

Fundação Instituto Oswaldo Cruz
Escola Nacional de Saúde Pública
Mestrado Profissional Gestão da Informação e Comunicação em saúde



Dissertação

**Buscando Interoperabilidade entre diferentes
bases de dados:
O caso da Biblioteca do Instituto Fernandes Figueira**

Por

Viviane Santos de Oliveira

Orientador:

Dr. Carlos Henrique Marcondes

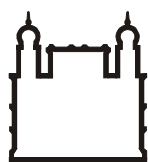
Rio de Janeiro
Março de 2005

**Buscando Interoperabilidade entre diferentes
bases de dados:
O caso da Biblioteca do Instituto Fernandes Figueira**

Dissertação apresentada ao Programa de Pós-Graduação em Saúde Pública da Escola Nacional de Saúde Pública Sérgio Arouca - FIOCRUZ, como requisito parcial para obtenção do título de mestre em Gestão da Informação e Comunicação em Saúde.

Orientador: Dr. Carlos Henrique Marcondes

Rio de Janeiro
Março de 2005



FIOCRUZ

Fundação Instituto Oswaldo Cruz
Escola Nacional de Saúde Pública
Mestrado Profissional Gestão da Informação e Comunicação em saúde



ENSP

Viviane Santos de Oliveira

**Buscando Interoperabilidade entre diferentes
bases de dados:
O caso da Biblioteca do Instituto Fernandes Figueira**

Trabalho de Conclusão de Curso defendido e aprovado em XX de março
de 2005, pela Banca Examinadora constituída pelos professores:

BANCA EXAMINADORA

Prof. Carlos Henrique Marcondes (Orientador)
IBICT/UFF -RJ

Prof. Luís Fernando Sayão
CNEN - RJ

Prof. Dr. Ilara Hammerli S. Moraes
ENSP-RJ

*Dedico esta dissertação ao meu amado
esposo Victor. Você é o presente de Deus
para mim.*

AGRADECIMENTOS

À Deus, por ser um grande Pai e meu Mestre maior. O autor da interoperabilidade, criador de um mundo com elementos tão diversos e singulares, mas todos interagindo e interdependentes, formando um todo maravilhoso.

À meu amado esposo Victor por me apoiar a todo momento, sendo compreensivo e “o companheiro para todas as horas”.

À minha querida mãe Neli, por me ensinar valores tão preciosos para vida e ser uma guerreira incansável na busca de um futuro melhor para sua filha.

Ao meu querido pai Paulo, por ser tão importante na minha formação e criação.

Às minhas queridas irmãs Renata e Ana Paula por serem amigas e companheiras.

Aos meus professores Eugenio Decourt e Beatriz Decourt (UNIRIO) pelo importante incentivo.

À toda equipe da Biblioteca do IFF (Diones, Neide, Márcia) pelo apoio, principalmente à Giulia, Giovania e Ciça, pelo apoio extra.

À todos os colegas da turma de mestrado pelos maravilhosos debates e discussões que tanto me auxiliaram.

Às minhas amigas que tive o prazer de conhecer nesta caminhada do mestrado, Rosane e Rejane, pela divisão de angústias, sugestões e apoio.

À minha amiga Ilza, pelo constante incentivo.

Ao corpo docente do Mestrado. Vocês foram fundamentais para a realização deste trabalho!!

Ao meu orientador, Dr. Carlos Henrique Marcondes, por me enriquecer compartilhando seu conhecimento e por aceitar este desafio!!

"Se penso que devo escrever um livro, todos os problemas de como esse livro deve ser e como não deve ser me bloqueiam e me impedem de ir adiante. Se, ao contrário, penso que estou escrevendo uma biblioteca inteira, sinto-me imediatamente aliviado: sei que qualquer coisa que eu escreva será integrada, contradita, equilibrada, amplificada, sepultada nas centenas de volumes que me resta escrever".

Ítalo Calvino

RESUMO

Este trabalho analisa as possibilidades de tornar interoperáveis bases de dados heterogêneas, de forma que possam ser pesquisadas através de uma única interface. A biblioteca do Instituto Fernandes Figueira, Unidade materno-infantil da FIOCRUZ é referência em tratamento de doenças de alta complexidade. Estes profissionais buscam a Biblioteca com necessidades de informações voltadas para a decisão clínica. As bases de dados mais pesquisadas por estes profissionais são as bases de dados LILACS, MEDLINE e ACERVOS ONLINE FIOCRUZ. Portanto estas bases norteiam este estudo e são analisadas em algumas facetas como: os procedimentos de descrição e as estruturas das bases, buscando semelhanças e divergências nos índices e nos campos de exibição das bases; a representação temática, caracterizada pela utilização dos Tesouros DeCS e MeSH; e tecnologias utilizadas e/ou compatíveis. Paralelamente busca-se apresentar algumas ferramentas tecnológicas atuais de interoperabilidade como os protocolos Z39.50 e o OAI-PMH, os metabuscadores, o conjunto de metadados Dublin Core e o MetaIAH. Através destas análises delineou-se três modelos conceituais para alcançar a interoperabilidade, são eles: **Compartilhamento de Esforços**, quando tanto a Interface de Consulta (IC) e os Recursos Informativos (RI) trabalham para garantir a interoperabilidade; **Esforço concentrado nos RIs**, quando os Recursos Informativos arcam com todo o esforço para possibilitar a interoperabilidade; e, **Esforço concentrado na IC**, quando a Interface de consulta se adapta para suprir as divergências de cada base proporcionando a interoperabilidade. Para finalizar comparam-se os modelos e as bases de dados estudadas destacando-se os modelos que se constituem como alternativas de interoperabilidade para a biblioteca do Instituto Fernandes e quais as vantagens e desvantagens de cada uma delas.

PALAVRAS-CHAVE: Interoperabilidade; Bases de dados; Metadados; Protocolos; Informação para tomada de decisão;

ABSTRACTS

This work analyses the options to achieve interoperability among different library databases so they can be searched from a single Web interface. The databases object of this research are the ones used in The Fernandes Figueira Institute Library, which is a branch of FIOCRUZ specialized in Maternal and Child Health. Library users generally look for medical decision support information. The Library users preferred databases are LILACS, MEDLINE and ACERVOS ONLINE FIOCRUZ. All of them use the DeCS and MeSH Thesaurus to subject description. Each of these databases are analyzed in aspects as access points, bibliographic description and display formats, in order to identify common elements. Technologies and standards to achieve interoperability as Z39.50 and OAI-PMH protocols, meta-search engines, Dublin Core Metadata Set BIREME's metaIaH meta-search engine are also analyzed. From this analysis emerged three conceptual models to achieve interoperability among the three databases, evolving two actors, the search interface – SI, and the information resources (the three databases) - IR: Effort Sharing, between SI and IRs; Concentrated Effort in the IRs; and Concentrated Effort in the SI to achieve interoperability. The three models are compared, outlining the advantages and disadvantages of each one to achieve interoperability, among the databases of interest to Fernandes Figueira Institute Library.

Keywords: Interoperability, Library Databases, Metadata, Protocols, Medical Decision Support Information

LISTA DE QUADROS

Quadro A <i>Formulários LILACS</i>	60
Quadro B <i>Formulários MEDLINE</i>	64
Quadro C <i>Formulários ACERVOS ONLINE</i>	68
Quadro D <i>Campos LILACS – Formato Detalhado</i>	70
Quadro E <i>Campos LILACS – Formato Longo</i>	71
Quadro F <i>Campos LILACS – Formato Título</i>	71
Quadro G <i>Campos LILACS – Formato Citação</i>	71
Quadro H <i>Campos LILACS – Formato Afiliação</i>	72
Quadro I <i>Campos MEDLINE – Formato Longo</i>	73
Quadro J <i>Campos MEDLINE – Formato Detalhado</i>	74
Quadro L <i>Campos MEDLINE – Formato Título</i>	75
Quadro M <i>Campos MEDLINE – Formato Citação</i>	75
Quadro N <i>Campos Acervos Online – Formato Padrão</i>	76
Quadro O <i>Campos Acervos Online – Formato Ficha Catalográfica</i>	77
Quadro P <i>Campos Acervos Online – Formato Ficha Reduzido</i>	77
Quadro Q <i>Campos Acervos Online – Formato Nomes MARC</i>	78

Quadro R	
<i>Índices das bases de dados LILACS, MEDLINE e Acervos Online Fiocruz.....</i>	79
Quadro S	
<i>Índices semelhantes das bases de dados LILACS, MEDLINE e Acervos Online Fiocruz.....</i>	80
Quadro T	
<i>Campos das bases de dados LILACS, MEDLINE e Acervos Online Fiocruz.....</i>	81
Quadro U	
<i>Campos semelhantes das bases de dados LILACS, MEDLINE e Acervos Online Fiocruz.....</i>	82

LISTA DE FIGURAS

Figura A	
<i>Interoperabilidade entre recursos informacionais utilizando o protocolo Z39.50.....</i>	42
Figura B	
<i>Base LILACS - formulário livre.....</i>	58
Figura C	
<i>Base LILACS - formulário básico.....</i>	58
Figura D	
<i>Base LILACS - formulário avançado.....</i>	59
Figura E	
<i>Base MEDLINE - formulário livre.....</i>	61
Figura F	
<i>Base MEDLINE - formulário básico.....</i>	62
Figura G	
<i>Base MEDLINE - formulário avançado.....</i>	63
Figura H	
<i>Base de Dados Acervos Online – pesquisa simples.....</i>	65
Figura I	
<i>Base de Dados Acervos Online – formulário multi-campo.....</i>	66
Figura J	
<i>Base de Dados Acervos Online – formulário multi-base.....</i>	66
Figura L	
<i>Base de Dados Acervos Online – formulário pesquisa avançada.....</i>	67
Figura M	
<i>Modelo 1 - Compartilhamento de Esforços.....</i>	86
Figura N	
<i>Modelo 2 - Esforço Concentrado nos RI's</i>	86
Figura O	
<i>Modelo 3- Esforço Concentrado na IC.....</i>	87
Figura P	
<i>Modelo 3- Esforço Concentrado na IC: Metabuscaror Próprio.....</i>	91

SUMÁRIO

1	INTRODUÇÃO	14
1.1	JUSTIFICATIVA.....	17
1.2	OBJETIVO.....	22
1.2.1	Objetivo Geral.....	22
1.2.2	Objetivos Específicos.....	22
1.3	METODOLOGIA.....	23
2	CAMPO EMPÍRICO	25
2.1	MEDLINE.....	26
2.2	LILACS.....	26
2.3	ACERVOS ONLINE FIOCRUZ.....	28
3	QUADRO CONCEITUAL	30
3.1	DEFININDO INTEROPERABILIDADE	31
3.2	PRINCIPAIS AUTORES.....	33
3.3	ÁREAS DO CONHECIMENTO.....	34
3.3.1	Área da Ciência da Informação.....	34
3.3.2	Área da Ciência da Computação.....	35
3.3.3	Área da Saúde.....	36
3.4	A COOPERAÇÃO/INTERAÇÃO: BREVE HISTÓRICO.....	36
3.5	ALTERNATIVAS DE INTEROPERABILIDADE.....	38
3.5.1	Protocolo Z39.50 e Protocolo OAI-PMH.....	39
3.5.1.1	Protocolo Z39.50.....	40
3.5.1.2	Protocolo OAI-PMH.....	43
3.5.2	Metabuscadores.....	45
3.5.3	Conjunto de Metadados Dublin Core.....	47
3.5.4	MetalAH.....	50
3.6	ONTOLOGIAS.....	53

4	ANÁLISE DAS BASES DE DADOS	56
4.1	ÍNDICES DISPONÍVEIS NOS FORMULÁRIOS DE PESQUISA ..	57
4.1.1	Formulários LILACS e MEDLINE.....	57
4.1.1.1	Formulários LILACS.....	57
4.1.1.2	Formulários MEDLINE.....	61
4.1.2	Formulários ACERVOS ONLINE FIOCRUZ.....	65
4.2	CAMPOS DE EXIBIÇÃO.....	68
4.2.1	Campos LILACS e MEDLINE.....	69
4.2.1.1	Campos LILACS.....	69
4.2.1.2	Campos MEDLINE.....	73
4.2.2	Campos ACERVOS ONLINEFIOCRUZ.....	75
4.3	BUSCANDO AS SEMELHANÇAS.....	78
4.3.1	Semelhanças nos índices.....	78
4.3.2	Semelhanças nos Campos de Exibição.....	80
5	MODELOS CONCEITUAIS	83
5.1	MODELO 1 – Compartilhamento de Esforços.....	85
5.2	MODELO 2 – Esforço concentrado nos RI's.....	86
5.3	MODELO 3 – Esforço concentrado na IC.....	87
5.4	OS MODELOS E AS BASES DE DADOS.....	87
6	CONSIDERAÇÕES FINAIS	92
	REFERÊNCIAS BIBLIOGRÁFICAS	96
	BIBLIOGRAFIA	103
	ANEXO	109

CAPÍTULO I

INTRODUÇÃO

“Por definição, a informação só adquire valor se utilizada pelo “receptor” para diminuir seu grau de incerteza(...) As tecnologias oriundas da eletrônica e da informática, em relação às informações, abrem possibilidades imensas para seu uso e sua aplicação na área de saúde, com alcances inimagináveis.”

Ilara Hammerli S. Moraes

1 INTRODUÇÃO

O campo saúde no Brasil perpassa várias dificuldades, e destaca-se neste trabalho a barreira informacional. O fenômeno da explosão informacional desencadeou uma grande quantidade de informações produzidas e disponibilizadas em diferentes formatos na Internet, dificultando sua identificação e acesso, e tornando o trabalho do pesquisador mais árduo. Na chamada “Sociedade da Informação”^{*}, caracterizada, por Capurro¹, pelo crescimento da Tecnologia da Informação e seus impactos globais, o valor da informação está diretamente ligado com seu potencial de orientar, com eficiência e eficácia a realização desta atividade. Para que este potencial seja concretizado, a informação relevante para um dado problema precisa ser identificada em tempo viável. “Platão, na Grécia antiga, indagava: ‘Como é possível que, com tão poucas informações, nós tenhamos chegado a saber tanto?’. No século 20, George Orwell inverteu a indagação: ‘Como é possível que, com tantas informações, nós tenhamos chegado a saber tão pouco?’”. (Konder²)

Diversas estratégias têm sido criadas para tentar minimizar as dificuldades geradas pela explosão informacional trazida pela Internet, dentre elas temos os mecanismos de buscas gerais (Google, Yahoo, Infoseek, Miner, e outros). As deficiências destes mecanismos já são bem conhecidas (Shneiderman³), e pode-se destacar a baixa qualidade da indexação, (feitas por programas robô), o que gera alta revocação (grande quantidade de informação recuperada), porém baixa precisão (pouca informação relevante).

A área da saúde, como todas as áreas do conhecimento, enfrenta estes problemas que acarretam dificuldades e morosidade no acesso à informação relevante para os profissionais da assistência. Na área da Saúde da Mulher e da Criança existem diversas fontes de informações, porém disponibilizadas de forma fragmentada sem articulação entre as mesmas. Atualmente os profissionais da área quando desejam obter informações sobre o assunto precisam conhecer as URL’s (Uniform Resource Locate)

^{*} Sociedade da informação: ”uma nova era em que a informação flui a velocidades e em quantidades há apenas poucos anos inimagináveis, assumindo valores sociais e econômicos fundamentais.” (Brasil⁵⁰, p. 3)

das bases de dados relevantes sobre o tema e então percorrer as bases de dados uma a uma.

O Instituto Fernandes Figueira (IFF), unidade materno-infantil da Fundação Oswaldo Cruz que realiza pesquisa, ensino e assistência – principalmente no nível terciário – no âmbito da saúde da criança, do adolescente e da mulher, sendo referência em tratamento de diversas doenças de alta complexidade, possui uma biblioteca cujo acervo (periódicos, livros e teses) é referência nacional na área da Saúde da Criança, do Adolescente e da Mulher.

A Biblioteca do Instituto Fernandes Figueira é uma das 10 bibliotecas que compõem a Rede de Bibliotecas da FIOCRUZ. O acervo de livros, teses e periódicos da Biblioteca do Instituto Fernandes Figueira é gerenciado pelo sistema de gerenciamento de bibliotecas Aleph, da empresa Ex-Libris. Este sistema está entre os mais avançados sistemas de gerenciamento de bibliotecas comercializados no mundo, e incorpora padrões como MARC* e Z39.50†. Este acervo, como os das demais bibliotecas integrantes do Centro de Informações Científicas e Tecnológicas (CICT) da FIOCRUZ, pode ser recuperado através do link *Acervos On-line* disponibilizado através do site da FIOCRUZ (www.fiocruz.br).

O acervo de periódicos tem cerca de 88 % de seus artigos indexados nas bases de dados MEDLINE e LILACS. Estas três bases se configuram como as mais consultadas entre os profissionais e pesquisadores que atuam no IFF.

A MEDLINE é uma base de dados da literatura internacional na área médica e biomédica, produzida pela US National Library of Medicine desde 1966. Esta base indexa cerca de 4.700 revistas científicas de 70 países sendo 24 brasileiras (Castro⁴).

A LILACS é uma base de dados da Literatura Latino Americana e do Caribe em Ciências da Saúde. Esta base é coordenada pela BIREME (Centro Latino Americano e do Caribe de Informação em Ciências da Saúde), promovido pela Organização Mundial da Saúde e Organização Panamericana da Saúde (OMS/OPAS) desde 1982. Ela indexa cerca de 650 revistas, de 18 países sendo 254 brasileiras. Inclui, em seu acervo, além de

* Padrão para intercâmbio de informações legível por computador. <<http://www.loc.gov/marc/>>

† Padrão para consulta à catálogos eletrônicos de bibliotecas. <<http://www.loc.gov/z3950>>

artigos de revistas, livros, capítulos de livros, trabalhos apresentados em eventos científicos, relatórios, teses, projetos, etc.

Estes recursos Web – as bases ACERVOS ONLINE FIOCRUZ, MEDLINE e LILACS - são heterogêneos, não integrados, desenvolvidos e mantidos de forma independente. Quando necessitam consultar estas bases, os usuários precisam percorrê-las uma a uma, para verem retornadas as informações procuradas.

A BIREME através de um convênio com a National Library of Medicine recebe em seu servidor a base de dados MEDLINE e através de uma ferramenta chamada MetaIAH possibilita a busca na base MEDLINE e LILACS através de uma única interface. Este nível de articulação será analisado posteriormente através da análise das bases e da ferramenta MetaIAH.

Apesar da possibilidade de pesquisar as bases MEDLINE e LILACS simultaneamente, o usuário ainda precisa pesquisar, em separado, a base ACERVOS ONLINE FIOCRUZ para poder consultar o acervo de livros e teses do IFF. Para esta busca é necessário abrir nova página WEB e dar entrada nas palavras chaves mais uma vez, provocando retrabalho. Este problema pode se agravar caso haja a necessidade de pesquisar ainda em outras bases, ou na incorporação de outros recursos informacionais, novas bases de dados ou bibliotecas digitais. A solução para este retrabalho seria a possibilidade de, através de uma única interface, poderem ser pesquisadas bases de dados diferentes, heterogêneas retornando resultados consolidados. Esta “capacidade de operar em conjunto” é conhecida na literatura como *interoperabilidade**. Este problema tem múltiplas dimensões: tecnológica, semântica, política, entre outras. (Miller⁵).

1.1 JUSTIFICATIVA

Existe uma grande relação entre desenvolvimento econômico e social e o estágio de desenvolvimento da ciência e tecnologia de um país. À ciência e tecnologia está reservado um papel fundamental na luta pelo desenvolvimento social do Brasil. Informação é o insumo fundamental para o desenvolvimento da ciência.

* Concise Oxford Dictionary, 9th Edition. *Apud* Miller⁵

Reconhecendo a importância da informação e da tecnologia, a Fiocruz tem como Missão:

“**Gerar, absorver e difundir conhecimentos científicos e tecnológicos** em saúde pelo desenvolvimento integrado em atividades de pesquisa e desenvolvimento tecnológico, ensino, produção de bens e insumos, de prestação de serviços de referência e assistência, informação e comunicação em C&T em Saúde, com a finalidade de **atender as demandas** do Ministério da Saúde através do apoio estratégico ao Sistema único de Saúde (SUS) e a melhoria da qualidade de vida da sociedade como um todo” (FIOCRUZ⁶) (grifo meu).

Da mesma forma, o CICT (Centro de Informação Científica e Tecnológica - unidade técnica da FIOCRUZ) tem como missão:

“... desenvolver estratégias e executar ações de informação e comunicação no campo da ciência e tecnologia em saúde, visando identificar e **atender as demandas** sociais, do SUS e de outros órgãos governamentais” (CICT⁷) (grifo meu).

Segundo o relatório de atividades do CICT de 1999, sua missão está voltada para:

“capacitação institucional, no tratamento e gerenciamento da informação científica e tecnológica, visando a integração interna e a articulação externa da FIOCRUZ, no campo da informação e comunicação em saúde”. (CICT⁸)

Analisando a missão da FIOCRUZ e a missão do CICT percebe-se como foco principal atender as demandas dos profissionais da área da saúde, através do gerenciamento da informação e promoção do conhecimento científico e tecnológico. Os estudos científicos na área da saúde têm crescido consideravelmente. A cada dia, novas descobertas são publicadas. Doenças, anteriormente incuráveis, aparecem com novos tratamentos que possibilitam a cura.

No Instituto Fernandes Figueira é comum o profissional (médico, enfermeiro...), antes de definir um diagnóstico, encaminhar um tratamento ou fazer uma cirurgia, buscar a Biblioteca do IFF para fazer um levantamento do assunto em questão. Esta procura se deve, principalmente, pela qualidade do acervo disponível na Biblioteca que se configura como referência Nacional para os estudos da área da Saúde da Criança, do Adolescente e da Mulher. Analisando também os profissionais desta área e suas dificuldades de acesso integrado à informação, percebe-se uma grande demanda para atuação desta Instituição.

Segundo Marcondes⁹ do ponto de vista do usuário, seria interessante e confortável poder submeter sua pesquisa e interagir com uma única interface e ter retornadas informações de diferentes fontes de informação de forma consolidada.

Como a Biblioteca do IFF está inserida em um contexto específico, a rotina de um hospital, e tem como público-alvo os profissionais que ali trabalham, percebe-se a importância de se disponibilizar informações consolidadas, ampliadas, que cooperem com a tomada de decisão clínica.

A necessidade de se utilizar informações registradas na decisão clínica tem se maximizado nas últimas décadas. Nos anos 80 a McMaster Medical School (Canadá) desenvolveu a expressão “Medicina Baseada em Evidências” para nomear a estratégia do saber clínico que o pessoal da faculdade desenvolveu durante mais de uma década. Segundo o DeCS¹⁰ – Tesouro utilizado pela Bireme como Descritores em Saúde – Medicina Baseada em Evidências se define como:

“o processo de procurar, avaliar e usar sistematicamente os achados de pesquisas contemporâneas como base para decisões clínicas. A medicina baseada em evidências faz perguntas, busca e avalia os dados relevantes, aproveitando as informações para a prática clínica diária, procurando seguir quatro passos: formular uma pergunta clínica clara a partir do problema de um paciente; procurar artigos clínicos relevantes na literatura; avaliar (criticamente) a validade e a utilidade das evidências; implementar os achados úteis na prática clínica”.

Algumas das dificuldades de se obter informações em saúde para a tomada de decisão são descritas na tese de doutorado da Dra. Rosany Bochner¹¹, quando ela aponta os problemas encontrados nos principais Sistemas de Informação em Saúde- SIS. Estes Sistemas tem o objetivo de fornecer “informações acerca da saúde da população”(Brasil¹², p.8). Alguns desses sistemas são gerenciados pelo Ministério da Saúde, porém de responsabilidade de órgãos distintos, e segundo Carvalho¹³, com precária, ou nenhuma coordenação central. Exemplos desses sistemas de saúde são o SIM (Sistema de Informações sobre Mortalidade), o SINASC (Sistema de Informações sobre Nascidos Vivos), o SINAN (Sistema de Informações de Agravos de Notificação), o SIH/SUS (Sistema de Informações Hospitalares) e o SIA/SUS (Sistema de Informações Ambulatoriais do SUS)

A análise de Bochner demonstra as dificuldades do profissional em encontrar informações relevantes através dos sistemas de informações em saúde devido a sua falta de articulação. A autora destaca que estes sistemas de informações foram criados

isoladamente com objetivos distintos buscando responder um conjunto de problemas isolados.

Na criação destes sistemas percebe-se a dificuldade de se ver a saúde como um todo e o sistema de informação como parte desse todo, como um subsistema deste grande sistema que é o Sistema Único de Saúde. Com uma visão holística se criariam sistemas de informação que além de responder aos objetivos específicos também se articularia com outros sistemas, constituindo um sistema geral que respondesse às suas demandas.

Apesar desses Sistemas não conterem informações bibliográficas, como no caso em estudo, é relevante destacar as dificuldades de informação deste campo em análise. No setor saúde é urgente a necessidade do conhecimento interdisciplinar. A saúde antes única foi se fragmentando e se especificando de tal forma que se perdeu a noção do corpo como um todo, uma unidade. E a informação em saúde demonstra esta divisão. Várias são as bases de dados que temos disponíveis na área da saúde, e em todo o mundo e a todo o momento surgem novas formas de se lidar com um dado problema ou doença. Porém, a falta de articulação entre os diversos setores e campos causa duplicidade de estudos, redundância e conseqüentemente, morosidade no desenvolvimento do setor. É de suma importância que os profissionais de saúde possam adquirir conhecimentos através de um mecanismo eficiente e eficaz.

Esta fragmentação do conhecimento na área da saúde como em outras áreas vem como herança da forma do pensamento ocidental na produção do conhecimento. Sua gênese histórica data da primeira metade do século XVII quando é formulada a filosofia mecanicista em que todos os fenômenos da natureza poderiam ser explicados pela matéria em movimento. Seguindo esta filosofia vem o método cartesiano de Descartes que afirma que **cada problema deve ser dividido em tantas partes separadas quanto possível** (Discurso do Método). Este paradigma cartesiano é considerado o paradigma da disjunção, onde tudo é condenado à separação: “corpo, alma e emoções, sujeito e objeto, ser humano e natureza, interioridade e exterioridade, eu e outro e assim sucessivamente” (Pellanda¹⁴). Na filosofia cartesiana a emoção, o subjetivo, o desejo, a experiência, são expulsos da investigação científica, pois atrapalham o conhecimento. Com esta repartição do conhecimento, vem a super-especialização, o conhecimento que anteriormente era universal, passa a ser fragmentado. E este conhecimento, antes unitário, passa a ser dividido em campos com regras próprias e atores comuns. O

homem não consegue mais saber de tudo, ou de todas as “artes” de forma universal. Cada campo se organiza com uma definição de valor, e estabelece objetos de disputas e interesses específicos. Sendo assim uma pessoa de um dado campo tem dificuldades em entender as demandas de um outro campo pois o objeto de interesse e o jogo disputado é diferente. Para conseguir um relacionamento entre campos é necessário que os dois se proponham a “jogar o mesmo jogo” possibilitando uma atuação intercampos. Nesta dicotomia perde-se a visão holística, e com isto perde-se conhecimento.

O esforço, no momento, está na necessidade de se unir o que era único e foi fragmentado, o conhecimento. Segundo Morin¹⁵ “uma disciplina tende naturalmente à autonomia pela delimitação de suas fronteiras, pela linguagem que ela constitui, pelas técnicas (...) e pelas teorias”. A transdisciplinaridade busca a troca, a cooperação, em uma tentativa de se tornar orgânica. Visa a integração de conceitos, métodos... para obter interação e geração de conhecimentos. É a busca da interação entre as teorias, a prática e a arte. E como nos explica Benjamin¹⁶, a informação para transformar precisa ser associada à experiência e às relações interpessoais, e, sendo assim, acrescentaria que o conhecimento para ser utilizado de forma ideal precisa ser otimizado pela relação ou interação entre os campos (intercampos), ou as disciplinas (interdisciplinas).

Konder² afirma que “o conhecimento depende da possibilidade da comparação”. Latour¹⁷ nos demonstra como pode ser revelador (e amplificador) a superposição de informações. Com um gráfico, desenvolvido pelo fisiologista Marey a partir da superposição de um mapa da Rússia, a medida das temperaturas durante a guerra, o percurso da Grande Armée, a data de seus deslocamentos e o número de soldados sobreviventes em cada bivaque, compreende-se, com clareza, a influência direta do clima nos resultados da guerra. Quanto a este gráfico ele diz que:

“Informações diferentes, procedentes de instrumentos separados, podem unificar-se numa só visão, porque suas inscrições possuem todas a mesma coerência ótica. Sem a superposição das inscrições móveis e fiéis, seria impossível apreender as relações entre os lugares, as datas, as temperaturas, os movimentos estratégicos e as vítimas do general Inverno. Neste “lugar-comum” oferecido pela roteirização do gráfico, cada dado se liga, por um lado, a seu próprio mundo de fenômenos, e, por outro lado, a todos aqueles com os quais se torna compatível.” p. 30

Atualmente, o volume de informações disponível é maior do que a capacidade de uma biblioteca em adquirir, processar e armazenar estas informações fisicamente. A tendência é a biblioteca disponibilizar para seus usuários além do que está fisicamente em seu acervo, as informações que estão distribuídas em outros acervos e bases. Lancaster¹⁸ (p. 9) afirma que as novas tecnologias mudaram a “filosofia da biblioteca”.

O novo papel da biblioteca não é controlar tudo que o usuário necessita, e sim “fornecer **acesso** aos recursos, em quaisquer formas que eles se apresentam...”. Segundo Latour ¹⁷ “uma biblioteca considerada como um laboratório não pode, é evidente, permanecer isolada, como se ela acumulasse, de modo maníaco, erudito e culto, milhões de signos. Ela serve antes de estação de triagem...” (p. 37)

1.2 OBJETIVO

1.2.1 – Objetivo Geral

Esta dissertação objetiva investigar as alternativas conceituais, técnicas, de padronização e metodológicas, de buscar a interoperabilidade entre as bases de dados MEDLINE, LILACS, ACERVOS ONLINE/FIOCRUZ de modo a permitir que um usuário possa consultá-las como se fossem uma única base.

Esta investigação busca delinear modelos conceituais que possibilitem apontar as alternativas para a implementação de uma consulta simultânea às três bases, contribuindo para a otimização da comunicação entre os profissionais e suas produções.

Este acesso integrado pretende possibilitar uma articulação das bases de dados e das informações nelas contidas, contribuindo para uma construção transversal e interdisciplinar do conhecimento na área da Saúde da Criança, do Adolescente e da Mulher, além de facilitar a descoberta na Internet destes recursos, encurtando o ciclo de comunicação científica entre seus pares, contribuindo para a otimização da tomada de decisão destes profissionais e para a promoção do desenvolvimento do conhecimento científico da área.

1.2.2 - Objetivos Específicos

Construir o estado da arte no conceito de interoperabilidade.

Comparar a estrutura das Bases de Dados MEDLINE, LILACS e ACERVOS ONLINE FIOCRUZ.

Analisar a dimensão da representação temática nas Bases de Dados Medline, Lilacs e Acervos Online Fiocruz.

Analisar as alternativas de interoperabilidade entre as Bases supra citadas.

1.3 METODOLOGIA

Com o intuito de analisar as alternativas conceituais, técnicas, de padronização e metodológicas, de buscar a interoperabilidade entre as bases de dados Medline, Lilacs e Acervos Online/Fiocruz, foi desenhada uma metodologia a partir das seguintes etapas e procedimentos que nortearam esta análise afim de responder aos objetivos propostos neste trabalho.

- ESTADO DA ARTE

Nesta primeira etapa se construiu o estado da arte em interoperabilidade, através de pesquisa bibliográfica e estudos dirigidos.

- CAMPO DE OBSERVAÇÃO

Nesta segunda etapa foi definido o campo de observação. Na proposta inicial se apresentavam como bases de dados relevantes para o público definido as bases de dados MEDLINE, LILACS, ACERVOS ONLINE FIOCRUZ e Biblioteca Cochrane. As três primeiras por, além da abrangência já citada, se configurarem as bases mais consultadas pelos profissionais do IFF, e a quarta base por apresentar informações relevantes para a tomada de decisão clínica. Porém, na qualificação deste projeto foi nos sugerido trabalhar apenas com as Bases de Dados MEDLINE, LILACS e ACERVOS ONLINE FIOCRUZ, pois a Base de Dados Biblioteca Cochrane é uma base com especificidades peculiares, possuindo uma formatação muito diversa da maioria das bases de dados bibliográficas, o que tornaria muito mais complexo o problema.

- ANÁLISE DAS BASES DE DADOS

Nesta fase foram analisados os procedimentos de descrição e análise das estruturas das bases de dados (Robredo¹⁹), através de um mapeamento destas estruturas, possibilitando a verificação de equivalências e diferenças entre as bases de dados.

- SIMULAÇÃO DE BUSCAS

Nesta etapa foram feitas simulações de buscas para considerar algumas alternativas de métodos e recursos de interoperabilidade, através da simulação de buscas.

- REPRESENTAÇÃO TEMÁTICA

Nesta fase foi feita a análise da dimensão da representação temática nas bases de dados, verificando possíveis superposições, equivalências e diferenças.

- ESTUDO DAS ALTERNATIVAS DE INTEROPERABILIDADE

Nesta etapa desenharam-se as alternativas disponíveis de interoperabilidade.

- MODELO CONCEITUAL

Nesta fase delineou-se modelos conceituais que permitam a interoperabilidade entre as bases de dados (MEDLINE, LILACS, ACERVOS ONLINE FIOCRUZ) e comparou-se os modelos às bases de dados em estudo.

CAPÍTULO II

CAMPO EMPÍRICO

“...de que nos serviriam todos os saberes parcelados, se nós não os confrontássemos, a fim de formar uma configuração que responda às nossas expectativas, às nossas necessidades e às nossas interrogações cognitivas?”

Edgar Morin

2 CAMPO EMPÍRICO

Este capítulo busca descrever as bases de dados MEDLINE, LILACS e ACERVOS ONLINE FIOCRUZ, que se configuram o campo empírico desta pesquisa.

2.1 – MEDLINE

A base de dados MEDLINE (MEDlars onLINE), versão eletrônica do Index Medicus, é uma base de dados bibliográficos, criada e mantida pela Biblioteca Nacional de Medicina dos Estados Unidos (National Library of Medicine, USA - NLM). Ela contém referências bibliográficas e resumos de mais de 4000 periódicos da literatura internacional, publicados nos Estados Unidos e em outros 70 países.

Esta Base de dados cobre as áreas de medicina, biomedicina, odontologia, veterinária e ciências afins, e contém aproximadamente 11 milhões de registros da literatura, desde 1966, com predominância de publicações em língua inglesa. A base MEDLINE utiliza o tesouro criado pela NLM para indexação de seus artigos, o MeSH - Medical Subject Headings da US National Library of Medicine. Ela é disponibilizada e operada por diversas agências. Apesar de ser gerenciada pela NLM, a MEDLINE é disponibilizada para os países latino-americanos e do Caribe através da BIREME, que recebe em seu servidor a base. Neste estudo estaremos analisando a versão MEDLINE Operada pela BIREME

Esta versão da base MEDLINE pode ser acessada via Internet através do site da Bireme (www.bireme.br) na Biblioteca Virtual em Saúde no *link* “Literatura científica”.

2.2 LILACS

A Organização Pan-Americana da Saúde (OPAS) opera no Brasil o Centro Latino-Americano e do Caribe de Informações em Saúde (BIREME) que foi criado especialmente para desenvolver o programa de informação em Ciências da Saúde da OPAS. A BIREME coordena o Sistema Latino-Americano de Informações em Ciências da Saúde que objetiva cooperar com o desenvolvimento da pesquisa, educação e atenção em saúde na América Latina e Caribe, colocando ao alcance da comunidade de profissionais da saúde informação científico-técnica produzida nacional e internacionalmente. Este Sistema Regional é composto pelos sistemas nacionais (abrange 37 países na América Latina e Caribe) que são produto da integração dos

centros cooperantes (bibliotecas e centros de documentação da área da saúde). A Base de Dados LILACS (literatura Latino-Americana e do Caribe em Ciências da Saúde) é produzida de forma cooperativa através do Sistema Latino-Americano e do Caribe de Informações em Ciências da Saúde, e é coordenada pela BIREME. Nesta Base de Dados se encontra literatura científico-técnica em saúde, produzida por autores latino-americanos e do Caribe, publicada a partir de 1982. A LILACS tem como objetivos principais o controle bibliográfico e a disseminação da literatura científico-técnica Latino-Americana e do Caribe na área da saúde.

O acervo da Base LILACS é composto de teses, livros, capítulos de livros, anais de congressos ou conferências, relatórios técnico-científicos, artigos de revistas, etc., relacionados à área da Saúde. Os centros cooperantes fazem a descrição bibliográfica e a indexação utilizando o LILDBI, um programa desenvolvido pela BIREME para facilitar o trabalho de descrição bibliográfica dentro do padrão LILACS. Este programa tem uma versão disponível via WEB a LILDBI-Web.

Outro componente importante é o DeCS (Descritores em Ciências da Saúde), vocabulário controlado, também criado pela BIREME, para indexação de recursos informacionais na área da saúde e para a pesquisa e recuperação de informações nas bases de dados LILACS, MEDLINE e outras. Este tesouro é utilizado na indexação dos artigos para a base de dados LILACS.

O DeCS foi criado a partir do MeSH com o objetivo de permitir o uso de terminologia comum para pesquisa em três idiomas. Portanto o DeCS é um vocabulário estruturado e trilingüe, tendo para um termo o seu correspondente em português, espanhol e inglês. Através desta metodologia pode-se indexar em qualquer uma das três línguas que a sua recuperação, através do assunto, poderá ser realizada por qualquer um dos três idiomas. Além dos termos originais do MeSH foram desenvolvidas as áreas específicas de Saúde Pública e Homeopatia. Atualmente, segundo a BIREME, o DeCS totaliza 26851 descritores, destes 3656 são da área de Saúde Pública e 1950 são da área de Homeopatia. O DeCS também participa do projeto UMLS – Unified Medical Language System da U.S. National Library of Medicine. Este projeto objetiva o desenvolvimento de terminologia única e rede semântica em saúde e o DeCS tem a responsabilidade da atualização e envio de termos em português e espanhol.

A base LILACS pode ser acessada via Internet através do site da Bireme (www.bireme.br) na Biblioteca Virtual em Saúde no *link* “Literatura científica”.

2.3 ACERVOS ONLINE FIOCRUZ

A base de dados Acervos Online Fiocruz foi desenvolvida através do sistema de automação de bibliotecas ALEPH (Automated Library Expandable Program). Ele é um dos produtos desenvolvidos pelo grupo Ex Libris, fornecedor mundial de soluções de software e serviços relativos a bibliotecas e centros de informação, com sede em Israel.

A primeira versão do ALEPH foi desenvolvida há cerca de 20 anos na Universidade Hebraica de Jerusalém. Atualmente, o Aleph é utilizado por 3 milhões de usuários em 700 instituições, distribuídas em 50 países e o sistema oferece interfaces em 20 idiomas que utilizam vários conjuntos de caracteres diferentes.

O Aleph busca gerenciar todos os aspectos de uma biblioteca e tem funções tanto para os usuários externos (pesquisadores, profissionais e público em geral) como para os usuários internos (bibliotecários, auxiliares e funcionários em geral). Para atender estas necessidades ele possui vários módulos que incluem Web e GUI OPAC (pesquisa a base), Circulação, Catalogação, Seriadados, Aquisição, EEB (Empréstimo Entre Bibliotecas) e DSI (Disseminação Seletiva da Informação). O Aleph se constitui em um sistema aberto, com interface transparente. Utiliza arquitetura cliente-servidor com múltiplas camadas sobre um banco de dados Oracle. Também foi construído com suporte API's (application program interfaces) usando técnicas de programação orientada a objetos (OOP). Esta estrutura permite ao cliente expandir o sistema Aleph500 sem custos de modificação do código fonte. Os recursos de informação podem ser registrados e compartilhados através do formato **MARC**. Também disponibiliza buscas em bibliotecas de todo o mundo, através de uma função de busca múltipla baseada no protocolo **Z39.50**.

A base ACERVOS ONLINE FIOCRUZ é a base de dados desenvolvida pela Biblioteca do Instituto Fernandes Figueira, Biblioteca de Manguinhos e Biblioteca da Escola Nacional de Saúde Pública Sérgio Arouca. Estas três bibliotecas fazem parte da Unidade do Centro de Informação Científica e Tecnológica da Fiocruz. Nesta base de dados estão registrados os acervos das três bibliotecas, utilizando o sistema Aleph.

Para a entrada de dados na base Acervos Online são utilizados o formato MARC e o vocabulário controlado DeCS.

CAPÍTULO III

QUADRO CONCEITUAL

“O grande problema é, portanto, encontrar o difícil caminho da articulação entre as ciências que têm, cada uma, sua linguagem própria e conceitos fundamentais que não podem passar de uma linguagem à outra”

Edgar Morin

3 QUADRO CONCEITUAL

Neste capítulo, desenvolveremos um quadro conceitual, onde possam ser identificados os principais conceitos utilizados neste trabalho. Para isto são apresentadas algumas definições de interoperabilidade, utilizadas por especialistas e governos. A seguir apresentaremos os principais autores que formam este quadro conceitual e suas contribuições, posteriormente descrevemos as áreas do conhecimento as quais passaremos durante esta busca a interoperabilidade e finalmente, fazendo uma ligação entre interoperabilidade e cooperação, conheceremos as principais alternativas técnicas para a interoperabilidade e possíveis ferramentas a serem utilizadas.

3.1 DEFININDO INTEROPERABILIDADE

Interoperável - “Capaz de operar em conjunto”.^{*} - é um adjetivo cuja utilização vem se expandindo. Segundo Miller⁵ seu conceito é cada vez mais utilizado na gestão da informação. Em maio de 2004, o Governo Federal lançou a versão 0 do documento E-PING (Padrões de Interoperabilidade do Governo Eletrônico) que define um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia da Informação e Comunicação (TIC) no Governo Federal, estabelecendo as condições de interação com os demais poderes e esferas de governo e com a sociedade em geral. Segundo este documento “um governo moderno e integrado exige sistemas igualmente modernos e integrados, interoperáveis, trabalhando de forma íntegra, segura e coerente” (Brasil²⁰).

O Governo do Reino Unido define interoperabilidade como “Intercâmbio coerente de informações e serviços entre sistemas. Deve possibilitar a substituição de qualquer componente ou produto usado nos pontos de interligação por outro de especificação similar, sem comprometimento das funcionalidades do sistema”.

O Governo da Austrália considera mais a consistência das informações do que meramente a parte física, tecnológica. Ele define interoperabilidade como “Habilidade de transferir e utilizar informações de maneira uniforme e eficiente entre várias organizações e sistemas de informação”.

Segundo a International Standart Organization (ISO) Interoperabilidade é a “habilidade de dois ou mais sistemas (de computadores, meios de comunicação, redes, software e outros componentes de tecnologia da informação) de interagir e de

intercambiar dados de acordo com um método definido, de forma a obter os resultados esperados”.

Segundo Lichum Wang – Instituto Europeu de Informática – CORBA Workshops “Interoperabilidade define se dois ou mais componentes de um sistema, desenvolvidos com ferramenta diferentes, de fornecedores diferentes, podem ou não atuar em conjunto”.

O Concise Oxford Dictionary define interoperabilidade como “a capacidade de um sistema ou produto trabalhar com outro sistema ou produto sem requerer esforço especial por parte do cliente”[†].

Segundo Arms²¹ o objetivo da interoperabilidade é desenvolver serviços coerentes para os usuários, a partir de recursos informacionais que são tecnicamente diferentes e gerenciados por diferentes organizações. Isto requer acordos de cooperação em três níveis: técnico, de conteúdo e organizacional. O nível técnico nos proporciona a interoperabilidade tecnológica, bem representada nas definições acima citadas. O nível de conteúdo nos remete a interoperabilidade semântica, onde a representação e organização do conhecimento são áreas chaves a serem estudadas. O nível organizacional se refere a interoperabilidade política, quando organizações se reúnem com o intuito de alcançar a interoperabilidade, implementando padrões e tecnologias que cooperem com este objetivo. A interoperabilidade política viabiliza a gestão articulada entre as organizações e os sistemas de informação, ela “depende fundamentalmente da criação de organizações detentoras de canais e fóruns adequados, nos quais a discussão e o consenso possam se estabelecer e as decisões possam ser tomadas endossadas pelo grau de representatividade dessas organizações” (Marcondes²², p. 50).

Estas duas últimas facetadas da interoperabilidade muitas vezes são minimizadas ou não encontradas em definições. Neste estudo a busca de delinear modelos conceituais que visem a interoperabilidade perpassa estes três níveis.

Baseado nestas definições, no caso estudado, o objetivo é tornar as três bases de dados, duas delas muito semelhantes por terem o mesmo formato e serem operadas pelo mesmo sistema (LILACS e MEDLINE, operadas pela BIREME, usando o sistema IAH),

* Concise Oxford Dictionary, 9th Edition. *Apud* Miller⁵

† Concise Oxford Dictionary, 9th Edition. *Apud* Miller 5.

mas uma terceira bastante distinta (Acervos Online FIOCRUZ, mantida pela FIOCRUZ e operada pelo sistema Aleph), passíveis de serem consultadas simultaneamente, a partir de uma mesma interface Web.

3.2 PRINCIPAIS AUTORES

Para atingir o objetivo proposto nesta pesquisa foram utilizados alguns autores que fundamentam este trabalho do ponto de vista teórico.

Do autor Paul Miller⁵ foi utilizado o conceito de interoperabilidade e suas múltiplas dimensões. Miller aponta para várias facetas da interoperabilidade, dentre estas nos apropriamos para este estudo da interoperabilidade técnica, interoperabilidade política ou organizacional e interoperabilidade semântica.

De William Arms²¹ foi apropriado o modelo para análise das dificuldades técnicas para atingir a interoperabilidade, a partir de recursos informacionais, padrões e tecnologias diversas.

Com autora Márcia Rosetto^{23 24 25 26} pode-se entender as formas de aplicação do protocolo Z39.50 e seu uso na recuperação da informação e, também, recolher informações sobre o Conjunto de Metadados Dublin Core, como forma de viabilizar a reutilização de informações e cooperação entre sistemas de informação.

Com as autoras Regina Meyer Branski²⁷ e Beatriz Valadares Cedon²⁸ foi adquirida uma visão geral das principais categorias de ferramentas de busca na Internet, como os diretórios, os motores de busca e os metabuscadores e suas formas de aplicação, destacando-se para esta pesquisa os metabuscadores.

Carlos Henrique Marcondes²² nos proporcionou uma análise das novas formas de cooperação entre os sistemas de informação, destacando-se as formas de aplicação do protocolo OAI-PMH (Open Archives Initiative Protocol Metadata Harvesting).

Maurício Almeida e Marcelo Bax²⁹ proporcionaram o estado-da-arte no estudo de ontologias* e sua utilização como alternativa para a organização do conhecimento, trazendo definições para o termo discussões sobre seu significado e propostas de aplicação em diferentes domínios de conhecimento.

* O termo ontologia tem origem no grego “*ontos*”, ser, e “*logos*”, palavra. Na Organização do conhecimento é definido por Sowa (1999) como “catálogo de tipos de coisas”. Sowa apud Almeida²⁸.

Maria Luiza de Almeida Campos^{30, 31} nos proporcionou uma análise e discussões sobre o conceito de ontologia utilizado na área da Ciência da Informação e sua ligação com o conceito utilizado na Ciência da Computação.

Estes autores, portanto, fundamentam este trabalho com suas contribuições que utilizadas de forma objetiva nos auxiliam no caminho em busca da interoperabilidade entre bases de dados heterogêneas

3.3 ÁREAS DO CONHECIMENTO

Este trabalho se encontra na interseção entre as áreas da Ciência da Informação, Ciência da Computação e da Saúde, por estar investigando as condições conceituais, técnicas, de padronização e metodológicas, que permitam aos profissionais e pesquisadores do IFF consultar diferentes bases de dados disponíveis na Instituição de forma integrada, como se fosse uma só base. Nas duas primeiras áreas do conhecimento estaremos abordando os aspectos teórico-metodológicos para a interoperabilidade de sistemas de informações e na área da Saúde as bases de dados mais utilizadas pelos profissionais e estudantes do Instituto Fernandes Figueira.

3.3.1 Área da Ciência da Informação

Um dos aspectos importantes para a interoperabilidade de informações é a organização do conhecimento e representação da informação. Para que duas pessoas, duas bases de dados ou até duas instituições possam trocar informações de forma eficaz é necessário o entendimento dos códigos utilizados por ambos e que eles tenham o mesmo entendimento quanto ao significado destes códigos. Para isto é necessária a implementação de padrões e normas que possibilitem o entendimento entre eles, rompendo barreiras trazidas pela hiperespecialização e pela fragmentação da informação, desenvolvendo e reforçando trabalhos cooperativos. Cooperação é um termo muito utilizado e estudado pela Ciência da Informação. Serviços cooperativos, com desenvolvimento de padrões e normas para reutilização da informação é algo amplamente utilizado nesta área. Dentro da área da saúde da criança, do adolescente e da mulher existem outras sub-áreas que a compõem e que precisam ser observadas

como um conhecimento unitário, para que o tratamento das doenças desta área não reflita esta divisão.

Para alcançar a interoperabilidade é preciso identificar e representar as características comuns às diferentes Bases de Dados, possibilitando que elas possam ser consultadas simultaneamente. Na ciência da Informação são estudados os tesouros, que são vocabulários controlados para representação da informação de um documento. Para que haja interoperabilidade semântica, é necessário que haja acordos quanto à utilização destes descritores, isto é, que um termo tenha o mesmo significado que o utilizado em outra base, fale a mesma língua.

Nesta dissertação, a apropriação do conceito de interoperabilidade pela Ciência da Informação está diretamente ligada com o seu conceito utilizado na Ciência da computação.

3.3.2 Área da Ciência da Computação

Para se unir informações das diversas fontes na *Web* é necessário um olhar tecnológico. Atualmente, existem vários recursos que possibilitam, mesmo com algumas limitações, a interoperabilidade entre as informações. Neste projeto o olhar tecnológico será utilizado como instrumento que possibilite a interoperabilidade informacional.

A busca da interoperabilidade é um objetivo perseguido com afinco na área da ciência da computação, principalmente depois da criação da WEB, que se configura como um ambiente informacional global, com múltiplas plataformas, sistemas operacionais, heterogêneo, que vai desde ensino e cultura a entretenimento e comércio eletrônico. A própria Web é um exemplo de interoperabilidade, embora em uma escala limitada: o protocolo http permite a qualquer usuário navegar por páginas disponibilizadas em computadores os mais diferentes, com os mais diversos sistemas operacionais, criando um ambiente interoperável. Um organismo internacional, o W3C*, se encarrega de propor padrões de TI que tornem a Web cada vez mais interoperável. Mais especificamente, a ciência da computação juntamente com a ciência

* World Wide Web Consortium. Disponível em: <<http://www.w3c.org>>

da informação criou padrões, já estabelecidos, de interoperabilidade entre bibliotecas como o MARC, o Z.39.50, o OAI-PMH e o Conjunto de Metadados Dublin Core, entre outros, descritos posteriormente.

3.3.3 Área da Saúde

As fontes de informações em análise para a busca da interoperabilidade são as da área da saúde, mais especificamente as fontes que obtenham informações relevantes e que são mais consultadas pelos profissionais e pesquisadores do IFF/Fiocruz: as bases de dados MEDLINE, LILACS e ACERVOS ONLINE FIOCRUZ. Estas bases estão disponíveis aos profissionais que atuam no Instituto Fernandes Figueira para auxiliá-los na decisão clínica, lembrando que o Instituto trabalha com doenças de alta complexidade. Neste processo a área da saúde contribui para esta dissertação com o conceito de Medicina Baseada em Evidências, citado anteriormente, que preconiza a utilização sistemática das descobertas e pesquisas contemporâneas para auxílio na decisão clínica. Na área da Saúde o acesso ao documento e às informações em Saúde é fundamental para decisão clínica, principalmente em se tratando de doenças de alta complexidade como o campo em estudo. Desta forma, o acesso consolidado à informações se torna suporte importante para a Medicina Baseada em Evidências.

3.4 A COOPERAÇÃO/INTERAÇÃO: breve histórico

“Trabalhar em conjunto”, definição de Miller⁵ para interoperabilidade, faz parte do cotidiano da maioria das bibliotecas. Cooperação e compartilhamento de recursos foram fundamentais iniciativas para que as bibliotecas pudessem fazer frente à explosão informacional. Para viabilizar a cooperação e o compartilhamento de informações, padrões e metodologias foram desenvolvidos, como o Marc e o Z39.50.

As limitações dos sistemas de informações desde muito tempo fizeram com que as bibliotecas buscassem formas criativas de cooperação. Várias foram estas iniciativas entre bibliotecas (Suaiden³²) objetivando disponibilizar informação mais ampla aos seus usuários.

Em 1694 (Suaiden³²) a Biblioteca Nacional de Paris iniciou a prática de permuta entre bibliotecas, trocando suas duplicatas por livros ingleses e alemães. A troca de

publicações acadêmicas entre universidades alemães, em 1770 também trouxe benefícios mútuos entre estas instituições. O primeiro Centro de Permutas foi fundado em 1851 em Washington sob o título de International Exchange Service da Smithsonian Institution. A partir deste, vários outros Centros de Permuta foram criados convergindo com a Convenção de Bruxelas em 15 de março de 1886 onde foram acordados entre 8 países, incluindo o Brasil, algumas normas para a Permuta Internacional de documentos oficiais e publicações de caráter científicos e literários. Dentre as normas acertadas está a adoção de “modelos e formulários uniformes para as notas detalhadas do conteúdo...” (Suaiden³², p. 4)

No 2º Seminário Nacional de Comutação Bibliográfica, em 1994, Antônio Miranda³³ diz que:

A biblioteconomia contemporânea é, em certa medida, a consequência de esforços do passado no sentido de **integração**, ou como era mais comum denominá-la, pelos ideais de cooperação e de normalização de sistemas de informação. A origem de organizações profissionais, tais como a **Library Association** e a **American Library Association**, assim como a FID – Federação Internacional das Associações de Bibliotecários (no final do século passado e a segunda há aproximadamente 70 anos), está indiscutivelmente associada à idéia da integração, estratégica e tática de serviços de informação. Instrumentos como a catalogação cooperativa e a criação de catálogos coletivos – involucrando iniciativas de normalização e padronização de dados e de metodologias – visaram, primordialmente, ao desenvolvimento de sistemáticas de intercâmbio e cooperação. Tanto com relação à troca de informações técnicas para organização interna das bibliotecas, quanto – e sobretudo – para propiciar o acesso a documentos em bases cooperativas aos usuários das bibliotecas. A palavra de ordem daqueles tempos pioneiros era o empréstimo interbibliotecário, atividade que pretendia transformar os estoques informacionais das organizações bibliotecárias – **portanto a disponibilidade documentária** – em fontes para uso da comunidade em geral, ou seja, pela montagem de esquemas de acesso ao documento primário. (Miranda³³ p.7) (grifo do autor)

Segundo Marcondes²² com o advento da explosão informacional na metade do século XX, “a saída encontrada pelas bibliotecas foi a cooperação (...) Desde a invenção do computador na década de 50 as tecnologias de informação passaram a ser usadas pelas bibliotecas para prover acesso não só a documentos dos seus próprios acervos, mas também aos armazenados em acervos de outras bibliotecas”. (p. 42)

Estas estratégias de cooperação possibilitavam que as bibliotecas fossem capazes de operar em conjunto, isto é, fossem interoperáveis, ampliando os serviços aos usuários.

A Library of Congress (LC) começou a publicar seu catálogo impresso em 1905, visando permitir seu aproveitamento por outras bibliotecas. Na década de 1960

desenvolveu o projeto MARC (Machine Readable-Cataloging – Catálogo legível por máquina), desta vez com o objetivo de publicar seu catálogo, não mais em papel, mas em meio legível por computador.

3.5 ALTERNATIVAS DE INTEROPERABILIDADE

Quando se trabalha com bases de dados com estruturas e recursos técnicos semelhantes, se torna mais simples alcançar uma interoperabilidade do ponto de vista técnico. Porém, quando estas bases de dados são distintas, utilizando padrões e tecnologias diferentes, o caminho para a interoperabilidade se torna mais complexo.

William Arms, no contexto do desenvolvimento do "The site of Science", discute um modelo para análise das dificuldades técnicas para atingir a interoperabilidade. Segundo Arms²¹ uma biblioteca digital como "The site of Science" é composta de recursos informacionais diversos, produzidos, gerenciados e operados de forma descentralizada e utilizando padrões e tecnologias diversas. O desafio de se buscar a interoperabilidade entre estes diversos recursos é fazer com que, para um usuário, eles se apresentem, de forma coerente e uniforme, como um único recurso. O autor apresenta um modelo para análise das opções de interoperabilidade em que são comparados custos de implementação e de adesão a um determinado patamar de interoperabilidade X funcionalidades comuns dos diferentes recursos funcionando como se fossem um só.

Atualmente, várias são as tentativas de se conseguir a interoperabilidade entre as diversas fontes de informações disponíveis na WEB. Dentre estas destacamos o Protocolo Z39.50, o Protocolo OAI-PMH, os Metabuscadores, e o Conjunto de Metadados Dublin Core.

3.5.1 Protocolo Z39.50 e Protocolo OAI-PMH

Para se analisar interoperabilidade entre bases de dados é necessário entender um pouco como funciona a comunicação entre computadores.

Vários conceitos de redes são desenvolvidos, em diferentes disciplinas, utilizando metáforas que indicam inter-relações, interações e vínculos não hierarquizados. Em todas elas o conceito de rede está associado à idéia de “malha ou trama de ligações simétricas entre nós, que conotam fluxos regulares de informação e relações equidistantes de comunicação (...). Rede também aparece associada à imagem de uma árvore, na qual a informação parte de uma “raiz” e se difunde ou dissemina através de “ramos” ou ramais, isto é, um processo comunicativo que se ramifica hierarquicamente até um certo limite (se for “podado”) ou pode se desdobrar indefinidamente com a agregação de novos participantes. Esse é o modelo adotado pelos sistemas computacionais, em que o “servidor” é a raiz da informação e os usuários, os ramos “distribuídos” pela rede.” (Aguiar³⁴, p.2). Rede de computadores se refere à ligação entre dois ou mais computadores que trocam informações e que mantêm relações equidistantes de comunicação. As redes de computadores podem ser classificadas dependendo dos serviços oferecidos e da distância física entre os nós como locais, remotas e internacionais. As redes locais (*Local área network* - LAN) são aquelas que se encontram em um mesmo prédio ou sala, são caracterizadas por estarem na mesma área geográfica. As redes remotas (*Wide Área Network* – WAN) utilizam microondas ou satélite ou linha telefônica para agruparem grandes regiões geográficas. As redes internacionais (*Global Área Network* – GAN) unem sistemas entre países.

Um tipo de arquitetura muito usada para a comunicação entre computadores é a Arquitetura Cliente-Servidor. Nesta arquitetura se apresenta a relação entre dois computadores. O primeiro, chamado de cliente, solicita serviços e informações ao segundo, chamado servidor. No servidor normalmente ficam os sistemas mais pesados da rede, como os bancos de dados. A maioria das aplicações na Internet é baseada na arquitetura cliente-servidor. O browser é um tipo de cliente que solicita dados a um computador funcionando como servidor Web. Neste tipo de arquitetura o relacionamento entre os computadores é feito quando necessário, isto é, não é necessário uma conexão permanente entre o cliente e o servidor.

A Internet (Rede Mundial de Computadores) é a Rede das redes, é a Rede constituída por várias redes. A Internet é uma Rede distribuída, isto é, não existe um nó central que controle os outros nós, os serviços são distribuídos conferindo mesmo nível de importância para todos os nós da Rede. A comunicação entre os nós na Internet é feita por vários caminhos, existindo rotas alternativas. Se um nó encontra-se com problemas outro caminho é percorrido, esta característica da Internet é chamada de *redundância*. Para que dois computadores, uma pequena rede de computadores ou a Internet consiga trocar dados entre os seus nós, isto é, consiga ser **interoperável**, é necessário, dentre outros elementos, que eles utilizem a mesma língua ou um “tradutor” entre eles, usando a mesma gramática, isto é, sejam estabelecidos protocolos de comunicação.

O termo protocolo, em informática, significa “um conjunto de regras que ditam os métodos e procedimentos usados para transportar os dados” (Stair³⁵, p. 156). Esse conjunto de regras, entre outros atributos, padroniza a sequência, o formato e o tamanho e opera eventuais correções dos dados transmitidos. Um exemplo de protocolo largamente conhecido é o *Transmission Control Protocol/Internet Protocol* (TCP/IP) – Protocolo de Controle de Transmissão/Protocolo da Internet. Este protocolo foi desenvolvido pelo governo americano, nos anos 70, para unir suas agências de pesquisa de defesa. Anos mais tarde ele se tornou o protocolo básico para troca de dados na Internet. Dentre os diversos protocolos existentes é de especial interesse para esta pesquisa os protocolos Z39.50 e o OAI-PMH, que serão descritos a seguir.

3.5.1.1 Protocolo Z39.50

O Z39.50, formalmente definido como *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification* é “um protocolo de comunicação entre computadores desenhados para permitir pesquisa e recuperação de informação – documentos com textos completos, dados bibliográficos, imagens, multimeios – em redes de computadores distribuídos” (Rosetto²⁴, p. 2)

A partir da criação de um protocolo experimental desenvolvido como parte do projeto “Linked System Project” da Library of Congress (LC) (1970-1980), a National

Information Standards Organization (NISO), visando o desenvolvimento de um protocolo que viabilizasse a troca de informações entre sistemas de computadores, nomeou, em 1979, um comitê específico para criação desta ferramenta. A 1ª versão deste protocolo data de 1988. Em 1989 a administração do Z39.50 passou para a LC, mais especificamente para a “Network Development and MARC Standards Office” da Library of Congress, tornando-se também responsável pela coordenação técnica do desenvolvimento do protocolo (Barbosa³⁶). Em 1990 se estabeleceu um Grupo de Implementadores Z39.50 (ZIG), formado por fabricantes de software, consultores, vendedores, universitários, entre outros interessados na continuidade do desenvolvimento do protocolo. Em 1991 duas normas complementares são aprovadas: Definição para serviços e aplicação de pesquisa e recuperação (ISO10162); e Especificação de protocolo para pesquisa e recuperação (ISO10161-1). No ano seguinte estas normas são incorporadas à 2ª versão do protocolo Z39.50, lançado em 1992. Em 1995 é lançada a 3ª versão do Z39.50 da Ansi(American National Standarts Institute)/Niso. Representando o consenso entre ZIG e LC (Barbosa³⁶).

O protocolo Z39.50 faz com que o usuário consultando através do programa-cliente rodando em um PC, tenha a impressão que os diversos catálogos que estão sendo consultados, gerenciados por diferentes programas-servidores em diferentes bibliotecas, apareçam como um único catálogo. O protocolo Z39.50 permite a consulta unificada a catálogos distribuídos, através de uma rede e gerenciados descentralizadamente (Marcondes³⁷). Este protocolo pode ser implementado em qualquer plataforma, permitindo a interoperabilidade entre “diferentes sistemas de computação com diferentes sistemas operacionais, equipamentos, formas de pesquisa, sistemas de gerenciamento de bases de dados(...). Este sistema é extremamente vantajoso para bibliotecas que querem uma interface única para o usuário, a fim de realizar pesquisas no catálogo on-line local e em bases de dados referenciais e remotas.” (Rosetto²⁴, p. 2-3)

Para a interoperabilidade entre recursos informacionais utilizando o protocolo Z39.50 é necessária a utilização deste protocolo em todas as máquinas diretamente envolvidas (figura A). O Z39.50 funciona com arquitetura cliente-servidor, enquanto as máquinas servidoras precisam utilizar um programa **servidor Z39.50**, a(s) máquina(s) cliente(s) precisam utilizar um programa **cliente Z39.50**. Esta utilização mútua é um pré-requisito para o funcionamento deste protocolo de sistema distribuído. Cabe

ressaltar que o protocolo Z39.50 executa busca imediata as bases de dados, “... a interface de busca distribui a consulta (broadcast search) a diferentes *sites*(...)identificados pela interface como capazes de fornecer respostas satisfatórias, e os resultados são consolidados e integrados.(Marcondes⁹, p. 27)

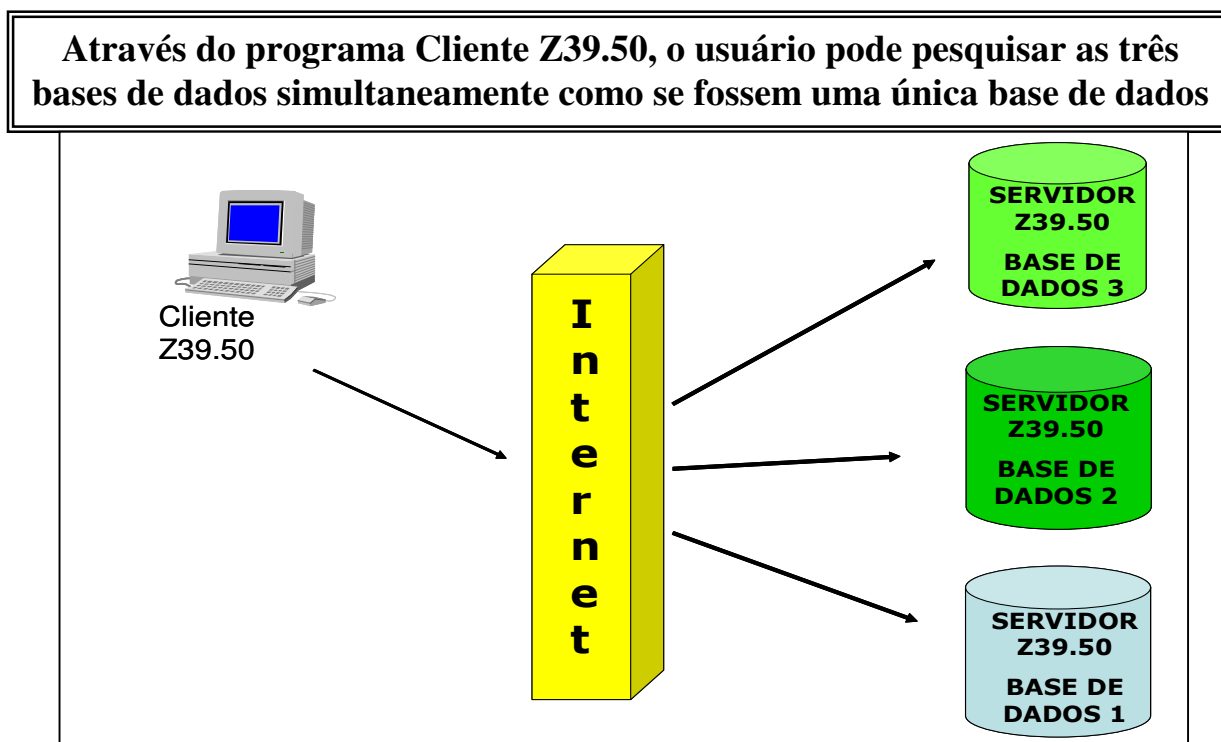


Figura A - interoperabilidade entre recursos informacionais utilizando o protocolo Z39.50

Uma evolução do Z39.50 é o SRW - Search Retrieval Webservice (<http://www.loc.gov/z3950/agency/zingsrw>) - um protocolo para a realização de consultas a várias bases de dados disponíveis via Internet. Este protocolo é conhecido também como Z39.50 "nova geração", na medida que se propõe a ser uma evolução deste protocolo, objetivando incorporar novas facilidades tecnológicas não existentes por ocasião do surgimento do Z39.50 no final da década de 80, como o protocolo SOAP* e a linguagem XML[†]. O SRW se propõe a ser um protocolo menos complexo e fácil de implementar. Baseia-se também na arquitetura cliente-servidor, onde um programa cliente faz solicitações de pesquisa a bases de dados disponíveis na Web,

* Simple Object Access Protocol

† Extended Markup Language

hospedadas em servidores compatíveis com este protocolo; o servidor responde a estas solicitações de pesquisa do cliente, enviando registros das suas bases de dados.

As solicitações de pesquisa e as respostas são codificadas usando XML. Entre os parâmetros da solicitação de pesquisa, um é especial, a consulta, formulada através de uma linguagem padronizada, a CQL - Common Query Language, que utiliza o conjunto de campos Dublin Core para especificar índices da base de dados.

3.5.1.2 Protocolo OAI-PMH

Em 1991, o físico Paul Ginsparg criou o ArXiv, primeiro arquivo eletrônico de preprints, os eprints, em Los Alamos National Laboratory (Van de Sompel³⁸). A partir desta iniciativa, “pesquisadores passaram a criar arquivos eletrônicos de preprints e posprints como alternativa para publicação direta de seus trabalhos em texto completo.” (Marcondes²², p. 43) Estes arquivos foram denominados “open archives”.

Desde então as comunidades internacionais vem criando repositórios de eprints. Estes repositórios tornaram-se uma alternativa de publicação para muitos pesquisadores que enfrentavam dificuldades para publicar em periódicos científicos impressos, controlados por grandes editores internacionais.

No open archives o autor do documento ao submeter seu trabalho, ele também o descreve através de um formulário de catalogação. Desta catalogação são retirados os metadados que representarão o documento e possibilitarão recuperá-lo.

Desde 1991, estes repositórios de eprints foram crescendo de forma considerável, principalmente pela visibilidade que um artigo em texto completo na Web adquire. “O maior retorno que a comunidade acadêmica almeja, publicando os resultados de suas pesquisas, é que estes possam servir de base a outras pesquisas(...) A citação é a medida clássica do prestígio e do valor de uma contribuição para a ciência em geral.(...) Estudos recentes confirmam que as publicações eletrônicas são muito mais citadas que as publicações em papel” (Marcondes²², p. 45). Segundo Lawrence³⁹ a média do número de citações de artigos publicados apenas em papel é 2,74, enquanto os artigos online tem uma média de citações de 7,03, com mais de 157% de citações.

Diante do crescimento destes repositórios, a comunidade científica internacional, em 1999, adotou “um conjunto de especificações técnicas e princípios organizacionais

bastante simples, porém potencialmente poderosos e de grande alcance, no objetivo de integração desses arquivos” (Marcondes²², p.46) . Esta iniciativa é conhecida como OAI – Open Archives Initiative ⁴⁰- que objetiva apoiar o desenvolvimento de publicação direta na Rede e tornar os vários repositórios de eprints interoperáveis.

No âmbito desta iniciativa foi criado o Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), um protocolo que oferece uma interoperabilidade não imediata entre informações digitais disponibilizadas através de formatos metadados. Segundo Marcondes²² “a possibilidade de coleta automática de metadados viabilizada pelo OAI PMH é a chave para uma nova prática de cooperação entre bibliotecas, desonerando os cooperantes do pesado ônus administrativo de gerenciar o envio de lotes, correções, atualizações, operações estas que permitem, num esquema de cooperação tão conhecido dos sistemas de informação brasileiros, manter um catálogo coletivo” (p. 52)

Ao contrário do Protocolo Z39.50, o OAI PMH não executa busca imediata e simultânea por vários servidores que hospedam catálogos de bibliotecas, buscando uma “solução mais simples e menos onerosa” (Troll⁴¹, p. 9). Dentre outras ferramentas utilizadas está o conjunto de metadados Dublin core (item 3.5.3). Ele define a troca de informações entre provedores de dados (*data providers*) e provedores de serviços (*service providers*). Um programa robô provedor de serviços coleta automaticamente metadados de diversos provedores de dados e os armazena em uma base de dados centralizada, onde são feitas as buscas. O OAI PMH estabelece o Dublin Core Metadata Element Set como “conjunto mínimo de metadados a ser suportado pelos provedores de dados em resposta a uma solicitação de um provedor de serviços; no entanto, o provedor de serviços pode, a seu critério, oferecer outros formatos de metadados, mais amplos e complexos, como o MARC.” (Marcondes²², p.48).

O provedor de dados pode utilizar até seis “verbos” previstos no OAI para coletar metadados de um documento do provedor de dados, são eles: *Identify* – recupera dados administrativos do provedor de dados; *ListSets* – lista as classificações utilizadas para organizar o documento no provedor de dados; *ListMetadataFormats* – lista os formatos de metadados utilizados para apresentar os documentos armazenados no provedor de dados; *ListIdentifiers* – lista de identificadores de registros armazenados no provedor de dados; *ListRecords* – Lista os metadados dos registros armazenados no

provedor de dados segundo um formato de metadados especificado: todos que pertencem a um *set* ou todos a partir de uma data; e *GetRecords* – obtém os metadados dos registros armazenados segundo um formato de metadados, dando um identificador de registro. (Marcondes²², p. 48)

3.5.2 Metabuscadores

Problema semelhante ao enfocado neste trabalho é o de buscar informações entre os milhões de *sites* diferentes na Internet. Existem vários recursos na Internet com o objetivo de auxiliar a busca de informações. Dentre estes recursos existem os mecanismos de busca que se subdividem em duas ferramentas básicas: os diretórios (também chamados de catálogo) e os motores de busca (também chamados de índices).

A primeira solução proposta para auxiliar as buscas na WEB foram os diretórios e o primeiro deles foi lançado em 1992 sob o nome de World Wide Web Virtual Library (<http://www.vlib.org/>). Os diretórios “organizam os sites que compõem sua base de dados em categorias, as quais podem conter subcategorias” (Cédon²⁸, p. 39), isto é, são organizados através de cabeçalhos de assunto, utilizando um vocabulário controlado. Um dos diretórios mais conhecidos atualmente é o Yahoo (www.yahoo.com).

Os motores de busca (também chamados de índices), por sua vez, não organizam suas bases de forma hierárquica, nem utilizam vocabulário controlado. Eles permitem a busca por palavras-chave ou linguagem natural. A coleta de sites para comporem suas bases é feita automaticamente através de softwares chamados de robôs, aranhas (spiders), viajantes, rastreadores ou vermes. Estes programas são lançados na Internet regularmente pelo computador hospedeiro da ferramenta de busca para obter dados sobre possíveis documentos a serem incorporados na base de dados. Alguns motores de busca coletam e indexam apenas páginas HTML, porém é cada vez maior o número de motores que indexam outros formatos como arquivos PDF, ASCII, imagens gráficos e vídeos.

Hoje as ferramentas de busca na Web, como Google, AltaVista, etc, são essenciais para fazer frente à explosão informacional trazida pela Web.

Destaca-se, para esta pesquisa, os chamados “metabuscadores”, programas ou “sites” como Metasearch, Metafind, Metacrawler, etc, que aceitam uma consulta e a submetem a vários mecanismos de busca simultaneamente, recebem os resultados, os analisam, alguns eliminam duplicatas, integram estes resultados e os mostram aos usuários. Desta maneira, os metabuscadores promovem a interoperabilidade entre vários mecanismos de busca.

Entretanto, existem páginas não indexadas pela maioria dos motores de busca. Esta parte da Internet, conhecida como Web invisível, abrange entre outros, os frames, image-maps e as páginas dinâmicas. As páginas dinâmicas, “geralmente são formadas de informações contidas nas bases de dados e são montadas no momento em que o usuário clica em um link, ou seja, as páginas são criadas no ato da busca”. (Cedón²⁸, p. 43) Os melhores mecanismos de busca, segundo relatório do NEC Research Institute⁴² não cobrem mais de 30% de todas as páginas Web. O que significa que cerca de 70% das informações na Web se encontra inacessível por estes programas. Neste universo encontram-se as Bases de Dados, que funcionam como páginas dinâmicas.

O conteúdo das fontes de informações, como bases de dados, está “escondido” atrás das interfaces de busca e estas interfaces são diferentes uma da outra. Gravano⁴³ diz que os algoritmos utilizados pelos motores de busca de cada fonte para fazer a pesquisa e a posição dos resultados desta pesquisa são incompatíveis entre as fontes. Por isso, construir metabuscadores é uma tarefa difícil devido a diferença e incompatibilidade entre os motores, dificultando a interoperabilidade.

Uma equipe da Universidade da Califórnia desenvolveu um projeto para construção de um protocolo para metabuscadores, denominado STARTS (Stanford Protocol Proposal for Internet Retrieval and Search). A idéia é construir um protocolo que viabilize a comunicação e a troca de dados entre várias fontes de informação, tornando-as interoperáveis. Na concepção deste projeto Gravano⁴³ afirma que para que haja interoperabilidade entre fontes de informações diferentes é necessário que o metabuscador, assim que receber uma questão a ser submetida à pesquisa, realize três tarefas para proporcionar uma interface única para várias fontes de informações diferentes: escolher as melhores fontes para interpretar a pergunta; interpretar a questão para cada fonte, conforme sua linguagem; e, consolidar os resultados destas fontes, apresentando-o de forma unificada.

3.5.3 Conjunto de Metadados Dublin Core

Com a explosão de documentos eletrônicos tornou-se imprescindível a criação de padrões para a descrição exata dos recursos de informação. Frente a esta necessidade, vários estudos sobre metadados tem sido realizados. O termo “meta” é um prefixo de auto-referência, então metadados, traduzindo literalmente, significa dado sobre o dado. Porém o termo metadados tem várias definições como, informação sobre informação, catalogação do dado ou descrição do recurso eletrônico. Segundo Rosetto²³ “metadados são um conjunto de dados-atributos, devidamente estruturados e codificados, com base em padrões internacionais, para representar informações de um recurso informacional em meio digital ou não-digital, contendo uma série de características e objetivos”. (p.9).

Rosetto relaciona 6 (seis) objetivos, são eles:

- 1- Localizar, identificar e recuperar dados de um recurso informacional.
- 2- Propiciar controles de ordem gerencial e administrativo permitindo conexões e remissões (links) para pontos internos e externos.
- 3- Possibilitar a interoperabilidade entre sistemas de informação, dentro de padrões.
- 4- Informar sobre as condições de acesso e uso da informação.
- 5- Ser legível tanto pelo homem como pela máquina.
- 6- Possibilitar a elaboração de índices

Os formatos metadados se configuram como “sumário de informações sobre a forma e o conteúdo de um recurso eletrônico, ou não (...) seu propósito primário é descrever, identificar e definir um recurso com o objetivo de modelar e filtrar o acesso, termos e condições para o uso, autenticação e avaliação, preservação e interoperabilidade.” (Rosetto²³, p. 4).

Dentre os vários formatos metadados, destaca-se nesta investigação o *Dublin Core* (DC)⁴⁴, por ser um formato menos estruturado e mais flexível, possibilitando a sua adaptação a particularidades de cada instituição/usuário. O *Dublin Core Metadata*

Element Set (Conjunto de elementos de metadados *Dublin Core*) é um conjunto de 15 elementos de descrição de um documento eletrônico e funciona como um mínimo múltiplo comum para troca de informações entre bases de dados distintas, possibilitando a interoperabilidade informacional entre elas. Esses elementos foram desenvolvidos com o intuito de facilitar a descrição dos recursos eletrônicos. Os elementos são:

- Título – O nome dado ao recurso
- Creator – A entidade responsável em primeira instância pela existência do recurso.
- Assunto – Tópicos do conteúdo do recurso.
- Descrição – Uma descrição do conteúdo do recurso.
- Editor – Uma entidade responsável por tornar o recurso acessível.
- Outro contribuinte – Uma entidade responsável por qualquer contribuição para o conteúdo do recurso.
- Data – Uma data associada a um evento do ciclo de vida do recurso.
- Tipo – A natureza ou gênero do conteúdo do recurso.
- Formato – A manifestação física ou digital do recurso.
- Identificador – Uma referência não ambígua ao recurso, definida num determinado contexto.
- Fonte – Uma referência a um recurso de onde o presente recurso possa ter derivado.
- Língua – A língua do conteúdo intelectual do recurso.
- Relação – Uma referência a um recurso relacionado.
- Cobertura – A extensão ou alcance do recurso.
- Direitos – Informação de direitos sobre o recurso ou relativos ao mesmo.

Todos os elementos Dublin Core são definidos por um conjunto de dez atributos da norma ISO/IEC 11179, para descrição de elementos de dados. São eles: Nome, Identificador, Versão, Autoridade de Registro, Língua, Definição, Obrigação, Tipo dos Dados, Máxima Ocorrência e Comentário.

Os metadados do recurso podem estar inseridos no próprio documento descrito, através de utilização das linguagens HTML, XML (Extensible Markup Language) e XML/RDF (Resource Description Framework), ou separada do recurso, armazenados em uma base de dados e associados ao documento eletrônico que descrevem.

Atualmente o Dublin Core vem sendo utilizado como padrão em vários projetos brasileiros como o Prossiga e a Biblioteca Digital de Teses do IBICT.

Como já mencionado o conjunto de metadados Dublin Core foi o padrão estabelecido pelo OAI-PMH como conjunto mínimo de metadados a ser utilizado pelo provedor de dados em resposta a uma solicitação do provedor de serviços. Esta interação entre o padrão de metadados Dublin Core e o OAI-PMH pode ser exemplificada com uma iniciativa brasileira no âmbito do Portal SciELO.

O SciELO – Scientific Eletronic Library Online – é fruto de um convênio entre FAPESP, BIREME e Instituições Nacionais e Internacionais ligadas a comunicação científica. O objetivo do SciELO é ampliar a visibilidade das publicações científicas dos autores da América Latina e do Caribe, através da publicação eletrônica em texto completo. No início do projeto, em 1997, o SciELO contava com 10 periódicos brasileiros. O projeto piloto foi bem sucedido, agregando novos títulos de periódicos e ampliando sua operação para outros países. Atualmente o SciELO Brasil disponibiliza mais de 10.000 artigos em texto completo de 90 periódicos científicos.

O SciELO utiliza o formato LILACS como formato padrão. Em 2002, se iniciou o projeto SciELO-Open Archives. Apesar do formato padrão do SciELO ser o LILACS, até o momento na implementação do SciELO-Open Archives só foi possível a utilização do formato metadados OAI-Dublin Core (Marcondes⁴⁵).

3.5.4 MetalAH

Com o intuito de otimizar a recuperação da informação em bases de dados ISIS, a BIREME desenvolveu uma interface de pesquisa, o Interface for Access Health information – IAH. Esta interface é escrita em IsisScript para ser executada com o WWWISIS XML IsisScript Server (WXIS), também desenvolvido pela BIREME. O WXIS, componente ativo da interface, é um servidor de acesso multiusuário a bases de dados Isis, e pode ser compilado para diversos ambientes operacionais como Windows, Unix e Linux. A Interface IAH precisa ser instalada junto a um servidor Web. A maioria das bases que integram o Sistema Latino Americano e do Caribe e são operadas pela BIREME utilizam a interface IAH, inclusive a versão da base MEDLINE operada pela BIREME. Os campos disponíveis através desta interface nas bases MEDLINE e LILACS foram descritos no capítulo sobre análise das bases. A forma de apresentação desta interface também pode ser visualizada no referido capítulo.

A partir desta interface foi criado um metabuscador chamado de metaIAH. Como dito anteriormente, o termo meta significa auto-referência, pode-se assim traduzir, metaIAH, como uma Interface de acesso às Interfaces de acesso à Informação em Saúde. Possibilitando o acesso simultâneo a várias bases de dados em saúde.

O metaIAH submete a pesquisa feita em sua interface a bases de dados que utilizam a interface IAH. Nesta distribuição, o metaIAH não submete a pesquisa de acordo com as especificidades de cada base (campos, índices). A busca através do metaIAH é feita basicamente no formulário livre em todas as bases, exceto nas bases LILACS e MEDLINE, nas quais foi montado uma página de pesquisa avançada que relaciona os campos destas bases. A pesquisa não é submetida a todos os índices das bases, pesquisando principalmente no campo *default* das bases de dados, que, na sua maioria, é o índice “Palavras do texto” (TW- Text Words). Este índice inclui título, autor e resumo. Os resultados da pesquisa são visualizados por bases de dados, não sendo consolidados os resultados, nem eliminadas possíveis duplicatas.

Esta ferramenta só está disponível para utilização dentro do âmbito das Bibliotecas Virtuais em Saúde. No site da Bireme, atualmente, o metaIAH é utilizado para busca em 14 bases de dados simultaneamente. O pesquisador/usuário submete a

sua pesquisa em uma única interface e visualiza respostas das 14 bases, na página de pesquisa. Além das bases MEDLINE e LILACS a busca é feita nas seguintes bases:

- ADOLEC – base de dados contendo citações relativas a Adolescência. O objetivo desta base é subsidiar a criação de uma sub-rede na região-latino-Americana, a qual se encarrega da coleta e processamento de informações pertinentes ao tema, atualizando a base constantemente. A criação desta sub-rede faz parte do projeto da OPAS, coordenado pelo Programa de Saúde Materno-Infantil.
- AdSaude – base de dados de literatura relativa à área de Administração de Serviços em Saúde. Esta base é de responsabilidade do Serviço de Biblioteca e Documentação da Faculdade de Saúde Pública da USP juntamente com outras Instituições da Sub-Rede AdSaude.
- BBO - base de dados de literatura nacional na área de saúde oral, de responsabilidade do Serviço de Documentação Odontológica da Faculdade de Odontologia da USP, desde 1986. Seu objetivo é preservar a memória nacional em odontologia, e para atingir sua missão, conta com acervos de livros, teses, folhetos, separatas e publicações periódicas, além de artigos de autores nacionais publicados em revistas estrangeiras e não especializadas.
- BDENF - base de dados especializada na área de Enfermagem, criada em 1988, com o objetivo de facilitar o acesso e a difusão das publicações da área. A manutenção da base é feita pela Sub-Rede Brasileira de Informação em Enfermagem - SURENF, da qual a Biblioteca J. Baeta Vianna do Campus da Saúde/UFMG é coordenadora. Esta base se desenvolveu através do patrocínio do PRODEN - Programa de Desenvolvimento da Escola de Enfermagem / UFMG e convênio estabelecido com o Centro Latino Americano e do Caribe de Informação em Ciências da Saúde – BIREME. Seu acervo abrange referências bibliográficas e resumos de documentos de livros, teses, manuais, folhetos, congressos, separatas e publicações periódicas, gerados no Brasil ou, escritos por autores brasileiros e publicados em outros países.
- HISA - base de dados referentes a estudos produzidos e/ou publicados, a partir de 1960, que tenham a Saúde Pública como tema central, os eventos e processos históricos como objetos das diversas disciplinas das Ciências Humanas, e a América Latina e o Caribe como dimensão geográfica. Desenvolvida pela Casa de Oswaldo Cruz (Fiocruz), conta com o apoio da Organização Panamericana de Saúde (OPAS)

e a orientação técnica do Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde -BIREME.

- LEYES - base de dados coordenada pelo Sistema de Documentação sobre Legislação Básica do Setor de Saúde na América Latina e Caribe - Desarrollo de Políticas de Salud (HSP/OPAS - Washington, USA). Seu acervo abrange referências bibliográficas da legislação básica em saúde vigente em mais de trinta países da América Latina e do Caribe.
- MEDCARIB - base de dados que reúne a literatura em Ciências da Saúde gerada principalmente nos países do Caribe de língua inglesa. Produzida pela Rede Caribenha e coordenada pelo Centro Coordenador da Rede MedCarib The Medical Library, University of the West Indies, Mona, Kingston - Jamaica. Seu acervo, com referência desde o século XVIII é composto de livros, capítulo de livros, teses, relatórios técnicos, anais de congressos e artigos de revistas.
- REPIDISCA - base de dados da literatura de Engenharia Sanitária e Ciências do Ambiente, coordenada pelo CEPIS (Centro Pan-americano de Engenharia Sanitária e Ciências do Ambiente localizado em Lima, Peru). A partir de 1994 incorporou os registros da base de dados ECO, sobre Ecologia Humana e Saúde.
- PAHO - base de dados de referências bibliográficas e resumos do acervo da Biblioteca da sede da Organização Pan-americana da Saúde em Washington, D.C., USA. Abrange a documentação sobre temas em Saúde indexada pela Biblioteca.
- WHOLIS é uma base de dados bibliográfica que contém publicações da sede da Organização Mundial da Saúde (OMS) e das Representações Regionais, artigos de periódicos, documentos técnicos e políticos, e publicações da OMS em colaboração com outros editores e organizações internacionais.
- DESASTRES - base de dados cujo acervo abrange referências bibliográficas resultantes de análises de publicações da OPAS ou outras agências das Nações Unidas, livros ou capítulos de livros, informes técnicos, apresentações de congressos, teses, planos de emergência, e artigos científicos extraídos de revistas especializadas. Esta base é produzida pelo Centro de Documentação de Desastres, do Programa de Preparativos para Situações de Emergência e Coordenação de Socorro para Casos de Desastres da Organização Pan-americana da Saúde (OPAS).

Face ao exposto, percebe-se, que através da ferramenta MetaIAH é possível tornar diversas bases interoperáveis desde que as mesmas utilizem a metodologia

LILACS. Das três bases em estudo, apenas as bases LILACS E MEDLINE possuem as configurações necessárias para a utilização do MetaIAH.

3.6 ONTOLOGIAS

Como vimos anteriormente (item 3.2), o termo ontologias tem origem no grego *ontos*, que significa “ser”, e *logos*, que traduzido é “palavra”. Este termo surgiu na transição da Idade Média para a Moderna, no âmbito da Filosofia, se referindo a “ciência do ser”. Na Ciência da Computação este termo é utilizado para representar a explicitação do significado das coisas no mundo, relevando a sua natureza. Guarino (1998) diz que “ontologia se refere a um artefato constituído por um vocabulário usado para descrever uma certa realidade, mais um conjunto de fatos explícitos e aceitos que dizem respeito ao sentido pretendido para as palavras do vocabulário. Este conjunto de fatos tem a forma da teoria da lógica de primeira ordem, onde as palavras do vocabulário aparecem como predicados unários ou binários” (Guarino apud Almeida²⁹, p. 9).

As ontologias possuem características e componentes básicos, porém não se apresentam sempre com a mesma estrutura. Segundo Gruber (1996) e Noy & Guinness (2001), os componentes básicos de uma ontologia são classes (organizadas em uma taxonomia), relações (representam o tipo de interação entre os conceitos de um domínio), axiomas (usados para modelar sentenças sempre verdadeiras) e instâncias (utilizadas para representar elementos específicos, ou seja, os próprios dados) (apud Almeida²⁹). “Uma ontologia é uma especificação de uma conceituação. É designada com o propósito de habilitar o compartilhamento e reuso de conhecimentos, de forma a criar ‘compromissos ontológicos’, ou definições necessárias à criação de um vocabulário comum” (Semantic Web⁴⁶)

A Inteligência Artificial (IA) se apropriou do termo ontologias, na IA as ontologias são guardadas em arquivos ou documentos onde são definidas “formalmente as relações entre termos e conceitos” (Souza⁴⁷, p.137). Trabalhar relações entre termos e conceitos é algo já consolidado na área de biblioteconomia, que se materializa nas

técnicas de construção de tesouros para vocabulários controlados. Porém, a construção de ontologias em Inteligência Artificial busca automatizar este relacionamento criando sistemas “inteligentes” entendidos por seres humanos e máquinas. Com este objetivo, várias são as tecnologias que estão sendo implantadas e analisadas para criação de ontologias, todas baseadas em XML.

XML (eXtensible Markup Language) é uma linguagem de marcação extensível criada em 1996 e recomendada formalmente pela W3C* (World Wide Web Consortium). Esta linguagem é considerada uma evolução do HTML (HyperText Markup Language – Linguagem de Marcação em Hipertexto). Tanto a linguagem HTML quanto a linguagem XML são derivadas do Padrão SGML (Standard Generalized Markup Language) uma linguagem utilizada para descrever outras linguagens. O padrão SGML “é baseado na idéia de que documentos contêm estrutura e outros elementos semânticos que podem ser descritos sem que se faça referência à forma com que estes elementos serão exibidos” (Souza⁴⁷, p. 134). Enquanto a linguagem HTML se preocupa apenas com o formato que o documento terá ao ser exibido na Web, o XML se preocupa com, além do formato de exibição, em descrever o conteúdo semântico e os significados contextuais. Por esta característica de valorizar a descrição dos documentos, este padrão tem sido muito utilizado para troca de dados na Web, porém para otimizar esta função é necessária a utilização de metadados baseados em padrões internacionais que possibilitem o entendimento universal do termo sem ocorrência de ambigüidade. Utilizando estas ferramentas como XML e o conjunto de metadados Dublin core é possível criar ontologias que contribuam para a interoperabilidade.

Na estrutura de uma base de dados estão inseridos, implicitamente, conhecimentos relativos a seus campos, semântica dos campos, possibilidades de conteúdo destes campos, pontos de acesso, formatos de exibição, conteúdos dos campos, etc. que precisariam ser explicitados, representados de forma comum, para poderem ser processados por uma eventual interface unificada e apresentados aos usuários. A Ciência da Informação junto com a Ciência da Computação abrange as ontologias, metodologias de representação do conhecimento, processáveis por máquinas e por seres humanos. (Almeida²⁹).

* W3C é responsável pela criação de padrões tecnológicos que regulam as empresas, instituições acadêmicas, profissionais, cientistas, dentre estes a Adobe, Microsoft, HP e Netscape.

Segundo Pacheco⁴⁸ para a integração de sistemas é necessária uma ontologia comum, isto é, que os sistemas falem a mesma língua, joguem o mesmo jogo, obedeçam as mesmas regras, com especificações explícitas de uma conceitualização. Esta formalização de conceitos em ciência e tecnologia permitiria que diversos aplicativos desta área compartilhassem um vocabulário comum.

Definindo uma Interface ela precisará ter uma descrição das características de cada base (índices, campos, formulários, endereço eletrônico, etc.) que podem torná-las interoperáveis. As ontologias, junto com as ferramentas descritas anteriormente, podem ser utilizadas para esta função, “guardando” as características de cada base.

CAPÍTULO IV

ANÁLISE DAS BASES DE DADOS

"É melhor tentar e falhar, que preocupar-se e ver a vida passar; é melhor tentar, ainda que em vão, que sentar-se fazendo nada até o final. Eu prefiro na chuva caminhar, que em dias tristes em casa me esconder. Prefiro ser feliz, embora louco, que em conformidade viver ..."

Martin Luther King

4 ANÁLISE DAS BASES DE DADOS

A análise dos procedimentos de descrição e o mapeamento das estruturas das bases de dados são fundamentais quando se quer torná-las interoperáveis. Com este intuito, analisaremos, neste capítulo, as bases de dados que são objeto desta pesquisa buscando suas semelhanças e divergências. Na primeira etapa desta análise faremos uma comparação entre as bases de dados referente aos pontos de acesso, isto é, quais campos podem ser utilizados como índices na pesquisa às bases de dados. Na segunda etapa faremos uma comparação entre as bases de dados referente aos campos que compõem o formato de exibição.

4.1 ÍNDICES DISPONÍVEIS NOS FORMULÁRIOS DE PESQUISA

Nos formulários de pesquisa são determinados quais pontos de acesso ou índices são passíveis de serem consultados. Geralmente estes formulários variam partindo de uma formulação para pesquisas mais simples que não necessitam cruzar vários índices até formulários mais complexos, normalmente utilizados por usuários mais experientes e que buscam uma informação mais precisa e específica, possibilitando cruzar vários índices.

4.1.1 Formulários LILACS e MEDLINE

Nas bases de dados LILACS e MEDLINE são disponibilizados 3 tipos de formulários para a pesquisa: Formulário Livre, Formulário Básico e Formulário Avançado.

4.1.1.1 Formulários LILACS

Na Base LILACS o formulário livre recupera dados buscando em até 5 índices, são eles: Palavras do título do artigo, Palavras do resumo, Nome de pessoas como assunto, Descritores de assunto e Nome de substâncias.

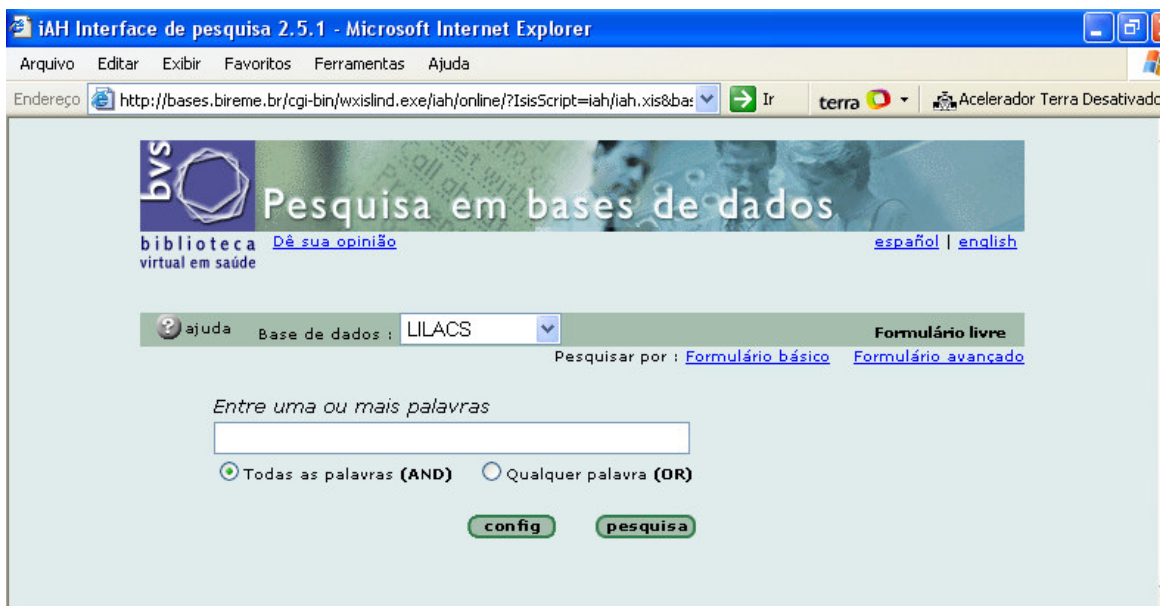


Figura B - Base LILACS - Formulário Livre

No Formulário Básico existe a possibilidade de cruzar até 3 índices entre os 16 disponíveis: Palavras, Descritor de assunto, Limites, Autor, Palavras do título, Revista, Monografia em série, Tipo de publicação, Idioma, Identificador único, País- Ano de publicação, Localização, Mês de entrada e Suporte eletrônico.

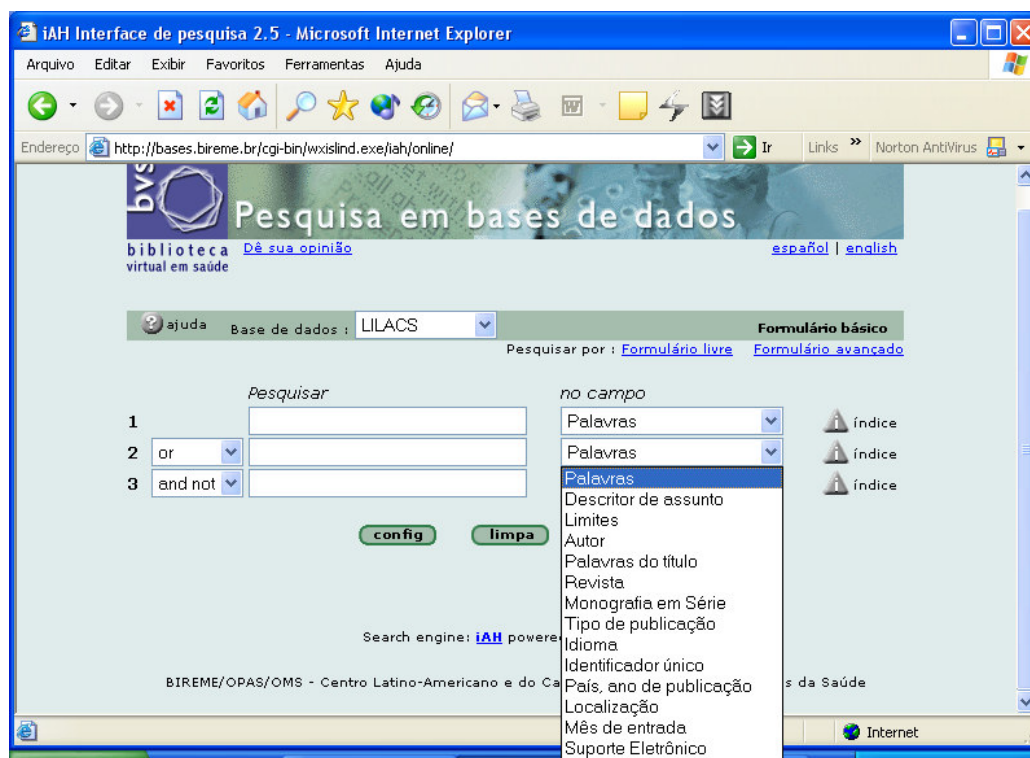


Figura C - Base LILACS - Formulário Básico

O Formulário Avançado disponibiliza a busca em até 3 índices entre os 25 índices disponíveis, são eles: Palavras, Descritor de assunto, Limites, Autor, Palavras do título, Revista, Monografia em série, Tipo de publicação, Idioma, Identificador único, País-Ano de publicação, Localização, Mês de entrada e Suporte eletrônico, Autor normalizado, Autor institucional, Palavras do resumo, Tipo de literatura, Conferência, Nível bibliográfico, Instituição, Cidade e País.

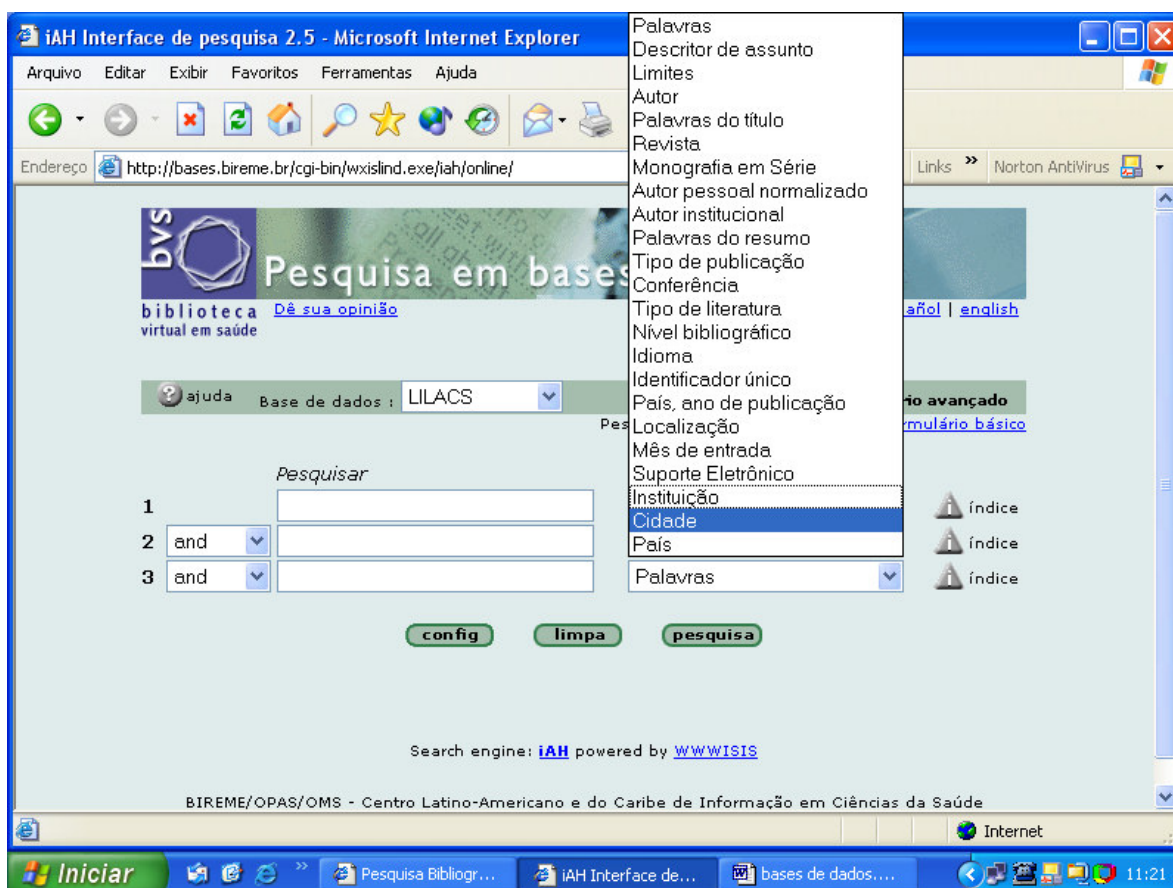


Figura D - Base LILACS - Formulário Avançado

O quadro a seguir relaciona todos os índices encontrados em cada formulário LILACS. Através deste quadro pode-se perceber as semelhanças e divergências entre os formulários.

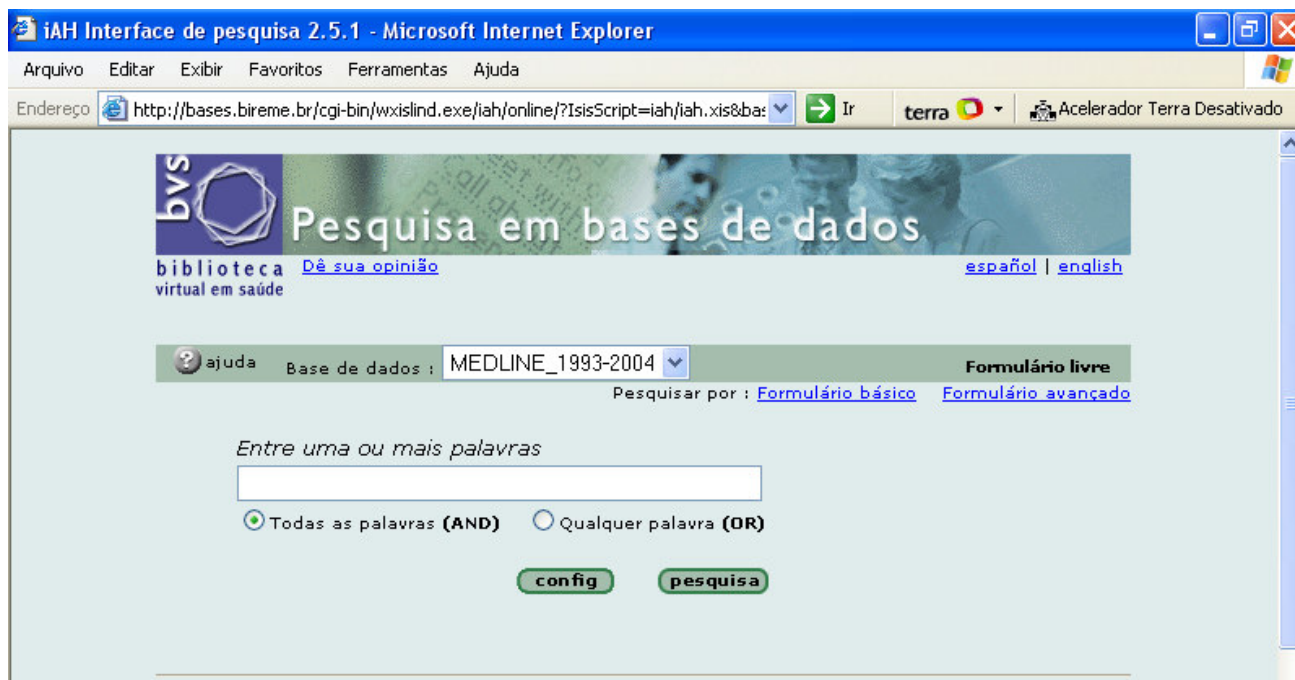
LILACS		
FORMULÁRIO LIVRE	FORMULÁRIO BÁSICO Pesquisa simultaneamente em até três campos entre os seguintes	FORMULÁRIO AVANÇADO Pesquisa simultaneamente em até três campos entre os seguintes
	Palavras	Palavras
Descritores de assunto	Descritor de assunto	Descritor de assunto
	Limites	Limites
	Autor	Autor
Palavras do título	Palavras do título	Palavras do título
	Revista	Revista
	Monografia em série	Monografia em série
	Tipo de publicação	Tipo de publicação
	Idioma	Idioma
	Identificador único	Identificador único
	País-Ano de publicação	País-Ano de publicação
	Localização	Localização
	Mês de entrada	Mês de entrada
	Localização	Localização
	Mês de entrada	Mês de entrada
	Suporte eletrônico	Suporte eletrônico
		Autor normalizado
		Autor institucional
Palavras do resumo		Palavras do resumo
Nome de pessoas como assunto		Tipo de literatura
Nome de substâncias		Conferência
		Nível bibliográfico
		Instituição
		Cidade
		País

Quadro A – Formulários LILACS

(URL: <http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&base=LILACS&lang=p>)

4.1.1.2 Formulários MEDLINE

Na Base de Dados MEDLINE o formulário Livre recupera dados em até 5 índices, são eles: Palavras do título do artigo, Palavras do resumo, Nome de substâncias, Nome de pessoas como assunto, e Descritores de assunto.



The screenshot displays the iAH Interface de pesquisa 2.5.1 in Microsoft Internet Explorer. The browser window shows the URL <http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&ba:> and the page title "Pesquisa em bases de dados". The page content includes a search form with a dropdown menu set to "MEDLINE_1993-2004", a search input field, and radio buttons for "Todas as palavras (AND)" and "Qualquer palavra (OR)". There are also buttons for "config" and "pesquisa".

Figura E - Base MEDLINE - Formulário Livre

O Formulário Básico disponibiliza 11 índices (Palavras, Descritores de assunto, Limites, Autor, Palavras do título, Revista, Tipo de publicação, Idioma, Identificador único, Ano de publicação, Texto completo SciELO) para recuperação de dados, dentre os quais podem ser escolhidos até 3 índices para cruzamento de dados.

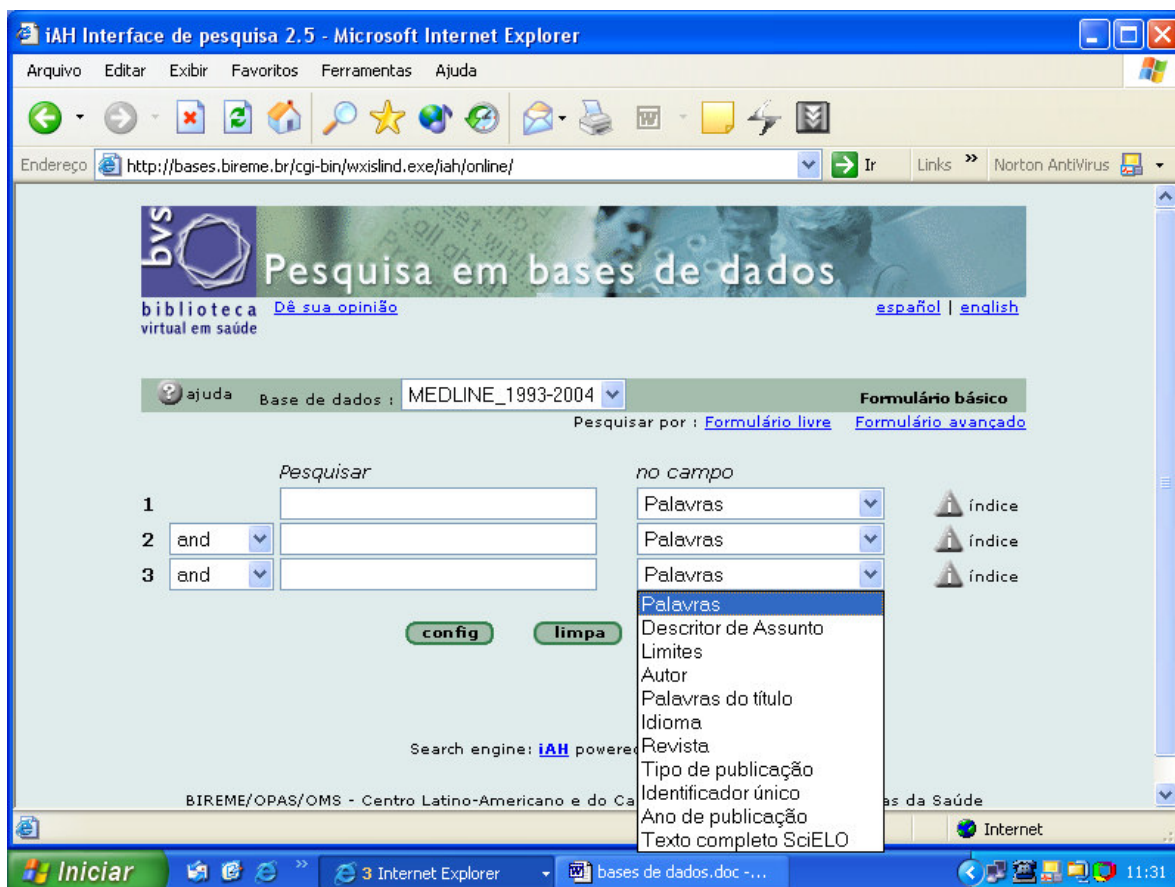


Figura F - Base MEDLINE - Formulário Básico

Já o Formulário avançado disponibiliza 18 índices para a recuperação de dados, dentre este podem ser escolhidos até 3 para cruzamento e recuperação de artigos. Os índices disponíveis neste formulário são: Palavras, Descritor de assunto, Limites, Autor, Palavras do título, Revista, Tipo de publicação, Idioma, Identificador único, Ano de publicação, Texto completo SciELO, País de publicação, Nome de substância, Símbolo de gene, Nome de pessoa como assunto, ISSN, Mês de entrada e Atualização por classe.

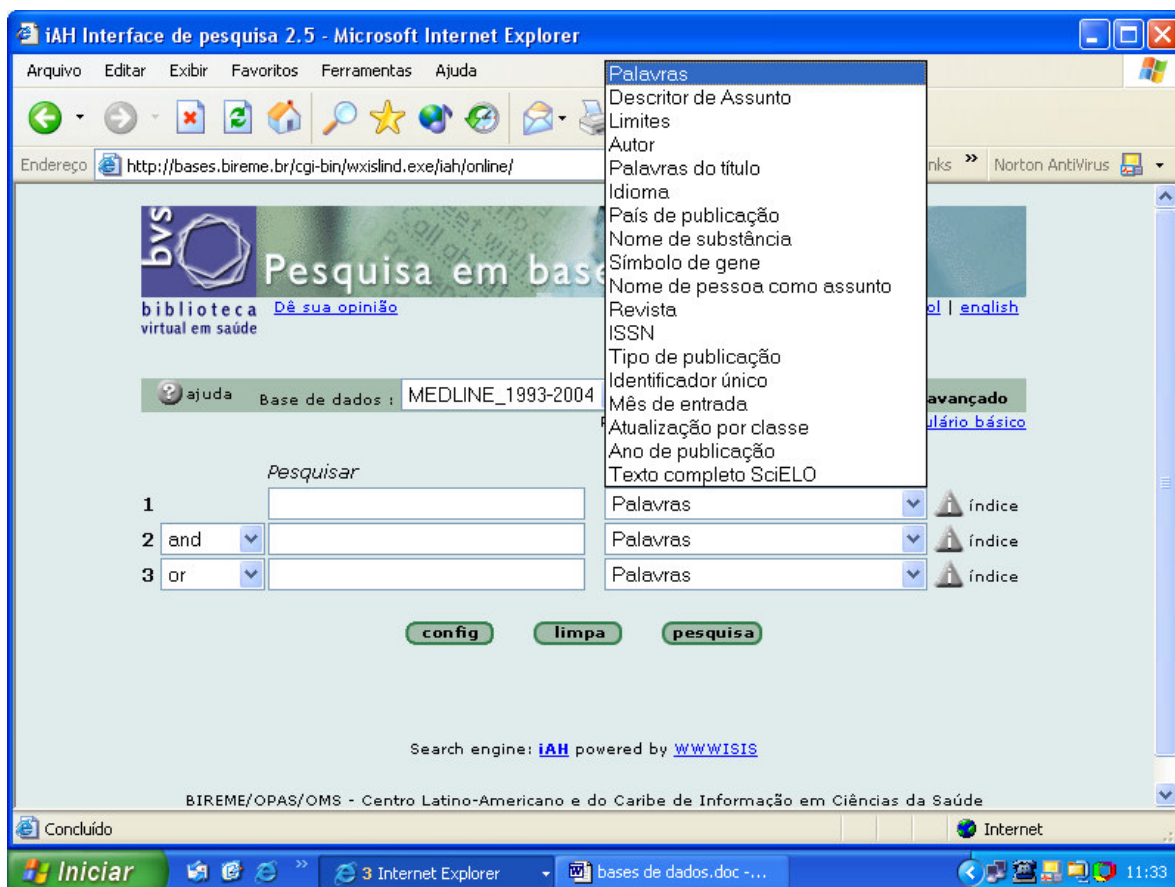


Figura G - Base MEDLINE - Formulário Avançado

O quadro a seguir apresenta os três formulários MEDLINE com seus diversos índices, para melhor compreensão de suas semelhanças e divergências.

MEDLINE		
FORMULÁRIO LIVRE	FORMULÁRIO BÁSICO Pesquisa simultaneamente em até três campos entre os seguintes:	FORMULÁRIO AVANÇADO Pesquisa simultaneamente em até três campos entre os seguintes:
palavra do resumo	Palavras	Palavras
Descritor de assunto	Descritor de assunto	Descritor de assunto
	Limites	Limites
	Autor	Autor
palavras do título	Palavras do título	Palavras do título
	Revista	Revista
	Tipo de publicação	Tipo de publicação
	Idioma	Idioma
	Identificador único	Identificador único
	Ano de publicação	Ano de publicação
	Texto completo SciELO	Texto completo SciELO
		País de publicação
Nome de substância		Nome de substância
		Símbolo de gene
Nome de pessoa como assunto		Nome de pessoa como assunto
		ISSN
		Mês de entrada
		Atualização por classe

Quadro B – Formulários MEDLINE

(URL: http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&base=MEDLINE_1993-2004&lang=p)

4.1.2 FORMULÁRIOS ACERVOS ONLINE

Na Base de Dados Acervos Online são disponibilizados 4 formulários para a pesquisa aos índices com as seguintes nomenclaturas: Pesquisa Simples, Pesquisa Multi-campo, Pesquisa Multi-base e Pesquisa Avançada.

No formulário Pesquisa Simples são disponibilizados 16 índices: Todos os campos, Título, Autor, Assunto, Sumário, ISSN, ISBN, Número de sistema, Código de barras, Número de controle, URL/856, Número de chamada, Ano de Publicação, Número de item GPO e Notas. Dentre estes índices o pesquisador precisa escolher um para submeter sua pesquisa.

Figura H - Base de Dados Acervos Online – Formulário Pesquisa Simples

No Formulário Multi-campo são disponibilizados 6 índices para a pesquisa, podendo ser cruzados todos estes índices, são eles: Assunto, Autor, Título (expressão), Título (palavras), Ano e Editor.

IFF - Aleph menu principal - Microsoft Internet Explorer

Arquivo Editar Exibir Favoritos Ferramentas Ajuda

Endereço http://157.86.8.8/ALEPH/EDCRUHM8Y1RIQUAN7CHRH9CQ59USAKAY91117YPSVK335BYPL-01652/file/start-0

Centro de Informação Científica e Tecnológica

Identificação Encerrar Sessão Base de Dados Preferências Ajuda

Percorrer Lista Pesquisar por Palavras Resultados Pesquisas Anteriores Usuário EEB Meus Docs

Pesquisa Multi-campo

Preencha os campos conforme desejado. Usando mais de uma linha, a pesquisa será mais específica. Se você escolhe Sim para "Palavras adjacentes" e o termo pesquisado for, por exemplo, *linguagem de programação*, o sistema recupera registros com a palavra *linguagem* PRÓXIMA à palavra *programação*. A adjacência não é executada se o conjunto contiver mais do que xxx registros.

Assunto

Autor

Título (expressão)

Título (palavras)

Ano

Editor

Palavras adjacentes? Não Sim

Buscar ou Limpar

Figura I - Base de Dados Acervos Online – Formulário Pesquisa Multi-campo

No formulário Pesquisa Multi-base são disponibilizados 4 índices para recuperação de dados: Todos os campos, Assunto, Autor e Título. Dentre estes índices podem ser selecionados até 3 índices para pesquisa simultânea e ainda selecionar sub-bases que integram a Base Acervos Online como: Acervos Online Fiocruz, Catálogo de Autoridades, Tesouros de Manguinhos e Periódicos de Manguinhos.

FIO01 - Aleph menu principal - Microsoft Internet Explorer

Arquivo Editar Exibir Favoritos Ferramentas Ajuda

Endereço http://157.86.8.8/ALEPH/R2LF6LHFM2YMT7456G2YD1A9NPX2HQ7461JERSU18A6514VBQ-00336/file/start-0

Acervos On line

Identificação Encerrar Sessão Base de Dados Preferências Ajuda

Percorrer Lista Pesquisar por Palavras Resultados Pesquisas Anteriores Usuário EEB Meus Docs

Simple Multi-campo Multi-base Avançada CCL Voltar Ajuda

Pesquisa Multi-base

Campo para pesquisa Digitar palavra ou expressão

Todos os campos

Campo para pesquisa Digitar palavra ou expressão

Todos os campos

Campo para pesquisa Digitar palavra ou expressão

Todos os campos

Selecionar a Base de Dados:

Acervos Online Fiocruz (FIO01)

Catálogo de Autoridades (FIO10)

Tesouros de Manguinhos (FIO11)

Periódicos de Manguinhos (FIO02)

Buscar ou Limpar

Figura J - Base de Dados Acervos Online – Formulário Pesquisa Multi-base

O formulário Pesquisa Avançada funciona com os mesmos recursos do formulário Pesquisa Multi-base exceto pela possibilidade de selecionar sub-bases.

Figura L - Base de Dados Acervos Online – Formulário Pesquisa Avançada

No quadro a seguir são apresentados os quatro tipos de formulários da base Acervos Online Fiocruz com seus respectivos índices, para melhor compreensão das semelhanças e diferenças entre os formulários.

ACERVOS ONLINE			
FORMULÁRIO PESQUISA SIMPLES Pesquisa em um destes campos selecionado pelo usuário	FORMULÁRIO PESQUISA MULTI-CAMPO Pesquisa simultaneamente entre os seguintes campos:	FORMULÁRIO PESQUISA MULTI-BASE Pesquisa simultaneamente em até três campos os seguintes campos:	FORMULÁRIO PESQUISA AVANÇADA Pesquisa simultaneamente em até três campos os seguintes CAMPOS
Todos os campos		Todos os campos	Todos os campos
Assunto	Assunto	Assunto	Assunto
Autor	Autor	Autor	Autor
titulo	Título (expressão)	Título	Título
Sumário	Título (palavras)		
Ano de Publicação	Ano		
ISBN	Editor		
Número de sistema			
Código de barras			
Número de controle			
URL/856			
Número de chamada			
ISSN			
Número de item GPO			
Notas			

Quadro C – Formulários ACERVOS ONLINE

(<http://157.86.8.8/ALEPH/R2LF6LHFM2YMT7456G2YD1A9NPX2HQA746IJER5U18A6514VBQ-00336/file/start-0>)

4.2 CAMPOS DE EXIBIÇÃO

Nesta segunda etapa faremos uma comparação entre as bases de dados referente aos campos que compõem o formato de exibição. Os formatos se referem a como o dado recuperado é exibido.

4.2.1 Campos LILACS E MEDLINE

Nas bases de Dados LILACS e MEDLINE estão disponíveis 5 formatos de exibição dos resultados: Formato Detalhado, Formato Longo, Formato Título, Formato Citação e Formato Afiliação. Cada um destes formatos possuem campos diferenciados.

4.2.1.1 Campos LILACS

Formato Detalhado

- Id – (Número de Identificação) Número único seqüencial, a ser atribuído e controlado pelo Centro Cooperante
- Autor – Nome da pessoa ou instituição responsável pelo conteúdo intelectual do documento
- Título - Título do documento no idioma e forma em que aparece no mesmo
- Fonte –Forma de identificar o documento. No caso de periódicos: título da revista volume número, paginação e ano do artigo. No caso de monografias descreve-se a
➤ imprensa.
- Idioma- Código que identifica o idioma do texto analisado, conforme a norma ISO-ST-R-639-1977 (novo nº da ISO-ST-8601-1988).
- Resumo- Resumo do conteúdo temático do documento
- Descritores- Descritores extraídos do DeCS para representar o conteúdo temático do documento
- Limites - Localização- Notação composta pelo Código do Centro que possui o documento e pelo número de localização física do documento (número de chamada) em seu acervo. O Código do Centro é composto pelo código ISO do país onde o Centro está localizado, seguido de um número que o identifica.

Id: 362998

Autor: Santoro Junior, Walter; Santoro, Andreza.

Título: Reações psicológicas e processo adaptativo de pais de recém-nascidos pré-termo e de muito baixo peso em Unidades de Terapia Intensiva Neonatal (UTIN) / Psychology reaction in parents of the newborns in Intensives Care Units

Fonte: Rev. paul. pediatr;20(2):95-101, abr. 2002.

Idioma: Pt.

Resumo: Vários autores têm estudado a condição da prematuridade e suas implicações no estabelecimento do vínculo mãe-filho nos últimos anos. O nascimento de uma criança criticamente enferma pela sua prematuridade precipita um evento estressante para os pais e promove uma inevitável e precoce separação materno-filial. Esta situação acarreta sentimentos de luto e reações de desajuste, que os pais deverão elaborar num processo adaptativo até a condição de equilíbrio. Os profissionais de saúde podem facilitar as tentativas paternas de elaboração destes sentimentos e reforçar seu senso de competência, de maneira a tornas os pais e seu filho uma unidade adaptada e funcional.(AU).

Descritores: Aleitamento Materno

Prematuro

Relações Mãe-Filho

Limites: Humano

Localização: BR584.1

Quadro D – Campos LILACS – Formato Detalhado

Formato Longo

- | | |
|----------|---------------|
| ➤ Id | ➤ Idioma |
| ➤ Autor | ➤ Resumo |
| ➤ Título | ➤ Localização |
| ➤ Fonte | |

Id: 362998

Autor: Santoro Junior, Walter; Santoro, Andreza.

Título: Reações psicológicas e processo adaptativo de pais de recém-nascidos pré-termo e de muito baixo peso em Unidades de Terapia Intensiva Neonatal (UTIN) / Psychology reaction in parents of the newborns in Intensives Care Units

Fonte: Rev. paul. pediatr;20(2):95-101, abr. 2002.

Idioma: Pt.

Resumo: Vários autores têm estudado a condição da prematuridade e suas implicações no estabelecimento do vínculo mãe-filho nos últimos anos. O nascimento de uma criança criticamente enferma pela sua prematuridade precipita um evento estressante para os pais e promove uma inevitável e precoce separação materno-filial. Esta situação acarreta sentimentos de luto e reações de desajuste, que os pais deverão elaborar num processo adaptativo até a condição de equilíbrio. Os profissionais de saúde podem facilitar as tentativas paternas de elaboração destes sentimentos e reforçar seu senso de competência, de maneira a tornas os pais e seu filho uma unidade adaptada e funcional.(AU).

Localização: BR584.1

Quadro E – Campos LILACS – Formato Longo

Formato Título

➤ Título

Reações psicológicas e processo adaptativo de pais de recém-nascidos pré-termo e de muito baixo peso em Unidades de Terapia Intensiva Neonatal (UTIN) / Psychology reaction in parents of the newborns in Intensives Care Units.

Quadro F – Campos LILACS – Formato Título

Formato Citação

➤ Autor

➤ Título

➤ Fonte

Santoro Junior, Walter; Santoro, Andreza. - Reações psicológicas e processo adaptativo de pais de recém-nascidos pré-termo e de muito baixo peso em Unidades de Terapia Intensiva Neonatal (UTIN) / Psychology reaction in parents of the newborns in Intensives Care Units. Rev. paul. pediatr; 20(2):95-101, abr. 2002.

Quadro G – Campos LILACS – Formato Citação

Formato Afiliação

- | | |
|---------------------|---------------|
| ➤ Id | ➤ Resumo |
| ➤ Autor e Afiliação | ➤ Descritores |
| ➤ Título | ➤ Limites |
| ➤ Fonte | ➤ Localização |
| ➤ Idioma | |

Neste formato no campo Autor vem detalhada a instituição a qual pertence (Afiliação).

Id: 362998

Autor: · Santoro Junior, Walter; USP. Departamento de Puericultura e Pediatria. Brasil · Santoro, Andreza; Faculdade de Filosofia, Ciências e Letras de Catanduva. Brasil

Título: Reações psicológicas e processo adaptativo de pais de recém-nascidos pré-termo e de muito baixo peso em Unidades de Terapia Intensiva Neonatal (UTIN) / Psychology reaction in parents of the newborns in Intensive Care Units

Fonte: Rev. paul. pediatr;20(2):95-101, abr. 2002.

Idioma: Pt.

Resumo: Vários autores têm estudado a condição da prematuridade e suas implicações no estabelecimento do vínculo mãe-filho nos últimos anos. O nascimento de uma criança criticamente enferma pela sua prematuridade precipita um evento estressante para os pais e promove uma inevitável e precoce separação materno-filial. Esta situação acarreta sentimentos de luto e reações de desajuste, que os pais deverão elaborar num processo adaptativo até a condição de equilíbrio. Os profissionais de saúde podem facilitar as tentativas paternas de elaboração destes sentimentos e reforçar seu senso de competência, de maneira a tornas os pais e seu filho uma unidade adaptada e funcional.(AU).

Descritores: Aleitamento Materno

Prematuro

Relações Mãe-Filho

Limites: Humano

Localização: BR584.1

4.2.1.2 CAMPOS MEDLINE

Formato Longo

- PMID – Número de Identificação único
- Autor - Nome da pessoa ou instituição responsável pelo conteúdo intelectual do documento
- Endereço - Endereço institucional do autor e endereço eletrônico
- Título - Título do documento no idioma e forma em que aparece no mesmo
- Fonte - Forma de identificar o documento. Título da revista volume número, paginação e ano do artigo.
- ISSN - Número que identifica internacionalmente uma série (International Standard Serial Number).
- País de publicação - Nome do país onde está localizada a editora do documento
- Idioma - Código que identifica o idioma do texto analisado, conforme a norma ISO-ST-R-639-1977 (novo nº da ISO-ST-8601-1988).
- Resumo - Resumo do conteúdo temático do documento
- Tipo de publicação Termos que definem o tipo ou os tipos de publicações indexadas
- Nome de substância - nome de substância que representa o conteúdo do documento

PMID: 15202311

Autor: Colic-Hadzic B; Tahirovic H

Endereço: Klinika za djecije bolesti, Univerzitetski klinicki centar Tuzla.

Título: [Importance of immunoglobulins and nutrition in infections in neonates]

Título: Znacaj imunoglobulina i ishrane u infekcijama novorodencadi..

Fonte: Med Arh; 58(2):79-81, 2004.

ISSN: 0350-199X

País de publicação: Bosnia and Hercegovina

Idioma: bos

Resumo: The purpose of the study was an estimation of the immunoglobulin serum and the way of nutrition of a new-born child at the beginning of bacterial infections during the first month of life. 120 new-born children (infantiles) of both sexes was prospectively studied, who were born in the Gynecology-obstetrician clinic in Tuzla, in the period from March-May 2000 year. On the base of mothers SES during the pregnancy, new-born children were separated in two groups. Group A was formed of 60 new-born children (32 female and 28 male), mothers had good SES during the pregnancy, and concentration of serum immunoglobulin G (IgG) examined from cords blood, from 10.02 to 17.58 g/L (X 14.48 g/L +/- 1.87). Group B was formed of 60 new-born children (27 female and 33 male), with bad mothers SES and with concentration of IgG from 6.80 to 13.01 g/L (X 10.62 g/L +/- 1.49). Immunoglobulins M and immunoglobulins A were negative in both groups of children. During the first month of life it was studied the way of nutrition and health condition of each new-born child. New-born children with bad mothers SES during the pregnancy, with low values of IgG and who were on the artificial food (nutrition), were showing with the significant difference ($p < 0.001$) in bacterial infections, comparing with the group who had good IgG. The new-born children with natural nutrition, despite the lower concentration of IgG, were protected from severe bacterial infections.

Tipo de publicação: JOURNAL ARTICLE

Nome de substância: 0 (Immunoglobulin G)

Formato Detalhado

- [PMID]
- Autor
- Endereço
- Título
- Fonte
- ISSN
- País de publicação
- Idioma
- Resumo
- Termos MeSH
- Tipo de publicação
- Nome de substância
- Mês de entrada
- Subgrupo de revista
- Data de entrada para processamento

[PMID]: 15202311

[Au] **Autor:** Colic-Hadzic B; Tahirovic H

[Ad] **Endereço:** Klinika za djecije bolesti, Univerzitetski klinicki centar Tuzla.

[Ti] **Título:** [Importance of immunoglobulins and nutrition in infections in neonates]

[Ti] **Título:** Znacaj imunoglobulina i ishrane u infekcijama novorodencadi..

[So] **Fonte:** Med Arh; 58(2):79-81, 2004.

[Is] **ISSN:** 0350-199X

[Cp] **País de publicação:** Bosnia and Hercegovina

[La] **Idioma:** bos

[Ab] **Resumo:** The purpose of the study was an estimation of the immunoglobulin serum and the way of nutrition of a new-born child at the beginning of bacterial infections during the first month of life. 120 new-born children (infantiles) of both sexes was prospectively studied, who were born in the Gynecology-obstetrician clinic in Tuzla, in the period from March-May 2000 year. On the base of mothers SES during the pregnancy, new-born children were separated in two groups. Group A was formed of 60 new-born children (32 female and 28 male), mothers had good SES during the pregnancy, and concentration of serum immunoglobulin G (IgG) examined from cords blood, from 10.02 to 17.58 g/L (X 14.48 g/L +/- 1.87). Group B was formed of 60 new-born children (27 female and 33 male), with bad mothers SES and with concentration of IgG from 6.80 to 13.01 g/L (X 10.62 g/L +/- 1.49). Immunoglobulins M and immunoglobulins A were negative in both groups of children. During the first month of life it was studied the way of nutrition and health condition of each new-born child. New-born children with bad mothers SES during the pregnancy, with low values of IgG and who were on the artificial food (nutrition), were showing with the significant difference ($p < 0.001$) in bacterial infections, comparing with the group who had good IgG. The new-born children with natural nutrition, despite the lower concentration of IgG, were protected from severe bacterial infections.

[Mh] **Termos MeSH:** Infecções Bacterianas/*IM

Alimentação Artificial

Aleitamento Materno/*

Resumo em Inglês

Feminino

Humano

Imunoglobulina G/*BL

Recém-Nascido

Masculino

Gravidez

[Pt] **Tipo de publicação:** JOURNAL ARTICLE

[Nm] **Nome de substância:**

0 (Immunoglobulin G)

[Em] **Mês de entrada:** 0407

[Sb] **Subgrupo de revista:** IM

[Da] **Data de entrada para processamento:** 040618

Formato Título

➤ Título

<p>[Breastfeeding among teenage and adult mothers in Brazil]. Amamentação entre mães adolescentes e não-adolescentes, Montes Claros, MG..</p>

Quadro L – Campos MEDLINE – Formato Título

Formato Citação

- Autor
- Título
- Fonte

<p>Frota DA, Marcopito LF. [Breastfeeding among teenage and adult mothers in Brazil] Amamentação entre mães adolescentes e não-adolescentes, Montes Claros, MG. Rev Saude Publica. 2004;38(1):85-92.</p>
--

Quadro M – Campos MEDLINE – Formato Citação

4.2.2 Campos ACERVOS ONLINE

Na base de Dados Acervos Online Fiocruz também estão disponíveis 5 formatos de exibição dos resultados, porém com nomenclaturas diferentes : Formato Padrão, Formato Ficha Catalográfica, Formato Ficha Reduzido, Formato Nomes Marc, Formato Campos Marc. Este último não será incluído nesta análise por se tratar de um formato que apenas descreve o número dos campos Marc não atendendo ao objetivo de buscar as semelhanças com vista a alcançar a interoperabilidade. Seguem abaixo os três primeiros formatos citados:

Formato Padrão

- No. Sistema – Identificador único
- Coleção geral- Não faz parte de nenhuma coleção específica
- Info. Biblioteca –Em qual biblioteca o documento está disponível
- No. chamada – código da biblioteca, seguido do número da classificação do documento
- Entrada Principal – Responsável pelo conteúdo intelectual do documento
- Título – título do documento
- Imprensa- informações sobre a publicação do documento
- Descrição Física – informações sobre a descrição física do documento como número de páginas, tamanho, ilustrações, etc.
- Idioma – idioma do documento
- Nota- Informações importantes para identificar o documento.
- Resumo: resumo do documento
- Assunto: representação temática do documento
- Secund. – autor: Autores secundários
- Secund. – entidade: Entidades co-autoras.

No. Sistema **000049303**

Coleção geral Todos os itens

Info. biblioteca IFF

No. chamada **BR663.1 649.33 A256p**

Entrada Principal Afonso, Viviane Weil.

Título Prevalência de aleitamento materno em Juiz de Fora, MG

Imprensa Rio de Janeiro : [s.n], 2003.

Descrição 127 f. tab.; graf.

Idioma por

Nota Tese/Diss Dissertação (mestrado) - Instituto Fernandes Figueira

Resumo Resumo: este trabalho teve como objetivo diagnosticar a situação do aleitamento materno e hábitos alimentares em crianças menores de um ano residentes em Juiz de Fora...

Assunto ALEITAMENTO MATERNO

Assunto LEITE HUMANO

Assunto DESMAME

Assunto HÁBITOS ALIMENTARES

Assunto NUTRIÇÃO INFANTIL

Formato Ficha Catalográfica

- No. sistema
- Autor
- Título
- Imprenta

No. sistema [000049303]

Afonso, Viviane Weil.

Prevalência de aleitamento materno em Juiz de Fora, MG. -- Rio de Janeiro : [s.n], 2003.

: 127 f. ; tab. ; graf..

Quadro O – Campos Acervos Online – Formato Ficha Catalográfica

Formato Ficha Reduzido

- Autor
- Título
- Imprenta

Afonso, Viviane Weil Prevalência de aleitamento materno em Juiz de Fora, MG Rio de Janeiro [s.n] 2003

Quadro P – Campos Acervos Online – Formato Ficha Reduzido

Formato Nomes Marc

- | | |
|----------------------|---------------------|
| ➤ Líder | ➤ Imprenta |
| ➤ Control No. ID | ➤ Descrição física |
| ➤ Data e hora atual | ➤ Nota de tese |
| ➤ Campo 008 | ➤ Nota de resumo |
| ➤ Fonte catalogadora | ➤ Assunto |
| ➤ Idioma | ➤ Secund - aut pess |
| ➤ Cód. país public. | ➤ Secund - entidade |
| ➤ No. Chamada | ➤ Base |
| ➤ Autor pessoal | ➤ No.sistema |
| ➤ Título | |

Líder **00634nam0000003010a04500**

Control No. ID BR-RjFOC

Data e hora atual 20031007144904.0

Physical Descr. FF ta

Campo 008 -----s2003----rj-dl--frm---#000-0#por-d

Fonte catalogadora BR-RjFOC

Idioma por

Cód. país public. BR

No.Chamada BR663.1 649.33 A256p

Autor pessoal Afonso, Viviane Weil.

Título Prevalência de aleitamento materno em Juiz de Fora, MG

Imprenta Rio de Janeiro : [s.n], 2003.

Descrição física 127 f. tab.; graf.

Nota de tese Dissertação (mestrado) - Instituto Fernandes Figueira

Nota de resumo Resumo: este trabalho teve como objetivo diagnosticar a situação do aleitamento materno e hábitos alimentares em crianças menores de um ano residentes em Juiz de Fora...

Assunto ALEITAMENTO MATERNO

Assunto LEITE HUMANO

Assunto DESMAME

Assunto HÁBITOS ALIMENTARES

Assunto NUTRIÇÃO INFANTIL

Secund - aut pess Dain, Sulamis orientadora

Secund - aut pess Teixeira, Maria Teresa Bustamante co-orientadora

Secund - entidade Instituto Fernandes Figueira

Base 99

Base 02

No.sistema 000049303

Quadro Q – Campos Acervos Online – Formato Nomes MARC

4.3 BUSCANDO AS SEMELHANÇAS

4.3.1 Semelhanças nos índices

A seguir veremos o quadro relacionando todos os índices de busca disponíveis pelas bases de dados nos seus vários formulários:

ÍNDICES DAS BASES DE DADOS		
FORMULÁRIOS LILACS	FORMULÁRIOS MEDLINE	FORMULÁRIOS ALEPH
Palavras	Palavras	Todos os campos
Descritor de assunto	Descritor de assunto	Assunto
Limites	Limites	Autor
Autor	Autor	titulo
Palavras do título	Palavras do título	Sumário
Revista	Revista	Ano de Publicação
Monografia em série	Tipo de publicação	ISBN
Tipo de publicação	Idioma	Número de sistema
Idioma	Identificador único	Código de barras
Identificador único	Ano de publicação	Número de controle
País-Ano de publicação	Texto completo SciELO	URL/856
Localização	País de publicação	Número de chamada
Mês de entrada	Nome de substância	ISSN
Localização	Símbolo de gene	Número de item
Mês de entrada	Nome de pessoa como assunto	GPO
Suporte eletrônico	ISSN	Notas
Autor normalizado	Mês de entrada	Editor
Autor institucional	Atualização por classe	
Palavras do resumo	palavra do resumo	
Tipo de literatura		
Conferência		
Nível bibliográfico		
Instituição		
Cidade		
País		
Nome de pessoas como assunto		

Quadro R – Índices das bases de dados LILACS, MEDLINE e ACERVOS ONLINE FIOCRUZ

Em cada uma destas bases são disponibilizados índices diferenciados, com alguma predominância. No quadro abaixo foram relacionados apenas os índices semelhantes. Alguns, apesar de utilizarem uma nomenclatura diferenciada, representam o mesmo conteúdo, conforme quadro a seguir:

ÍNDICES SEMELHANTES DAS BASES	MEDLINE	LILACS	ACERVOS ONLINE FIOCRUZ
IDENTIFICADOR	Identificador único	Identificador único	Número de sistema
AUTOR	Autor	Autor	Autor
TÍTULO	Palavras do título	Palavras do título	título
ASSUNTO	Descritor de assunto	Descritor de assunto	Assunto
ANO DE PUBLICAÇÃO	Ano de publicação	País-Ano de publicação	Ano de Publicação

Quadro S – Índices semelhantes das bases de dados LILACS, MEDLINE e ACERVOS ONLINE FIOCRUZ

4.3.2 Semelhanças nos Campos de Exibição

A seguir é apresentado o quadro relacionando todos os campos de exibição disponíveis pelas bases de dados nos seus vários formatos:

CAMPOS DE EXIBIÇÃO DAS BASES DE DADOS		
MEDLINE	LILACS	ACERVOS ONLINE FIOCRUZ
PMID	Id	No. Sistema
Autor	Autor	Coleção geral
Endereço	Título	Info. biblioteca
Título	Fonte	No. chamada
Fonte	Limites	Entrada Principal
ISSN	Idioma	Título
País de publicação	Resumo	Imprensa
Idioma	Localização	Descrição Física
Resumo	Descritores	Idioma
Tipo de publicação		Nota / Nota de tese
Nome de substância		Resumo / nota de resumo
Termos MeSH		Assunto
Tipo de publicação		Secund - aut pess
Nome de substância		Secund - entidade
Mês de entrada		Base
Subgrupo de revista		Autor
Data de entrada para processamento		Líder
		Control N° ID
		Campo 008
		Fonte Catalogadora
		Cód. País de Publicação
		Autor pessoal

Quadro T – Campos das bases de dados LILACS, MEDLINE e ACERVOS ONLINE FIOCRUZ

Em cada um destes formatos são exibidos campos diferenciados, com alguma predominância. No quadro abaixo foram relacionados apenas os campos semelhantes. Alguns apesar de utilizarem uma nomenclatura diferenciada, representam o mesmo conteúdo, conforme quadro a seguir:

CAMPOS DE EXIBIÇÃO	MEDLINE	LILACS	ACERVOS ONLINE FIOCRUZ
IDENTIFICADOR	PMID	Id	No. Sistema
AUTOR	Autor	Autor	Entrada Principal
TÍTULO	Título	Título	Título
IMPRESSÃO	Fonte	Fonte	Impressão
IDIOMA	Idioma	Idioma	Idioma
RESUMO	Resumo	Resumo	Resumo
ASSUNTO	Termos MeSH	Descritores	Assunto

Quadro U – Campos semelhantes das bases de dados LILACS, MEDLINE e ACERVOS ONLINE FIOCRUZ

A maioria das diferenças entre os índices das bases de dados se dá devido a diferença de documentos tratados nestas bases. A base MEDLINE indexa apenas artigos científicos e seus índices refletem a descrição deste material, já a base LILACS indexa além de artigos científicos, teses, livros, capítulos de livros, anais de congressos ou conferências, relatórios técnico-científicos etc., a base Acervos Online Fiocruz, por sua vez, indexa livros, teses e anais de congresso. Enquanto a base MEDLINE e LILACS tem os artigos de periódicos como o principal acervo de suas bases, a base Acervos Online é voltada para a catalogação de livros.

Apesar das diferenças, os índices fundamentais para a descrição de um documento são contemplados nas três bases de dados, bem como os campos de exibição, o que contribui para a interoperabilidade entre elas.

CAPÍTULO V

MODELOS CONCEITUAIS

“Uma verdadeira viagem de descoberta não é procurar novas terras, mas ter um olhar novo”

Marcel Proust

5 MODELOS CONCEITUAIS

Como já referido, a interoperabilidade tem muitas facetas e envolve acordos em três níveis: organizacional, técnico e de conteúdo. Estes níveis estão relacionados com a Interoperabilidade tecnológica, política e semântica. Entendendo estas facetas da interoperabilidade analisaremos alguns modelos conceituais para o alcance da interoperabilidade entre as bases de dados envolvidas neste estudo.

As Bases de Dados analisadas utilizam tesouros para a representação do conteúdo de seus recursos informacionais. A base MEDLINE utiliza o MESH, a Base LILACS utiliza o DeCS, e a Base de dados Acervos Online também utiliza o DeCS. Como o DeCS foi desenvolvido a partir do MESH (como descrito no capítulo Campo Empírico), e até hoje é atualizado com os novos termos ao MESH introduzidos, os tesouros são praticamente os mesmos sendo utilizado o mesmo termo para representar as informações nas três bases. Sendo assim, pode-se afirmar que as bases de dados em questão possuem interoperabilidade semântica, isto é, a representação temática nas três bases, por utilizarem o mesmo tesouro para representar seus conteúdos, não é uma barreira para alcançar a interoperabilidade entre as três bases de dados.

Diferente do que acontece na área semântica, na área tecnológica e política estas bases de dados são heterogêneas. Um exemplo de interoperabilidade tecnológica é a Internet. A Web utiliza um sistema cliente/servidor para alcançar a interoperabilidade, fazendo com que máquinas completamente distintas consigam operar em conjunto, consigam interagir. Quando acessamos páginas na Web estas páginas não estão fisicamente gravadas em nosso computador, mas em máquinas funcionando como servidoras e o conteúdo destas páginas é disponibilizado em nosso computador (que tem um browser funcionando como cliente) que solicita as informações deste servidor. Quando fazemos este acesso não precisamos saber onde fisicamente esta página está instalada e qual é o servidor, basta inserir sua URL ou clicar em um link e a página aparece na máquina cliente como se o conteúdo estivesse instalado fisicamente em sua máquina. Este relacionamento é possível, pois as máquinas estão utilizando o mesmo protocolo, TCP/IP, para troca de dados.

Quando se trata da busca em Bases de Dados o processo é o mesmo. O usuário faz uma consulta na base que é submetida ao servidor que possui o conteúdo da base e responde para o usuário.

Para melhor compreensão, chamaremos de **Recursos Informativos (RI)** as bases de dados LILACS, MEDLINE E ACERVOS ONLINE FIOCRUZ, pois nelas se encontram os recursos informativos procurados pelos usuários. Com o mesmo objetivo chamaremos de **Interface de Consulta (IC)** o site que o usuário utilizará para solicitar estas informações.

Nesta dissertação buscamos a interoperabilidade entre estes atores. A partir destes atores e do entendimento adquirido sobre interoperabilidade traçamos a seguir alguns modelos para se alcançar a interoperabilidade, mostrando as possibilidades de se atingir a interoperabilidade entre os recursos informativos de interesse da Biblioteca do IFF. Ressaltamos que “a simulação de modelos digitais seria antes operacional, provisional, ou mesmo normativa. Ela responde melhor à pergunta ‘como?’ do que à pergunta ‘por quê?’” Lévy⁴⁹ p.121.

5.1 MODELO 1 - Compartilhamento de Esforços

O primeiro modelo traçado para alcançar a interoperabilidade é de **Compartilhamento de Esforços**. Neste caso os Recursos informativos se esforçam para utilizar um padrão comum que possibilite o interagir das Bases de dados e a Interface de consulta também se adapta adotando o padrão estabelecido, assim o trabalho é compartilhado, dividido. Este é o modelo utilizado nos protocolos Z39.50 e OAI-PMH, que dividem os esforços de interoperabilidade entre a IC (Cliente) e os RIs (Servidores).

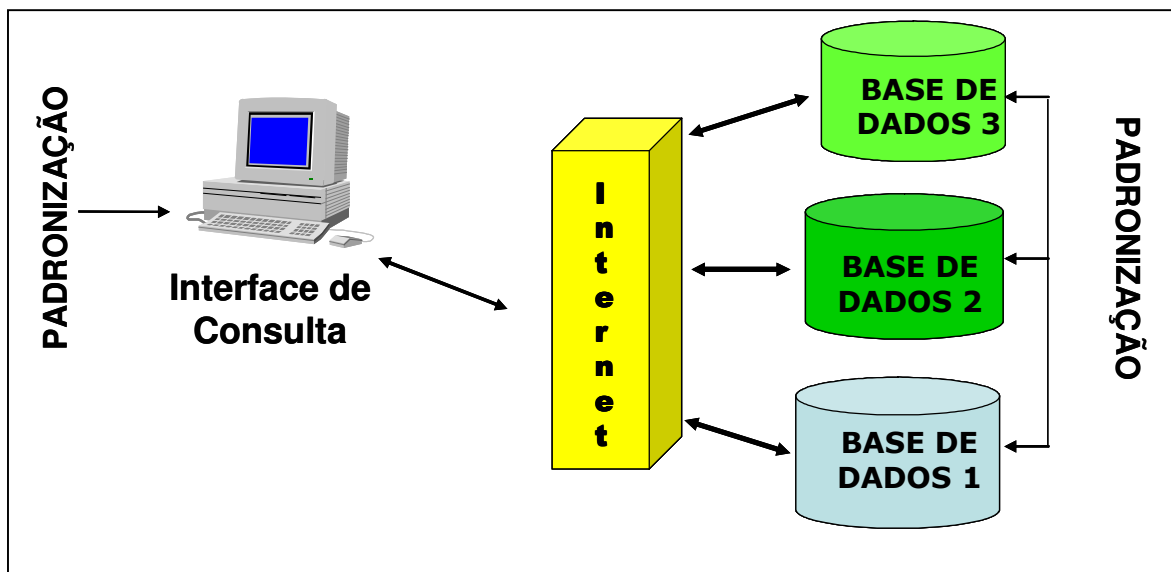


Figura M – Modelo 1- Compartilhamento de Esforços

5.2 MODELO 2 - Esforço Concentrado nos RI's

O segundo modelo é o **Esforço Concentrado nos RI's**. Neste modelo é adotada uma padronização apenas dos Recursos Informacionais, neste caso as Bases de Dados arcariam com todos os esforços para se tornarem interoperáveis e a Interface de consulta não precisaria fazer nenhum esforço. Neste segundo modelo todo o trabalho é realizado por parte do Servidor (Recursos Informacionais) não necessitando esforço por parte do Cliente (Interface de Consulta). Este modelo se aplica em bases de dados idênticas, com mesmos formatos e índices.

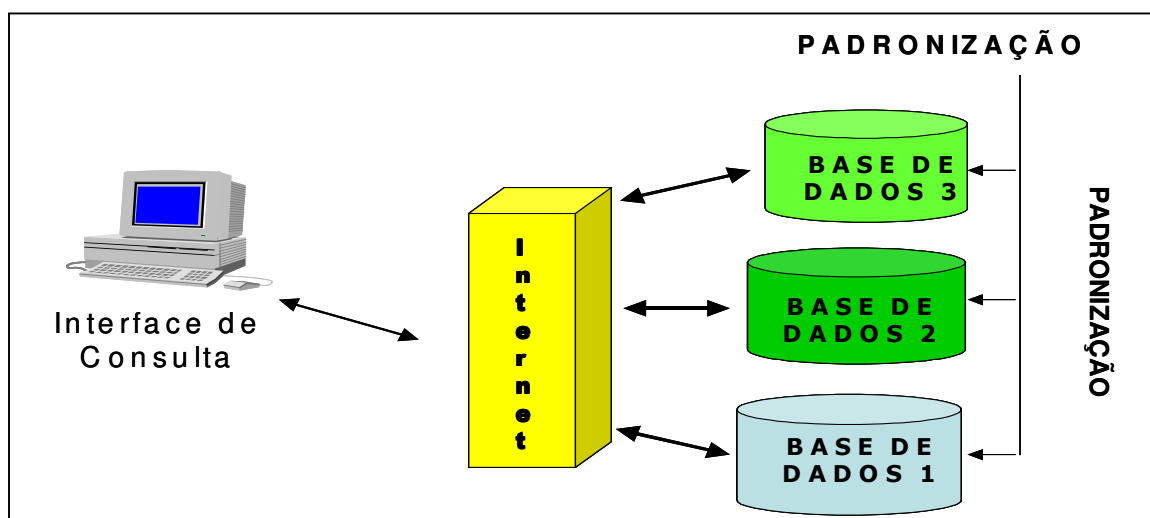


Figura N – Modelo 2 - Esforço Concentrado nos RI's

5.3 MODELO 3 – Esforço Concentrado na IC

O terceiro e último modelo é o **Esforço Concentrado na IC**. Este modelo concentra todos os esforços para alcançar a interoperabilidade apenas na Interface de Consulta (Cliente), que neste caso arcaria com todo o trabalho de viabilizar a interoperabilidade entre os Recursos Informacionais. São os modelos utilizados pelos metabuscadores.

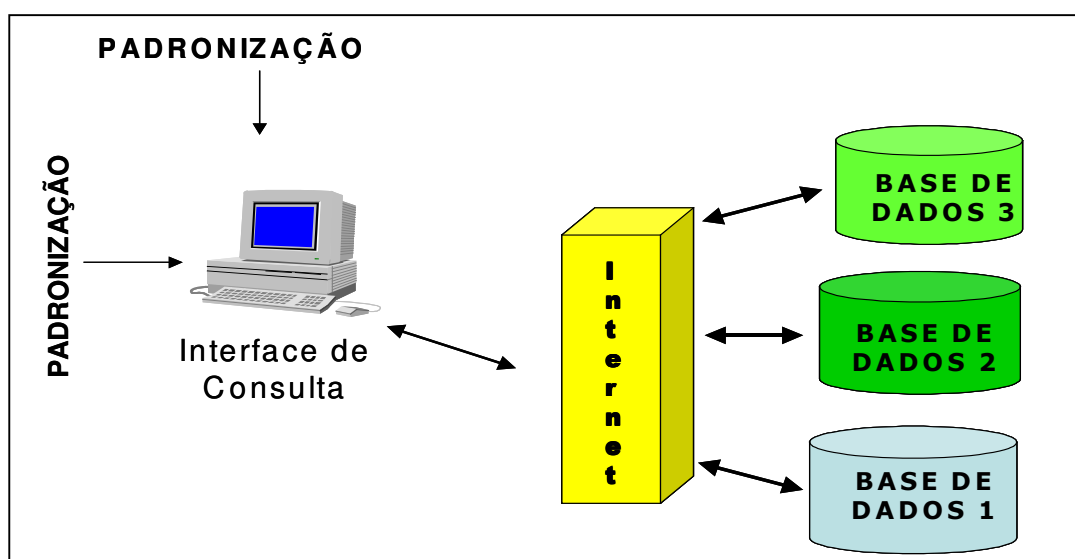


Figura O – Modelo 3- Esforço Concentrado na IC

5.4 OS MODELOS E AS BASES DE DADOS

O primeiro modelo pode ser aplicado utilizando os protocolos Z39.50 e OAI-PMH. O Z39.50 compreende servidores de dados e servidores de serviço, necessitando alto grau de padronização entre os atores. Cada máquina servidora precisa utilizar o protocolo Z39.50 em sua máquina para viabilizar a comunicação entre as máquinas. A Base de Dados Acervos Online, conforme analisamos, utiliza o protocolo Z39.50, portanto é compatível com outras BD's que utilizam o mesmo protocolo. Porém, as outras duas bases (MEDLINE E LILACS) não utilizam este protocolo. Para a utilização desta alternativa de interoperabilidade a FIOCRUZ precisaria ter gerência sobre as

outras duas bases de dados para torná-las servidores Z39.50. Como vimos anteriormente a Base de dados LILACS é gerenciada pela Bireme e a MEDLINE, apesar de ser gerenciada pela NLM (National Library of Medicine, EUA) é disponibilizada para os países latino-americanos e do Caribe através da BIREME, que recebe em seu servidor a base MEDLINE. Portanto a Biblioteca do IFF não tem gerência sobre estas bases para poder promover estas modificações.

Uma outra aplicação deste modelo pode ser feita utilizando o protocolo Open Archives Initiative Protocol Metadata Harvesting (OAI-PMH). Nesta aplicação um programa robô provedor de serviços coleta automaticamente metadados de diversos provedores de dados e os armazena em uma base de dados centralizada, onde são feitas as buscas. A utilização do protocolo OAI-PMH também precisa ser compartilhada entre as BD's. A BIREME já realizou uma experiência no uso deste protocolo, através da Interface SciELO-OAI (Marcondes⁴⁵).

A Base de Dados Acervos Online que utiliza o software ALEPH atualmente não possui este recurso. Apesar do Aleph ser um programa aberto que possibilita para a instituição que o adquire e utiliza desenvolver várias modificações, a definição do uso deste protocolo precisa ser feita pela Sede da Ex Libris, em Israel. A partir da versão 16 o Aleph se torna compatível com o protocolo OAI (*Anexo*). Apesar de a Base Acervos Online ainda utilizar a versão 14, existe uma expectativa de que brevemente será implementada esta nova versão o que possibilitaria pensar a utilização desta ferramenta como instrumento para alcançar a interoperabilidade.

Segundo esta informação, o uso do protocolo OAI poderia, num futuro breve, se tornar uma alternativa de interoperabilidade entre as três bases. Deve-se observar, no entanto, que o protocolo OAI se destina originalmente, a tornar interoperáveis bases de dados de texto completo, o que não é exatamente o caso das bases em questão.

No segundo modelo, Esforço Concentrado nos RI's, busca-se a conversão da base divergente para que se torne semelhante com as demais, pode ser aplicado convertendo a Base Acervos Online em LILACS. Pode-se utilizar para isto ferramentas como o ISIS-MARC (programa da UNESCO de entrada de dados em formato MARC para bases em CDS/ISIS, <http://www.unesco.org/isis/>), os utilitários CISIS

(<http://www.bireme.br/produtos/>) e as FSTs (*Field Select Table* – Tabela de seleção de campos) de reformatação da linguagem de formatação do CDS/ISIS. Esta conversão faria com que as 3 bases “falassem a mesma língua” utilizando a mesma metodologia e sendo compatíveis com a Interface de pesquisa META-IAH. A vantagem para esta alternativa é a relativa simplicidade com que poderia ser implementado e a rapidez com que este processo poderia ser efetuado, pois o MARC, formato utilizado pelo Aleph, é facilmente convertido para ISIS que por sua vez pode ser, sem esforço, exportado para LILACS. Por outro lado esta conversão também tem desvantagens como duplicidade de dados; para cada novo registro inserido na base Acervos Online precisaria ser dada nova entrada na Base do Acervo utilizando a metodologia LILACS, ou periodicamente precisaria atualizar a base com uma nova conversão dos dados. Outra desvantagem é a perda de informação. O MARC é um padrão amplo para descrição de documentos, portanto utiliza muitos campos. Estes campos não são todos contemplados na metodologia LILACS o que poderia causar uma perda de informação durante a conversão das bases. Uma terceira desvantagem está relacionada com implementação de novas bases de dados. Caso a Biblioteca do IFF, futuramente queira incluir outros Recursos Informativos que não utilizassem a metodologia LILACS, teria dificuldades.

No terceiro modelo, esforço concentrado na IC, onde todos os esforços se concentram apenas na Interface de Consulta (Cliente), é necessária a incorporação de um poderoso cliente que possa, sozinho, viabilizar a interoperabilidade, recuperando informações das diferentes bases de dados e retornando-as de forma consolidada para o usuário. Este modelo se aplica com a utilização de um metabuscador para bases de dados. Dentro desta perspectiva temos duas alternativas.

A primeira alternativa deste modelo é a utilização de um metabuscador comercial (ex.: Akwan - <http://www.akwan.com.br>). Esta solução foi utilizada no projeto da Biblioteca Digital Brasileira (BDB). Este tipo de metabuscador consulta as bases de dados uma a uma, repetindo os termos, entrados na pesquisa pelo usuário, inalterados em cada base. Esta alternativa tem algumas desvantagens, uma delas é o custo para contratação do serviço junto a empresa. Outra desvantagem é a pouca precisão, pois quando este tipo de metabuscador submete a pesquisa às bases de dados ele não consulta

os índices e não solicita a pesquisa de forma a se adequar às especificidades de cada Base de Dados.

A Segunda alternativa deste modelo é o desenvolvimento de um metabuscador próprio. Nesta alternativa os dados sobre cada formulário, cada campo e cada índice das bases de dados estariam guardados (registrados) através de ontologias, assim como outras informações específicas de cada base. Esta ontologia conteria informações do tipo: nome da base, URL(endereço eletrônico) da base, formulários de cada base, formas de entrada de pesquisa de cada base, formas de entrada de pesquisa de cada formulário, índices de cada base, seus campos, etc. Esta Interface construída consultaria as ontologias antes de submeter a pesquisa do usuário às três bases.

Desta forma, a Interface Web desenvolvida proporcionaria a interoperabilidade entre as bases de dados, “entendendo” as diferenças e semelhanças e direcionando a pesquisa para cada base de acordo com suas especificidades. As desvantagens desta alternativa são: necessidade de desenvolver este programa, tornando-a mais complexa; e, o tempo para implementação (longo prazo).

Por outro lado, algumas vantagens podem ser relacionadas como a garantia de precisão, pois a Interface Web construída utilizaria cada base de forma diferenciada, utilizando seus índices, formulários e campos próprios. Outra vantagem percebida é a possibilidade de trabalhar com recursos heterogêneos, facilitando futuras incorporações de novos Recursos Informativos heterogêneos. (Conforme figura a seguir)

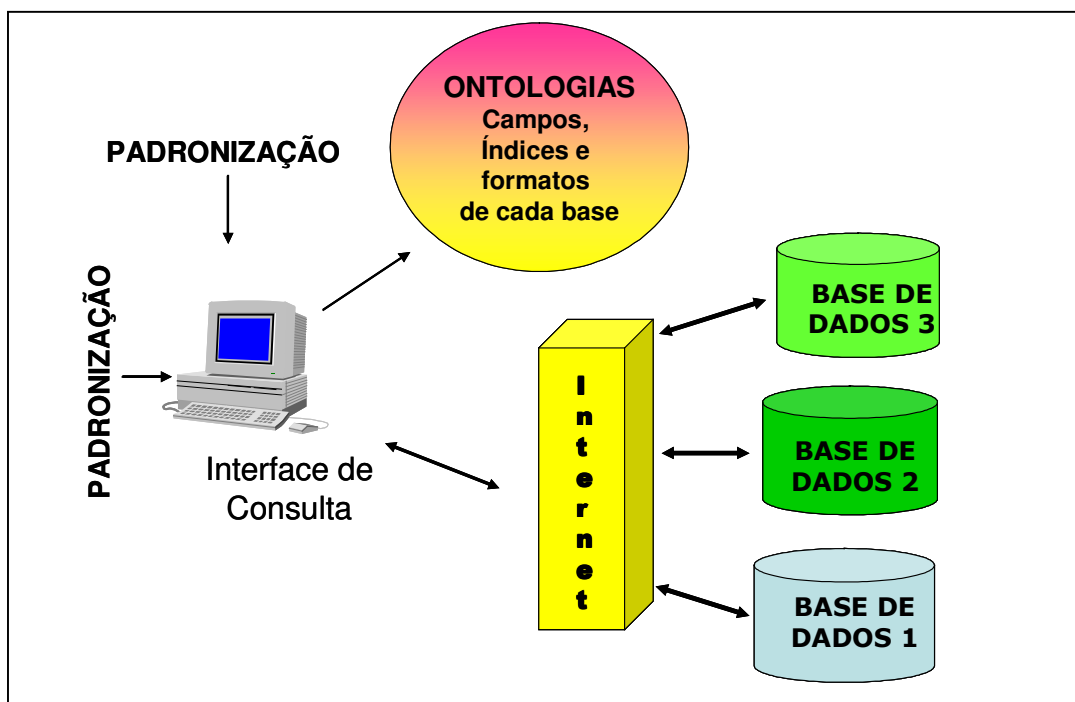


Figura P – Modelo 3 - Esforço Concentrado na IC: Metabusador Próprio

CAPÍTULO VI

CONSIDERAÇÕES FINAIS

“Em vez de considerar, a biblioteca como uma fortaleza isolada ou como um tigre de papel, pretendo pintá-la como o nó de uma vasta rede onde circulam não signos, não matérias, e sim matérias tornando-se signos”.

Bruno Latour

6 CONSIDERAÇÕES FINAIS

O trabalho cooperativo faz parte do cotidiano da maioria das bibliotecas há muito tempo. Como mencionado, desde 1694 que a Biblioteca Nacional de Paris, entendendo a necessidade de trabalhar em conjunto, trocava suas duplicatas por livros ingleses e alemães. Como esta várias foram as iniciativas destacadas nesta pesquisa demonstrando a tradição das bibliotecas em serviços cooperativos. Com este conhecimento adquirido ao longo dos últimos séculos a biblioteconomia pôde auxiliar no desenvolvimento de ferramentas e serviços que contribuem para a interoperabilidade, e fazer frente à explosão informacional das últimas décadas.

Atualmente, o volume de informações produzidas por dia, é maior do que a capacidade de processarmos. Manter-se atualizado e bem informado é um grande desafio. A missão da biblioteca em proporcionar o acesso à informação continua a mesma, porém, não mais tentando adquirir, ter posse de todos os documentos necessários aos seus usuários, mas funcionando como “estação de triagem”.

Se por um lado, a cada dia novas fontes de informação são desenvolvidas e de forma heterogênea, por outro lado aumenta a necessidade do usuário em obter informações consolidadas de forma eficiente e eficaz.

Como exposto nesta dissertação, na área da saúde, mais especificamente na Biblioteca do Instituto Fernandes Figueira, percebe-se a necessidade de informações consolidadas para a tomada de decisão clínica. A morosidade na busca de informações pode custar a vida de um “paciente”, e a eficiência desta busca pode salvá-la. O estudo das bases de dados MEDLINE, LILACS e ACERVOS ONLINE FIOCRUZ, demonstrou-nos as possibilidades de tornar interoperáveis estas bases e assim otimizar o serviço da Biblioteca do IFF. As semelhanças entre os campos, os formatos e índices, conforme mostrado anteriormente, aponta várias alternativas possíveis de serem implementadas.

Porém, como nos ensinou Miller⁵ a interoperabilidade não está associada apenas a parte tecnológica, mas envolve várias facetas como a interoperabilidade política e semântica. No caso da biblioteca do IFF a faceta da interoperabilidade semântica é facilmente resolvida pois já houve acordos entre as organizações neste ponto, com a utilização dos tesouros DeCS, pela LILACS e ACERVOS ONLINE FIOCRUZ e a utilização do MeSH pela MEDLINE, que podem ser considerados, como vimos, o mesmo tesouro devido suas gêneses e atualizações conjuntas.

Para a faceta interoperabilidade técnica/tecnológica foram desenvolvidas alternativas neste trabalho que podem ser implementadas como o modelo de Esforço concentrado nos RI's, que, para a Biblioteca do IFF, concluímos seria necessário a conversão da Base ACERVOS ONLINE FIOCRUZ para a Metodologia LILACS. Outro modelo que também contatamos como alternativa para a interoperabilidade na biblioteca foi o modelo de esforço concentrado na IC, onde utiliza-se um metabuscador poderoso que possa suprir as divergências entre as RI's.

Na faceta interoperabilidade política constatamos que as outras facetas muitas vezes se esbarram nesta. A decisão da Bireme de desenvolver o DeCS a partir de um tesouro já existente e reconhecido mundialmente (MeSH), foi uma decisão política que contribuiu para a interoperabilidade semântica. A adoção do Descritor em Ciências da Saúde (DeCS) pela Biblioteca do IFF foi uma decisão política que, também, cooperou para a interoperabilidade semântica. Da mesma forma, a escolha pela utilização de uma ferramenta como o Z39.50 ou o OAI-PMH também perpassa por decisão política das mantenedoras das Bases de Dados, que pode contribuir ou não para a interoperabilidade tecnológica.

Para que haja interoperabilidade entre os sistemas de informação na área da saúde é preciso que haja, primeiramente, interoperabilidade política para conciliação de interesses. As organizações da área da saúde precisam criar fóruns específicos onde estes assuntos possam ser debatidos e acordos serem estabelecidos, promovendo articulação e cooperação entre estas organizações e otimizando o acesso à informação para os profissionais da área.

No caso específico das bibliotecas que compõem a Base Acervos Online FIOCRUZ percebemos um avanço, pois o acervo destas bibliotecas, como dito anteriormente, estão articulados, podendo ser acessados através de uma única interface. Porém, para a própria FIOCRUZ há um grande caminho ainda a ser percorrido, pois ainda falta buscar soluções para tornar as 10 bibliotecas da FIOCRUZ interoperáveis e, assim, otimizar o trabalho de pesquisa da grande comunidade interna da FIOCRUZ e externa. Percebe-se, assim, a necessidade da criação de fóruns onde estes assuntos possam ser discutidos e esclarecidos para a comunidade interna buscando soluções conjuntas.

Portanto, sugere-se, neste trabalho, a criação de fóruns internos que viabilizem a discussão e a construção conjunta de soluções para estudar as alternativas de se tornarem interoperáveis as bases de dados das bibliotecas da FIOCRUZ.

Por fim, percebemos que muito já avançamos em termo de cooperação na área da Biblioteconomia/Ciência da Informação, porém percebemos que, atualmente, mais do que nunca, precisamos estreitar laços, somar forças para que o campo da informação realmente continue a avançar, porém, com eficiência e eficácia, de forma interoperável, e na área da Saúde, salvando vidas.

REFERÊNCIAS BIBLIOGRÁFICAS

REFERÊNCIAS BIBLIOGRÁFICAS

- 1- Capurro, R, Hjørland, B. The concept of information. *Annual Review of information Science & Technology*. 2003; 37 (8): 343-401.
- 2- Konder, L. Informação e conhecimento. *Jornal do Brasil* 2003 maio 17; Caderno Idéias.
- 3- Shneiderman, B, Byrd, D; Croft; WB. Clarifying search: a user-interface framework for text searches. *D-Lib Magazine*, Jan. 1997. Disponível em: <<http://www.dlib.org/dlib/January97/retrieval/01sneiderman.html>>. Acesso em: 22 nov. 2001.
- 4- Castro, RCF. Sistemas e fontes de informação em saúde. São Paulo, 2004. [apostila apresentada no Curso de Mestrado Profissional em Saúde Pública, ENSP/FIOCRUZ, Rio de Janeiro, 19 de janeiro de 2004]
- 5- Miller, P. Interoperability: what is it and why should I want it? Disponível em: <<http://www.ariadne.ac.uk/issue24/intro.html>>. Acesso em: 04 abr. 2004.
- 6- FIOCRUZ. Diretrizes para a formulação do plano quadrienal 2001-2005. [Rio de Janeiro]: s.n, 2001.
- 7- CICT. Centro de Informação Científica e Tecnológica. *Missão do CICT*. Disponível em: <www.cict.fiocruz.br>. Acesso em: 10 jun. 2003.
- 8- CICT. Centro de Informação Científica e Tecnológica. Relatório de atividades do CICT 1999. Rio de Janeiro: CICT/FIOCRUZ, 2000.
- 9- Marcondes, CH, Sayão, LF. Integração e interoperabilidade no acesso a recursos informacionais

eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. *Ci Inf*, Brasília 2001; 30 (3): p. 24-33

10- DeCS. Descritores em Ciências da Saúde. Disponível em: <www.bireme.br> Acesso em: 20 jan 2004.

11- Bochner, R. Acidentes por animais peçonhentos: aspectos históricos, epidemiológicos, ambientais e sócio-econômicos. [Tese de Doutorado]. Rio de Janeiro: Escola nacional de Saúde Pública, Fundação Oswaldo Cruz. 2003.

12- Brasil. Ministério da Saúde. Sistemas de Informação sobre Mortalidade (SIM) e Nascidos Vivos (SINASC) para os profissionais do Programa Saúde da Família. 2. ed. rev. e atual. Brasília, 2004. Disponível em: <http://dtr2001.saude.gov.br/bvs/publicacoes/importancia_sist_inform.pdf> <Acesso em 08 nov. 2004>.

13- Carvalho, DM. Grandes sistemas nacionais de informação em saúde: revisão e discussão da situação atual. *IESUS*, v. 4, out./dez., 1997.

14- Pellanda, NMC. Novos Paradigmas do conhecimento em início de milênio. Disponível em: <<http://pessoal.portoweb.com.br/pellanda/StaTosa.htm>>. Acesso em: 02 mar. 2005.

15- Morin, E. Articular os saberes. In: Alves, N, Garcia, KL, organizador. O sentido da escola. 3. ed. Rio de Janeiro: DP&A, 2000. p. 65-80.

16- Benjamin, W. O Narrador. In: Obras Escolhidas: Magia e Técnica, Arte e Política. - 6. ed. São Paulo: Brasiliense, 1993.

17- Latour, B. Redes que a razão desconhece: laboratórios, bibliotecas, coleções. In: Baratin, M, Jacob, C. (Organizadores). O poder das bibliotecas: a memória dos livros no Ocidente. Rio de Janeiro: UFRJ, 2000.

- 18- Lancaster, FW. Ameaça ou oportunidade? O futuro dos serviços bibliotecários à luz das inovações tecnológicas. *Revista de Biblioteconomia da UFMG*, 1994; 23 (1): p. 7-27, jan./jul.
- 19- Robredo, J, Cunha, MB. Documentação de hoje e de amanhã: uma abordagem automatizada da biblioteca e dos sistemas de informação. 2. ed. São Paulo: Global; 1994.
- 20- Brasil. Comitê Executivo de Governo Brasileiro. E-PING: Padrões de interoperabilidade do Governo Eletrônico: documento de referência versão 0. Brasília, 2004.
- 21- Arms, W. A spectrum of interoperability: the site for science prototype for the NSDL. *Dlib Magazine*, v. 8, n. 1, jan. 2002.
- 22- Marcondes, CH, Sayão, LH. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. *Ci Inf*, Brasília, 2002; 31 (3): 42-54, set./dez.
- 23- Rosetto, Márcia. Metadados e recuperação da informação: padrões para bibliotecas digitais. *Anais da 2ª CIBERÉTICA: Simpósio Internacional de Propriedade Intelectual, Informação e Ética*; 2003 nov 12-14; Florianópolis, Brasil, 2003.
- 24- Rosetto, Márcia. Metadados: novos modelos para descrever recursos de informação digital. *Anais do 1º INTEGRAR: Congresso Internacional de Arquivos, Bibliotecas, Centros de Documentação e Museus*; 2002 mar 17-22; São Paulo, 2002.
- 25- Rosetto, Márcia. O uso do protocolo Z39.50 para a recuperação de informação em redes eletrônicas. *Ci Inf.*, Brasília, 1997; 2 (2), maio/ago.

26- Rosetto, M, Nogueira, AH. Aplicação de elementos metadados dublin core para descrição de dados bibliográficos on-line da biblioteca digital de teses da USP. In: 12º Seminário Nacional de Bibliotecas Universitárias; 2002, Recife. Disponível em: <<http://www.sibi.ufrj.br/snbu/snbu2002/oralpdf/82.a.pdf>>. Acesso em: 20 fev. 2004

27- Branski, RM. Localização de informações na Internet: características e formas de funcionamento dos mecanismos de busca. Transinformação, 2000; 12 (1): 11-19, jan./jun.

28- Cedon, BV. Ferramentas de busca na Web. Ci Inf, Brasília, 2001; 30 (1): 39-49, jan./abr. 2001.

29- Almeida MB, Bax MP. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. Ci Inf, Brasília, 2003; 32 (3): 7-20, set./dez.

30- Campos, MLA. Souza, RF, Campos, MLM. Organização de unidades de conhecimento em hiperdocumentos: o modelo conceitual com espaço comunicacional para a realização da autoria. Ci Inf, Brasília, 2003; 32 (2):7-16, maio/ago.

31- Campos, ML. Organização de unidades de conhecimento em hiperdocumentos: o modelo conceitual com espaço comunicacional para a realização da autoria. [Tese de Doutorado] - CNPQ/IBICT - UFRJ/ECO, Rio de Janeiro, 2001.

32- Suaiden, EJ. O intercâmbio em bibliotecas e centros de documentação. 2. ed. Brasília: [s. n.], 1976. p. 3

33- Miranda, A. A integração de serviços bibliotecários e de informação e o acesso ao documento primário: evolução dos conceitos e situação atual no Brasil. Anais do 2º Seminário Nacional de Comutação Bibliográfica, 1994; Campinas; Brasil, 1994.

- 34- Aguiar, S. Produção compartilhada e socialização do conhecimento em rede: uma abordagem exploratória. Anais do 2º Seminário Nacional do Programa de Pós-graduação em Educação da UFF - Produção do conhecimento e Educação: história, utopias, 2002, Niterói.
- 35- Stair, RM. Princípios de sistemas de informação: uma abordagem gerencial. 2. ed. Rio de Janeiro: LTC, 1998.451p.
- 36- Barbosa, EBM. Uma ferramenta para disseminação de dados científicos do CPTEC/INPE através de um banco de metadados. Faculdade de Engenharia de Guaratinguetá, SP [monografia], 2002.
- 37- Marcondes, Carlos Henrique. *Conceitos básicos*: protocolo de consulta Z39.50. [mimeo].
- 38- Van De Sompel, H, Lagoze, C. "The Santa Fe Convention of the Open Archives Initiative" D-Lib Magazine, v. 6, n. 2, February 2000. Disponível em: <doi:10.1045/february2000-vandesompel-oai>. Acesso em: 20 out 2004.
- 39- Lawrence, S. Free online availability substantially increases a paper's impact: nature web debates. Disponível em: <<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>>. Acesso em: 20 ago.2004.
- 40- The Open Archives Initiative. Disponível em: <<http://www.openarchives.org/>> Acesso em: 17 maio 2004.
- 41- Troll, D, Moen, B. Report to the DLF on the Z39.50 Implementers' Group. DLF: 2001. Disponível em <<http://www.diglib.org/architectures/zig0012.htm>>, Acesso em: 13 mar. 2002.

42- NEC Research Institute. *Finding a needle in the web*. Publicado no Boletim EDUPAGE, em português, (RNP) 05 abr. 1998.

43- Gravano, L, Chang, K, Garcia-Molina, H, Paepcke, Q. Starts: Stanford Protocol for Internet Retrieval and Search, ACM SIGMOD Record 1997; 26 (2): 207-218 jun. Disponível em: <<http://portal.acm.org/citation.cfm?id=253262.253299>>. Acesso em 05 set. 2004.

44- Dublin Core Metadata Initiative (DCMI). Disponível em: <<http://dublincore.org/>>. Acesso 18 maio 2004.

45- Marcondes, CH, Sayão, LF. The Scielo Brazilian Scientific Journal Gateway and Open Archives. D-Lib Magazine 2003; 9(3) mar. Disponível em: <www.slib.org/dlib/march03/marcondes/03marcondes.html> Acesso em: 05 set 2004.

46- Semantic Web. Disponível em <<http://www.sticweb.org/about.html>>. Acesso em: 03 maio 2004.

47- Souza, RR, Alvarenga, L. A web semântica e suas contribuições para a ciência da informação. Ci Inf., Brasília 2004; 33 (1): 132-141 jan./abr.

48- Pacheco, RCS, Kern, VM. Uma ontologia comum para a integração de bases de Ci Inf 2001; 30 (3): 56-63, set. / dez.

49- Lévy, P. Tecnologias da inteligência: o futuro do pensamento na era da informática. São Paulo: Ed. 34, 1993.

50- Brasil. Ministério da Ciência e Tecnologia. Sociedade da Informação: livro verde. Brasília, 2000.

BIBLIOGRAFIA

BIBLIOGRAFIA

Angeloni, MT. Elementos intervenientes na tomada de decisão. *Ciência da Informação*, 2003; 32(1): p. 17-22, jan./abr.

Araújo, JW. Ciência e senso comum: a divulgação do conhecimento no campo da saúde. *Perspectivas em Ciência da Informação*, Belo Horizonte, 2003; n. especial, p. 72-93, jul./dez.

Arellano, MAM. Serviços de referência virtual. *Ciências da Informação*, Brasília, 2001; 30(2): p. 7-15, maio/ago.

Aun, MP. A construção de políticas nacional e supranacional de informação: desafio para os Estados nacionais e blocos regionais. *Ciência da Informação*, Brasília, 1999; 28(2): p. 115-123, maio/ago.

Bergman, MK. The deep web: surface hidden value. *Jal of Eletronic Publishing*, 2001; 7(1), Aug. Disponível em <<http://ww.press.umich.edu/jep/07//bergman.html>>, Acesso em 15 jun. 2002.

BRASIL, Ministério da Saúde. ABRASCO. Uso e disseminação de informações em saúde: subsídios para elaboração de uma política de informações em saúde para o SUS, oficina de trabalho: relatório final. Brasília, 1993.

BRASIL. Ministério da Ciência e Tecnologia. Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT. Programa de Informação e comunicação para pesquisa - PROSSIGA. Biblioteca Digital Brasileira: Programa de implantação biênio 2001-2002. Brasília, 2001.

Carvalho, JM. Cidadania no Brasil: o longo caminho. 4. ed. Rio de Janeiro: Civilização Brasileira, 2003.

Castro, RC. A sociedade em rede. *Perspectivas em Ciência da Informação*, Belo Horizonte, 2003; n. especial, p. 134-145, jul./dez.

Chataignier, MC Pragana, Silva, Margareth Prevot da. Biblioteca digital: a experiência do Impa. . *Ciência da Informação*, Brasília, 2001; 30(3), p. 7-12, set./dez.

Ciência, tecnologia e inovação: desafio para a sociedade brasileira: livro verde. Brasília: Ministério da Ciência e Tecnologia. Academia Brasileira de Ciências, 2001.

CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL: promulgada em 5 de outubro de 1988, atualizada até a Emenda

Constitucional n. 39 de 19 de dezembro de 2002. 1. ed. Rio de Janeiro: América Jurídica, 2003.

Cuenca, AMB, Alvarez, MCA, Ferraz, MLEF, Abdalla, ERF. Capacitação no uso das bases Medline e Lilacs: avaliação de conteúdo, estrutura e metodologia. *Ciência da Informação*, Brasília, 1999; 28(3), p. 340-346, set./dez.

Cunha, MB. As tecnologias de informação e a integração das bibliotecas brasileiras. *Ciência da Informação*, Brasília, 1994; 23(21), p. 182-189, maio/ago.

Cunha, MB. Para saber mais: fontes de informação em ciência e tecnologia. Brasília: Briquet de Lemos/Livros, 2001. 168p.

Davis, JR. Creating a networked computer science technical report library. *D-Lib Magazine*, Sept. 1995. Disponível em: <<http://www.dlib.org/dlib/september95/09davis.html>>. Acesso em: 16 jul. 2001.

Dias, GA. Periódicos eletrônicos: considerações relativas à aceitação deste recurso pelos usuários. *Ciência da Informação*, Brasília, 2002; 31 (3), p. 18-25, set./dez.

DUBIN Core: elementos do núcleo de metadados Dublin Core, versão 1.1: descrição de referência. Disponível em: <www.dublincore.org/documents/1999/07/02/dces/>. Acesso em: 13 de janeiro de 2004.

FIOCRUZ. Diretrizes para a formulação do plano quadrienal 2001-2005. [Rio de Janeiro]: s.n, 2001.

Greenberg, J. Metadata Generation: processes, people and tools. *Bulletin of the American Society for Information Science and Technology*, p. 16- 28, dec./jan. 2003.

Japiassu, H. Interdisciplinaridade e patologia do saber. Rio de Janeiro: Imago, 1976.

Lancaster, FW. Indexação e resumos: teoria e prática. Tradução de Antônio Agenor Briquet de Lemos. Brasília: Briquet de Lemos/Livros, 1993. 347p.

Liu, X. et al. Arc: na OAI service provider for digital library federation. *D-lib magazine*, v. 7, n.4, Apr.2001. Disponível em: <<http://www.dlib.org/dlib/apri01/liu/04liu.html>>. Acesso em: 20 jan 2002

Lopes, IL. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. *Ciência da Informação*, Brasília, 2002; 31(1), p. 41-22, jan./abr.

Lopes, IL. Novos paradigmas para avaliação da qualidade da informação em saúde recuperada na Web. *Ciência da Informação*, Brasília, 2004; 33(1), p. 81-90, jan./abr.

Lynch, C. Metadata Harvesting and the Open Archives Initiative. *ARL Bimonthly Report*, n. 217, Aug. 2001, p.1-9. Disponível em <<http://www.arl.org/newsltr/217/mhp.html>>. Acesso em 18 nov. 2003.

Marcondes, CH, Sayão, LF. The Scielo Brazilian Scientific Journal Gateway and Open Archives: a report on the development of the scielo-open archives data provider Server. *D-lib Magazine*, 2003; 9 (3), mar.

Marconi, MA, Lakatos, EM. *Técnicas de pesquisa*. 2. ed. São Paulo: Atlas, 1990.

Marteleto, RM. Informação da sociedade na Sociedade da Informação. *Perspectivas em Ciência da Informação*, Belo Horizonte, 2003; n. especial, p.4-7, jul./dez.

Masiero, PC, et al. A Biblioteca Digital de teses e dissertações da Universidade de São Paulo. *Ciência da Informação*, Brasília, 2001; 30(3), p. 34-41, set./dez. 2001.

Mattelart, A. *História das teorias da comunicação*. 6. ed. São Paulo: Loyola, 2003.

Mattelart, A. *Comunicação mundo: história das idéias e das estratégias*. 4. ed. Rio de Janeiro: Vozes, 2001.

Minayo, MCS, Deslandes, SF (Org). *Caminhos do pensamento: epistemologia e método*. Rio de Janeiro: Fiocruz, 2002.

Moraes, IHS. *Política, tecnologia e informação em saúde: a utopia da emancipação*. Salvador, BA: Casa da Qualidade, 2002.

Moresi, EAD. *Inteligência organizacional: um referencial integrado*. *Ciência da Informação*, Brasília, 2001; 30 (2), p. 35-46, maio/ago.

Moris, C. Are search engines dead? Disponível em <http://www.stars.com/Internet/Dead_SearchEngines/>. Acesso 17 maio 2004. Search tools: information, guide and news. Disponível em <<http://www.searchtools.com>>. Acesso 17 maio 2004.

Paepke, A. et al. Search middleware and the simple digital library interoperability protoco. *D-Lib Magazine*, v. 6, n. 3, mar. 2000. Disponível em: <<http://www.dlib.org/dlib/march00/paepck/03paepck.html>>. Acesso em: 26 jul. 2001.

Payette, S. et al. Z39.50: the user's perspective. *Dlib Magazine*. Abr. 1997. Disponível em <http://www.dlib.org/dlib/april97/content/04payette.html> . Acesso em: 15 maio 2004.

Portocarrero, V (org.). *Filosofia, História e Sociologia das Ciências: abordagens contemporâneas*. Rio de Janeiro: Fiocruz, 2002.

Rivera, FJU. *Análise estratégica em saúde e gestão pela escuta*. Rio de Janeiro: Fiocruz, 2003.

Sena, NK. Open Archives: caminho alternativo para a comunicação científica. *Ciência da Informação*, Brasília, v. 29, n. 3, p. 71-78, set./dez. 2000.

Severino, AJ. *Metodologia do trabalho científico*. 20. ed. rev. ampl. São Paulo: Cortez, 1996.

Silveira, HFR. Internet, governo e cidadania. *Ciência da Informação*, Brasília, v. 30, n. 2, p.80-90, maio/ago, 2001.

Souza, MIF, Vendrusculo, LG, Melo, GC. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ciência da Informação*, Brasília, 2000; 29(1), p. 93-102, jan./abr.

Souza, RR, Alvarenga, L. A web semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, 2004; 33(1), p. 132-141, jan./abr.

Souza, TB, Catarino, ME, Santos, PC. Metadados: catalogando dados na Internet. *Transinformação*, Campinas, 1997; 9 (2), p.93-105, maio/ago.

Texeira, CMS, Schiel, U. A Internet e seu impacto nos processos de recuperação da informação. *Ciência da Informação*, 1997; 26(1), p.65-71, jan./abr.

THE IMESH toolkit: na architecture and toolkit for distributed subject gateways. *International Digital Libraries*, 1999; n. 19, jan.

Triska, R, Café, L. Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. *Ciência da Informação*, Brasília, 2001; 30(3), p.92-96, set./dez.

Vieira, S. *Metodologia científica: para a área da saúde*. São Paulo: Sarvier, 1984.

Warner, S. Exposing and Harvesting Metadata Using the OAI Metadata Harvesting Protocol: A Tutorial. *High Energy Physics Libraries Webzine*, v. 4, june 2001. Disponível em

<<http://library.cern.ch/HEPL/4/papers/3/>>. Acesso em 18 nov. 2003.

XML schema part 0: primer. W3C Working Draft, 7 April 2000. Disponível em:<<http://www.w3.org/TR/2000WD-xmlschema0-20000407>>. Acesso em: 26 maio 2003.

Z39.50 PROFILES. Disponível em:
<<http://Icweb.Ioc.gov/z39.50/agency>>. Acesso em: 15 maio 2004.

ANEXO



USER DOCUMENTATION

OAI DATA PROVIDER

Ex Libris

© Ex Libris Ltd., 2002, 2004

Version 16

Document Revision 1.2

Last Update: July 8, 2004

Table of Contents

OVERVIEW	3
OAI DATA PROVIDER CONFIGURATION.....	3
HOW TO RUN THE OAI DATA PROVIDER	4
PROTOCOL REQUESTS AND RESPONSES.....	4
GetRecord	4
Identify	4
ListIdentifiers	5
ListMetadataFormats	5
ListRecords	6
ListSets.....	6
PROTOCOL ISSUES	6
Datestamp Granularity	6
Sets	6
Metadata Formats.....	6

Overview

The Ex Libris OAI (Open Archive Initiative) Data Provider implements the OAI Protocol Version 2.0. It supports 6 OAI-PMH requests specified by the OAI protocol.

For more information, refer to the OAI Web site: <http://www.openarchives.org/>.

OAI Data Provider Configuration

The configuration files are in the `./alephe/tab/oai` directory.

The files are:

- `oaiconf.xml` – an XML file with OAI Data Provider settings
- `oaiconf.dtd` – the DTD of `oaiconf.xml`

oaiconf.xml

The `oaiconf.xml` file contains the following data:

1. `sets available` (shown in ListSets response, can be used in ListRecords and ListIdentifiers requests)
2. `timeZone` (for example, `-00.00`, used in protocol response to specify `responseDate`)
3. `repositoryName` (used in Identify response)
4. `baseURL` (used in Identify response)
5. `adminEmail` (used in Identify response)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE oaiconf SYSTEM "oaiconf.dtd">
<oiroot>
<set>
  <setSpec>USM01</setSpec>
  <setName>USM01 Test library</setName>
</set>
<set>
  <setSpec>UNI01</setSpec>
  <setName>UNI01 Test library</setName>
</set>
<timeZone>-00:00</timeZone>
<repositoryName>ExLibris Israel</repositoryName>
<baseURL>http://ram43:8991/OAI-script</baseURL>
<adminEmail>Irina.Dijour@exlibris.co.il</adminEmail>
</oiroot>
```

oaiconf.dtd

```
<!-- DTD for oaiconf.xml -->

<!ELEMENT oiroot set*>
<!ELEMENT set (setName, setSpec)>
<!ELEMENT timeZone (#PCDATA)>
<!ELEMENT repositoryName (#PCDATA)>
<!ELEMENT baseURL (#PCDATA)>
<!ELEMENT adminEmail (#PCDATA)>
```

How to Run the OAI Data Provider

The OAI data provider is part of our Web server.

To access OAI data, type the following in the address bar of your Web browser
<www host>:<apache port>/OAI-script?<oai request>

where <www host> is the Web server host, <apache port> is the apache port used by the Web server and <oai request> is the OAI protocol request.

For example:

<http://ram43:8991/OAI-script?verb=ListSets>

Note: if the OAI configuration in `./alephe/tab/oai` is changed, the Web server should be restarted.

Protocol Requests and Responses

GetRecord

Summary

Retrieves a specified record in a specified format from the repository.

Arguments

- **identifier** (required): has the form <base>-<doc number> (for example, USM01-500)
- **metadataPrefix** (required): supported values – oai_dc, marc21

Request Example

http://ram43:8991/OAI-script?verb=GetRecord&identifier=USM01-500&metadataPrefix=oai_dc

Identify

Summary

Retrieves information about the repository

Arguments

None

Request Example

<http://ram43:8991/OAI-script?verb=Identify>

ListIdentifiers

Summary

Retrieves identifiers of records in the repository.

Arguments

- ❑ **until** (optional): record modification date (for example, 2001-01-01)
- ❑ **from** (optional): record modification date (for example, 2001-01-01)
- ❑ **metadataPrefix** (required): supported values – oai_dc, marc21; only identifiers for records which can be retrieved in the specified format are shown.
- ❑ **set** (optional): set (e.g. USM01), default is USM01
- ❑ **resumptionToken** (exclusive)

Request Examples

<http://ram43:8991/OAI-script?verb=ListIdentifiers>

http://ram43:8991/OAI-script?verb=ListIdentifiers&from=1997-04-02&until=1999-02-01&set=UNI01&metadataPrefix=oai_dc

<http://ram43:8991/OAI-script?verb=ListIdentifiers&resumptionToken=USM01@036983001000000030>

ListMetadataFormats

Summary

Retrieves available metadata formats from the repository.

Ex Libris supports:

oai_dc (http://www.openarchives.org/OAI/2.0/oai_dc/oai_dc.xsd)

marc21 (<http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd>)

Arguments

- ❑ **identifier** (optional): retrieve metadata formats available for specified record

Request Examples

<http://ram43:8991/OAI-script?verb=ListMetadataFormats>

<http://ram43:8991/OAI-script?verb=ListMetadataFormats&identifier=USM01-500>

ListRecords

Summary

Retrieves records from the repository.

Arguments

- ❑ **until** (optional): record modification date (for example, 2001-01-01)
- ❑ **from** (optional): record modification date (for example, 2001-01-01)
- ❑ **metadataPrefix** (required): supported values – oai_dc, marc21; only records that can be retrieved in the specified format are shown.
- ❑ **set** (optional): set (for example, USM01), default is USM01
- ❑ **resumptionToken** (exclusive):

Request Examples:

http://ram43:8991/OAI-script?verb=ListRecords&metadataPrefix=oai_dc

http://ram43:8991/OAI-script?verb=ListRecord&from=1997-04-02&until=1999-02-01&set=UNI01&metadataPrefix=oai_dc

<http://ram43:8991/OAIscript?verb=ListIdentifiers&resumptionToken=UNI01@036984001000000030>

ListSets

Summary

Retrieves the list of sets from the repository.

Arguments

- ❑ **resumptionToken** (exclusive)

Request Examples:

<http://ram43:8991/OAI-script?verb=ListSets>

Protocol Issues

Datestamp Granularity

Ex Libris supports YYYY-MM-DD granularity for Datestamps

Sets

Ex Libris supports sets. We have flat set organization. Any physical library (specified in `tab_base`) can be used as a set.

Metadata Formats

oai_dc (http://www.openarchives.org/OAI/2.0/oai_dc/oai_dc.xsd)

marc21 (<http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd>)