área para indexar os documentos e, com isso, permitir sua efetiva recuperação, tendo, como ponto de partida, as linguagens documentárias efetivamente utilizadas pela rede Fiocruz, que é o vocabulário DeCS – Descritores em Ciências da Saúde.

O vocabulário estruturado e trilingue DeCS foi criado pelo Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME), centro especializado da Organização Pan-Americana da Saúde, escritório regional para as Américas da Organização Mundial da Saúde, doravante “BIREME/OPAS/OMS”, a partir do *MeSH - Medical Subject Headings* da *U.S. National Library of Medicine*, com objetivo de permitir o uso de uma terminologia controlada, tanto na indexação, como na pesquisa de assuntos da literatura técnica e científica nas bases de dados LILACS, MEDLINE e outras da Biblioteca Virtual em Saúde (BVS), proporcionando um meio consistente de recuperação de informação em ciências da saúde na região, independente de idioma.

Com a implantação do metadado assunto, a busca não será realizada no conteúdo dos arquivos no ARCA, mas, sim, nos descritores dos documentos.

A utilização do vocabulário controlado irá trazer vários benefícios, como por exemplo, a maior consistência na entrada de dados, com a diminuição de erros.

4 OBJETIVO

4.1 GERAL

Contribuir para a utilização do vocabulário controlado DeCS para indexar (o metadado assunto) e recuperar efetivamente os documentos depositados no ARCA (Memória Digital Institucional Informação Científica e Tecnológica em Saúde).

4.2 ESPECÍFICOS

A realização do estudo inclui os seguintes objetivos específicos:

- a. Levantar o estado da arte sobre recuperação da informação em estruturas de metadados e o uso das ferramentas de informação e comunicação necessárias;
- b. Customizar os metadados para descrição, identificação e recuperação de informações;
- c. Utilizar ferramentas tecnológicas que garantam a implementação do DeCS no ARCA como ferramenta de indexação e recuperação;
- d. Extrair os descritores do DeCS com avaliação da relevância destes e classificação dos documentos do ARCA usando estes descritores para gerar índices;
- e. Construir arquivos de autoridade e afiliação;
- f. Avaliar a indexação e recuperação da informação no ARCA.

5 METODOLOGIA

A metodologia deste trabalho consiste em duas abordagens: uma teórica que consistiu na revisão da literatura especializada sobre o assunto proposto e a obtenção de resultados expressivos por outros autores, e a outra empírica onde estão em desenvolvimento diversas fases, a saber:

1. Levantar o estado da arte sobre recuperação da informação em estruturas de metadados e o uso das ferramentas de informação e comunicação necessárias;
2. Customizar os metadados para descrição, identificação e recuperação de informações;
3. Utilizar ferramentas tecnológicas que garantam a implementação do DeCS no ARCA como ferramenta de indexação e recuperação;
4. Extrair os descritores do DeCS com avaliação da relevância destes e classificação dos documentos do ARCA usando estes descritores para gerar índices;
5. Construir arquivos de autoridade e afiliação;
6. Avaliar a indexação e recuperação da informação no ARCA.

6 RESULTADOS ESPERADOS

Os resultados esperados envolvem implementação de uma ferramenta de recuperação da informação e indexação dos documentos depositados no ARCA (Memória Digital Institucional Informação Científica e Tecnológica em Saúde), objetivando melhorar o mecanismo de busca atual por meio da implementação de um vocabulário controlado e otimizar a gestão da informação técnico-científica gerada pelos pesquisadores do ICICT.

A expectativa, também, é oferecer uma metodologia de fácil implementação e um mecanismo ágil para o armazenamento, a disseminação e o acesso à produção científica em texto completo, ampliando assim, a visibilidade do mesmo, e ainda, garantindo a preservação da memória científica.

7 REFERÊNCIAS CONSULTADAS

1. ALBERTS, B. **Engaging in a worldwide transformation**: our responsibility as scientists for the provision of global public goods, 2002. Disponível em: <<http://www.nasonline.org/site/DocServer/speech2002.pdf?docID=121>> Acesso em: 07 mar. 2008.
2. BACK, H. B. What information dissemination studies imply concerning the design of on-line reference retrieval-systems. **Journal of the American Society For Information Science**, Washington, v. 23, n. 3, p. 156-163, 1972.
3. BERNAL, J. D. **The social function of science**. London: George Routledge, 1946. Cap. 11, p. 292-308.
4. BORBINHA, J. L. **Biblioteca Nacional**, 2000. Disponível em: <<http://purl.pt/201/1/>>. Acesso em: 21 out. 2008.
5. BRIQUET DE LEMOS, Antônio. **Periódicos eletrônicos: problema ou solução?** Disponível em: <http://www.briquetdelemos.com.br/briquet/briquet_lemos7.htm>. Acesso em: 26 mar. 2008.
6. BOORSTIN, Daniel J. **Os Descobridores**: de como o homem procurou conhecer-se a si mesmo e ao mundo. Rio de Janeiro: Civilização Brasileira, 1989.
7. BUSH, Vannevar. **As we may think** (1945). Disponível em: <<http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush.shtml>>. Acesso em: 28 jul. 2004.
8. CAFÉ, Lígia; MELO, Bianca Amaro de; BARBOSA, Elza Maria Ferraz et al. **Repositórios institucionais**: nova estratégia para publicação científica na Rede. In: INTERCOM - CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 26., 2003, Belo Horizonte. **Anais...** Belo Horizonte: [s.n.], 2003.
9. CAMPOS, M. L. A.; GOMES, HAGAR E. Tesouro e normalização terminológica: o termo como base para intercâmbio de informações. **Datagrama zero**, Rio de Janeiro, v. 5, n. 6, 2004. Disponível em: <www.dgz.org.br/dez04/Art_02.htm>. Acesso em: 15 ago. 2007.
10. CROW, Raym. **The Case for Institutional Repositories**: A SPARC position paper. The Scholarly Publishing & Academic Resources Coalition 21 Dupont Circle. Washington, DC 2003. Disponível em: <<http://www.arl.org/sparc/IR/ir.html>>. Acesso em: 15 set. 2007.
11. GARVEY, W. D. **Communication, the essence of science**: facilitating information among librarians scientists, engineers and students. Oxford: Pergamon Press, 1979.
12. LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Brasília: Briquet de Lemos, 1993.
13. LAWRENCE, S. Online or Invisible? **Nature**, London, v. 411, n. 6837, p. 521, 2001.

14. LYNCH, C. A. **Institutional repositories: essential infrastructure for scholarship in the digital age.** ARL Bimonthly Report, 26, 2003. Disponível em: < <http://www.arl.org/newsltr/226/ir.html> > Acesso em 29/09/2007.
15. MEADOWS, A J. **A comunicação científica.** Brasília: Briquet de Lemos, 1999.
16. MURAMAKI, Tiago Rodrigo Marçal. Tesouros e a World Wide Web. 2005. 92f. Trabalho de Conclusão de Curso - (Escola de Comunicações e Artes) - Universidade de São Paulo, 2005.
17. ORTELLADO, P.; MACHADO, J. A. Direitos autorais e o acesso às publicações científicas. **Revista ADUSP**, São Paulo, p. 6 -15, ago. 2006. Disponível em: < <http://adusp.org.br/revista/37/r37a01.pdf> > Acesso em: 15 ago. 2008
18. RUSSELL, Jane M. La comunicación científica a comienzos del siglo XXI. **Revista Internacional de Ciencias Sociales**, Paris, n. 168, jun. 2001. Disponível em: <<http://www.campus-oei.org/salactsi/rusell.pdf>>. Acesso em: 27 dez. 2004.
19. SHINTAKU, M.; BRÄSCHER, M. Dspace versão 1.4: uma análise das facilidades relacionadas ao assunto. In: SEMINÁRIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS BRASIL, 2007, São Paulo. **Anais...** São Paulo, [s.n.], 2007.
20. SOARES, G. A. D. O Portal de periódicos da Capes: dados e pensamentos. **RBPG: Revista Brasileira de Pós-graduação**, Brasília, v.1, n.1, p. 10-25, 2004.
21. STUMPF, Ida Regina Chitto. Passado e futuro das revistas científicas. **Ciência da Informação**, Brasília, v. 25, n. 3, p. 383-386, set./dez. 1996.
22. TARGINO, M. G. O óbvio da informação científica: acesso e uso. **TransInformação**, Campinas, 19(2):95-105, maio/ago., 2007
23. TENOPIR, Carol. Information metrics and user studies. **Aslib Proceedings**, London, v. 55, n. ½, p. 13-17, 2003.
24. TOMAÉL, Maria Inês et al. Fontes de informação na Internet. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS DA AMÉRICA LATINA, 2000, Florianópolis. **Anais...** Florianópolis: Universidade Federal de Santa Catarina, 2000.
25. VIDOTTI, S. A. B. G.; SANCHES, S. A. S. Arquitetura da informação em Web site. In: II SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 2004, Campinas. **Anais....** Disponível em: <<http://libdigi.unicamp.br/document/?view=8302>>. Acesso em: 23/06/2008.
26. WEISMAN, H. M. **Information Systems, services and centers.** New York: Becker & Hayes, 1972.
27. WEITZEL, S. R. E-prints: modelo da comunicação científica em transição. In: FERREIRA, S.M.S.P., TARGINO, M.G. **Preparação de revistas científicas: teoria e prática.** São Paulo: Reichman, 2005. p.161-194.

8 CRONOGRAMA

ETAPA /MÊS	1	2	3	4	5	6
Aprofundamento da revisão de literatura						
Customização dos metadados para descrição, identificação e recuperação de informações						
Utilização de ferramentas tecnológicas que garantam a implementação do DeCS no ARCA como ferramenta de indexação e recuperação						
Extração dos descritores do DeCS com avaliação da relevância destes e classificação dos documentos do ARCA usando estes descritores para gerar índices						
Construção de arquivos de autoridade e afiliação						
Avaliação da indexação e recuperação da informação no ARCA						

9 ORÇAMENTO

O orçamento abaixo relaciona os recursos financeiros a serem utilizados ao longo de todo o projeto. Os itens relacionados ao material de consumo serão utilizados aqueles disponíveis no ICICT.

PROPOSTA DETALHADA DE ORÇAMENTO

DESCRIÇÃO	ITEM	QUANTIDADE	VALOR (R\$)	JUSTIFICATIVA
Bolsa de Pesquisa	Bolsa	1 (6 meses)	7.200,00	Bolsa por 6 meses, para pesquisador que irá desenvolver a pesquisa
Bolsa de Estágio	Bolsa	2 (6 meses)	6.240,00	Bolsa por 6 meses, para 2 estagiários de nível superior: 1 de biblioteconomia e 1 analista de sistemas
Computador	Material Permanente	2	4.800,00	Aquisição de 2 computadores para o processamento técnico
Impressora	Material Permanente	1	300,00	Aquisição de 1 impressora
Banner	Serviço de Terceiros	2	300,00	Impressão de banner para divulgação e andamento do projeto, bem como, para ser usado em trabalhos científicos
Passagens	Passagem (total)	1	1.100,00	Passagens para apresentar resultados em evento científico.
Diárias	Diárias Totais	2	247,40	Diárias para apresentar resultados em evento científico