

PATRÍCIA DE ANDRADE BUENO GARCIA

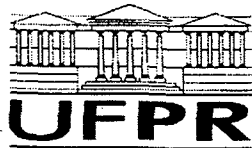
**PROVEDORES DE DADOS DE BAIXO CUSTO:
PUBLICAÇÃO DIGITAL AO ALCANCE DE TODOS**

Dissertação apresentada como requisito parcial
à obtenção do grau de Mestre em Informática,
Curso de Pós-Graduação em Informática, Setor
de Ciências Exatas, Universidade Federal do
Paraná.

Orientador: Prof. Dr. Marcos S. Sunye

CURITIBA

2003




Ministério da Educação
Universidade Federal do Paraná
Mestrado em Informática


PARECER

Nós, abaixo assinados, membros da Banca Examinadora da defesa de Dissertação de Mestrado em Informática, da aluna *Patrícia de Andrade Bueno Garcia*, avaliamos o trabalho intitulado, "*Provedores de Dados de Baixo Custo: Publicação Digital ao Alcance de Todos*", cuja defesa foi realizada no dia 24 de julho de 2003, às quatorze horas, no Auditório da Informática da Universidade Federal do Paraná. Após a avaliação, decidimos pela aprovação do candidato.

Curitiba, 24 de julho de 2003.


Prof. Dr. Marcos Sfair Sunyê
DINF/UFPR (Orientador)


Prof. Dr. Luis Fernando Sayão
CNEN


Profª. Dra. Laura Sanchez Garcia
DINF/UFPR



Ao meu marido André e ao meu filho Victor, pela felicidade de todos os dias.

Aos meus pais, meu irmão e meus avós, fontes infinitas de amor.

AGRADECIMENTOS

Ao Prof. Dr. Marcos Sunye, pela orientação precisa em tempo recorde.

Ao Prof. Dr. Michel Gagnon, pela orientação até o ano passado.

Ao Prof. Dr. Alexandre Direne, por seu incentivo que fez diferença.

Ao Leonardo Boiko, pela ajuda na instalação do Eprints.

Aos funcionários e professores do Departamento de Informática, pela atenção inestimável e pela contribuição.

À comunidade Eprints/OAI no Brasil, que embora pequena, mostrou-se gigante na boa-vontade: Hernani e Evelyn (USP), Eustáquio (IBICT), Profs. Marcondes e Sayão (RJ).

A todos aqueles que direta ou indiretamente contribuíram para a realização deste trabalho.

*O conhecimento compartilhado é uma semente lançada
nos campos férteis das mentes abertas.*

Patrícia de Andrade Bueno Garcia

SUMÁRIO

LISTA DE FIGURAS	VIII
LISTA DE TABELAS.....	IX
LISTA DE SIGLAS	X
RESUMO	XII
ABSTRACT	XIII
1. INTRODUÇÃO	1
1.1. CONCEITOS	1
1.1.1. <i>Bibliotecas Digitais</i>	1
1.1.2. <i>Interoperabilidade</i>	1
1.1.3. <i>Open Archive</i>	2
1.2. MOTIVAÇÃO	2
1.3 OBJETIVO	1
1.4. METODOLOGIA	6
2. A INICIATIVA OPEN ARCHIVES	8
2.1. INTRODUÇÃO	8
2.2. OS PROVEDORES DE DADOS.....	9
2.3. OS PROVEDORES DE SERVIÇOS.....	10
2.4. O PROTOCOLO OAI-PMH.....	12
2.4.1. <i>Histórico</i>	12
2.4.2. <i>Características</i>	13
2.4.2.1. <i>A Linguagem XML</i>	13
2.4.2.2. <i>Harvesting</i>	14
2.4.2.3. <i>Verbos do OAI-PMH</i>	15
2.5. Os PROGRESSOS DA INICIATIVA OPEN ARCHIVES.....	21
2.5.1. <i>Principais Programas e Projetos de Pesquisa</i>	22
2.6. O FUTURO DA OAI	23
2.7. OUTRAS INICIATIVAS DA ÁREA.....	23
2.7.1. <i>Self-Archiving Initiative</i>	23
2.7.2. <i>BOAI (Budapest Open Access Initiative)</i>	24
2.7.3. <i>Dublin Core Metadata Initiative</i>	24
3. AVALIAÇÃO DOS SOFTWARES PARA PUBLICAÇÃO DIGITAL BASEADOS NA OAI.....	27
3.1. GNU EPRINTS 2.....	28
3.1.1. <i>Finalidade</i>	28
3.1.2. <i>Disponibilidade</i>	28
3.1.3. <i>Utilização</i>	28
3.1.3.1 <i>Eprints nas Instituições</i>	28
3.1.4. <i>Arquitetura do Eprints</i>	30
3.1.4.1. <i>Estrutura do Registro (eprint)</i>	30

3.1.5. <i>Funcionalidade</i>	31
3.1.5.1 <i>Áreas Públicas</i>	33
3.1.5.2. <i>Áreas Restritas</i>	39
3.1.5.3. <i>Administração</i>	43
3.1.6. <i>Protocolo OAI-PMH</i>	44
3.1.7. <i>Custos</i>	44
3.1.7.1. <i>Custos de Software</i>	44
3.1.7.2. <i>Custos de Hardware</i>	44
3.1.7.3. <i>Custos de Mão-de-Obra</i>	45
3.2. KEPLER	45
3.2.1. <i>Finalidade</i>	45
3.2.2. <i>Arquitetura</i>	46
3.2.2.1. <i>A Arquitetura Inicial do Kepler</i>	46
3.2.2.2. <i>O Projeto EKF</i>	48
3.2.2.3. <i>Por que não usar o EKF agora?</i>	50
3.2.3. <i>Funcionalidade</i>	50
3.2.3.1. <i>Biblioteca Digital Comunitária</i>	50
3.2.4. <i>Disponibilidade</i>	51
3.2.5. <i>Utilização</i>	51
3.2.6. <i>Protocolo OAI-PMH</i>	51
3.2.6.1. <i>Modelos Push/Pull e Híbrido</i>	51
3.2.6.2. <i>Extensão ao OAI-PMH</i>	53
3.2.7. <i>Custos</i>	54
3.3. DSPACE	55
3.3.1. <i>Finalidade</i>	55
3.3.2. <i>Disponibilidade</i>	55
3.3.3. <i>Utilização</i>	56
3.3.4. <i>Arquitetura</i>	56
3.3.5. <i>Funcionalidade</i>	56
3.3.6. <i>Protocolo OAI-PMH</i>	58
3.3.7. <i>Custos</i>	58
3.4. CONCLUSÕES	59
3.4.1 <i>Resumo</i>	59
4. UFPR EPRINTS	62
4.1. PROCEDIMENTOS ANTERIORES À INSTALAÇÃO	62
4.1.1. <i>Finalidade do Repositório (Archive)</i>	62
4.1.2. <i>Definição dos Tipos e Formatos de Documentos</i>	63
4.1.3. <i>Procedimento de Submissão / Supervisão</i>	63
4.1.4. <i>Metadados</i>	64
4.1.4.1. <i>Regulamentação do Uso do Padrão Dublin Core no Eprints</i>	64
4.1.5. <i>Texto Completo</i>	65
4.2. A IMPLANTAÇÃO DO EPRINTS	65
4.2.1. <i>Obtenção do Software</i>	65
4.2.2. <i>Documentação</i>	65
4.2.3. <i>Listas de Discussão</i>	65
4.2.4. <i>Requisitos de Hardware</i>	66
4.2.5. <i>Requisitos de Software</i>	66
4.2.6. <i>Instalação</i>	67

4.2.7. Início.....	67
4.2.8. Tradução e Customização.....	67
4.2.9. Os Registros nas Comunidades Eprints e OAI.....	67
4.2.9.1. Registro na Eprints.org.....	68
4.2.9.2. Validação do Repositório.....	68
4.2.9.3. Registro na Iniciativa Open Archives.....	68
4.2.10. O Maior Desafio: dar "vida" ao repositório.....	69
4.2.10.1. O Envolvimento.....	70
4.2.10.2. Apresentação e Treinamento.....	71
4.2.10.3. Adesão de líderes da comunidade.....	71
4.2.10.4. Suporte aos usuários.....	71
4.2.10.5. Operacionalização e Divulgação.....	71
4.2.11. Custos.....	72
4.2.12. Conclusão.....	72
5. SOLUÇÕES EPRINTS/OAI NO BRASIL.....	74
5.1. BDB/IBICT.....	74
5.1.1. Biblioteca Digital de Teses e Dissertações (BDTD).....	75
5.2. SBG/IBICT.....	76
5.3. CORUJA.....	76
5.4. SCIELO.....	77
5.5. EDIÇÃO ELETRÔNICA DE ANAIS DE EVENTOS.....	78
5.5.1. Funcionalidades.....	79
5.5.2. Interfaces.....	81
6. CONCLUSÃO.....	83
7. TRABALHOS FUTUROS.....	85
7.1. ESPECIFICAMENTE EM RELAÇÃO AO EPRINTS.....	85
7.2. DENTRE AS TRÊS FERRAMENTAS.....	85
REFERÊNCIAS BIBLIOGRÁFICAS.....	87
ANEXO I - OAI DATA PROVIDERS.....	91
ANEXO II - REPOSITÓRIOS EPRINTS.....	94
ANEXO III - UFPR EPRINTS: TESTE DE CONFORMIDADE OAI.....	98
ANEXO IV - UFPR EPRINTS: REGISTRO OAI.....	102
ANEXO V - RECOMENDAÇÕES PARA USO DO DUBLIN CORE NO EPRINTS.....	104

LISTA DE FIGURAS

FIGURA 1 - ENCURTAMENTO DO CICLO DA COMUNICAÇÃO CIENTÍFICA.....	5
FIGURA 2 - HARVESTING	14
FIGURA 3 - OAI-PMH - RESPOSTA AO VERBO IDENTIFY	16
FIGURA 4 - OAI-PMH - RESPOSTA AO VERBO LISTMETADATA FORMATS.....	17
FIGURA 5 - OAI-PMH - RESPOSTA AO VERBO GETRECORD	18
FIGURA 6 - OAI-PMH - RESPOSTA AO VERBO LISTRECORDS.....	19
FIGURA 7 - OAI-PMH - RESPOSTA AO VERBO LISTIDENTIFIERS	20
FIGURA 8 - OAI-PMH - RESPOSTA AO VERBO LISTSETS	21
FIGURA 9 - EPRINTS - PRINCIPAIS FUNCIONALIDADES.....	32
FIGURA 10 - EPRINTS - TELA INICIAL.....	34
FIGURA 11 - EPRINTS - OPÇÕES DE LISTAGEM DE REGISTROS	34
FIGURA 12 - EPRINTS - LISTAGEM POR ASSUNTO.....	35
FIGURA 13 - EPRINTS - LISTAGEM POR ANO	35
FIGURA 14 - EPRINTS - PESQUISA SIMPLES.....	36
FIGURA 15 - EPRINTS - PESQUISA AVANÇADA.....	37
FIGURA 16 - EPRINTS - RESULTADO DA PESQUISA.....	37
FIGURA 17 - EPRINTS - METADADOS DO REGISTRO	38
FIGURA 18 - EPRINTS - TEXTO COMPLETO DO REGISTRO	39
FIGURA 19 - EPRINTS - CADASTRO DE NOVO USUÁRIO	40
FIGURA 20 - EPRINTS - ÁREA DO USUÁRIO	41
FIGURA 21 - EPRINTS - AGENDAMENTOS	42
FIGURA 22 - KEPLER - ARQUITETURA INICIAL.....	47
FIGURA 23 - KEPLER - ARQUITETURA EKF	49
FIGURA 24- KEPLER - MODELOS DE IMPLEMENTAÇÃO DE SINCRONISMO ENTRE SERVIDORES	53
FIGURA 25 - DSPACE - MODELO DE DADOS	57
FIGURA 26- OAISTER - PROVEDOR DE SERVIÇOS MOSTRANDO UFPR EPRINTS	69
FIGURA 27- INTERFACES DO SISTEMA DE EDIÇÃO DE ANAIS DE CONGRESSO	82

LISTA DE TABELAS

TABELA 1 - PROVEDORES DE SERVIÇOS OAI	11
TABELA 2 - PADRÕES DE ACESSO XML	14
TABELA 3 - DSPACE - EXEMPLOS DE OBJETOS DO MODELO DE DADOS	58
TABELA 4 - RESUMO DA ANÁLISE DAS FERRAMENTAS	60
TABELA 5 - ANEXO I - PROVEDORES DE DADOS OAI	93
TABELA 6 - ANEXO II - REPOSITÓRIOS EPRINTS 2.....	96
TABELA 7 - ANEXO II - REPOSITÓRIOS EPRINTS 1.....	97

LISTA DE SIGLAS

ASCII	<i>American Standard Code for Information Interchange</i>
BD	Banco de Dados
BDB	Biblioteca Digital Brasileira
BDTD	Biblioteca Digital de Teses e Dissertações
BIREME	Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde
BOAI	<i>Budapest Open Access Initiative</i>
BSD	<i>Berkeley Software Distribution</i>
CIN	Centro de Informações Nucleares
CNEN	Comissão Nacional de Energia Nuclear
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DCMI	<i>Dublin Core Metadata Initiative</i>
DL	<i>Digital Library</i>
DTD	<i>Document Type Definition</i>
EKF	<i>Enhanced Kepler Framework</i>
GNU	<i>GNU is Not Unix</i>
GPL	<i>General Public License</i>
HP	<i>Hewlett-Packard</i>
HTML	<i>Hiper-Text Markup Language</i>
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
LILACS	<i>Latin America and Caribbean Health Science Literature</i>
MIT	<i>Massachussets Institute of Technology</i>
NSDL	<i>The National Science Digital Library</i>
OAI	<i>Open Archives Initiative</i>
OAI-PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>

OCLC	<i>Online Computer Library Center</i>
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-Americana da Saúde
OSI	<i>Open Society Institute</i>
PDF	<i>Portable Document Format</i>
PERL	<i>Practical Extraction and Report Language</i>
SBG	Sociedade Brasileira de Genética
SciELO	<i>Scientific Electronic Library Online</i>
UFPR	Universidade Federal do Paraná
USP	Universidade de São Paulo
XML	<i>eXtended Markup Language</i>
W3C	<i>World Wide Web Consortium</i>

RESUMO

Atualmente, para uma Instituição de Ensino e/ou Pesquisa, tão importante quanto ter uma produção acadêmica/científica de qualidade, é ter também uma maneira eficiente de divulgar estes trabalhos. Até há pouco tempo, isto era privilégio de poucas Instituições, dado o custo de uma biblioteca digital onde pudessem ser publicadas as teses, dissertações, monografias ou os artigos destas comunidades.

Com o advento de iniciativas voltadas para a divulgação científica, dentre elas a Iniciativa Open Archives (OAI - *Open Archives Initiative*), surgiram intensos debates e pesquisas na área, e, recentemente, programas para a implementação de repositórios digitais baseados em software livre, onde os próprios usuários podem publicar seus trabalhos de maneira padronizada e interoperável, baseada no protocolo da OAI (OAI-PMH).

Neste trabalho é apresentada a Iniciativa Open Archives, o protocolo OAI-PMH, e é feita uma análise dos programas baseados na OAI para a implementação de repositórios de dados (também chamados provedores de dados) para verificar os benefícios, funcionalidades, e custos efetivos de implantação, tendo por objetivo gerar um provedor de dados de baixo custo. Dentre os programas disponíveis (*Kepler*, *DSpace* e *Eprints*), o *Eprints* foi implantado como repositório da produção científica do Departamento de Informática da UFPR (denominado **UFPR Eprints**) com a finalidade de comprovar as informações levantadas. Com a experiência desta implantação, também são sugeridas ações para envolver e incentivar a comunidade do *Eprints* a utilizar o programa e publicar seus trabalhos.

Este trabalho também pode ser usado como referência para as Instituições que desejam implementar repositórios digitais para abrigar artigos ou teses e dissertações, com baixo custo e interoperabilidade.

Palavras-Chaves: Bibliotecas Digitais, Eprints, Iniciativa Arquivos Abertos, Provedores de Dados.

ABSTRACT

Nowadays, for a University or Research Institute, an efficient way to show the academics/scientific production is so important as quality. Until few years ago, these was a privilege for some Institutions, because of development costs for a digital library, place where articles, theses and dissertations could be published.

By the emerging initiatives for scientific publishing, i.e. Open Archives Initiative (OAI), many debates and researches has appeared in this area, and recently, softwares for creating digital repositories (free software based), where users can publish their works in standardized and interoperable way by themselves, based on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

At this work, the Open Archives Initiative and its Protocol are presented, and an analysis of the softwares for data providers OAI based was done, to verify the benefits, functionalities and effective implementation costs, aiming to obtain a low-cost data provider. Among the softwares available (Kepler, DSpace and Eprints), the Eprints was installed as a UFPR Computer Science Department's Repository (called **UFPR Eprints**) to validate the informations obtained. With this experience, some actions are suggested to involve the Eprints' community by using the data provider and publishing their documents.

Also, this work can be used as a reference for Institutions who wants to implement digital repositories for their articles, theses or dissertations, with low-costs and interoperability.

Key-Words: Digital Libraries, Eprints, Open Archives Initiative, Data Providers.

1. INTRODUÇÃO

Muitas idéias desenvolvem-se melhor quando transplantadas para outras mentes do que naquelas em que foram concebidas.
Oliver Wendell Holmes
1809-1894

Algumas pessoas, ao serem informadas do tema deste trabalho (Bibliotecas Digitais), demonstravam estranheza, e uma delas interpelou: "Onde ficam as Bibliotecas Digitais?". Excetuando um pequeno grupo de pessoas ligadas à informática e/ou à biblioteconomia, as pessoas quase não conhecem o assunto, ainda pouco difundido em nossa cultura.

1.1. CONCEITOS

Por este motivo, antes de começar a discutir o problema e a solução proposta, é importante apresentar os principais conceitos para melhor compreensão do contexto deste trabalho. Após pesquisa em vários artigos, trabalhos e páginas oficiais de entidades, as definições mais claras e suficientes para este propósito foram encontradas na tese de Hussein Suleman [48]:

1.1.1. Bibliotecas Digitais

"Uma Biblioteca Digital é um sistema eletrônico de armazenamento de informações com foco na localização das informações necessárias a seus usuários."
[48]

1.1.2. Interoperabilidade

"Habilidade de um sistema utilizar as partes ou equipamentos de outro sistema". [48]

No contexto de bibliotecas digitais, a interoperabilidade refere-se à habilidade de uma biblioteca digital trabalhar cooperativamente com outras com o intuito de oferecer melhores serviços aos usuários. Existem várias maneiras de se atingir um grau de interoperabilidade, e uma delas é a criação e uso de *Open Archives* (Arquivos Abertos).

1.1.3. *Open Archive*

"É uma interface de computador para acessar uma coleção de dados, que esteja em conformidade com as especificações ditadas pelo Protocolo para Colheita de Metadados da Iniciativa Open Archives". [48]

A Iniciativa *Open Archives* é uma organização formada por pesquisadores, bibliotecários, e outros profissionais cujo objetivo é criar padrões simples para promover a interoperabilidade entre sistemas. É a criadora do OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting* - Protocolo para Colheita de Metadados), que especifica como dois sistemas de computadores podem se comunicar através de um conjunto de registros estruturados e de maneira contínua. Um sistema que contém dados para compartilhamento e está em conformidade com este protocolo é chamado de *Open Archive*.

1.2. MOTIVAÇÃO

Os professores e pesquisadores de países em desenvolvimento, que contam com recursos escassos para a manutenção de suas Universidades e Instituições de Pesquisa, têm dificuldade de acesso à literatura científica tradicional de sua área, sob a forma de periódicos, muitas vezes altamente especializados e com assinaturas caras. Recentemente, com o surgimento dos periódicos eletrônicos, esse problema ainda persiste, pois, muitos só oferecem acesso aos artigos mediante pagamento.

A cada dia surgem bibliotecas digitais e bases de dados públicas que se constituem de importante fonte de informação para pesquisadores. O problema é o tempo que se gasta para reunir informações relevantes a um dado assunto de pesquisa, seja visitando diversos portais de bibliotecas virtuais, seja utilizando a busca convencional oferecida por *sites* como *Google*, *AltaVista*, *Cadê*, etc.

A pesquisa através de portais de busca tradicionais é imprecisa e atinge apenas as páginas HTML, ignorando as bases de dados que se encontram por trás de algumas destas páginas.

Por outro lado, ao tentar divulgar seus trabalhos através dos periódicos, em busca do impacto de suas pesquisas, os pesquisadores esbarram em um processo burocrático e demorado, onde, desde a submissão até a publicação, devido à lenta arbitragem que por vezes ocorre, pode haver um espaço de tempo tão longo que os efeitos daquela pesquisa já não tenham valor quando de sua publicação.

Ao citarmos todas estas dificuldades em relação aos periódicos, nem entramos no mérito do direito autoral, onde, para conseguir contribuir com a Ciência, os autores abrem mão dos direitos sobre seus trabalhos em benefício dos editores.

As Universidades e Centros de Pesquisa são grandes geradores de conhecimento, no entanto, a disponibilização do conteúdo está sujeita a uma infraestrutura básica, muitas vezes, de difícil acesso.

Em resposta a todos estes problemas, em 1991 surge o arXiv [11], um repositório de artigos inicialmente da área de Física, onde pesquisadores do mundo todo começaram a depositar e compartilhar artigos, e que se tornou uma referência em consultas nesta área. Em paralelo, vários pesquisadores começaram a publicar seus trabalhos na Internet, culminando na organização da Iniciativa *Self-Archiving* [1], por parte de Stevan Harnad.

Com a evolução da Internet, várias bibliotecas digitais começaram a surgir, baseadas em grandes orçamentos e longos projetos, algumas com a finalidade de expor a produção de teses e dissertações das grandes universidades dos Estados Unidos e Europa.

Tudo isso resultou em um grande avanço, onde informações científicas e acadêmicas já poderiam ser obtidas livremente pela Internet, e disponibilizadas, através da publicação nas páginas de seus autores. Os mecanismos de publicação direta na internet passaram a ser uma alternativa de pesquisa e visibilidade para a comunidade acadêmica.

Para um pesquisador, seu maior reconhecimento vem em forma de citações, que demonstram a utilização de seu trabalho pela comunidade científica. Steve Lawrence [47] realizou uma pesquisa com quase 120 mil publicações e concluiu que "*o número médio de citações de artigos offline é 2.74, enquanto o número médio de citações de artigos online é 7.03, um aumento de 157%*". Isto demonstra a importância e a força das publicações online, porém, no mesmo artigo, Lawrence nos lembra que existem mais de um milhão de artigos científicos disponíveis na rede, e o acesso a estes artigos pode ser um problema devido à falta de padrões para a publicação e pesquisa.

Inúmeras discussões sobre esta falta de padrões para disponibilização e pesquisa de informações científicas na Internet, levaram à criação da Iniciativa *Open Archives* (Arquivos Abertos) e ao desenvolvimento de um protocolo com o intuito de oferecer simplicidade e eficiência na tarefa de unificar as consultas a bases de dados científicas/acadêmicas.

Com os recursos oferecidos pela Iniciativa *Open Archives*, é possível melhorar significativamente a precisão das consultas eletrônicas e reduzir o tempo de procura, graças ao compartilhamento de informações (metadados) entre os participantes da Iniciativa. Isto proporciona um encurtamento do ciclo da comunicação científica, conforme apresenta a Figura 1 de Marcondes e Sayão [42]:

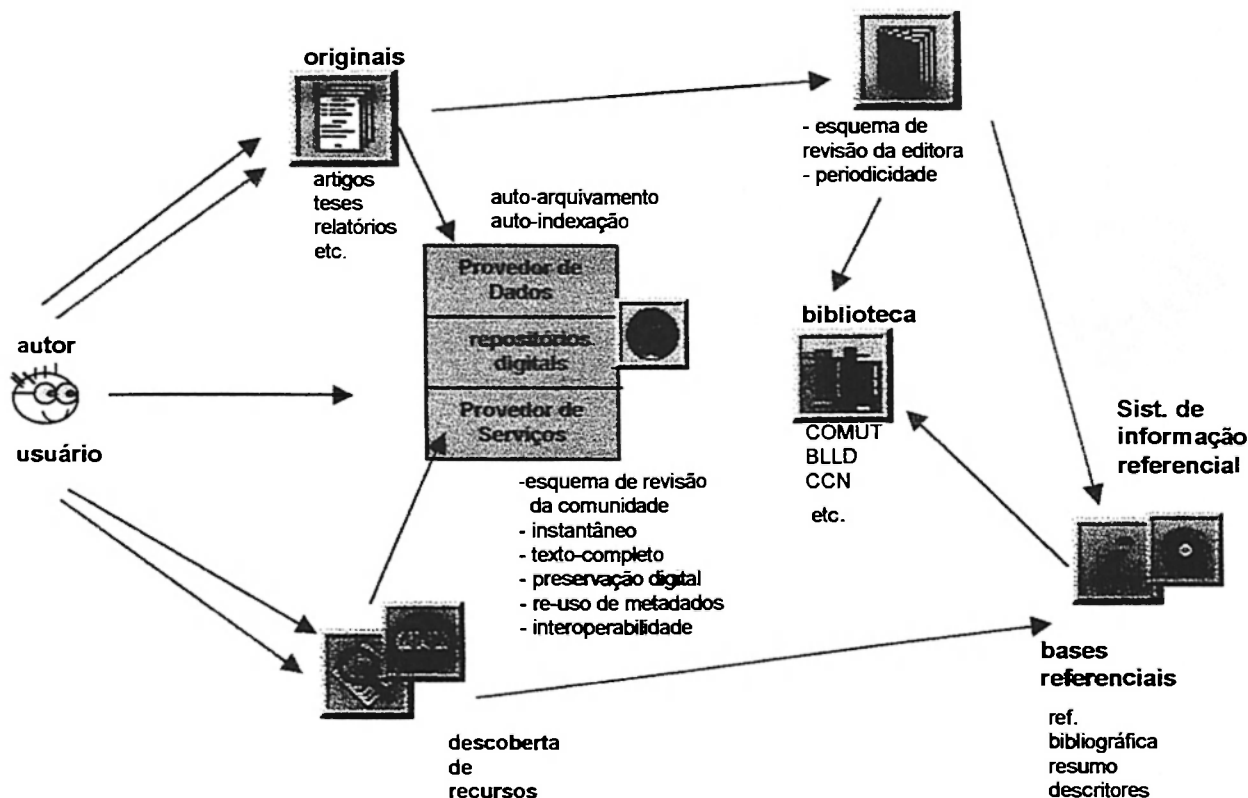


Figura 1 – Encurtamento do Ciclo da Comunicação Científica

Anteriormente à OAI, o usuário/autor escrevia seus documentos (artigos, teses, etc), e, em busca de reconhecimento, enviava-o para uma revisão e aguardava a sua publicação. Após a publicação, o documento normalmente era arquivado em uma biblioteca e, para acessá-lo era necessário dirigir-se à biblioteca e localizá-lo. Com a nova tecnologia proporcionada pela OAI, o autor, imediatamente após a conclusão de seu documento, pode (ele mesmo) depositar o arquivo em um provedor de dados e o documento torna-se disponível quase imediatamente para revisão pela comunidade, em texto completo e de fácil consulta através dos provedores de serviço da Iniciativa *Open Archives*.

1.3 OBJETIVO

Apesar da abrangência da Iniciativa *Open Archives* e da relativa estabilidade da tecnologia empregada no protocolo de coleta, há pouca literatura que oriente as instituições nos riscos e custos associados a esta iniciativa. Não há critérios de análise do protocolo ou das ferramentas para criação de provedores de dados, nem tampouco sobre os custos embutidos na sua implantação.

O presente trabalho tem por objetivo analisar as soluções baseadas na OAI para a construção de provedores de dados, visando identificar opções de baixo custo viáveis como alternativa de divulgação de trabalhos para a área acadêmica/científica. A efetiva implementação permitiu uma análise detalhada do processo de implantação de um provedor de dados baseado em software livre, onde buscou-se comprovar a viabilidade da construção de bibliotecas digitais de baixo custo.

1.4. METODOLOGIA

A metodologia para o levantamento de informações baseou-se na bibliografia existente, tal como artigos, manuais e periódicos científicos, e também nas informações constantes das páginas oficiais das entidades citadas.

Para a efetiva comprovação da viabilidade, custos e demais informações da ferramenta Eprints, foi realizada sua implantação sob a forma do projeto-piloto **UFPR Eprints**, repositório do Departamento de Informática da UFPR, onde foi possível validar o que foi levantado, bem como realizar análises e sugerir medidas para a divulgação dos repositórios, e aumento das adesões da comunidade.

O trabalho está estruturado da seguinte maneira:

O capítulo 2 apresenta a Iniciativa *Open Archives* (OAI), seu histórico e evolução, o que são provedores de dados e provedores de serviços, o protocolo OAI-PMH, e outras Iniciativas da área de publicação digital.

Em seguida, o capítulo 3 analisa as ferramentas para criação de provedores de dados: *Eprints*, *Kepler* e *DSpace*. São descritas suas principais funcionalidades, arquitetura, utilização, custos e outros critérios de análise.

No capítulo 4 é apresentada a experiência de instalação do *Eprints* como repositório do Departamento de Informática da UFPR (**UFPR Eprints**), com o intuito de se validar as informações levantadas no capítulo anterior.

O capítulo 5 apresenta os projetos brasileiros que implementam o protocolo da Iniciativa *Open Archives*.

No capítulo 6 são apresentadas as conclusões do presente trabalho e no capítulo 7, os trabalhos futuros.

2. A INICIATIVA OPEN ARCHIVES

Nosso conhecimento é a coleção de pensamentos e experiências de inúmeras mentes.
Ralph Waldo Emerson
1803-1882

2.1. INTRODUÇÃO

A Iniciativa *Open Archives* (Arquivos Abertos) teve suas raízes na comunidade *E-Prints* e surgiu a partir da Convenção de Santa Fé, no final de 1999 com o intuito de "*desenvolver e promover soluções de interoperabilidade que facilitem uma disseminação eficiente do conteúdo*" [12].

Estes padrões e estruturas que estão sendo continuamente desenvolvidos pela OAI têm se tornado muito importantes para implementar a interoperabilidade e disponibilizar acesso a uma grande quantidade de artigos, teses, dissertações e outros documentos arquivados em repositórios na Internet que se utilizam deste padrão. Os padrões desenvolvidos independem do conteúdo que é disponibilizado, bem como de questões econômicas que possam restringir o acesso aos documentos armazenados.

O termo "*Archive*" é usado no sentido de repositório para armazenamento de informações, indo além da proposta de auto-publicação da comunidade *Eprints* [1, 18]. Os *Eprints* podem ser interpretados tanto como qualquer documento digital ou como, "*textos digitais de artigos revisados, antes e depois da arbitragem e publicação. Antes da arbitragem e publicação, o documento é chamado de 'preprint'. Após a arbitragem e publicação, esta versão final é chamada 'postprint'*" [18].

Já o termo "*Open*" significa "aberto" do ponto de vista da arquitetura da proposta, não significando acesso gratuito ou ilimitado às informações constantes dos repositórios que fazem parte da Iniciativa. Em resumo, o protocolo da *Open Archives* é que é aberto.

O livre acesso a documentos e questões envolvendo cobrança de valores, direitos autorais e outras questões, têm sido amplamente discutidos em outras Iniciativas (ver 2.7. - Outras Iniciativas).

A OAI é formada por um comitê diretor, que é composto por conceituadas Universidades e outras entidades com interesse na disponibilização digital de informações e por um comitê técnico presidido por Carl Lagoze e Herbert Van de Sompel [23]. A Iniciativa é patrocinada por *Digital Library Federation*, *Coalition for Networked Information* e, *Natural Science Foundation*.

A base da Iniciativa é o protocolo OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*), que faz com que os participantes da Iniciativa possam compartilhar seus metadados. Metadados significam "dados sobre os dados", ou seja, são informações que descrevem o conteúdo dos registros dos repositórios, que por sua vez são documentos eletrônicos (artigos, teses, e outros). São exemplos de metadados: título, autor, resumo, palavras-chaves, e outros.

Os participantes da Iniciativa são divididos em Provedores de Dados (*Data Providers* ou *DP*) e Provedores de Serviços (*Service Providers* ou *SP*). Os provedores de dados mantêm repositórios de documentos digitais que implementam o protocolo OAI-PMH como forma de expor os metadados de seus documentos. Já os provedores de serviço oferecem buscas a estes metadados ou outros serviços que visam agregar valor à Iniciativa.

Os metadados do protocolo OAI-PMH seguem o padrão Dublin Core (ver 2.7.3. Iniciativa *Dublin Core*).

2.2. OS PROVEDORES DE DADOS

Os provedores de dados são bases de dados (chamadas de repositórios) de documentos digitais que implementam o protocolo OAI-PMH para exposição dos metadados dos seus registros. Isso significa que o conteúdo dos registros (texto

completo) não é necessariamente exposto, a menos que o repositório inclua o endereço digital (link) do documento em um dos campos dos metadados.

Os provedores de dados podem ser criados já com o suporte ao protocolo OAI-PMH através de programas desenvolvidos para apoiar a Iniciativa *Open Archives* (*Eprints*, *Kepler*, *DSpace*), ou, podem ser uma base de dados pré-existente que realizou uma implementação de interface com o protocolo (por exemplo: *arXiv*, *SciELO*, *Virginia Tech Image Base*).

A vantagem de se participar da OAI, além da adoção de um padrão que está se consolidando na comunidade científica para a disseminação do conhecimento é também de aumentar a exposição dos trabalhos publicados pela instituição de maneira barata e eficiente. Em 13/06/2003, haviam 96 repositórios cadastrados na Iniciativa *Open Archives* (principais deles listados no Anexo I). Porém, vale ressaltar que muitos repositórios compatíveis não constam da lista da página da Iniciativa [20] por não terem migrado para a versão 2 do OAI-PMH, mas podem ser vistos no provedor de serviços *OAIster* [8], onde a quantidade de repositórios passa de 180.

2.3. OS PROVEDORES DE SERVIÇOS

Os participantes da OAI registram seus repositórios ou serviços na página da *Open Archives* [20], como forma de explorar a capacidade de exposição dos conteúdos dos mesmos. Na página da OAI são encontradas as listas de instituições que atuam como Provedores de Dados e outras que são Provedores de Serviços.

São os Provedores de Serviços que coletam, organizam e disponibilizam os metadados dos repositórios da OAI. Ao acessar um Provedor de Serviços e realizar uma pesquisa, a cada dia se tem respostas de melhor qualidade, dada a crescente adesão à Iniciativa por parte da comunidade científica.

Em 27/04/2003, a OAI já contava com diversas ferramentas de busca/suporte à Iniciativa, apresentadas na **Tabela 1 - Provedores de Serviços OAI [20]**.

Provedor de Serviço	Implementado por	Descrição	Versão OAI
Arc	Old Dominion University	Provedor de serviços ainda em estágio experimental que realiza buscas em diversos repositório compatíveis com OAI, apresentando no resultado os repositórios que contêm o assunto pesquisado, quantidade e os metadados agrupados por repositório..	1.1, 2.0
citebase Search	Southampton University	Oferece pesquisa em vários repositórios com os resultados baseados em vários critérios, dentre eles data de criação e citações.	1.1, 2.0
DP9	Old Dominion University	DP9 é um serviço que permite ferramentas de busca (como Google), indexarem repositórios compatíveis com OAI. Situa-se entre o <i>crawler</i> e o repositório, intercepta a requisição do <i>crawler</i> , redireciona aos repositórios e traduz os resultados de XML para HTML. "Permite que os repositórios OAI ocultos nas profundezas da internet sejam indexados pelas ferramentas de busca." [3]	1.1, 2.0
iCite	ICite	Sistema de indexação automática de citações em periódicos da área de física.	2.0
my OAI	my.OAI	Oferece busca em repositórios pré-selecionados da OAI. Usuários registrados têm acesso a facilidades tais como: salvar pesquisas para uso posterior, manter histórico de pesquisas, setar preferências, e outros.	1.1, 2.0
NCSTRL (pronuncia-se "ancestral")	Old Dominion U. Univ. of Virginia Virginia Tech	Acesso unificado a artigos e eprints da área de ciência da computação. É a implementação OAI do projeto NCSTRL [15].	1.1, 2.0
OAIster	Univ. of Michigan Libraries DL Production Service	Poderoso provedor de serviços de onde é possível pesquisar os mais de 160 provedores de dados registrados na OAI de um único ponto.	1.1, 2.0
Perseus	Tufts College	Pesquisa o conteúdo dos repositórios OAI registrados e incorpora as informações à interface de pesquisa.	1.1, 2.0
Public Knowledge Harvester	Univ. of British Columbia	Serviço de coleta de metadados direcionado a apenas alguns repositórios (12 em 27/04/2003).	1.1, 2.0
Repository Explorer	Virginia Tech	Ferramenta interativa para teste de conformidade com o Protocolo da Iniciativa Open Archives.	1.0, 1.1, 2.0
Scirus	Scirus	Scirus destaca-se por seu conteúdo essencialmente científico, composto por sites e bases de dados muitas vezes proprietárias. Atualmente também acessa dados da OAI.	1.1, 2.0
TORII	Intern. School for Adv. Studies, Trieste, Italy	Acesso unificado a vários repositórios (Física e Ciência da Computação). Oferece personalização e filtros.	1.1, 2.0

Tabela 1 – Provedores de Serviços OAI

2.4. O PROTOCOLO OAI-PMH

2.4.1. Histórico

Os primeiros seis meses da Versão 1 do OAI-PMH (1º semestre de 2001) foram basicamente experimentais. A intenção deste período era oferecer uma plataforma razoavelmente estável para os recentes adeptos testarem os conceitos da coleta de metadados e para a construção de alguns serviços básicos. Ao final dos seis meses iniciais (Versão 1), já havia a lista de discussão *OAI-Implementers*, que foi importante fonte de informação desde os desenvolvedores do OAI-PMH (dando *feedbacks*) quanto aos recém-chegados interessados na implementação do protocolo. Esta lista está ativa até hoje e pode ser assinada em [20].

A Versão 2 do OAI-PMH começou a ser delineada em junho de 2001 com a criação de um novo Comitê Técnico OAI (*OAI Technical Committee - OAI-Tech*). Seus integrantes eram implementadores que haviam trabalhado na criação e nos testes *alpha* da versão 1.0 do OAI-PMH. O objetivo do OAI-Tech era revisar os detalhes das versões 1.x do protocolo e suas funcionalidades, fazendo as correções necessárias, e disponibilizar uma nova versão do protocolo para meados de 2002.

Após muita discussão, reuniões, implementação e testes, surgiu a versão *alpha* do protocolo em 1º de março de 2002. Em 1º de Maio de 2002 a versão OAI-PMH 2.0 *beta* foi liberada apenas entre os implementadores, e, finalmente, em 1º de Junho de 2002, a Versão 2.0 é oficialmente disponibilizada, um mês antes do prazo inicial estipulado pelo *OAI-Tech*.

Para estimular a migração para a versão 2.0, a Iniciativa *Open Archives* não aceita mais inscrições na versão 1.x do protocolo desde 1º de setembro de 2002, e, desde 1º de dezembro de 2002, todos os repositórios listados no página da Iniciativa [20] estão em conformidade com a versão 2.0 (os não-conformes foram retirados da lista, até sua migração à nova versão).

2.4.2. Características

O Protocolo OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) consiste de 6 verbos ou *requests*, e as respostas a todas as requisições estão codificadas em XML. É a base da interoperabilidade entre os repositórios.

2.4.2.1. A Linguagem XML

XML (*eXtended Markup Language* ou linguagem de marcação de textos estendida) é uma linguagem que foi apresentada pelo W3C (*World Wide Web Consortium*) [46] pela primeira vez em 1996, e passou a ser recomendada em 1998.

A linguagem XML especifica dados em um formato legível por humanos, similar ao HTML. Sendo uma meta linguagem, pode ser utilizada para definir outros domínios ou linguagens específicas de um domínio (ex. matemático, químico, biológico) [35].

A estrutura e o tipo de informação de um documento XML podem ser precisamente armazenados em arquivos de definição do documento, DTDs (*Document Type Definition*), os quais podem ser passados, junto com o documento XML, de uma aplicação para outra.

A linguagem XML é independente de mídia de apresentação (navegador, aplicativo, telefone celular, etc.), pois, descreve dados e não apresentação visual, é interoperável, sendo independente tanto de plataforma quanto de fornecedor, já que XML é, basicamente, um arquivo texto [36].

Para trabalhar com XML, a W3C criou alguns padrões de acesso, resumidamente descritos na **Tabela 2** [35, 36]:

PADRÃO	DESCRIÇÃO
XML	A linguagem em si
XLL (<i>Xlink and Xpointer</i>)	Para ligação entre documentos e partes de documentos, hipertextos
XSL (<i>eXtensible Style Language</i>) e XSLT (<i>eXtensible Style Language Transformations</i>)	Permitem a transformação de um documento em outro, para fins de apresentação ou intercâmbio.
DOM (<i>Document Object Model</i>)	Estrutura criada com a finalidade de possibilitar a navegação, escrita e leitura, em documentos XML
XML Namespaces	Para resolver conflitos de nomes.
XML Schemas	Para especificação de tipos de dados, subclasses e melhores modelos de conteúdo.
XML Query	Permite exprimir consultas de forma semelhante à utilizada na linguagem SQL.

Tabela 2 - Padrões de Acesso XML

2.4.2.2. Harvesting

O protocolo OAI-PMH introduz o conceito de *Harvesting* (colheita), um processo unilateral (representado graficamente na Figura 2), que, a partir da lista de repositórios cadastrados na Iniciativa (os provedores de dados), periodicamente faz uma busca a estes repositórios, por meio do envio de solicitações (*requests*), coletando os metadados para exibição pelos provedores de serviços.

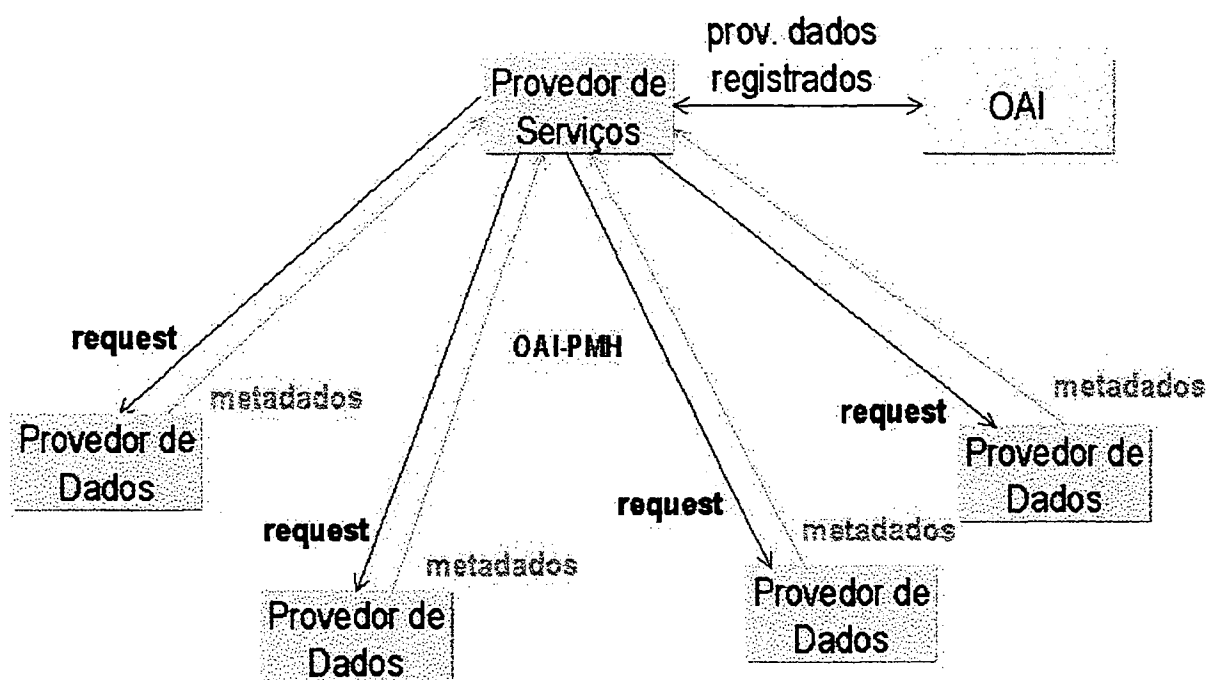


Figura 2 - Harvesting

Esta coleta de metadados pode ser integral ou baseada em critérios. Os critérios de coleta do *Harvesting* são:

- *Date-based* (baseado em data) - Serão coletados apenas os metadados incluídos/alterados após a data especificada.
- *Set-based* (baseado em conjuntos) – o protocolo define *set* como “*uma estrutura opcional para agrupar itens num repositório para o propósito de uma coleta seletiva de registros*” [16]. Esta estrutura é uma árvore de assuntos que representa a hierarquia do repositório.

2.4.2.3. Verbos do OAI-PMH

Os verbos do OAI-PMH são explicados a seguir, com exemplos de solicitações (*requests*) e suas respectivas respostas em XML. Estes exemplos são simulações do *Harvesting*, ou seja, como se a solicitação fosse enviada ao provedor de dados por parte do provedor de serviços. Nestes exemplos, está sendo simulada a atuação do provedor de serviços através de um comando *http* num *browser* (utilizou-se o Internet Explorer). Este processo é transparente para o usuário final, que receberá a informação já coletada e tratada pelo provedor de serviços, de maneira amigável.

O provedor de dados ao qual são submetidas as solicitações é o **UFPR Eprints** [8], repositório digital do Departamento de Informática da Universidade Federal do Paraná (<http://eprints.c3sl.ufpr.br>). O **UFPR Eprints** é um provedor de dados registrado na OAI e em conformidade com o protocolo OAI-PMH 2.0.

2.4.2.3.1. Identify

Este verbo é utilizado para recuperar as informações sobre o provedor de dados. A resposta mínima esperada consiste em: nome do provedor de dados, endereço do repositório (URL), versão do protocolo implementada e endereço eletrônico (e-mail) do administrador do repositório (Figura 3).

Solicitação: <http://eprints.c3sl.ufpr.br/perl/oai2?verb=Identify>

```

<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2003-07-08T07:35:45Z</responseDate>
  <request verb="Identify" resumptionToken="">http://eprints.c3sl.ufpr.br/perl/oai2</request>
  <Identify>
    <repositoryName>UFPR EPrints</repositoryName>
    <baseURL>http://eprints.c3sl.ufpr.br/perl/oai2</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>mailto:admin@eprints.c3sl.ufpr.br</adminEmail>
    <earliestDatestamp>0001-01-01</earliestDatestamp>
    <deletedRecord>persistent</deletedRecord>
    <granularity>YYYY-MM-DD</granularity>
    <description>
      < oai-identifier xmlns="http://www.openarchives.org/OAI/2.0/oai-identifier"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier
          http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
        <scheme>oai</scheme>
        <repositoryIdentifier>GenericEPrints.OAI2</repositoryIdentifier>
        <delimiter>:</delimiter>
        <sampleIdentifier>oai:GenericEPrints.OAI2:23</sampleIdentifier>
      </oai-identifier>
    </description>
    <description>
      < eprints xmlns="http://www.openarchives.org/OAI/1.1/eprints"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.openarchives.org/OAI/1.1/eprints
          http://www.openarchives.org/OAI/1.1/eprints.xsd">
        <content>
          <text>OAI Site description has not been configured.</text>
        </content>
        <metadataPolicy>
          <text>No metadata policy defined. This server has not yet been fully
            configured. Please contact the admin for more information, but if
            in doubt assume that NO rights at all are granted to this
            data.</text>
        </metadataPolicy>
        <dataPolicy>
          <text>No data policy defined. This server has not yet been fully
            configured. Please contact the admin for more information, but if
            in doubt assume that NO rights at all are granted to this
            data.</text>
        </dataPolicy>
        <submissionPolicy>
          <text>No submission-data policy defined. This server has not yet been
            fully configured.</text>
        </submissionPolicy>
        <comment>This system is running eprints server software (EPrints 2.2.1
          (pepper) [Born on 2002-11-14]) developed at the University of
          Southampton. For more information see
          http://www.eprints.org/</comment>
        </eprints>
      </description>
    </Identify>
  </OAI-PMH>

```

Figura 3 - Resposta do provedor de dados ao verbo Identify do OAI-PMH

2.4.3.2.2. ListMetadataFormats

Este verbo recupera os formatos de metadados disponíveis no repositório (Figura 4). Opcionalmente, pode-se definir o identificador de um registro como argumento, que listará os metadados deste registro em particular. O formato padrão é *Dublin Core* (representado no protocolo por *oai_dc*).

Solicitação: <http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListMetadataFormats>

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2003-07-31T22:52:41Z</responseDate>
  <request verb="ListMetadataFormats"
    resumptionToken="">http://eprints.c3sl.ufpr.br/perl/oai2</request>
  <ListMetadataFormats>
    <metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>
      <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc</metadataNames
        pace>
      </metadataFormat>
    </ListMetadataFormats>
  </OAI-PMH>
```

Figura 4 - Resposta do provedor de dados ao verbo *ListMetadataFormats* do OAI-PMH

2.4.3.2.3. GetRecord

Recupera um único registro do repositório. Deve-se especificar o formato dos metadados (*metadataPrefix*), cujo padrão é *oai_dc* (*Dublin Core*) e o identificador do registro (*Identifier*), que é único na comunidade OAI (Figura 5).

Solicitação:

http://eprints.c3sl.ufpr.br/perl/oai2?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:GenericEPrints.OAI2:49

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
  instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2003-08-27T12:55:43Z</responseDate>
  <request verb="GetRecord" identifier="oai:GenericEPrints.OAI2:49" metadataPrefix="oai_dc"
    resumptionToken="">http://eprints.c3sl.ufpr.br/perl/oai2</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:GenericEPrints.OAI2:49</identifier>
        <timestamp>2003-08-01</timestamp>
        <setSpec>7374617475733D707562</setSpec>
        <setSpec>7375626A656374733D646174616261736573</setSpec>
        <setSpec>7375626A656374733D696E666F726D6174696F6E5F72657472696576616C</setS
          pec>
      </header>
```

```

<metadata>
  <oai_dc:dc xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
      http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <dc:title>Provedores de Dados de Baixo Custo: Publicacao Digital ao Alcance de
      Todos</dc:title>
    <dc:creator>Garcia, Patricia A. B.</dc:creator>
    <dc:subject>Databases</dc:subject>
    <dc:subject>Information Retrieval</dc:subject>
    <dc:description>The availability of information by the electronic media has been
      improving the visibility of scientific researches, which was restricted to
      traditional periodicals. Research centers and academic databases, and
      more recently, digital libraries and e-prints archives have been contributed
      to a new scenery that offers fast and free access to scientific information.
      Now, several initiatives encourages the creation of open archives and
      digital libraries. The Open Archives Initiative (OAI) promotes the
      communication standardization between archives through OAI-PMH (Open
      Archives Initiative Protocol for Metadata Harvesting), improving the
      interoperability among themselves. There are software available at OAI
      which purpose is to facilitate the implementation of data providers
      (electronic content archives). However, kind of works that, through a
      Computer Science vision, analyses cost, complexity and the range of these
      solution are still rare. This work analyses the presented solutions,
      proposing a criteria for its evaluation and confirming the viability of
      installing a low-cost interoperable digital library.</dc:description>
    <dc:date>2003-07-01</dc:date>
    <dc:type>Thesis</dc:type>
    <dc:identifier>http://eprints.c3sl.ufpr.br/archive/00000049/</dc:identifier>
    <dc:format>pdf
      http://eprints.c3sl.ufpr.br/archive/00000049/01/dissertacao_patricia.pdf</dc:fo
      rmat>
  </oai_dc:dc>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

Figura 5 - Resposta do provedor de dados ao verbo GetRecord do OAI-PMH

2.4.3.2.4. ListRecords

É o verbo que "colhe" os metadados do repositório. Através de argumentos opcionais pode-se estabelecer a coleta seletiva baseada em data (*date-based*) ou em conjuntos (*set-based*). É obrigatório especificar o *metadataPrefix*, cujo padrão é *oai_dc* (Figura 6).

Solicitação:

http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListRecords&metadataPrefix=oai_dc

```

<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2003-07-31T23:07:07Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc"
    resumptionToken="">http://eprints.c3sl.ufpr.br/perl/oai2</request>
  <ListRecords>
    <record>
      <header>
        <identifier>oai:GenericEPrints.OAI2:36</identifier>
        <timestamp>2003-06-27</timestamp>
        <setSpec>7374617475733D707562</setSpec>

```



```

<setSpec>7375626A656374733D646174616261736573</setSpec>
</header>
<metadata>
  <oa_dc:dc xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
      http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
    xmlns:oa_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <dc:title>Mapeamento de atributos complexos e multivalorados na extração de
      esquemas utilizando XML</dc:title>
    <dc:creator>PINTO, José Simão de Paula</dc:creator>
    <dc:creator>SUNYE, Marcos Sfair</dc:creator>
    <dc:subject>Databases</dc:subject>
    <dc:description>This work describes a way to map complex and / or multivalued
      attributes from a visual query environment to relational databases, using XML.
      These attributes are usually produced after the application of integration
      methodologies, like ERC+, whenever exists in original databases. The mapping
      was produced in databases already integrated, using XML DTD files to retain
      knowledge about pre-integration databases, and a tool, called VIQUEN, that
      generates a schema visualization by use the ERC+ model. Queries and their
      results are both expressed in a visual environment. Express queries in a visual
      environment enables common users to create your own queries, since the tool
      hide complex database details and don't demand user knowledge about SQL
      statements.</dc:description>
    <dc:date>2001-08-01</dc:date>
    <dc:type>Thesis</dc:type>
    <dc:identifier>http://eprints.c3sl.ufpr.br/archive/00000036/</dc:identifier>
    <dc:format>pdf
      http://eprints.c3sl.ufpr.br/archive/00000036/01/DissertacaoFinal.pdf</dc:format>
    </oa_dc:dc>
  </metadata>
</record>
.
.
  (devido à extensão da resposta completa, alguns registros foram omitidos)
.
.
<record>
  <header>
    <identifier>oai:GenericEPrints.OAI2:49</identifier>
    <datestamp>2003-08-01</datestamp>
    <setSpec>7374617475733D707562</setSpec>
    <setSpec>7375626A656374733D646174616261736573</setSpec>
    <setSpec>7375626A656374733D696E666F726D6174696F6E5F72657472696576616C</setSpec>
  </header>
  <metadata>
    <oa_dc:dc xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
        http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
      xmlns:oa_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:dc="http://purl.org/dc/elements/1.1/">
      <dc:title>Provedores de Dados de Baixo Custo: Publicacao Digital ao Alcance de
        Todos</dc:title>
      <dc:creator>Garcia, Patricia A. B.</dc:creator>
      <dc:subject>Databases</dc:subject>
      <dc:subject>Information Retrieval</dc:subject>
      <dc:description>The availability of information by the electronic media has been
        improving the visibility of scientific researches, which was restricted to
        traditional periodicals. Research centers and academic databases, and more
        recently, digital libraries and e-prints archives have been contributed to a new
        scenery that offers fast and free access to scientific information. Now, several
        initiatives encourages the creation of open archives and digital libraries. The
        Open Archives Initiative (OAI) promotes the communication standardization
        between archives through OAI-PMH (Open Archives Initiative Protocol for
        Metadata Harvesting), improving the interoperability among themselves. There
        are software available at OAI which purpose is to facilitate the implementation of
        data providers (electronic content archives). However, kind of works that,
        through a Computer Science vision, analyses cost, complexity and the range of
        these solution are still rare. This work analyses the presented solutions,
        proposing a criteria for its evaluation and confirming the viability of installing a
        low-cost interoperable digital library.</dc:description>
      <dc:date>2003-07-01</dc:date>
      <dc:type>Thesis</dc:type>
      <dc:identifier>http://eprints.c3sl.ufpr.br/archive/00000049/</dc:identifier>
      <dc:format>pdf
        http://eprints.c3sl.ufpr.br/archive/00000049/01/dissertacao_patricia.pdf</dc:format>
      >
    </oa_dc:dc>
  </metadata>
</record>
<resumptionToken>0475038/oai_dc</resumptionToken>
</ListRecords>
</OAI-PMH>

```

Figura 6 - Resposta do provedor de dados ao verbo ListRecords do OAI-PMH

2.3.4.2.5. ListIdentifiers

O *ListIdentifiers* é uma versão abreviada do verbo *ListRecords*, que retorna apenas os cabeçalhos dos registros. Através de argumentos opcionais pode-se estabelecer a coleta seletiva baseada em data (*date-based*) ou em conjuntos (*set-based*). É obrigatório especificar o *metadataPrefix* (*oai_dc*) (Figura 7).

Solicitação:

http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc

```
<?xml version="1.0" encoding="UTF-8" ?>
: <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/
XMLSchema-Instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2003-07-31T23:18:21Z</responseDate>
  <request verb="ListIdentifiers" metadataPrefix="oai_dc"
    resumptionToken="">http://eprints.c3sl.ufpr.br/perl/oai2</request>
: <ListIdentifiers>
:   <header>
:     <identifier>oai:GenericEPrints.OAI2:36</identifier>
:     <datestamp>2003-06-27</datestamp>
:     <setSpec>7374617475733D707562</setSpec>
:     <setSpec>7375626A656374733D646174616261736573</setSpec>
:   </header>
:
:   (devido à extensão da resposta, alguns itens foram omitidos)
:
:   <header>
:     <identifier>oai:GenericEPrints.OAI2:49</identifier>
:     <datestamp>2003-08-01</datestamp>
:     <setSpec>7374617475733D756E707562</setSpec>
:     <setSpec>7375626A656374733D6172746966696369616C5F696E74656C6C6967656E6365</
:       setSpec>
:   </header>
:   <resumptionToken />
: </ListIdentifiers>
</OAI-PMH>
```

Figura 7 - Resposta do provedor de dados ao verbo *ListIdentifiers* do OAI-PMH

2.3.4.2.6. ListSets

Este verbo lista a estrutura do conjunto de um repositório (também conhecida como árvore de assuntos ou hierarquia do repositório), útil para o *harvesting* seletivo (Figura 8).

Solicitação: <http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListSets>

```

<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2003-07-31T23:29:30Z</responseDate>
  <request verb="ListSets" resumptionToken="">http://eprints.c3sl.ufpr.br/perl/oa2</request>
  <ListSets>
    <set>
      <setSpec>7375626A656374733D646174616261736573</setSpec>
      <setName>Subject = Databases</setName>
    </set>
    <set>
      <setSpec>7375626A656374733D6F7065726174696E675F73797374656D73</setSpec>
      <setName>Subject = Operating Systems</setName>
    </set>
    <set>
      <setSpec>7375626A656374733D68756D616E5F636F6D70757465725F696E74657261637469
6F6E</setSpec>
      <setName>Subject = Human Computer Interaction</setName>
    </set>
    .
    . (devido à extensão da resposta, alguns itens foram omitidos)
    .
    <set>
      <setSpec>7375626A656374733D696E666F726D6174696F6E5F72657472696576616C</setSp
ec>
      <setName>Subject = Information Retrieval</setName>
    </set>
  </ListSets>
</OAI-PMH>

```

Figura 8 - Resposta do provedor de dados ao verbo ListSets do OAI-PMH

2.5. OS PROGRESSOS DA INICIATIVA OPEN ARCHIVES

Ao final de 2002, após quase dois anos da primeira versão do Protocolo OAI-PMH (Jan/2001), as seguintes razões podem ser utilizadas como medidas do sucesso da Iniciativa Open Archives e do Protocolo OAI-PMH [17]:

1. O OAI-PMH consolidou-se como base para a interoperabilidade das bibliotecas digitais.
2. A aceitação geral do OAI-PMH está baseada na sua característica intencional "*low-barrier*", que explora tecnologias largamente disseminadas como HTML e XML. A Iniciativa também está calcada no aproveitamento de muito anos de prática em metadados, que resultou no conjunto *Dublin Core*, da Iniciativa *Dublin Core* [24].

3. Estabelecimento de várias comunidades e/ou projetos de pesquisa que adotam o protocolo como base para a federação de informações.
4. Crescimento dos provedores de serviços que promovem a criação de mais e mais provedores de dados, agregando valor aos metadados expostos.

2.5.1. Principais Programas e Projetos de Pesquisa

- *National Science Digital Library (NSDL)* - Projeto financiado pela *National Science Foundation* que intenciona ser a maior e mais variada biblioteca digital da atualidade. Nos próximos cinco anos, a biblioteca espera servir milhões de usuários acessando dezenas de milhões de recursos digitais [10].
- *Metadata Harvesting Initiative of the Mellon Foundation* - Iniciada em agosto de 2001, com 1,5 milhão de dólares de verba da Fundação Andrew W. Mellon, deu origem ao *site* OAISTER um provedor de serviços da Universidade de Michigan de onde é possível pesquisar os mais de 160 provedores de dados registrados na OAI de um único ponto [8].
- *Eprints - Software* de publicação digital desenvolvido pela Universidade de Southampton (UK) para implementação de provedores de dados [21].
- *Kepler* - ferramenta de auto-publicação de documentos acadêmicos/científicos desenvolvida pelo Departamento de Ciência da Computação da Universidade Old Dominion em Norfolk, Virginia. Implementa "provedores de dados pessoais" [34].

- *Dspace* - um sistema de código aberto para coleta, preservação, indexação e distribuição de material científico e acadêmico (criação de provedores de dados), oriundo de uma parceria do MIT (*Massachusetts Institute of Technology*) e dos Laboratórios Hewlett-Packard. Aceita diversos formatos de dados [27].

2.6. O FUTURO DA OAI

Espera-se que em breve o OAI-PMH torne-se parte da infra-estrutura básica das bibliotecas digitais em todo o mundo[17]. Isto vem se consolidando ao longo dos últimos meses com o envolvimento de diversas comunidades da área, discutindo o protocolo e sua utilização, bem como promovendo-o em larga escala.

Atualmente, as discussões dos comitês da Iniciativa *Open Archives* giram em torno de estratégias de garantir a manutenção do sucesso do OAI-PMH e sua evolução.

2.7. OUTRAS INICIATIVAS DA ÁREA

2.7.1. *Self-Archiving Initiative*

Liderada por Stevan Harnad (Universidade de Southampton, UK), tem como pressuposto o incentivo a autores de texto científico para publicação de seus trabalhos completos na Internet (em repositórios digitais) para oferecer livre acesso aos interessados [1]. Os autores também são incentivados a publicar seus artigos apenas em revistas ou jornais cujas versões eletrônicas sejam livremente acessíveis, ou caso isto não seja possível, que o autor mantenha seu direito de publicar seus trabalhos em outros locais da *web*, ao invés de conceder totalmente os direitos autorais à instituição que está publicando seu trabalho. Esta iniciativa incentiva a publicação dos documentos desde suas primeiras versões (chamados

pre-prints) até aqueles que já passaram por arbitragem e publicação (chamados *post-prints*).

2.7.2. BOAI (*Budapest Open Access Initiative*)

A BOAI [22] surgiu de um encontro do *Open Society Institute (OSI)* em Budapeste, em Dezembro de 2001. O propósito do encontro era aumentar os esforços internacionais para tornar os artigos de pesquisa em todos os campos acadêmicos livremente disponíveis na internet. Como resultado, surgiu a BOAI, que prega o livre acesso aos documentos acadêmicos através da auto-publicação (*self-archiving*) e/ou da criação de novos jornais e revistas acadêmicas cujos conteúdos sejam livremente acessíveis (*Open Access Journals*). Isto será conseguido através do oferecimento de subsídios monetários para a criação de novos softwares, ferramentas e jornais ou repositórios de artigos na internet. O foco da BOAI são documentos que já tenham passado por arbitragem, como forma de garantir a qualidade do material publicado.

A BOAI tem recebido adesões de um crescente número de pessoas e organizações ao redor do mundo (pesquisadores, universidades, laboratórios, bibliotecas, fundações, publicações e outras iniciativas similares).

2.7.3. *Dublin Core Metadata Initiative*

A Iniciativa de Metadados *Dublin Core (Dublin Core Metadata Initiative - DCMI)* é uma organização que surgiu em Dublin, Ohio (EUA) em 1995 e que dedica-se a promover a adoção de padrões de interoperabilidade em metadados [24].

O Conjunto de Elementos *Dublin Core (Dublin Core Element Set)*, atualmente na versão 1.1 [25] é um conjunto de 15 elementos:

1. *Title* - Nome pelo qual o recurso é formalmente conhecido.
2. *Creator* - Entidade primariamente responsável pelo conteúdo do recurso (pessoa, organização ou serviço).
3. *Subject* - Palavras-chaves, frases-chaves ou códigos de classificação, preferencialmente selecionados de um vocabulário controlado.
4. *Description* - Resumo (*abstract*), sumário ou texto livre sobre o conteúdo.
5. *Publisher* - Entidade responsável pela disponibilização do recurso (pessoa, organização ou serviço)
6. *Contributor* - Entidade responsável por contribuições ao conteúdo do recurso (pessoa, organização ou serviço)
7. *Date* - Data da criação ou disponibilização do recurso, no formato YYYY-MM-DD.
8. *Type* - Natureza ou gênero do conteúdo do recurso (existe um vocabulário padrão DCMI para este elemento em *DCMI Type Vocabulary*(<http://dublincore.org/documents/dcmi-type-vocabulary>)
9. *Format* - Tipo de mídia do recurso. Usado para identificar o *hardware* ou *software* necessário para exibir ou executar o recurso. É recomendado o uso de um vocabulário controlado como uma lista de Tipos de Mídia Internet (*Internet Media Types* - <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>)
10. *Identifier* - Identificador único do recurso dentro de um dado contexto.
11. *Source* - Referência a um recurso do qual o presente recurso é derivado.
12. *Language* - Linguagem do conteúdo intelectual do recurso.

13. *Relation* - Referência a um recurso relacionado.

14. *Coverage* - Extensão ou escopo do conteúdo do recurso.

15. *Rights Management* - Informações sobre os direitos inerentes ao recurso (propriedade intelectual, direitos autorais e outros direitos de propriedade).

No próximo capítulo são analisados os programas *Eprints*, *Kepler* e *DSpace*, que são compatíveis com a Iniciativa *Open Archives* e que têm por finalidade facilitar a publicação digital.

3. AVALIAÇÃO DOS SOFTWARES PARA PUBLICAÇÃO DIGITAL BASEADOS NA OAI

O conhecimento deseja sempre se expandir; é como o fogo, que primeiro necessita de um agente externo para iniciar, mas, em seguida, se propaga sozinho.
Samuel Johnson
1709-1784

Foram selecionadas para análise as três ferramentas de uso gratuito existentes atualmente para a criação de provedores de dados baseadas na Iniciativa *Open Archives (Eprints, Kepler e DSpace)* tomando por base os critérios:

- Finalidade
- Disponibilidade
- Utilização dentro da comunidade OAI
- Arquitetura
- Funcionalidade
- Protocolo OAI-PMH
- Custos

O *Eprints* tem uma aplicação institucional (acadêmica/científica), onde os membros da instituição podem se cadastrar e submeter *eprints* para publicação. A publicação pode ser direta ou passar por um "editor".

Já o *Kepler* pode ser classificado como uma ferramenta de publicação pessoal, onde o próprio usuário baixa, instala, publica e controla o conteúdo publicado. Institui o conceito de "provedor de dados pessoal".

O *DSpace* oferece um ambiente altamente configurável, de forma a atender todo tipo de fluxo de trabalho inerente a um repositório digital. Pode-se implementar desde um pequeno repositório, com submissões e consultas livres, até um complexo ambiente de submissão, avaliação, arbitragem e gerenciamento.

3.1. GNU EPRINTS 2

3.1.1. Finalidade

O software GNU Eprints 2 foi desenvolvido pela Universidade de *Southampton* (Inglaterra) por Robert Tansley e Christopher Gutteridge, colaboradores de Stevan Harnad. Propõe-se a construir repositórios de *eprints* (documentos eletrônicos) de maneira rápida e fácil, comparado com a construção dos mesmos desde o início. Oferece uma interface *web* para gerenciamento, *submissão* e *download* de documentos.

3.1.2. Disponibilidade

O GNU Eprints está disponível desde novembro/2000, é um software livre distribuído sob licença GPL (*General Public License*), e pode ser baixado diretamente do *site* <http://software.eprints.org>

3.1.3. Utilização

O objetivo inicial do Eprints era facilitar a auto-publicação de documentos por seus autores [18], porém, sua utilização tem ultrapassado este objetivo, ocorrendo a partir de repositórios pequenos até repositórios com grande volume de informações em instituições de renome internacional (Anexo II).

3.1.3.1 Eprints nas Instituições

Duas instituições internacionais relatam suas experiências com o Eprints através de publicações [33, 9]:

3.1.3.1.1. Instituto de Tecnologia da Califórnia

O Caltech (*Califórnia Institute of Technology – Caltech*) utiliza Eprints desde 2001, tendo criado vários repositórios que podem ser acessados pelo link <http://library.caltech.edu/digital>

A conclusão do Caltech em relação ao Eprints é:

“uma ferramenta poderosa e barata para compartilhamento de trabalhos acadêmicos com o mundo. O processo de submissão baseado na web é intuitivo e requer esforço mínimo por parte dos autores. (...) Possibilita aos pesquisadores disponibilizar gratuitamente os resultados de seus trabalhos com qualquer um, aumentando o impacto de sua pesquisa no processo.” [33]

3.1.3.1.2. Universidades de Edinburgh e Nottingham

A experiência destas Universidades no uso do Eprints [9] relata a necessidade de avaliação de diversos pontos antes da própria instalação do software. Pontos como tipos de dados, formatos, supervisão, metadados e participação dos usuários devem ser amplamente discutidos junto com a implantação do repositório.

Também são relacionadas as vantagens de se utilizar o Eprints e as ações realizadas para "vender a idéia" do uso do repositório internamente. Algumas destas ações são:

- Inicialmente - Localizar onde as publicações dos usuários em potencial estavam publicadas e pedir-lhes autorização para colocá-las no Eprints.
- Promover o Eprints através de rápidas apresentações em comitês da universidade, das bibliotecas, reuniões departamentais, e outros eventos. Divulgar a aplicação nas revistas especializadas.

Os principais custos apontados foram os de mão-de-obra, já que não existem custos de software e o custo de hardware é baixo, por se tratar de ambiente Linux.

3.1.4. Arquitetura do Eprints

O *Eprints* é desenvolvido em linguagem PERL, e roda em sistema operacional Linux e servidor *web* Apache. Há uma série de comandos para construir e manter um *archive*. É possível criar mais de um *archive* (repositório) em uma única instalação de Eprints [6, 19]. Os metadados dos registros e usuários são armazenados em banco de dados MySQL. Os arquivos com textos completos (*formatos PDF, HTML, ASCII, Postscript*) ficam armazenados em diretórios com o mesmo ID (identificador) do documento. A seguir, pode-se observar o exemplo de endereçamento de dois documentos formato PDF num repositório Eprints (**UFPR Eprints**), um com ID 37 e outro, com ID 38:

```
./archives/ufprprints/documents/disk0/00/00/00/37/01/pedrini_phdthesis_rpi.pdf
./archives/ufprprints/documents/disk0/00/00/00/38/01/db_interop.pdf
```

Se houvesse mais de um formato para o mesmo registro (por exemplo, em *postscript*), o formato seria:

```
./archives/ufprprints/documents/disk0/00/00/00/38/02/db_interop.ps
```

3.1.4.1. Estrutura do Registro (*eprint*)

Um “*eprint*” é um registro único no sistema, constituído por:

- *System Metadata Fields* – Campos de metadados do sistema, ou seja, campos que descrevem o registro no âmbito do sistema. São eles: *eprint-id* e *depositing-user-id* (identificação do documento e do depositante).
- *Archive Metadata Fields* – Campos de metadados do repositório, ou seja, os campos que descrevem o *eprint*. São exemplos de metadados: título, autor e ano. São as informações para pesquisa e visibilidade do *archive*. As informações de metadados podem ser

customizadas (acrescentar ou excluir campos) na criação do *archive*.

- Um ou mais documentos. Pode aceitar zero ou mais, conforme decisão do administrador. Cada registro pode conter documentos em vários formatos (PDF, HTML, ASCII, PostScript) e um *eprint* também pode ser composto por mais de um arquivo.

Um “*eprint*” sempre está em um dos quatro estágios [6] abaixo:

- *User Work-Area* (área de trabalho do usuário) – informações do usuário, sem necessidade de submissão de documentos.
- *Editor Buffer ou Submission Buffer* (buffer de submissão ou do editor) – após a submissão, o *eprint* aguarda aprovação por um editor (administrador do *archive* ou usuário com poderes de editor).
- *Archive* – contém os registros aprovados, que aparecem no *website* público, em pesquisas e são compatíveis com OAI.
- *Deleted* – para os registros que já estiveram no buffer *Archive*. São listados como “deletados” no *website* e via OAI.

3.1.5. Funcionalidade

Após a instalação do software *Eprints*, o *site* de seu *archive* é disponibilizado no endereço definido na instalação, e conta com áreas públicas e restritas (que necessitam de registro por parte do usuário). As principais funcionalidades do *Eprints* estão representadas na Figura 9:

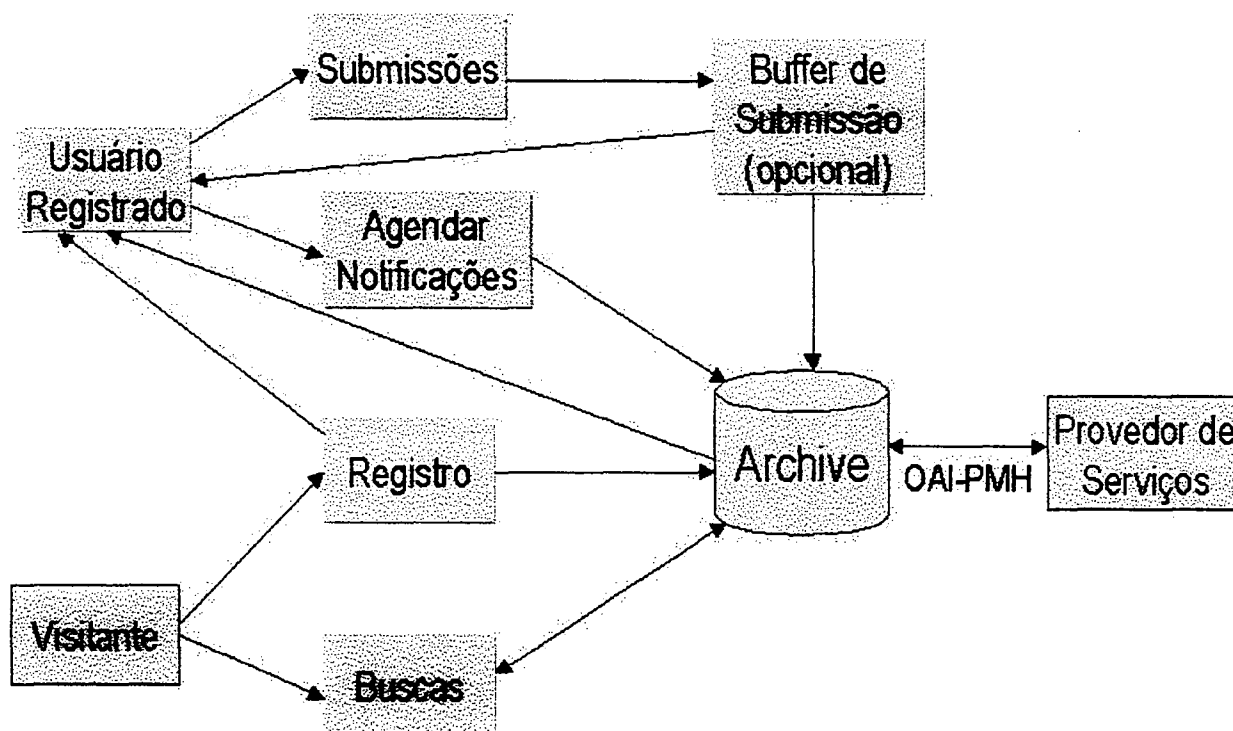


Figura 9 - Eprints - Principais Funcionalidades

Os retângulos com letras verdes (Visitante e Provedor de Serviços - PS) representam entidades externas ao Eprints, que podem acessá-lo através do site (no caso do Visitante) ou através do protocolo OAI-PMH, no caso do Provedor de Serviços da OAI, ao realizar o Harvesting. Existem dois tipos de usuários do sistema: o usuário Visitante e o usuário Registrado. O Visitante é qualquer pessoa que acesse a página do Eprints. Ele pode realizar buscas simples ou avançadas na base de dados ou optar pelo registro, tornando-se um usuário Registrado. Normalmente, os registros são efetuados por pessoas da comunidade-alvo do Eprints, pois, a partir do registro, elas passam a poder submeter arquivos à base de dados. Pessoas externas à comunidade também podem efetuar o registro para agendar Notificações. Estas Notificações são assinaturas (*subscriptions* na linguagem do sistema) de serviços de mensagem eletrônica que informam diária, semanal ou mensalmente a inclusão de registros nas áreas de interesse do usuário Registrado.

O usuário Registrado pode efetuar submissões de documentos, completando os metadados solicitados e carregando um ou mais arquivos por documento, sempre nos formatos PDF, PostScript (PS), ASCII ou HTML. Estas submissões seguem para um Buffer de Submissão (estrutura opcional, selecionada na instalação do Eprints), onde ficam à espera de revisão e aprovação por parte do Administrador do Eprints ou de um usuário específico, com função de Editor. O Editor pode editar os metadados e efetuar correções nos mesmos, e editar os arquivos para verificação do conteúdo. Após a aprovação, o documento é enviado para o Arquivo Principal (*Main Archive*), onde passa a ser visível por meio de consultas ou listagens do conteúdo da base de dados. Em caso de recusa do documento, este é devolvido ao usuário, através do próprio sistema.

3.1.5.1 Áreas Públicas

Para ilustrarmos a explicação da estrutura do *site*, utilizaremos o *Eprints* criado para o Departamento de Informática da UFPR (**UFPR Eprints**), disponível em <http://eprints.c3sl.ufpr.br>.

Para consultar as publicações disponíveis através de pesquisa (*Search*) ou listagem por título ou ano (*Browse*), bem como acessar a Ajuda (*Help*) ou ter informações sobre os Créditos da página (*About*), não é necessário ser um usuário registrado.

Da tela inicial (Figura 10) já é possível fazer uma consulta simples, utilizando o campo disponível, que pesquisa em títulos, resumos e palavras-chaves ao mesmo tempo.

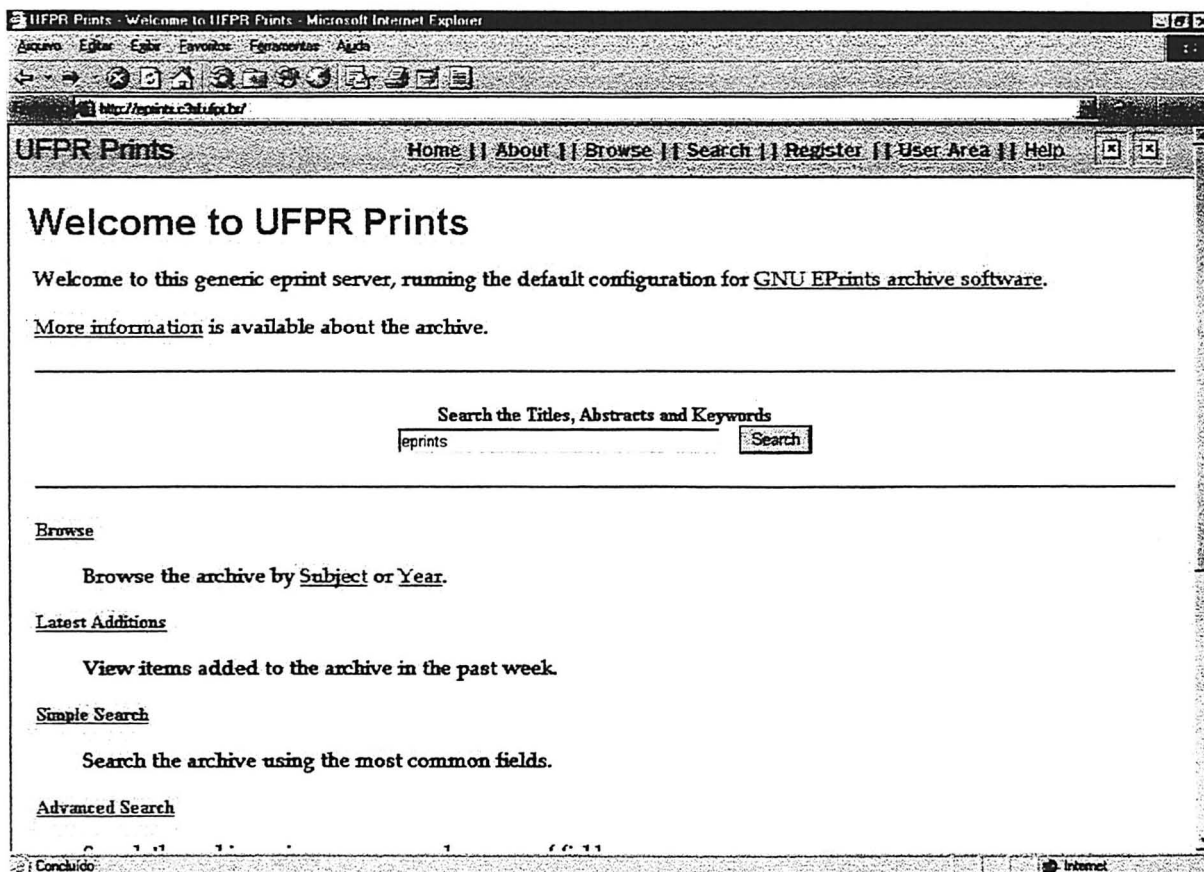


Figura 10 - Tela Inicial do UFPR Eprints

Também é possível vasculhar o repositório através da opção *Browse* (listagem) - Figura 11.

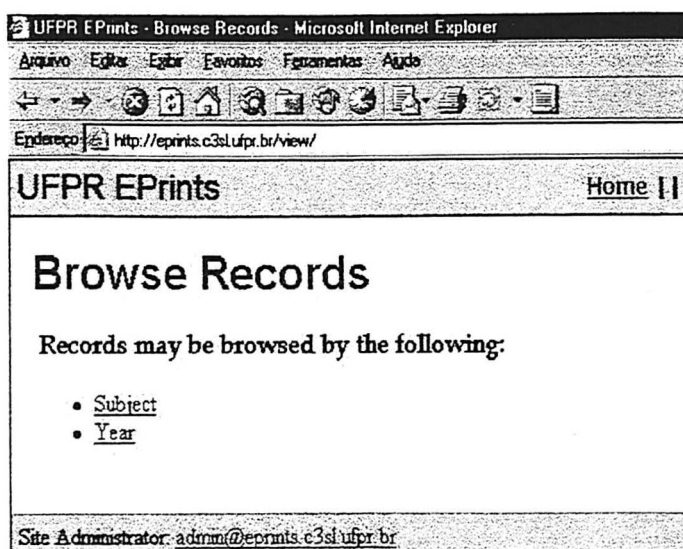


Figura 11 - Opções de listagem de registros do Eprints

O conteúdo pode ser visualizado por Assunto (*Subject*) - Figura 12 - ou Ano (*Year*) - Figura 13.

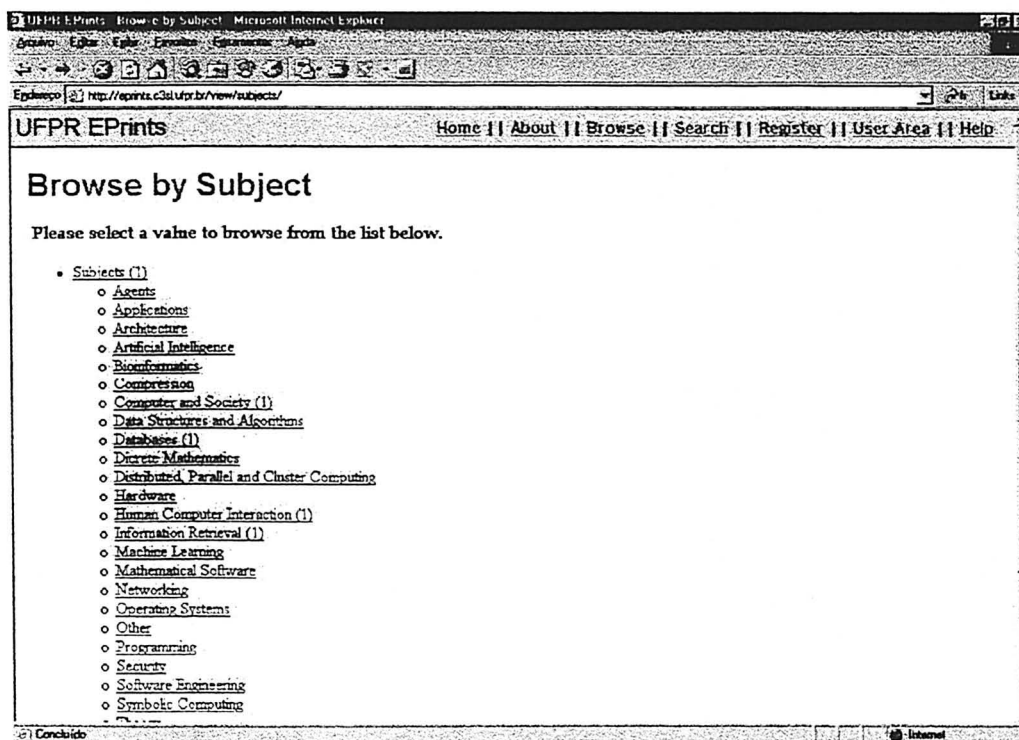


Figura 12 - Listagem por assunto

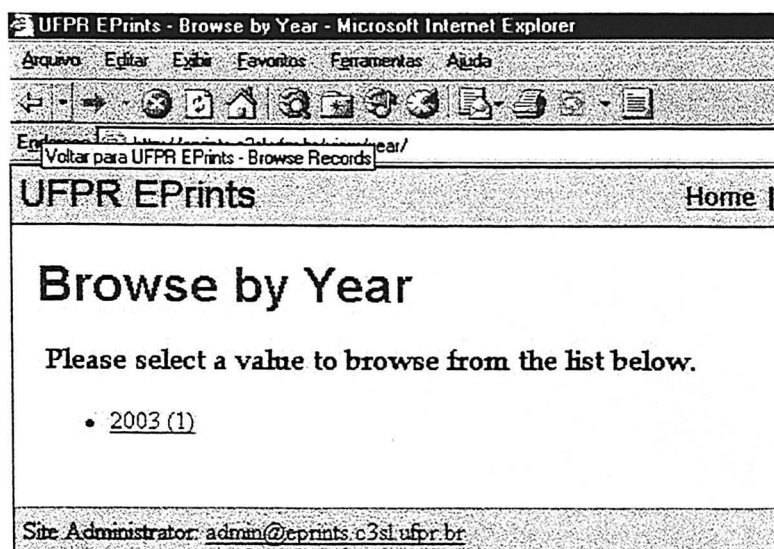
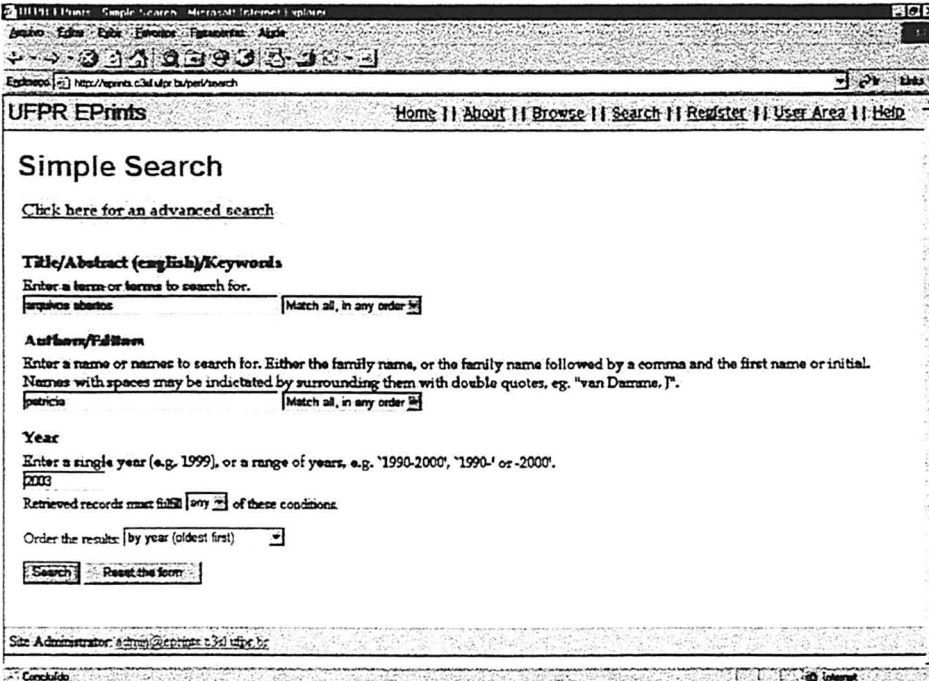


Figura 13 - Listagem por ano

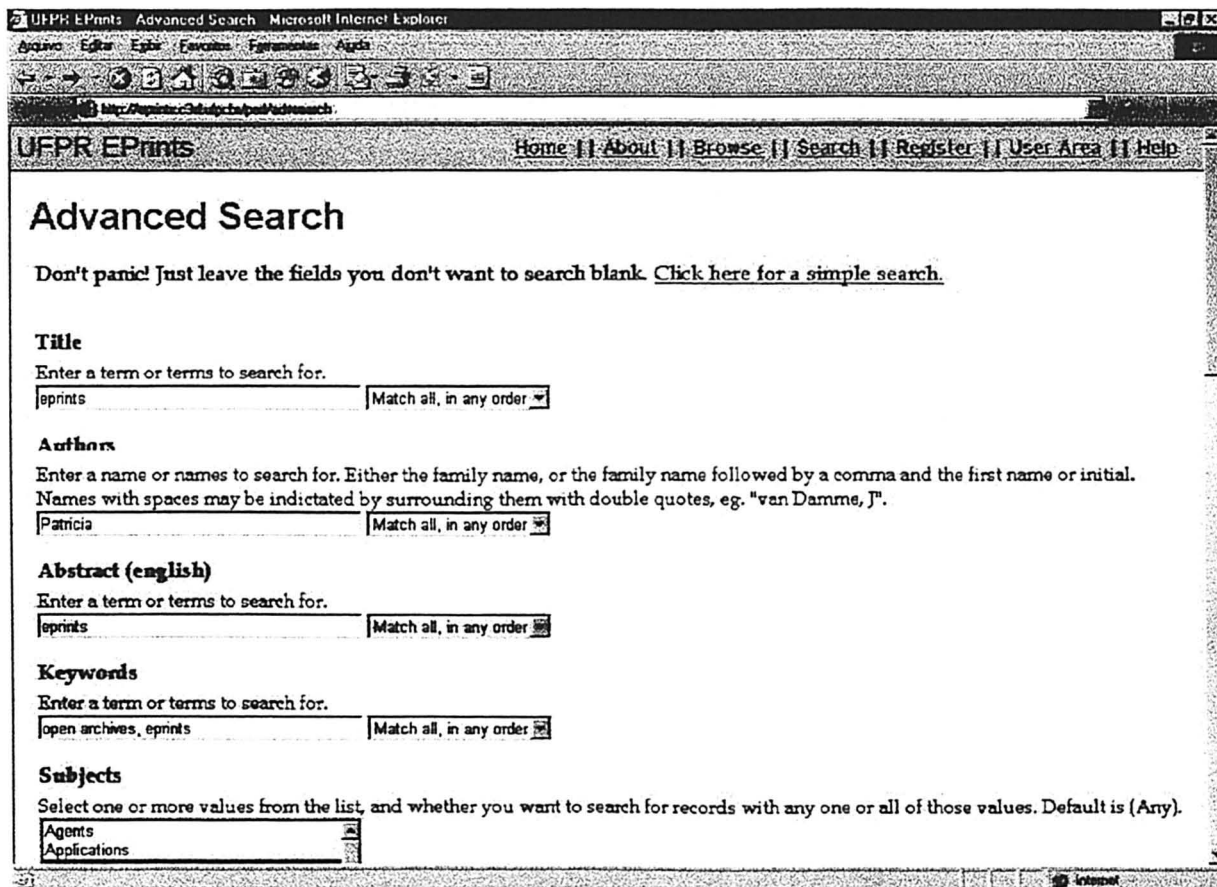
O usuário pode optar pela opção de Pesquisa no repositório (*Search*), que o direciona para Pesquisa Simples (*Simple Search*), onde é possível pesquisar os campos Título/Resumo/ Palavras-Chaves, Autor/Editor e/ou Ano (Figura 14).



The screenshot shows a web browser window titled "Simple Search" from UFPR EPrints. The browser's address bar shows "http://www.c3d.ufpr.br/epet/search". The page has a navigation menu with links: Home, About, Browse, Search, Register, User Area, and Help. The main content area is titled "Simple Search" and includes a link for an advanced search. There are three search criteria sections: 1. "Title/Abstract (english)/Keywords" with a text input field containing "arquivos abertos" and a "Match all, in any order" checkbox. 2. "Author/Editor" with a text input field containing "patricia" and a "Match all, in any order" checkbox. 3. "Year" with a text input field containing "2003". Below these is a note: "Retrieved records must fulfil any of these conditions." There is a dropdown menu for "Order the results:" set to "by year (oldest first)". At the bottom of the form are "Search" and "Reset the form" buttons. The footer of the page includes "Site Administrator: admin@epet.c3d.ufpr.br" and a "Concluido" status indicator.

Figura 14 - Pesquisa simples

Caso deseje uma pesquisa ainda mais restrita, há um link para a Pesquisa Avançada (*Advanced Search*) logo abaixo do título *Simple Search* que direciona o usuário para uma longa tela onde é possível entrar valores para diversos campos tornando a pesquisa altamente seletiva (Figura 15).



UFPR EPrints Advanced Search Microsoft Internet Explorer

UFPR EPrints Home | About | Browse | Search | Register | User Area | Help

Advanced Search

Don't panic! Just leave the fields you don't want to search blank. [Click here for a simple search.](#)

Title
Enter a term or terms to search for.
eprints Match all, in any order

Authors
Enter a name or names to search for. Either the family name, or the family name followed by a comma and the first name or initial. Names with spaces may be indicated by surrounding them with double quotes, eg. "van Damme, J".
Patricia Match all, in any order

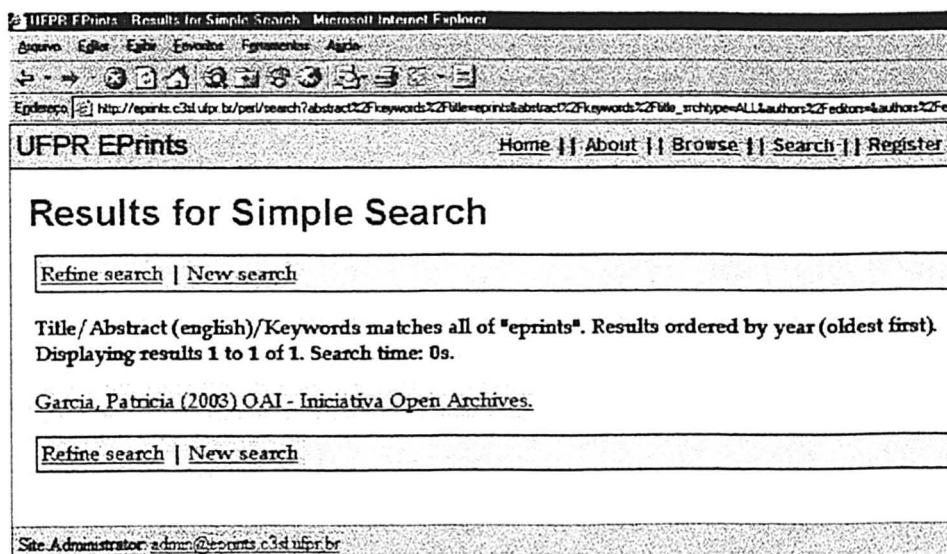
Abstract (english)
Enter a term or terms to search for.
eprints Match all, in any order

Keywords
Enter a term or terms to search for.
open archives, eprints Match all, in any order

Subjects
Select one or more values from the list, and whether you want to search for records with any one or all of those values. Default is (Any).
Agents
Applications

Figura 15 - Pesquisa Avançada

Seja qual for o tipo de pesquisa escolhida, o Eprints apresenta os registros coincidentes com os critérios estabelecidos (Figura 16).



UFPR EPrints Results for Simple Search Microsoft Internet Explorer

UFPR EPrints Home | About | Browse | Search | Register

Results for Simple Search

[Refine search](#) | [New search](#)

Title/ Abstract (english)/Keywords matches all of "eprints". Results ordered by year (oldest first).
Displaying results 1 to 1 of 1. Search time: 0s.

[Garcia, Patricia \(2003\) OAI - Iniciativa Open Archives.](#)

[Refine search](#) | [New search](#)

Site Administrator: admin@eprints.c3d.ufpr.br

Figura 16 - Resultado da pesquisa

Ao clicar em um dos registros retornados pela pesquisa, são mostrados os metadados do *eprint* (Figura 17).

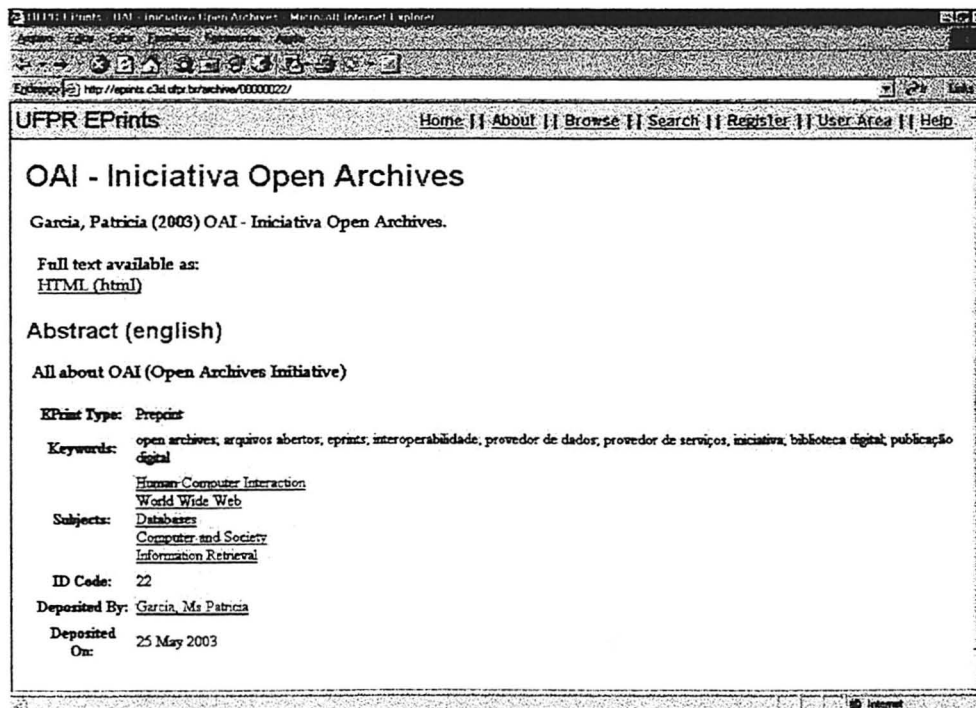


Figura 17 - Metadados do registro

Se o texto completo está disponível, há um link que também indica o formato do arquivo, neste caso, HTML. Ao clicar no link, o texto completo é aberto (Figura 18).

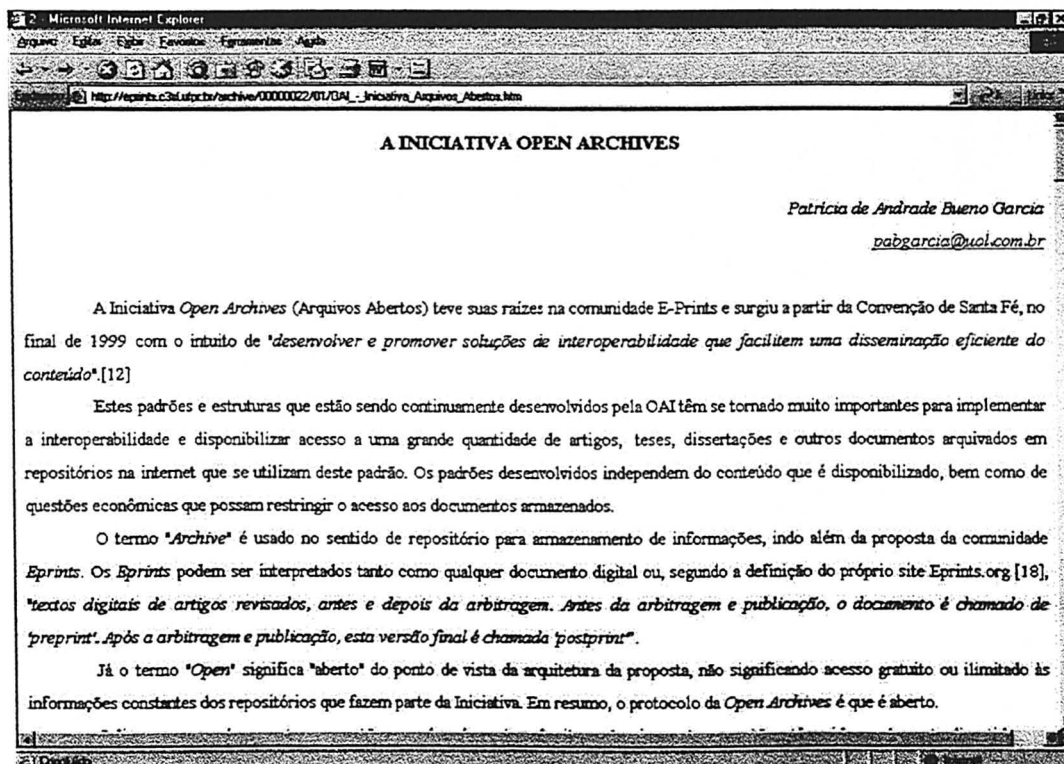


Figura 18 - Texto completo do registro

3.1.5.2. Áreas Restritas

As áreas restritas do Eprints compreendem a publicação e agendamentos (assinaturas), e é necessário estar registrado no repositório para se ter acesso a elas.

O papel do registro dentro do Eprints não fica claro de início para a maioria dos usuários. Ao acessar a interface padrão, não há informações sobre quem pode se cadastrar e as finalidades deste registro.

Existem dois papéis para os usuários do Eprints, que serão chamados usuário-publicador (aquele que deve pertencer à comunidade do Eprints e que publicará seus trabalhos no repositório) e usuário-assinante (qualquer usuário que deseje receber informações sobre inclusões no repositório, através de agendamentos de mensagens). Um usuário pode ser apenas publicador, assinante, ou publicador e assinante ao mesmo tempo.

3.1.5.2.1. Usuário-publicador

Para incluir suas contribuições no *Eprints*, o usuário deve estar registrado. O registro é efetuado mediante um formulário (que pode ser customizado na instalação), onde, dentre outras informações, o usuário cria seu nome-de-usuário (*username*) e senha (*password*) - Figura 19.

The screenshot shows a web browser window displaying the 'Record for Garcia, Patricia' registration form. The browser's address bar shows 'http://eprints.c3at.ufpr.br/real/users/record'. The page title is 'UFRPR Eprints' and the navigation menu includes 'Home', 'About', 'Browse', 'Search', 'Register', 'User Area', and 'Help'. The form content includes:

Record for Garcia, Patricia

Please enter correct information about yourself for our records. This information will be useful to us and readers of your items. You don't have to supply all this information if you don't want to; you need only fill out those fields marked with a * to start using the archive.

For instructions on how to change your e-mail address, [click here](#).

Password
Leaving this field blank will not affect the password.

Hide Email
Select this item if you do not wish your email to be displayed on public pages. The Eprints administrators will still be able to see it.

Name *

Title	Given Name(s)/Initials	Family Name(s)	Lineage
	Patricia	Garcia	

Department
Computer Science

Organisation
UFRPR

Address
Centro Politecnico

Figura 19 - Cadastro de novo usuário

Em seguida, o *Eprints* envia um e-mail com um código de confirmação de inscrição, para validação do e-mail e garantia de que o endereço pertence ao solicitante.

Após o cadastro, o usuário encontra as opções disponíveis ao seu perfil de usuário normal na Área do Usuário (Figura 20). Usuários com perfis de Editores ou Administradores terão mais opções à disposição.

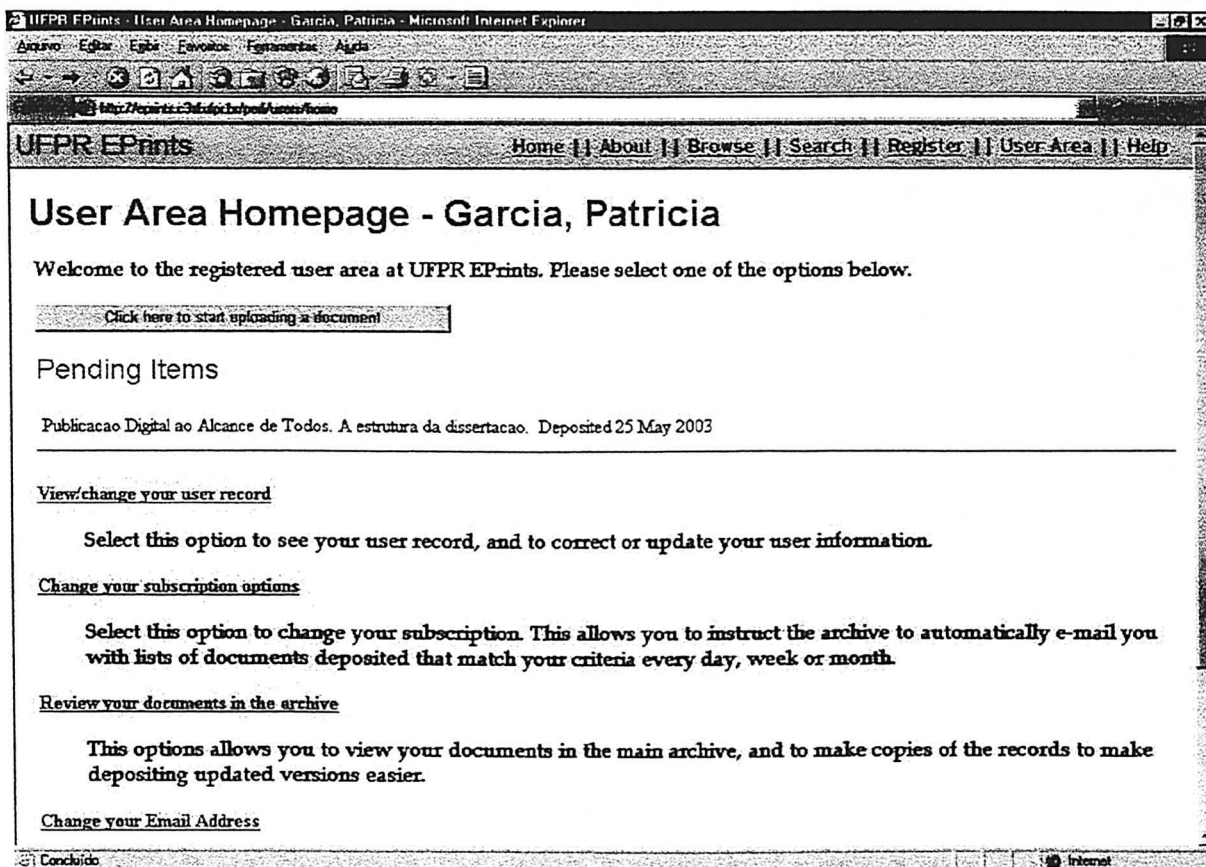


Figura 20 - Área do usuário

Uma destas opções é o botão de envio de documento, que representa a criação de um novo registro de Eprint (situado no alto da *User Area*, logo abaixo do nome do usuário).

Inicialmente, deve-se informar o tipo do documento (tal como livro, artigo, etc.) para determinar quais os metadados necessários para preenchimento. Após preenchê-los, é possível adicionar um ou mais documentos ao registro. Concluído este procedimento, o registro é “depositado” e fica no *buffer* editorial, aguardando aprovação.

O usuário com privilégio de editor pode visualizar o buffer de submissão. Um editor pode mover os registros deste buffer para o *archive* principal, modificá-los ou retorná-lo ao usuário com um comentário (via *e-mail*) sobre os problemas encontrados.

3.1.5.2.2. Usuário-assinante

O usuário registrado pode assinar um serviço de boletins diários, semanais ou mensais (*subscriptions*). Ele receberá um e-mail contendo a lista de novos registros que atendam os requisitos da pesquisa agendada. Por exemplo, é possível criar um agendamento (Figura 21) para receber semanalmente um e-mail informando as inclusões de textos de qualquer assunto, que já passaram por arbitragem. Também é possível criar outro com periodicidade diária, que informe as publicações da área de Banco de Dados, independente de arbitragem. Outra opção a ser escolhida é se, em caso de conteúdo vazio de e-mail (nenhuma publicação do critério solicitado), o e-mail deve ou não ser enviado.

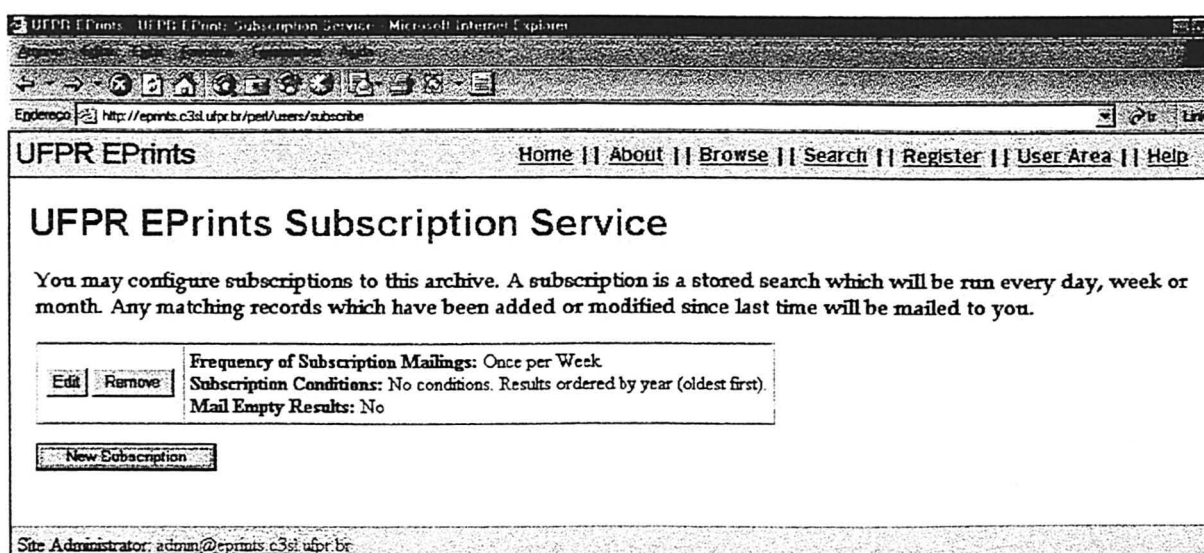


Figura 21 - Agendamentos

3.1.5.2.3. Registro Automatizado

Os dois tipos de usuários vistos anteriormente são tratados de maneira igual dentro do *Eprints*, e isto dificulta a administração. Todo usuário que completa seu registro tem ambas as opções: publicação e agendamento (assinaturas). Caso um estranho à comunidade publique algum documento, cabe ao Editor filtrar a

anomalia. Neste caso, o Editor teria que conhecer todas as pessoas da comunidade, o que é inviável.

Para evitar este problema, o registro de usuários via *web* pode ser substituído por outro método [6]: o Repositório de Eprints do Departamento de Eletrônica e Ciência da Computação da Universidade de *Southampton* tem um *script* que atualiza, de hora em hora, os usuários registrados no *Eprints* com base no banco de dados do Departamento. Também há a opção de importação de usuários. Estas duas opções não acompanham a versão-padrão do *Eprints*, mas, podem ser solicitadas diretamente ao seu desenvolvedor, Christopher Gutteridge (cjg@ecs.soton.ac.uk).

Os recursos de automatização do cadastro de usuários podem ser utilizados para evitar o cadastramento de pessoas estranhas à comunidade *Eprints*. Neste caso, por meio de customização, a opção *Register* (Registro) poderia ser retirada da página, com a desvantagem de não permitir mais a inclusão dos usuários-assinantes.

Outro tipo de customização poderia manter a opção *Register* (registro) na página, permitindo o registro de usuários em geral apenas para agendamentos. Na Área do Usuário (*User Area*), as opções referentes às publicações seriam retiradas.

3.1.5.3. Administração

Um administrador do *Eprints* pode pesquisar e modificar todas as informações do banco de dados, incluindo os registros de usuários. Os administradores também são os responsáveis por conceder o privilégio de editor a quem desempenhar esta função.

3.1.6. Protocolo OAI-PMH

Os repositórios *Eprints* são compatíveis com a Iniciativa *Open Archives* (OAI), podendo ser vasculhados pelo *Harvesting*, desde que o repositório esteja cadastrado como provedor de dados na OAI.

O Eprints 2 implementa o Protocolo OAI-PMH nas versões 1.1 e 2.0.

3.1.7. Custos

3.1.7.1. Custos de Software

O Eprints é distribuído sob *GNU Public License (GPL)*. Isto não significa que o software tem custo zero, mas, que ele pode ser livremente modificado e distribuído. A Universidade de Southampton detém todos os direitos autorais e a qualquer momento pode lançar o software sob uma nova licença, adicional à GPL. Todos os softwares sobre os quais o Eprints2 é construído estão disponíveis sem custo.

3.1.7.2. Custos de Hardware

Para finalidades de testes, qualquer máquina com GNU/Linux (ou outro Unix) pode ser utilizada. Para um sistema em produção, é recomendado [6]:

- PC rodando servidor GNU/Linux ou Unix equivalente com 1 GHz ou mais;
- 512Mb de RAM ou mais;
- 20Gb ou mais de espaço em disco com disponibilidade de expansão, à medida que o *archive* aumente.

3.1.7.3. Custos de Mão-de-Obra

Em teoria, um administrador Unix experiente, com algum conhecimento de Apache e PERL pode instalar e colocar no ar um *archive* em um dia. Na prática, a maioria dos sites que rodam o Eprints necessitam alguma customização, sejam pequenas modificações no *lay-out* das páginas, ajustes de metadados, ou traduções.

Concluídas estas tarefas, os únicos custos de mão-de-obra restantes serão dos usuários que depositarão registros e dos editores que validarão e aprovarão (ou não) os registros depositados. Neste estágio uma pequena administração Unix será necessária em forma de realização de eventuais *upgrades* e aplicações de *patches*.

3.2. KEPLER

3.2.1. Finalidade

O Kepler é uma ferramenta de auto-publicação de documentos acadêmicos/científicos desenvolvida pelo Departamento de Ciência da Computação da Universidade Old Dominion em Norfolk, Virginia, pelos pesquisadores Kurt Maly, Mohammad Zubair e Xiaoming Liu.

Tem como objetivo “*satisfazer a necessidade dos pesquisadores de universidades de publicar e disseminar resultados de maneira rápida, conveniente, e sob seu controle. Proporcionar acesso público a todo o material publicado.*” [34]

Segundo os autores, os fatores críticos para o sucesso da ferramenta são a simplicidade de uso e o controle das publicações, que se traduzem nas seguintes características:

- ser baixada pela internet
- independente de plataforma
- pacote que possa ser instalado em estações e PCs, por qualquer pessoa
- interface amigável
- compatível com OAI
- arquivos armazenados no próprio *file system*

A proposta do Kepler visa implementar vários pequenos provedores de dados (publicadores), chamados de "*archivelets*". O objetivo é promover a rápida disseminação de artigos técnicos de autores individuais.

3.2.2. Arquitetura

3.2.2.1. A Arquitetura Inicial do Kepler

A idéia inicialmente apresentada "ferramenta Kepler para auto-publicação" foi criada em 2001, para facilitar o processo da auto-publicação [1]. A arquitetura original (Figura 22 [34]) é muito similar a uma rede P2P (*peer-to-peer*). Os *archivelets* registram-se a um servidor centralizado e um provedor de serviços contacta este servidor para receber a lista dos *archivelets* para coleta de informações. Existem dois tipos de usuários: os publicadores individuais (usuários do *archivelet*) e os usuários em geral, interessados na recuperação dos documentos publicados. Os publicadores individuais interagem com a ferramenta de publicação do *archivelet*, enquanto que os usuários em geral interagem com um provedor de serviços e com os *archivelets*, através de um *browser*.

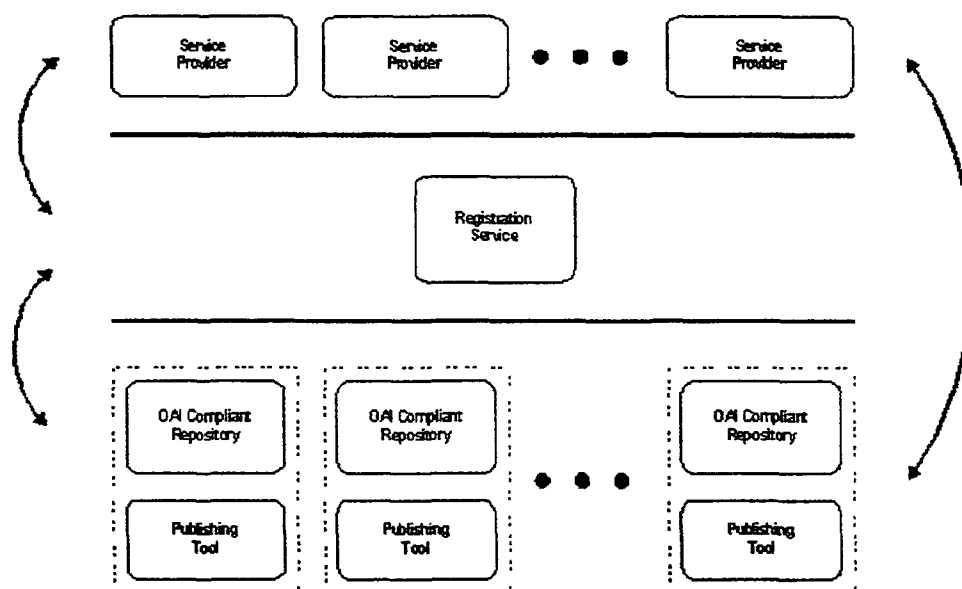


Figura 22 - Kepler - Arquitetura da Versão Inicial

A experiência de mais de um ano no desenvolvimento, utilização e suporte do Kepler revelaram algumas limitações de seu uso [13], que deram origem à nova versão, o *Extended Kepler Framework (EKF)*.

3.2.2.1.1. Limitações da Versão Inicial

- Servidor de Registro Centralizado - ponto frágil em caso de falha.
- Protocolo - o *archivelet* usa protocolo não-padronizado para registro, o que dificulta o desenvolvimento de ferramentas por terceiros.
- Segurança e NAT (*Network Address Translation*) - Quando um *archivelet* encontra-se sob um *firewall* ou *NAT proxy*, o provedor de serviço tem dificuldade de acessá-lo.
- Disponibilidade - A máquina que contém o *archivelet* é tipicamente um micro doméstico ou de trabalho, o qual não fica permanentemente *online* ligado, além de utilizar endereçamento dinâmico de IP. Tudo isso dificulta a tarefa de coleta (*harvesting*).

- Atualização - O provedor de serviços usa o protocolo OAI-PMH para coletar informações dos *archivelets* e periodicamente contacta os provedores de dados em busca de novos dados. Este modelo é mais indicado para Bibliotecas Digitais, que possuem muita alteração/entrada de dados, proveniente de fontes limitadas. Já o modelo do Kepler representa o oposto: são diversas fontes (*archivelets*) que realizam pouca alteração/inclusão. Além do mais, os provedores do Kepler (*archivelets*) são instáveis, e uma boa porção deles podem estar *offline* no momento da coleta de dados. Estas poucas alterações realizadas pelo *archivelets* devem ser imediatamente disponibilizadas ao provedor de serviço devido ao pouco tempo em que um *archivelet* normalmente fica no ar.
- Texto Completo X Metadado - Como um *archivelet* não está no ar todo o tempo, é importante coletar os textos completos (ao invés de apenas os metadados) para aumentar a disponibilidade destes textos para os usuários finais.

3.2.2.2. O Projeto EKF

A nova versão do Kepler, ou, EKF (*Enhanced Kepler Framework*), disponibilizada em 26/01/2003 [14], surge para resolver os problemas encontrados na primeira versão do software.

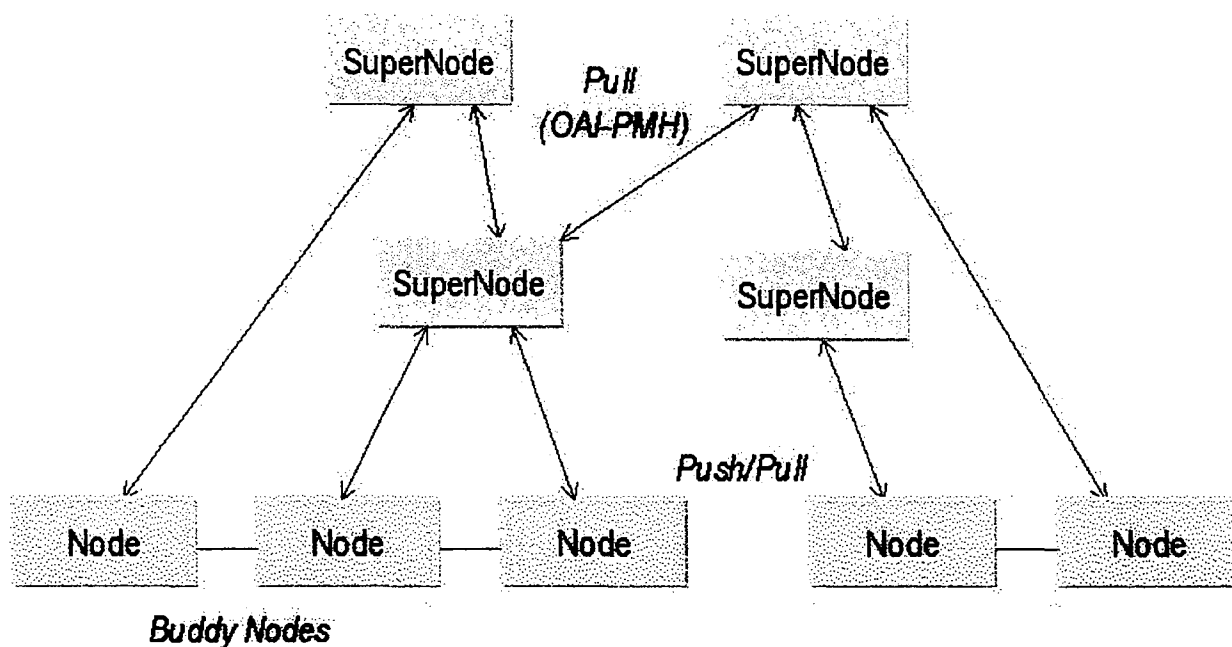


Figura 23 - Kepler - Arquitetura EKF

Ao contrário do modelo "pull" usado pelo OAI-PMH (que "puxa" as informações dos provedores de dados periodicamente), sugere-se um modelo "push" e um modelo híbrido "push/pull" para atender as questões de escalabilidade, segurança e atualização. Através destes novos modelos, o *archivelet* pode trocar documentos e metadados com outros *archivelets* (chamados *buddy nodes*), resultando em replicação de metadados e textos completos e aumento da disponibilidade (Figura 23 [14]). Surge também o conceito de Super-Nó (*SuperNode*), que coleta informações de um número de *archivelets* e provê serviços de consulta além das funcionalidades básicas de um *archivelet*. O protocolo do EKF é aberto e espera-se que possa incentivar o desenvolvimento de ferramentas Kepler por terceiros.

O EKF é um esforço conjunto entre *US Geological Survey (USGS)*, *Old Dominion University (ODU)*, *Los Alamos National Laboratory (LANL)* e *Open Language Archives Community (OLAC)*.

3.2.2.3. Por que não usar o EKF agora?

O EKF é um projeto de 36 meses, de acordo com o cronograma apresentado em [14] e iniciado em Janeiro/2003, o que extrapola o prazo para conclusão deste trabalho. Infelizmente, neste momento só podemos apresentá-lo de maneira teórica de acordo com seu projeto.

3.2.3. Funcionalidade

No EKF, cada *archivelet* passa de mero provedor de dados na versão inicial a provedor de dados e *harvester* (coletor de dados) nesta nova versão. Ele será um provedor de dados OAI quando uma solicitação de protocolo OAI-PMH chegar, e ao mesmo tempo poderá coletar registros de outros *archivelets* para aumentar a disponibilidade dos dados. Esta funcionalidade caracteriza os *Buddy Nodes* (Nós-Companheiros), *archivelets peer-to-peer* que poderão replicar metadados para aumentar a disponibilidade de documentos.

Já os Super-Nós terão todas as funcionalidades de um *archivelet*, mas, coletarão (*harvests*) um maior número de *archivelets* e terão serviços agregados adicionais. Um Super-Nó seria um nó mais estável implementado a nível institucional. O Super-Nó também pode ser coletado por outro Super-Nó de nível superior, caracterizando uma estrutura escalável e flexível.

3.2.3.1. Biblioteca Digital Comunitária

Conceitualmente, o Novo Kepler (EKF) também oferece uma nova proposta, a de Biblioteca Digital Comunitária (*Communal Digital Library - CDL*) como uma federação de pequenos grupos de Bibliotecas Digitais [14]. Neste modelo, uma comunidade representa os usuários que desejam compartilhar objetos digitais de interesse comum. Um grupo é um sub-conjunto desta comunidade. Por exemplo, os

professores/pesquisadores da área de Banco de Dados do Departamento de Informática da UFPR formam um grupo e junto com outros grupos de professores e pesquisadores de outras Universidades e outras Entidades formariam uma comunidade.

3.2.4. Disponibilidade

O Kepler pode ser obtido gratuitamente mediante simples cadastro na página <http://kepler.cs.odu.edu:8080/kepler/downloads.htm>

3.2.5. Utilização

Desde sua criação em 2001, a versão inicial do Kepler teve centenas de *downloads*, e vários *archivelets* foram registrados, porém, em caráter de testes, sem conteúdo significativo.

3.2.6. Protocolo OAI-PMH

O Kepler implementa o protocolo OAI-PMH, mas, devido à característica individual (não-institucional), os "provedores de dados" (*archivelets*) criados ficam pouco tempo no ar, por se tratarem de computadores domésticos e não de servidores todo o tempo à disposição dos usuários. Isto gerou um questionamento quanto ao modelo de sincronismo do protocolo, e à sugestão de uma extensão ao mesmo.

3.2.6.1. Modelos Push/Pull e Híbrido

O problema de sincronismo entre provedores de dados e serviços, não é um problema do Kepler, isoladamente, mas, do protocolo OAI-PMH de maneira geral.

Existem três modelos de implementação do sincronismo entre os provedores de dados e serviços [13]:

- *Pull* (Puxar) - Recuperação unilateral por parte do provedor de serviços, sem informação prévia (usada pelo OAI-PMH).
- *Hybrid Push/Pull* (Empurrar/Puxar Híbrido) - Recuperação pelo provedor de serviços, após notificação por parte do provedor de dados.
- *Push* (Empurrar ou Enviar) - Notificação seguida de um envio pelo provedor de dados.

O maior problema do modelo *Pull* é a varredura constante que deve ser realizada para sincronizar os PSs (provedores de serviços) e PDs (provedores de dados), e que se mostra ineficiente se apenas um pequeno percentual de dados são alterados/inseridos, além de gerar sobrecarga nas redes e servidores. Isto ocorre quando o repositório já fez o cadastro inicial de suas obras, e outras vão se juntando gradativamente, com frequência irregular. Em alguns repositórios, após a carga inicial, não há mais cadastramento, ficando o acervo estático. É claro que esta não é uma situação ideal, mas, ocorre em vários repositórios da Iniciativa. É possível visualizar a quantidade de registros por repositório através do OAISter [8].

Se a frequência das varreduras do protocolo também for espaçada, corre-se o risco dos PSs tornarem-se inconsistentes com os PDs.

Para minimizar estes problemas, são propostos os modelos Híbrido e *Push*. No modelo Híbrido, o provedor de dados notifica o provedor de serviços quando há uma inclusão/alteração e só então é feita a coleta de metadados pelo protocolo. O protocolo não necessita fazer varreduras desnecessárias até que haja outra notificação pelo provedor de dados.

Já o modelo *Push* descaracteriza o conceito de *Harvesting* (colheita), pois, entrega diretamente ao provedor de serviços as alterações, instituindo assim, um

novo conceito: o de Entrega ou *Delivery*. Este modelo, além de notificar o provedor de serviços de que há inclusão/alteração, envia o(s) registro(s) afetado(s) ao provedor de serviços, dispensando a varredura do protocolo.

Na arquitetura do Kepler, os *archivelets* (PDs) são numerosos e as alterações são esparsas e devem aparecer rapidamente no PS devido à natureza instável do *archivelet*.

3.2.6.2. Extensão ao OAI-PMH

O EKF cria uma extensão ao OAI-PMH, adicionando três verbos ao provedor de serviço (Figura 24) [13].

Para o modelo Híbrido são adicionados os verbos: *AddFriend* e *Notify*. O verbo *AddFriend* informa o provedor de serviço da existência de um PD. O verbo *Notify* informa o PS de que o PD está no ar / fora do ar, ou que um novo dado está disponível.

Já o verbo *Push Metadata* implementa o modelo *Push* e permite ao PD enviar metadados diretamente ao PS.

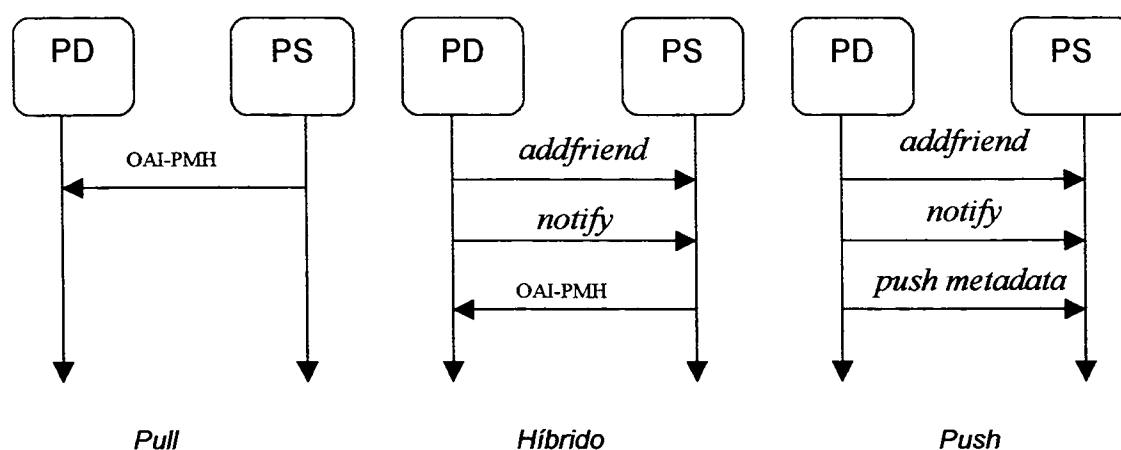


Figura 24 - Modelos de Implementação de Sincronismo entre os Provedores.

Um *archivelet* (como também é um *harvester*) pode trabalhar com a combinação dos três modelos, por exemplo, coletando (*pull*) uma vez por semana mas respondendo imediatamente quando uma solicitação do modelo Híbrido ou *Push* chegar. Espera-se que os modelos Híbrido e *Push* sejam mais utilizados no nível dos *archivelets* e que o modelo *Pull* seja mais frequentemente usado no nível *SuperNode* [13].

3.2.7. Custos

A versão inicial do Kepler não tinha qualquer custo de obtenção. Tinha uma instalação muito simples, rodava em qualquer máquina com configuração mínima que suporte acesso razoável à internet, e é compatível com ambiente Windows e Unix/Linux, podendo ser feita pelo próprio usuário final sem problemas, o que elimina custos de mão-de-obra especializada ou de hardware específico.

A nova versão (EKF) promete manter os mesmos princípios que, segundo os criadores do Kepler, são a base para o sucesso da ferramenta:

- Gratuidade
- Facilidade de Uso
- Obtenção via Internet
- Auto-Instalação
- Total Controle por parte do Usuário (Auto-Publicação)

3.3. DSPACE

3.3.1. Finalidade

Da parceria entre as bibliotecas do MIT (*Massachusetts Institute of Technology (MIT)*) e da *Hewlett-Packard Labs* há pouco mais de dois anos atrás, e com o investimento de 1,8 milhões de dólares, foi disponibilizado em 4 de novembro de 2002 o *DSpace*, um sistema para coleta, preservação, indexação e distribuição de material científico e acadêmico. Além de textos, o *DSpace* está preparado para gerenciar diversos formatos, tais como: áudio, vídeo, imagens, conjuntos de dados, programas de computador, simulações, e outros formatos.

O *DSpace*, apesar de desenvolvido para o MIT, é planejado para servir a outras instituições, sob licença de código aberto.

O modelo do *Dspace* gira em torno de unidades chamadas Comunidades (faculdades, departamentos, laboratórios, e outros centros do Instituto). Cada comunidade pode adaptar o sistema de acordo com suas necessidades e gerenciar o seu processo de submissão através da configuração de *workflows*.

3.3.2. Disponibilidade

O *DSpace* é um programa de código aberto, distribuído livremente sob licença BSD (*Berkeley Software Distribution*), que pode ser obtido através de qualquer um dos endereços abaixo:

<http://dspace.org/technology/download.html>

<http://sourceforge.net/projects/dspace/>

3.3.3. Utilização

As instalações atuais do DSpace são:

- *CTPID (Center for Technology, Policy, and Industry Development)*
- *Department of Ocean Engineering*
- *LIDS (Laboratory for Information and Decision Systems)*
- *MIT Press*
- *Sloan School of Management*

3.3.4. Arquitetura

O DSpace é um software de código aberto, desenvolvido para rodar em plataformas Unix, com codificação em Java. Utiliza SGBD PostgreSQL, Apache, Tomcat, Jena (HP), OAICat (OCLC) e outras bibliotecas, todos softwares livres.

3.3.5. Funcionalidade

O principal diferencial do *DSpace* é sua proposta de acomodar diferentes estruturas de submissão, necessárias a um sistema multidisciplinar. Isto quer dizer que cada comunidade, seja ela faculdade, departamento, laboratório de pesquisa ou outras, pode decidir que materiais devem ser submetidos, por quem, e sob quais restrições. Quem pode depositar os itens, que tipos de itens, quem deverá revisar e aprovar a submissão, em quais coleções poderão depositar materiais, quem poderá ver os itens depositados, e outras. Todas estas questões serão definidas pelos representantes da comunidade em conjunto com a equipe das bibliotecas DSpace e modelarão um fluxo de trabalho para cada coleção de acordo com suas decisões.

Em relação a metadados, o DSpace usa o padrão *Dublin Core*, com apenas três campos obrigatórios: título, idioma e data de submissão. Porém, outros metadados podem ser criados.

Pode-se definir contextos de trabalho completamente diferentes dentro do Dspace. Por exemplo, podemos criar uma coleção de artigos de altíssima qualidade, onde teremos papéis de “publicadores” para a maioria dos usuários, um pequeno grupo de “revisores”, um “editor de metadados”, e ainda um “coordenador” do grupo. O item depositado deverá passar por uma revisão e receber as aprovações necessárias antes de aparecer no acervo *DSpace*. Em contrapartida, podemos ter uma outra coleção totalmente simplificada, onde todos os membros da faculdade/instituto poderiam depositar itens (que apareceriam no acervo diretamente) e onde qualquer pessoa – público em geral – poderia ter acesso a estas informações.

Esta flexibilidade é possível através da estrutura configurável do Modelo de Dados do DSpace (Figura 25 [49]). Cada objeto está exemplificado na Tabela 3 [49].

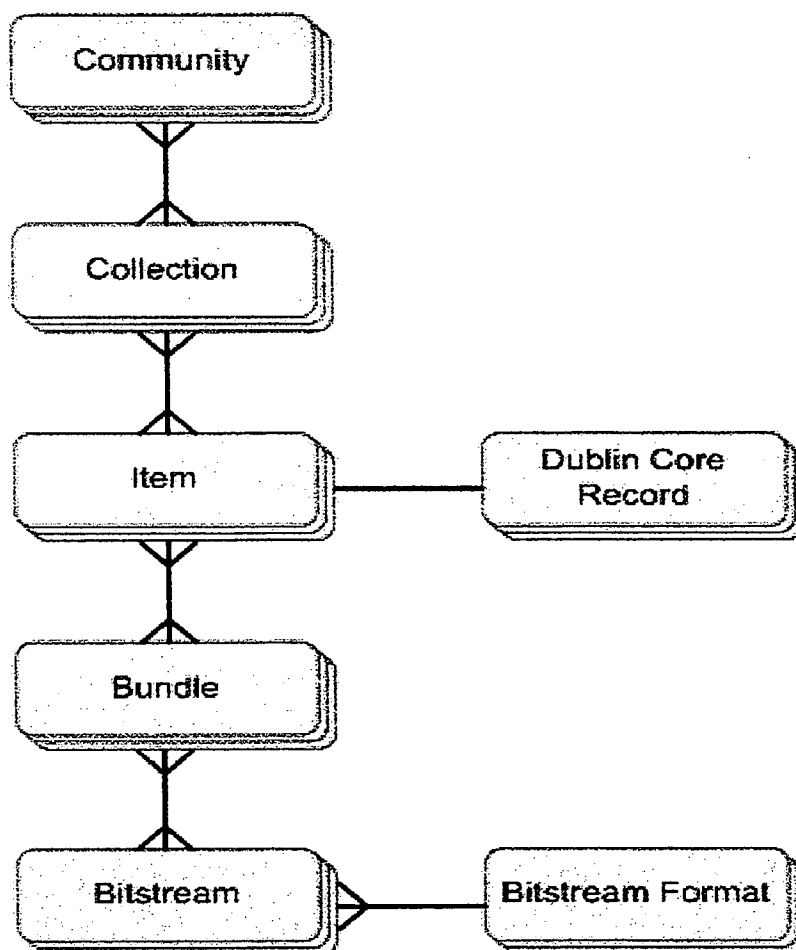


Figura 25 - DSpace - Modelo de Dados

Objeto	Exemplo
<i>Community</i>	Dpto. Informática, Centro de Pesquisas Geológicas
<i>Collection</i>	Artigos Dpto. Inform., Estatísticas Centro Pesq. Geológ.
<i>Item</i>	Um artigo, uma planilha, um vídeo sobre uma pesquisa
<i>Bundle</i>	Um grupo de <i>bitstreams</i> no formato de imagens e HTML que compõem um documento HTML
<i>Bitstream</i>	Um único arquivo HTML, uma imagem, um código-fonte
<i>Bitstream Format</i>	Microsoft Word 6.0, formato imagem JPEG, formato PDF

Tabela 3 - DSpace - Exemplos de Objetos do Modelo de Dados

3.3.6. Protocolo OAI-PMH

Para garantir a interoperabilidade com outros parceiros DSpace, e com outros repositórios digitais, o Dspace utiliza o Protocolo OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*), através da ferramenta OAICat. O OAICat é um software de código aberto desenvolvido pelo OCLC - *Online Computer Library Center* que implementa suporte ao protocolo OAI-PMH em repositórios de dados já existentes [45].

3.3.7. Custos

Apesar de não haver custos de software na instalação do DSpace, haverá no mínimo a necessidade de um administrador Unix para sua instalação, e, na maioria dos casos, de um programador Java para localização/customização e melhorias desejadas.

Devido ao seu suporte a estruturas mais complexas, pode ser necessária também a criação de uma comissão de gerenciamento do fluxo dos documentos (revisores, editor de metadados, coordenador, e outros), além da equipe técnica.

Sua instalação e configuração dependerão do contexto a ser definido pela instituição, que pode variar de um simples repositório com livre acesso até o mais controlado acervo.

Com estas características, o Dspace torna-se um software mais complexo que o Eprints, acarretando maiores custos de instalação e manutenção.

3.4. CONCLUSÕES

Das três ferramentas apresentadas, a mais estável e em utilização comprovada pelo maior número de instituições é o *Eprints*. Tem caráter institucional, implementa um ambiente completo de submissão/publicação/consulta, é de fácil aprendizado e tem sido largamente utilizado pela comunidade acadêmica mundial. Pode-se optar por repositórios com ou sem o papel de Editor (realiza a aprovação do material para publicação). Como é o próprio usuário que realiza os cadastramentos e submissões, não há o ônus da alocação de um profissional de biblioteconomia/gestão da informação para a criação dos registros. O usuário também é responsável pela manutenção (alterações, novas versões) de seus documentos.

O EKF deve apresentar os primeiros resultados concretos dentro de alguns meses. No momento, proporciona auto-instalação, porém, devido à sua utilização pessoal, implementa repositórios instáveis e de maneira informal, sem garantia de qualidade de conteúdo.

Já o DSpace possui arquitetura mais completa, requer um planejamento prévio com o envolvimento da comunidade em questão junto ao pessoal técnico, para a definição das características do repositório a ser implementado, e, quanto mais complexo, maiores serão seus custos de implantação/gestão.

3.4.1 Resumo

Com o objetivo de facilitar a visão geral das três ferramentas, é apresentado um resumo dos critérios avaliados neste capítulo na **Tabela 4**, a seguir:

	EPRINTS	KEPLER	DSPACE
<i>Finalidade</i>	Criação de repositórios de eprints (documentos eletrônicos)	Ferramenta auto-publicação de documentos científicos/acadêmicos	Sistema para coleta, preservação, indexação e distribuição de material científico e acadêmico em diversos formatos, inclusive multimídia.
<i>Disponibilidade para Download</i>	Desde novembro de 2000. Atualmente na versão 2.2.1 (nov/2002).	Versão inicial: desde 2001 EKF: projeto de 36 meses a partir de janeiro/2003	Desde novembro/2002.
<i>Utilização</i>	71 repositórios em diversos países (Anexo II)	Versão inicial: centenas de downloads e publicações em caráter experimental.	Até junho/2003, por 5 instituições americanas (ver 3.2.3.)
<i>Arquitetura</i>	Linguagem Perl, roda como módulo Apache/ Linux. Usa MySQL para armazenar dados.	Inicial: rede P2P (peer-to-peer), protocolo de registro não-padronado. EKF: protocolo de registro padronizado, extensão ao OAI-PMH	Linguagem Java, plataforma Unix. Utiliza BD PostgreSQL, Apache, OAICat (OCLC) e outras bibliotecas, tudo software livre.
<i>Funcionalidade</i>	Interface Web para cadastro de usuários, pesquisa, gerenciamento e submissão de documentos.	Archivelet baixado pela internet, multi-plataforma, auto-instalável, implementa ambiente de submissão / gerenciamento de documentos. Consultas via provedor de serviços. EKF: novos conceitos (buddy nodes, super nodes)	Proposta de estrutura flexível, de modo a acomodar diferentes modelos de submissão e gerenciamento, de acordo com o tipo de instituição (faculdade, laboratório, etc).
<i>OAI-PMH</i>	Implementa versões 1.1 e 2.0 do protocolo.	Implementa as versões do protocolo. EKF: propõe criação de 3 novos verbos e novos modelos de sincronismo.	Implementa a versão 2.0 do protocolo.
<i>Custos Sw</i>	Não há.	Não há.	Não há.
<i>Custos Hw</i>	Pode-se usar qualquer estação que rode Linux. Em caso de aquisição, custo básico de R\$ 1.500,00.	Qualquer estação com capacidade de acesso à internet.	Varia de acordo com o projeto a ser definido pela instituição.
<i>Custos Mão-de-Obra</i>	Necessita profissional com conhecimentos de administração Linux e Apache para instalação.	Não requer habilidades específicas	Requer equipe de planejamento da estrutura a ser implementada. Para implantação, um administrador Unix e um programador Java para customização.

Tabela 4 - Resumo da Análise das Ferramentas

Tomando por base estas informações, o *Eprints* foi selecionado como a opção com melhor custo/benefício para a criação de repositórios ou bibliotecas digitais em

Universidades e Instituições de Pesquisa. De maneira a fundamentar estas conclusões sobre as características e baixo custo do *Eprints*, foi realizada uma instalação do mesmo com a finalidade de implementar um repositório voltado ao armazenamento da produção científica e acadêmica do Departamento de Informática da UFPR.

Esta experiência, que originou o **UFPR Eprints**, encontra-se detalhada no capítulo 4, a seguir.

4. UFPR EPRINTS

Se você tem conhecimento, permita que outros acendam suas velas nele.
Margaret Fuller
1810-1850

Com o objetivo de fundamentar as conclusões sobre as características e baixo custo do *Eprints*, foi realizada uma instalação do mesmo com a finalidade de implementar um repositório voltado ao armazenamento da produção científica e acadêmica do Departamento de Informática da UFPR. Esta análise visa comprovar na prática, o desempenho do programa nos critérios estabelecidos no capítulo anterior.

4.1. PROCEDIMENTOS ANTERIORES À INSTALAÇÃO

Antes de partir para a instalação efetiva do *software Eprints*, é preciso atentar para alguns pontos que precisam ser definidos com antecedência, e algumas políticas que precisarão ser implementadas para o bom funcionamento do ambiente [5, 9, 10, 38]. Para tanto, é preciso responder às seguintes questões antes de se iniciar a instalação do programa:

4.1.1. Finalidade do Repositório (*Archive*)

Qual o escopo do repositório a ser instalado? Ele é voltado para artigos departamentais, teses/dissertações ou para todo tipo de artigo científico de uma determinada área? Esta é a decisão inicial que vai orientar a identidade de seu repositório.

O **UFPR Eprints** abrigará a produção científica do Departamento de Informática, seja ela em forma de artigos, teses ou dissertações.

4.1.2. Definição dos Tipos e Formatos de Documentos

De acordo com a decisão anterior, que tipos de documentos seu repositório vai aceitar? No caso de artigos, serão aceitos apenas *post-prints* (após a publicação) ou também *pre-prints* (antes da publicação) ?

E quanto ao formato? A instalação padrão do *software Eprints* aceita os formatos

- Postscript
- PDF
- ASCII
- HTML

O **UFPR Eprints** aceitará os formatos-padrão do Eprints como forma de avalia-los. Serão aceitos artigos do tipo *pre-prints* (para fins de revisão/complementação) e *post-prints*.

4.1.3. Procedimento de Submissão / Supervisão

Apesar da corrente "*Eprint*" [1] e do próprio *software Eprints* [21] pregar e facilitar a auto-publicação, é consenso geral que uma certa intermediação é necessária no processo de submissão de um *eprint* a um repositório, visando garantir a qualidade das publicações ali contidas.

É preciso decidir quem vai fazer a submissão (o próprio usuário ou um bibliotecário) e, quem vai fazer o papel de Supervisor do repositório, analisando os metadados, os documentos anexados e fazendo sua complementação quando necessário, bem como publicando definitivamente (em caso de aceitação) ou devolvendo-o ao usuário (em caso de rejeição/correção).

O **UFPR Eprints** terá um usuário "Editor" que fará a supervisão dos *eprints* submetidos. A execução desta tarefa futuramente deverá ser delegada a um usuário treinado em metadados e na necessidade da manutenção de padrões.

4.1.4. Metadados

Os metadados são as informações sobre as publicações que estão armazenadas e têm papel crucial num repositório *eprints*, dado que serão exportados pelo protocolo OAI e coletados pelos provedores de serviços para serem pesquisados. Garantir a obediência aos padrões Dublin Core e a qualidade destes metadados é dever do Supervisor do repositório.

No momento, o preenchimento dos metadados do Eprints é totalmente aberto, não havendo nenhum controle interno sob a forma de restrições (*constraints*) automáticas dos conteúdos. Isto torna o trabalho de revisão de metadados altamente subjetivo e sujeito a equívocos.

O **UFPR Eprints** fará uso dos metadados originais do Eprints e, posteriormente, serão levantadas as necessidades de novos campos para identificação das especificidades dos documentos publicados.

4.1.4.1. Regulamentação do Uso do Padrão Dublin Core no Eprints

Sendo o Conjunto de Elementos Dublin Core o padrão recomendado pelo Protocolo da Iniciativa Arquivos Abertos (*Open Archives Initiative Protocol for Metadata Harvesting - OAI-PMH*), a Universidade de Bath (Reino Unido) publicou um artigo recentemente (12/03/2003), com recomendações de uso do *Dublin Core* para descrever os documentos armazenados nos repositórios *Eprints* [28].

Estas recomendações são uma tentativa de padronizar o conteúdo dos metadados, facilitando assim suas recuperações. Podem ser usadas na construção de uma política interna para cadastramento de documentos no repositório e ser divulgada aos usuários potenciais. A tradução destas recomendações encontra-se no Anexo V.

4.1.5. Texto Completo

Anexar ou não o texto completo ao *eprint*, complementando o registro (metadado + texto completo)? Esta é uma questão muito discutida e o consenso diz que anexar o texto completo não deve ser obrigatório, mas, bastante incentivado, dada a natureza "*Self-Archiving*" da ferramenta.

O **UFPR Eprints**, embora não torne obrigatória a submissão do texto completo, incentivar-á com bastante ênfase esta prática.

4.2. A IMPLANTAÇÃO DO EPRINTS

4.2.1. Obtenção do Software

O GNU Eprints é distribuído sob licença GPL (*General Public License*), e pode ser baixado diretamente do site <http://software.eprints.org>

4.2.2. Documentação

A documentação do *Eprints* vem junto com o programa e também pode ser obtida em vários formatos no endereço <http://software.eprints.org/documentation.php>. É imprescindível uma boa leitura antes de se iniciar a instalação dos softwares necessários e do próprio *Eprints*.

4.2.3. Listas de Discussão

Dada a natureza cooperativa das comunidades de software livre, no *Eprints* não poderia ser diferente: caso alguma dúvida não seja sanada através da documentação do programa, pode-se recorrer às listas de discussão (<http://software.eprints.org/maillist.php>) que são parte fundamental na comunicação dos usuários e no aperfeiçoamento da ferramenta.

A comunidade *Eprints* mantém as seguintes listas [26]:

- *eprints-announce* - informa a disponibilização de novas versões do software *Eprints*.
- *eprints-underground* - voltada aos interessados na implantação do *Eprints* em suas instituições. Trata das funcionalidades do software, configuração, melhores práticas, políticas relacionadas, metadados, casos de implantação, etc. Não é uma lista técnica, que discute "como fazer", mas, "o que fazer".
- *eprints-tech* - aqui sim, pode se discutir o lado técnico do *Eprints*, como dúvidas de instalação, configuração e novas versões.

4.2.4. Requisitos de Hardware

Qualquer computador do tipo PC capaz de rodar o sistema operacional GNU/Linux ou similar. É óbvio que quanto mais rápido e melhor for a configuração da máquina, melhor desempenho o sistema terá.

4.2.5. Requisitos de Software

É possível escolher entre obter todos os *softwares* necessários para a instalação do *Eprints* diretamente na página <http://www.eprints.org/files/tools/> (não é necessário instalar tudo, apenas os listados a seguir) ou diretamente das páginas oficiais de cada *software*. Fica a critério de quem estiver fazendo a instalação.

Os *softwares* necessários para a instalação do *Eprints* são:

- Sistema Operacional GNU/Linux
- MySQL - instalar a versão mais recente do MySQL 3.
- Servidor Web Apache com módulo mod_perl (recomendado: apache 1.3.14 com mod_perl 1.25)
- Linguagem de programação Perl (recomendado: perl 5.6.1)

4.2.6. Instalação

A instalação padrão do Eprints é relativamente fácil para quem tem experiência em administração Unix/Linux e conhecimentos de Apache. O capítulo "*How to Install Eprints (and get started)*" da documentação do sistema detalha todos os passos necessários para a instalação e primeiros passos no sistema. Também é possível encontrar um guia de instalação do *Eprints* em português na monografia de Hernani Vila [7], que relata sua experiência na instalação do Coruja.

4.2.7. Início

Após a instalação do software é necessário agendar no sistema operacional o processo "*generate_views*" que cria as listas das consultas por assunto (*Browse/Subjects*) e por ano (*Browse/Year*).

4.2.8. Tradução e Customização

Para o projeto-piloto deste trabalho, o ambiente do *Eprints* permanece na interface original, sem qualquer alteração no fluxo de telas, *templates* ou tradução. As alterações para melhorias de interface, bem como a tradução do software serão objeto de trabalhos futuros.

4.2.9. Os Registros nas Comunidades Eprints e OAI

Após testar e comprovar a perfeita operação do *archive*, é hora de inscrevê-lo na lista de repositórios que implementam *Eprints* e como Provedor de Dados na Iniciativa *Open Archives*.

4.2.9.1. Registro na *Eprints.org*

A inclusão na lista da comunidade *Eprints* é feita mediante comunicação da implementação diretamente ao desenvolvedor do programa: Christopher Gutteridge (cjg@ecs.soton.ac.uk).

4.2.9.2. Validação do Repositório

O registro na Iniciativa *Open Archives* deve ser precedido por uma verificação da conformidade do repositório com o protocolo, através do serviço *Repository Explorer*, disponível na opção *Tools* da página da OAI [20]. O *Repository Explorer* realiza quase 30 acessos ao repositório que está sendo avaliado e retorna a conformidade (ou não) aos verbos do protocolo. Os resultados da verificação do **UFPR Eprints** encontram-se no Anexo III.

4.2.9.3. Registro na Iniciativa *Open Archives*

Para solicitar o registro na OAI através da opção "*Register as Data Provider*" [20], é preciso que haja pelo menos um item (um "*eprint*" cadastrado) no provedor de dados candidato. Feita a solicitação, a OAI realiza testes de validação do repositório, e após algumas horas retorna uma mensagem ao endereço do Administrador do repositório com o resultado (aprovação ou não) da solicitação de registro. Para obter o registro e tornar-se um provedor de dados, o repositório não pode apresentar nenhum erro. Periodicamente são realizados testes de conformidade nos provedores de dados, para manutenção do registro.

A mensagem da OAI informando o registro como provedor de dados do **UFPR Eprints** encontra-se no Anexo IV.

A partir do ingresso na comunidade OAI, o *Harvester* passa a coletar os metadados do provedor de dados e seu conteúdo passa a ser acessível através dos

provedores de serviços, como o *OAIster* [8]. A Figura 26 mostra os registros do *UFPR Eprints* disponibilizados em pesquisa ao *OAIster*.

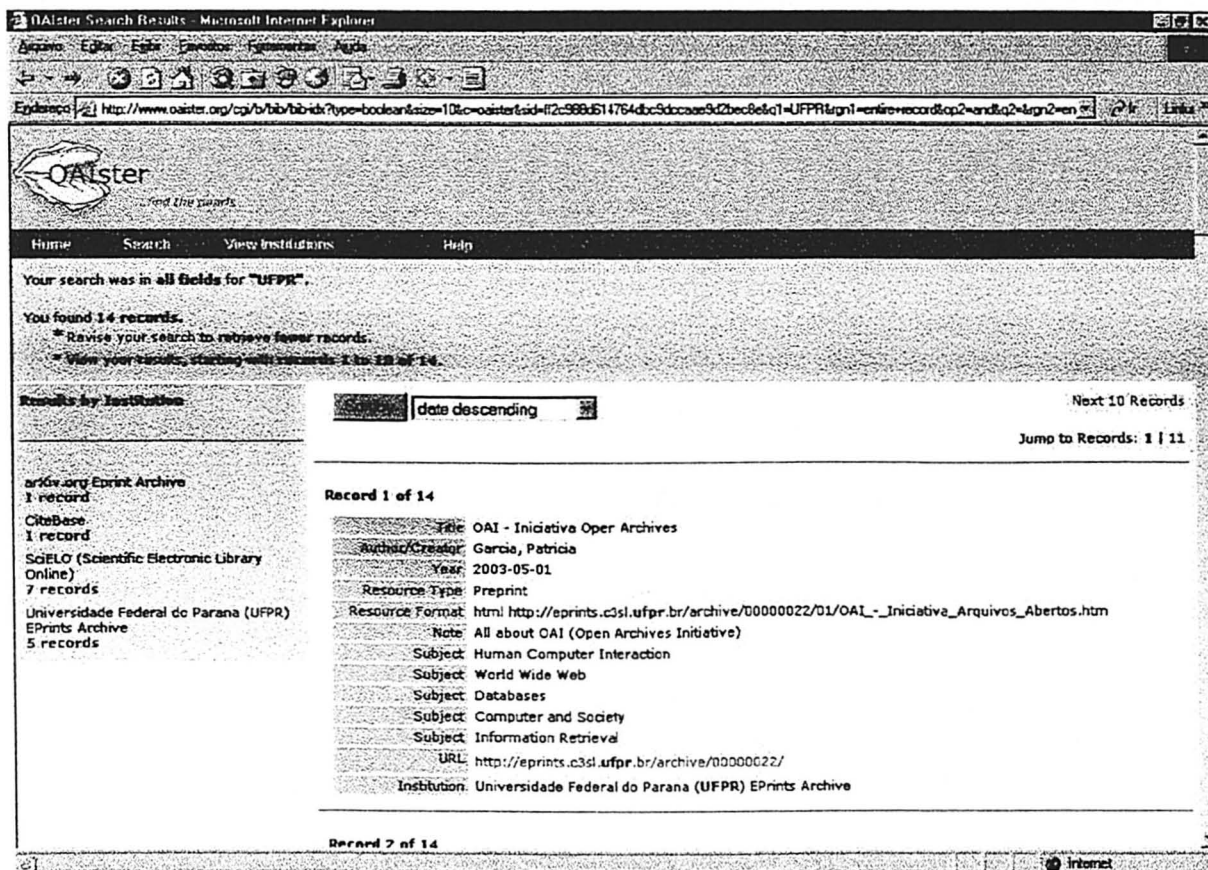


Figura 26 - Provedor de Serviços mostrando registros do *UFPR Eprints*

4.2.10. O Maior Desafio: dar "vida" ao repositório

Instalar e configurar o *Eprints*, apesar de não ser tarefa trivial, consome cerca de um dia de trabalho de um profissional experiente. Porém, colocar o repositório no ar não é o maior desafio.

Sem conteúdo, o repositório não tem valor, não atinge seu objetivo, que é compartilhar conhecimento e oferecer visibilidade ao trabalho de pesquisadores/cientistas.

A seguir são apresentadas algumas sugestões de ações que visam incrementar a adesão aos repositórios.

4.2.10.1. O Envolvimento

O sucesso de um repositório começa antes da instalação do software, quando a comunidade que o partilhará é envolvida na idéia de divulgar suas idéias para um público direcionado, que busca informações de qualidade, pois, o acesso aos repositórios é feito através de serviços de busca específicos, aqueles que implementam o *harvesting*, o que evita consultas em vão por um público que não as deseja (como ocorrem em portais de busca genéricos tipo *Google*, *AltaVista*, *Cadê* e outros).

Apresentar as entidades participantes da Iniciativa *Open Archives* e as que usam o software *Eprints* também ajuda na compreensão da relevância do repositório, pois, os trabalhos serão expostos no mesmo nível das mais conceituadas universidades e instituições de pesquisa do mundo. Para se conseguir publicar trabalhos nos periódicos convencionais da mesma maneira seria muito difícil.

Outra questão a ser enfocada na conscientização da comunidade é o pioneirismo e a oportunidade que estes repositórios representam. Existem poucos repositórios fora do eixo Estados Unidos - Europa e, na América Latina, o Brasil é o único participante na instalação de *Eprints* até o momento. Quanto à oportunidade, há que se ressaltar que antes do protocolo OAI-PMH e do surgimento do *Eprints*, um repositório digital departamental só poderia alcançar alguma projeção parecida caso fosse agraciado com um orçamento grandioso, para a realização de um projeto extenso e custoso em termos de tempo e recursos (o que é sabido ser muito difícil em instituições públicas de países em desenvolvimento).

4.2.10.2. Apresentação e Treinamento

Após a instalação do *Eprints* e com o repositório no ar, é importante apresentá-lo à comunidade por meio de uma palestra ou um seminário em que as funcionalidades do software sejam apresentadas para que se quebre a barreira do desconhecimento no uso da ferramenta. Questões como consulta, criação dos usuários, processo de submissão, aprovação e publicação devem ser esclarecidas de maneira interativa. Após o treinamento, o repositório deve entrar em um breve período de exploração, onde os usuários ainda estarão pouco à vontade para navegar e submeter seus trabalhos.

4.2.10.3. Adesão de líderes da comunidade

Neste momento será de fundamental importância o apoio de pessoas-chave da instituição/departamento onde se está implantando o *Eprints*. A adesão destas pessoas ao repositório, cadastrando-se e publicando seus trabalhos, bem como comentando sobre o mesmo com seus colegas e pessoas de outras instituições/ departamentos fará com que pouco a pouco a barreira da nova tecnologia seja quebrada.

4.2.10.4. Suporte aos usuários

Um bom "suporte técnico" a ser oferecido pelos responsáveis ou administradores do repositório também será de grande valia para que as pessoas possam tirar suas dúvidas.

4.2.10.5. Operacionalização e Divulgação

Com algum conteúdo relevante (teses, dissertações, artigos - dependendo da finalidade do repositório), o repositório poderá ser apresentado às outras áreas

que poderão usufruir de seu conteúdo. No caso do **UFPR Eprints**, poderá ser apresentado aos alunos de graduação e pós-graduação como fonte de consultas, e aos demais Departamentos da Universidade como uma alternativa para a divulgação de seus trabalhos, originando novos repositórios departamentais na Instituição, ou até a criação de um repositório para toda a Universidade.

Quanto mais ampla for a divulgação do repositório, maior a chance de sua finalidade de divulgação do conhecimento ser alcançada, o que justifica esforços para a promoção do repositório em outras instituições de ensino e em eventos da área científica e acadêmica. Uma boa maneira de divulgar o repositório sem custos financeiros é criando *links* para o mesmo dentro das páginas da Instituição e de eventos.

4.2.11. Custos

Após a criação do **UFPR Eprints** foi possível atestar o baixo custo de instalação do software *Eprints*. Foi utilizado um servidor Linux já existente no Departamento de Informática, e mão-de-obra interna, o que resultou em custo zero do ponto de vista de aquisição/contratação de recursos. Como já foi citado, todos os softwares necessários à instalação também são softwares livres como o *Eprints*, e portanto, também não houve custo algum em sua aquisição.

4.2.12. Conclusão

Os esforços dispendidos para a instalação, configuração e promoção do repositório podem ser realizados pela própria equipe interna à instituição, onde os custos serão apenas das horas dos profissionais alocados para estas atividades.

Os esforços de promoção do repositório visam agregar-lhe conteúdo e visibilidade, fatores críticos para o sucesso de sua criação. Muitos repositórios *Eprints* listados no site *Eprints.org* [21], inclusive os brasileiros (ver Capítulo 5)

sofrem uma "crise de conteúdo". São instalações ativas do software *Eprints*, algumas com traduções e/ou interfaces customizadas, mas, sem o conhecimento a ser compartilhado.

A utilização do *Eprints* no Brasil começou em 2001. Na América Latina, o Brasil é o único país a possuir instalações deste programa, de acordo com a lista de repositórios da página *Eprint.org* (Anexo II). As experiências brasileiras na área de *Eprints* e Arquivos Abertos estão descritas no próximo capítulo.

5. SOLUÇÕES EPRINTS/OAI NO BRASIL

A imaginação é mais importante que o conhecimento.
Albert Einstein
1879-1955

No Brasil, desde 2001 já existem alguns repositórios com interface OAI, onde o IBICT foi o pioneiro, através do projeto da BDB (Biblioteca Digital Brasileira).

5.1.BDB/IBICT

O projeto Biblioteca Digital Brasileira - BDB, propõe-se a *"integrar em um único portal os mais importantes repositórios de informação digital do país, de forma a permitir consultas simultâneas e unificadas aos conteúdos informacionais destes acervos"* [37].

O modelo de interoperabilidade da BDB baseia-se em dois elementos: (1) mecanismos de submissão de consultas, a partir da interface única do portal, aos diferentes recursos informacionais que comporão a BDB e (2) conjunto de metadados que descreverão e fornecerão uma visão unificada dos diferentes conjuntos de documentos. Os detalhes deste modelo encontram-se em [39, 40].

A idéia básica da BDB, apresentada no relatório "Biblioteca Digital Brasileira - Programa de Implantação - Biênio 2001/2002" [44], era de integrar os principais acervos digitais do país. Segundo este documento, a BDB será dividida em duas grandes bibliotecas (C&T, Histórico-Cultural) com as participações das entidades:

Biblioteca de C&T

- MEC
- CNPq
- IBICT/Prossiga
- CNEN

- BIREME (SciELO)
- Outras instituições do Brasil e do Exterior - através do OAI-PMH
- IMPA e IBICT - através do OAI-PMH

Biblioteca Histórico-Cultural

- Biblioteca Nacional
- Arquivo Nacional

Embora a previsão do relatório do IBICT/Prossiga fosse a implantação da estrutura das duas grandes bibliotecas até o final de 2002, em maio/2003 encontramos na página da BDB apenas as opções :

- Biblioteca Digital de Teses e Dissertações (BDTD)
- Diálogo Científico (página da SBG, ver 5.2)
- Revista Ciência da Informação (apresentava erro em 29/05/2003 e não abria nenhuma página)

5.1.1. Biblioteca Digital de Teses e Dissertações (BDTD)

A Biblioteca Digital de Teses e Dissertações (BDTD), é resultado dos esforços de especialistas da BIREME, CNPq, IBICT, USP, PUC-Rio e UFSC e de consultores contratados pelo IBICT, com o objetivo de criar um ponto único de acesso à produção de teses e dissertações de nosso país, bem como de promover a divulgação destes trabalhos.

Há atualmente cerca de 121.000 registros na BDTD, provenientes de 17 instituições cooperantes. São metadados enviados pelas bibliotecas e inseridos na base da BDTD. Não é possível fazer submissão de teses/dissertações diretamente.

5.2. SBG/IBICT

Na lista de repositórios que rodam a versão 1 do *Eprints* no site *Eprints.org* [21], encontramos menção ao site da Sociedade Brasileira de Genética / IBICT (<http://www.sbg.ibict.br/>).

Trata-se de um repositório de documentos eletrônicos nas áreas de ensino, genética de microorganismo, evolução animal, evolução vegetal, genética humana, melhoramento vegetal e animal e mutagênese.

Apesar de estar no ar há mais de um ano, conta apenas com 20 documentos depositados.

5.3. CORUJA

O Instituto de Matemática e Estatística da USP utiliza o software *Eprints* desde dezembro de 2002 como solução para um repositório de documentos digitais, que atende os professores e alunos do Departamento de Ciência da Computação.

O portal de nome "Coruja" foi desenvolvido por Hernani Vila, sob orientação do Prof. Imre Simon [2] e encontra-se listado no site *Eprints.org* [21] como repositório que roda a versão 2 do *Eprints*. Sua publicação sobre a implantação do site [7] descreve a trajetória da instalação da "Coruja" e dá dicas de tradução do *Eprints* para o português. Hernani contribuiu gentilmente para o projeto **UFPR Eprints**, enviando arquivos traduzidos para o português que muito auxiliarão na tradução das páginas do repositório.

Infelizmente, o portal Coruja conta com poucas publicações (29/05/2003), apesar de estar no ar desde o final de 2002 e contar com interface customizada e traduzida.

5.4. SCIELO

O SciELO é uma Biblioteca Científica Eletrônica Online (*Scientific Electronic Library Online*), que disponibiliza artigos científicos integrais (texto completo). Vários *gateways* SciELO estão em operação, no Brasil, América Latina e Caribe, oferecendo acesso a publicações acadêmicas destes países e também de Portugal e Espanha.

O SciELO possui um papel importante na disseminação da literatura técnica e científica publicada em países em desenvolvimento, aumentando a visibilidade deste tipo de literatura.

A metodologia do SciELO foi desenvolvida pela BIREME (*Latin America and Caribbean Center on Health Sciences Information*), uma organização pertencente ao PAHO (*PanAmerican Health Organization*) e ao WHO (*World Health Organization*). Esta metodologia utiliza o software ISIS para o banco de dados de metadados. Este software foi desenvolvido pela UNESCO, e junto com outros software, forma as bases para sistemas de informações, bancos de dados e redes STM (*Science, Technology and Medicine*) em vários países em desenvolvimento.

Aqui no Brasil está sendo desenvolvida uma interface com o banco de dados de metadados ISIS compatível com o protocolo da Iniciativa *Open Archives* [41].

O *gateway* SciELO no Brasil tem cerca de 10.000 artigos, em mais de 90 publicações acadêmicas. Foi o pioneiro, e hoje existem *sites* no Chile, Cuba e outros países estão em processo de adoção/treinamento no SciELO.

A metodologia SciELO incorpora novos serviços, *links* com bases de dados bibliográficos como LILACS e MEDLINE e, recentemente, o sistema CVLattes (currículos de pesquisadores brasileiros) mantido pelo CNPq e com a Biblioteca Digital de Teses e Dissertações, mantida pelo IBICT.

No momento a implementação do projeto *SciELO-Open Archives* conta com os seis verbos do protocolo OAI-PMH implementados (ver Capítulo 2), e implementa o conjunto de metadados *Dublin Core* (padrão OAI). O próximo passo é desenvolver um *schema XML* para o formato LILACS (*Latin America and Caribbean Health Science Literature*), o conjunto de metadados usado pelo SciELO, onde cada registro inclui título em inglês, português e espanhol.

5.5. EDIÇÃO ELETRÔNICA DE ANAIS DE EVENTOS

A Coordenação de Informática do Centro de Informações Nucleares (CIN), órgão da CNEN (Comissão Nacional de Energia Nuclear) está lançando um projeto que visa [43] :

"a criação de um conjunto de metodologias e tecnologias que permitam criar um ambiente de submissão, avaliação e publicação de anais de eventos diretamente na Internet. Através deste ambiente, toda a tramitação dos trabalhos (submissão, revisão, correções, discussões e aceitação) seria feita por meio de ferramentas de comunicação web. Ao final desse processo de tramitação, e junto com a realização do evento, os trabalhos estarão disponíveis on-line, em texto completo para ampla consulta e discussão."

A proposta é uma implantação em duas etapas, onde inicialmente seria desenvolvido um protótipo com as principais facilidades do sistema, e em seguida, o desenvolvimento do sistema pleno, com as novas características, conforme necessidades específicas dos eventos científicos.

- Protótipo - a ser desenvolvido pela equipe do CIN, utilizando a metodologia Eprints/OAI. Terá como função testar as metodologias de submissão, arquivamento, busca e disponibilização de metadados para "*harvesting*" automático. Deverá incorporar um único congresso e será parte integrante do projeto, constituindo-se na materialização das principais funções nele descritas.

- Sistema Pleno - Abrigará a base de dados de metadados e a biblioteca digital com os textos completos dos diversos anais de eventos, que poderão ser fonte para outros serviços de informação.

5.5.1. Funcionalidades

Quando o sistema estiver totalmente implementado, oferecerá as seguintes opções referentes à gestão do evento:

1. Inscrição no evento: submissão à comissão científica do pedido de inscrição na qualidade de Autor, mediante formulário eletrônico. O sistema enviará confirmação de inscrição por e-mail.
2. Submissão de contribuições: permitirá aos usuários inscritos submeter documentos para avaliação nos formatos mais comuns. Na submissão o Autor deverá fornecer os dados cadastrais e bibliográficos do documento, que mais tarde comporão a base de dados de metadados do evento. Através de login no sistema, o autor poderá acompanhar a avaliação de seu trabalho e executar as modificações que possam vir a ser sugeridas pelos avaliadores.
3. Ambiente de avaliação: as comissões de avaliação poderão trabalhar diretamente na rede, on-line. O sistema preservará o anonimato dos trabalhos e possibilitará avaliações parciais e avaliação final (sempre notificando automaticamente o autor). Haverá também um módulo de gerência de avaliação destinado ao coordenador da comissão/comitê científico do evento. Em caso de aprovação, o sistema emitirá um modelo de declaração de cessão de direitos autorais em formato PDF que deverá ser assinado pelo autor e devolvido via fax ou correio convencional.

4. Armazenamento de documentos: cada documento receberá uma URL padronizada e única, que o identificará e este endereço deverá ser preservado através da implementação de esquemas padronizados de endereços permanentes. O armazenamento dos documentos deverá ser fisicamente independente da base de dados de metadados, sendo o relacionamento entre eles implementado via hiperlink.
5. Base de dados de metadados: será definido um conjunto de metadados capaz de descrever as características dos trabalhos de congresso, e ao mesmo tempo, compatível com a BDB (Biblioteca Digital Brasileira), tendo como base o padrão Dublin Core.
6. Base de dados da biblioteca digital de anais: os documentos em texto completo de cada evento serão armazenados de acordo com o esquema de identificação definido pelo protocolo OAI-PMH. A biblioteca digital de anais de congresso deverá ser gerenciada por um software livre para biblioteca digital a ser selecionado. Servirá de fonte de dados para diversos serviços a serem agregados.
7. Interoperabilidade: através da implementação do protocolo OAI-PMH, os metadados poderão ser coletados automaticamente pela BDB ou outro provedor de serviços que se utilize do processo de "harvesting". É desejável que o sistema também interaja com o sistema de currículo Plataforma Lattes do CNPq.
8. Módulo de edição de CD-ROM e versão impressa: o sistema deverá permitir a geração de versão em meio eletrônicos dos anais, e, opcionalmente de versão impressa.
9. Interface web de busca: o sistema deverá apresentar opções de buscas simples (caixa única, pesquisando em vários campos) ou avançada (possibilidade de escolha de campos e conectores

lógicos). O sistema também deverá oferecer acesso aos índices de forma direta (*browser*).

10. Portal web de busca unificada aos anais de congresso: será mantida uma base de dados unificada de anais de congresso, possibilitando busca integrada em todos os congressos incluídos nesta base. As *home-pages* originais dos eventos já encerrados deverão ser preservadas. Será estudada a possibilidade de integração com o Catálogo Coletivo de Anais de Congresso.

11. Padrões e Protocolos:

- OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*).
- Metadados: Padrão Dublin Core mais as extensões necessárias à descrição de trabalhos de congresso.
- Formato dos textos: PDF
- Endereço persistente: PURL da OCLC (a ser definido em comum acordo com a coordenação da BDB, Bireme e CNPq).

5.5.2. Interfaces

As possíveis interfaces do Sistema de Edição de Anais de Congresso estão apresentadas na Figura 27 [43]:

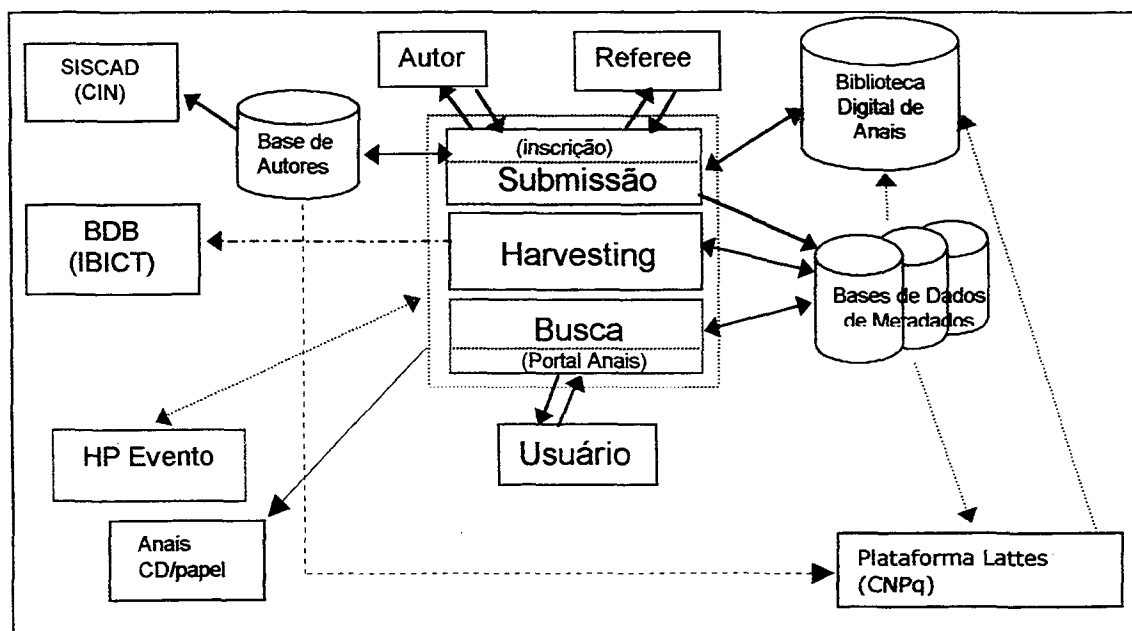


Figura 27 - Interfaces do Sistema de Edição de Anais de Congresso

No próximo capítulo são apresentadas as conclusões deste trabalho.

6. CONCLUSÃO

Tudo tem seu tempo e até certas manifestações mais vigorosas e originais entram em voga ou saem de moda. Mas a sabedoria tem uma vantagem: é eterna.
Baltasar Gracián
1601-1658

O protocolo da Iniciativa *Open Archives* tem se consolidado a cada dia como padrão de interoperabilidade entre bibliotecas e repositórios científicos/acadêmicos como se pode comprovar com o constante surgimento de novos projetos que o adotam como padrão e da criação de interfaces para importantes bases de dados e bibliotecas já existentes (por exemplo, *SciELO* e *arXiv*).

Os programas para publicação digital desenvolvidos com o protocolo OAI-PMH (*Eprints*, *DSpace*, *Kepler*) proporcionam uma oportunidade inédita às Instituições ao oferecerem mecanismos para a implantação de provedores de dados com baixo custo.

Como foi possível observar através da análise das ferramentas, atualmente existem soluções para várias necessidades:

- a da publicação pessoal, onde um pesquisador independente, que não esteja ligado a nenhuma Instituição pode publicar seus trabalhos usando o protocolo OAI-PMH sem qualquer necessidade de programação, em sua própria casa, independente de plataforma. Este é o segmento do *Kepler*.
- a da publicação institucional, porém, utilizando um programa que já traz implementado um ambiente *web* de submissão, gerenciamento, armazenamento e consulta a documentos, com o protocolo OAI-PMH implementado. Este "pacote" é o *Eprints*, onde, caso não haja necessidade de tradução/customização, após a instalação está pronto para o uso.

- e por último, a necessidade de armazenamento e gerenciamento de elementos multimídia, tais como textos, sons, programas, imagens, etc, porém, de acordo com a necessidade de cada Instituição/Departamento. Onde seja possível especificar diferentes níveis de segurança e hierarquia, de acordo com as características de cada coleção a ser gerenciada. Este será o público do DSpace, a mais recente das ferramentas.

O objetivo inicialmente proposto, de se analisar as soluções disponíveis baseadas na OAI e verificar a viabilidade da instalação de um repositório de dados/biblioteca digital de baixo custo foi atingido e comprovado através dos resultados da instalação do *Eprints* no projeto-piloto **UFPR Eprints**. Sua funcionalidade e baixo custo o qualificam como uma oportunidade para a criação de Provedores de Dados nas Universidades e Instituições de Pesquisa, principalmente naquelas onde os recursos para investimento em tecnologia são escassos.

Está é a chance destas instituições brasileiras e de países em desenvolvimento *mostrarem seus trabalhos* e de terem acesso à produção científica sem barreiras.

7. TRABALHOS FUTUROS

A educação é aquilo que permanece depois que tudo o que aprendemos foi esquecido.
Burrhus Frederic Skinner
1904-1990

A Iniciativa Open Archives, o protocolo OAI-PMH e os programas que implementam provedores de dados constituem um vasto campo a ser explorado, dado serem tecnologias recentes e em desenvolvimento. Com base no presente trabalho, poderão ser realizados os seguintes trabalhos futuros:

7.1. ESPECIFICAMENTE EM RELAÇÃO AO EPRINTS

Desenvolvimento de melhorias no cadastro dos usuários do Eprints, de maneira que os papéis de publicadores e assinantes sejam distinguidos tanto pela administração do sistema, quanto pelos próprios usuários. Implementação de gerenciamento de usuários.

Validação das estratégias para adesão da comunidade propostas na implantação do **UFPR Eprints**;

Desenvolvimento de uma extensão aos metadados do Eprints de maneira a atender as características das publicações brasileiras, bem como preservar informações no idioma inglês para exposição dos metadados.

7.2. DENTRE AS TRÊS FERRAMENTAS

Desenvolvimento de critérios de avaliação do desempenho dos programas *Eprints*, *DSpace* e *EKF* em produção, de maneira a identificar quais as melhores soluções a médio e longo prazo;

Desenvolvimento de metodologias de customização / tradução que facilitem a implantação do Eprints ou do DSpace em países onde o inglês não seja o idioma oficial.

Avaliação dos modelos de sincronismo entre provedores de dados e serviços e das extensões aos protocolo OAI-PMH, propostas pelo projeto EKF [13].

Introdução de controles automatizados de qualidade e padronização dos metadados.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] HARNAD, S. **Nature Debates: The Self-Archiving Initiative**. Disponível em <http://www.nature.com/nature/debates/e-access/Articles/harnad.html>. Acesso em 15 de abril de 2003.
- [2] SIMON, I.. **O Oráculo Bibliográfico: Sonhos de um Pesquisador**. Disponível em <http://www.ime.usp.br/~is/papir/oraculo> . Acesso em 25 de março de 2003.
- [3] DP9. <http://arc.cs.odu.edu:8080/dp9/index.jsp> . Acesso em 27 de abril de 2003.
- [4] SMITH, M. et al. **DSpace – An Open Source Dynamic Digital Repository**. D-Lib Magazine, Janeiro/2003. Disponível em <http://www.dlib.org/dlib/january03/smith/01smith.html> . Acesso em 04 de junho de 2003.
- [5] GUTTERIDGE, C.; HARNAD, S. **Application, Potential Problems and a Suggested Policy for Institutional E-Print Archives**. Disponível em <http://eprints.ecs.soton.ac.uk/archive/00006768/> . Acesso em 02 de março de 2003.
- [6] GUTTERIDGE, C. **GNU Eprints2 Overview**. Disponível em <http://eprints.ecs.soton.ac.uk/archive/00006840/> . Acesso em 07 de maio de 2003.
- [7] VILA, H. **Coruja: Um Provedor de Dados para a Iniciativa Open Archives**. Instituto de Matemática e Estatística da USP. Disponível em http://www.linux.ime.usp.br/~hvila/mac499/mono_final.htm. Acesso em 12 de maio de 2003
- [8] OAISter. Disponível em <http://oaister.umdl.umich.edu/o/oaister/>. Acesso em 08 de julho de 2003
- [9] PINFIELD, S.; GARDNER, M.; MACCOLL, J. **Setting Up an Institutional E-Print Archive**. Disponível em <http://www.ariadne.ac.uk/issue31/eprint-archives/> . Acesso em 15 de maio de 2003.
- [10] NSDL - The National Science Digital Library. Disponível em <http://www.nsdlib.org> . Acesso em 26 de junho de 2003.
- [11] arXiv.org. Disponível em: <http://www.arxiv.org> Acesso em 18 de junho de 2003.
- [12] LAGOZE, C. / VAN DE SOMPEL, H. **The Open Archives Initiative: Building a low-barrier interoperability framework**. Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, Roanoke VA, June 24-28, 2001, pp. 54-62. Disponível em <http://www.cs.cornell.edu/lagoze/papers/oai-icdl.pdf> . Acesso em 22 de fevereiro de 2003.
- [13] MALY, K.; ZUBAIR, M.; LIU, X. **Enhanced Kepler Framework for Self-Archiving**. ICPP-02, pp. 455-461, Vancouver, August 2002. Disponível em <http://kepler.cs.odu.edu/publications/kepler.pdf> Acesso em 27 de janeiro de 2003.

- [14] MALY, K.; ZUBAIR, M.; LIU, X. **Kepler Proposal and Design Document**. January 2003. Disponível em <http://kepler.cs.odu.edu/publications/finaldes.doc> Acesso em 27 de janeiro de 2003.
- [15] NCSTRL. Disponível em <http://www.ncstrl.org> Acesso em 27 de maio de 2003.
- [16] LAGOZE, C. et al. **The Open Archives Initiative Protocol for Metadata Harvesting**. <http://www.openarchives.org/OAI/openarchivesprotocol.html> . Acesso em 17 de fevereiro de 2003.
- [17] LAGOZE, C.; VAN DE SOMPEL, H. **Notes from the Interoperability Front: A Progress Report on the Open Archives Initiative**. 6ª Conferência de Pesquisa e Tecnologias Avançadas para Bibliotecas Digitais – Roma – Setembro/2002. Disponível em <http://www.openarchives.org/documents/ecdl-oai.pdf>. Acesso em 18 de fevereiro de 2003.
- [18] Self-Archiving FAQ for the Budapest Open Access Initiative (BOAI). Disponível em <http://www.eprints.org/self-faq/>. Acesso em 27 de fevereiro de 2003.
- [19] GUTTERIDGE, C.. **Eprints 2.2 Documentation**. Disponível em <http://software.eprints.org/docs/> . Acesso em 25 de fevereiro de 2003.
- [20] Open Archives Initiative. Disponível em <http://www.openarchives.org> . Acesso em 02 de fevereiro de 2003.
- [21] Eprints.org. Disponível em <http://www.eprints.org> . Acesso em 12 de fevereiro de 2003.
- [22] Budapest Open Access Initiative. Disponível em <http://www.soros.org/openaccess> . Acesso em 17 de março de 2003.
- [23] OAI-Tech. Disponível em <http://www.openarchives.org/organization/index.html> . Acesso em 09 de março de 2003.
- [24] Dublin Core Metadata Initiative. Disponível em <http://www.dublincore.org> . Acesso em 12 de abril de 2003.
- [25] Dublin Core Metadata Element Set, V.1.1:Reference Description. Disponível em <http://www.dublincore.org/documents/dces> . Acesso em 12 de abril de 2003.
- [26] Software GNU Eprints. Disponível em <http://software.eprints.org> . Acesso em 04 de junho de 2003.
- [27] DSpace.org. Disponível em <http://www.dspace.org> . Acesso em 26 de junho de 2003.
- [28] POWELL, A.; DAY, M.; CLIFF, P. **Using Simple Dublin Core to Describe Eprints**. UKOLN, University of Bath, 2003. Version 1.2. Disponível em www.rdn.ac.uk/projects/eprints-uk/docs/simpledc-guidelines . Acesso em 12 de abril de 2003.

- [29] W3C Encoding Rules for Date and Times. Disponível em <http://www.w3.org/TR/NOTE-datetime> . Acesso em 12 de abril de 2003.
- [30] Internet Assigned Numbers Authority, *IANA registered Media Types (MIME types)*. Disponível em <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types> . Acesso em 12 de abril de 2003.
- [31] *RFC 3066 Tags for the Identification of Languages*. Disponível em <http://www.ietf.org/rfc/rfc3066.txt> . Acesso em 12 de abril de 2003.
- [32] Getty Thesaurus of Geographic Names, version 3.0-Web. Disponível em <http://www.getty.edu/research/tools/vocabulary/tgn/> . Acesso em 12 de abril de 2003.
- [33] SPONSLER, E.; VAN DE VELDE, E. **Eprints.org Software: A Review**. Disponível em <http://arl.org/sparc/core/index.asp?page=g20#6> . Acesso em 06 de maio de 2003.
- [34] MALY, K.; ZUBAIR, M.; LIU, X. **KEPLER – An OAI Data/Service Provider for the Individual**. Disponível em <http://www.dlib.org/dlib/april01/maly/04maly.html> . Acesso em 04 de dezembro de 2002.
- [35] PASQUAL, J. **Uso de XML para Interoperabilidade entre Bases Heterogêneas**. Dissertação de Mestrado. Curitiba, 2002. Setor de Ciências Exatas. Universidade Federal do Paraná. Disponível em <http://eprints.c3sl.ufpr.br/archive/00000038/> . Acesso em 05 de julho de 2003.
- [36] PINTO, J.S.P. **Mapeamento de Atributos Complexos e Multivalorados na Extração de Esquemas utilizando XML**. Dissertação de Mestrado. Curitiba, 2001. Setor de Ciências Exatas. Universidade Federal do Paraná. Disponível em <http://eprints.c3sl.ufpr.br/archive/00000036/> . Acesso em 05 de julho de 2003.
- [37] IBICT. Disponível em <http://www.ibict.br> . Acesso em 15 de março de 2003.
- [38] ECS Research Self-Archiving Policy. Disponível em <http://www.ecs.soton.ac.uk/~lac/archpol.html> . Acesso em 12 de abril de 2003.
- [39] MARCONDES, C. H. / SAYÃO, L. F. **Integration and Interoperability Accessing Electronic Information Resources in Science and Technology: the proposal of Brazilian Digital Library**. Disponível em <http://arxiv.org/abs/cs.DL/0210029> . Acesso em 10 de maio de 2003.
- [40] MARCONDES, C. H. / SAYÃO, L. F. **Integração e Interoperabilidade no Acesso a Recursos Informacionais Eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira**. Revista CI Online setembro/dezembro-2001, Brasília: IBICT, 2001. Disponível em <http://www.ibict.br/cionline/300301/> . Acesso em 10 de maio de 2003.
- [41] MARCONDES, C. H. / SAYÃO, L. F. **The SciELO Brazilian Scientific Journal Gateway and Open Archives - A Report on the Development of SciELO - Open Archives Data Provider Server**. D-Lib Magazine, Março/2003. Disponível em <http://www.dlib.org/dlib/march03/marcondes/03marcondes.html> . Acesso em 10 de maio de 2003.

- [42] MARCONDES, C. H. / SAYÃO, L. F. **Documentos Digitais e Novas Formas de Cooperação entre Sistemas de Informação em C&T.** Reunión Internacional de Especialistas en Información Científica Digital (BIREME /OPS/OMS/UNESCO). São Paulo, 26-27/03/2002. Disponível em http://eventos.bvsalud.org/unesco/docs/es/doc/sayao_marcondes.doc . Acesso em 10 de maio de 2003.
- [43] SAYÃO, L. F. **Rascunho da proposta "Edição Eletrônica de Anais de Evento em C&T"**. Rio de Janeiro: Março, 2003. Coordenação Geral de Informática - Centro de Informações Nucleares (CIN). Documento fornecido pelo autor.
- [44] IBICT/Prossiga. **Biblioteca Digital Brasileira - Programa de Implantação Biênio 2001-2002.** Brasília: Fevereiro, 2001. Disponível em <http://www.ibict.br/bdb/bdb.pdf> Acesso em 10 de maio de 2003.
- [45] OAIcat. Disponível em <http://www.oclc.org/research/software/oai/cat.shtm> . Acesso em 28 de junho de 2003.
- [46] W3C (World Wide Web Consortium). Disponível em <http://www.w3.org> . Acesso em 30 de junho de 2003.
- [47] Lawrence, S. **Free Online Availability Substantially Increases a Paper's Impact.** Disponível em <http://www.nature.com/nature/debates/e-access/articles/lawrence.html>. Acesso em 25 de julho de 2003.
- [48] Suleman, H. **Open Digital Libraries.** Tese de Doutorado. Blacksburg, Novembro de 2002. Virginia Polytechnic Institute. Disponível em <http://www.husseinsspace.com/research.htm> Acesso em 25 de julho de 2003.
- [49] Dspace.org. **DSpace System Documentation: Functional Overview.** Disponível em: <http://dspace.org/technology/system-docs/functional.html> Acesso em 25 de julho de 2003.

ANEXO I - OAI DATA PROVIDERS

Listamos a seguir alguns dos repositórios registrados na Iniciativa *Open Archives*, em ordem alfabética, coletados em 09/06/2003. A lista completa e atualizada encontra-se em [20].

Nome/Descrição do Repositório	URL
Archives in London and the M25 Area (5607 records)	http://www.aim25.ac.uk:80/
Articles en Ligne, Jean Nicod Institut (234 records)	http://jeanNicod.ccsd.cnrs.fr/
arXiv.org Eprint Archive (226264 records)	http://arXiv.org/
Auburn University Digital Library (13 records)	http://diglib.auburn.edu/
BioMed Central (BMC) (1333 records)	http://www.biomedcentral.com/
Books by Caltech Authors (1 record)	http://library.caltech.edu/digital/
California Digital Library (CDL) (87414 records)	http://www.cdlib.org/
Caltech Computer Science Technical Reports (400 records)	http://library.caltech.edu/digital/
Caltech Control and Dynamical Systems Technical Reports (21 records)	http://library.caltech.edu/digital/
Caltech Earthquake Engineering Research Laboratory Technical Reports (292 records)	http://library.caltech.edu/digital/
Caltech Library System Papers and Publications (19 records)	http://library.caltech.edu/digital/
Caltech Parallel and Distributed Systems Group Technical Reports (49 records)	http://library.caltech.edu/digital/
Carnegie Mellon University Informedia Public Domain Video Archive (1075 records)	http://www.informedia.cs.cmu.edu/
CCSD Archive Ouverte en Sciences de l'Information et de la Communication (119 records)	http://archiveSIC.ccsd.cnrs.fr/
A Celebration of Women Writers (195 records)	http://digital.library.upenn.edu/women/
Central Connecticut State University (CCSU) Digital Archive (20 records)	http://fred.ccsu.edu:8000/
CERN Document Server (16907 records)	http://cdsweb.cern.ch/
Chemistry Preprint Server (262 records)	http://www.chemweb.com/CPS/
CiteBase (229103 records)	http://citebase.eprints.org/
CogPrints Archives (1593 records)	http://cogprints.ecs.soton.ac.uk/
Computer Science Preprint Server (410 records)	http://www.compsci-preprints.com/comp/Preprint/show/index.htm

Nome/Descrição do Repositório	URL
conoZe: Intelligere ut Credas, Credere ut Intelligas (1270 records)	http://www.conoze.com/
Digitale Hochschulschriften der Ludwig-Maximilians-Universität (LMU) München (505 records)	http://edoc.ub.uni-muenchen.de/
DSpace at CWRU (8 records)	http://www.cwru.edu/
DSpace at MIT (373 records)	http://www.dspace.org/
Duisburger Elektronische Texte (DUETT) (426 records)	http://www.ub.uni-duisburg.de/ETD-db/
Electronics and Computer Science (ECS) EPrints Database, University of Southampton (873 records)	http://eprints.ecs.soton.ac.uk/
Elektronischen Publikationen der Universität München (1 record)	http://epub.ub.uni-muenchen.de/
Elektronisches Dokumenten-, Archivierungs- und Retrievalsystem der Universität Dortmund (ELDORADO) (674 records)	http://eldorado.uni-dortmund.de/
E-LIS: E-Prints in Library and Information Science (72 records)	http://eprints.rclis.org/
ePub-WU OAI Archive, Vienna University of Economics and Business Administration (253 records)	http://epub.wu-wien.ac.at/
EXYSTENCE ePrints Archive (1 record)	http://www.complexityscience.net/
Institute of Mathematics and Statistics, University of São Paulo (4 records)	http://coruja.arca.ime.usp.br/
Langley Technical Reports Server (LTRS), NASA (1274 records)	http://techreports.larc.nasa.gov/ltrs/
Mathematics Preprint Server (826 records)	http://www.mathpreprints.com/
National Advisory Committee for Aeronautics Reports, NASA (7549 records)	http://naca.larc.nasa.gov/
National Digital Archive of Datasets (NDAD) (26 records)	http://ndad.ulcc.ac.uk/
The National Science Digital Library (22952 records)	http://www.nsdlib.org/
Open Language Archives Community (OLAC) (4644 records)	http://www.language-archives.org/
OpenVideo Project (1841 records)	http://www.open-video.org/
Perseus Digital Library (1433 records)	http://www.perseus.tufts.edu/
PhysiologieAnimale (8 records)	http://phy043.tours.inra.fr:8080/
Project Euclid, Cornell University (5203 records)	http://projecteuclid.org:80/Dienst/UI/1.0/Home
Research Papers in Economics (RePEc) (91426 records)	http://www.repec.org/
Schoenberg Center for Electronic Text and Image (54 records)	http://dewey.library.upenn.edu/sceti/
State and University Library, Bremen (SUUB) (490 records)	http://elib.suub.uni-bremen.de/
L'Université Lumière Lyon 2 Archive (8 records)	http://eprints.univ-lyon2.fr:8050/

Nome/Descrição do Repositório	URL
UFPR Eprints - Universidade Federal do Paraná, Brasil	http://eprints.c3sl.ufpr.br/
University of California eScholarship Repository (1433 records)	http://repositories.cdlib.org/
University of Melbourne ePrints Repository (155 records)	http://eprints.unimelb.edu.au/
The University of Michigan, University Library, Digital Library Production Service Collections (133854 records)	http://www.umdl.umich.edu/
University of Oslo, Digital Library (433 records)	http://www.digbib.uio.no/
Virginia Tech ImageBase (29595 records)	http://oai.dlib.vt.edu/OAI_VT_ImageBase/

Tabela 5 - Provedores de Dados OAI

ANEXO II - REPOSITÓRIOS EPRINTS

Lista dos repositórios (*archives*) que rodam o software *Eprints* mundo afora (em 02/06/2003). A lista atualizada por ser obtida em [16].

Repositórios EPrints 2

AKT Prints 94 records (c) http://eprints.aktors.org/	IUBio: An archive of biology data and software maintained at Indiana University Biology department 517 records (c) http://iubio.bio.indiana.edu:7780/
AMS Acta 6 records (c) http://amsacta.cib.unibo.it/	Indian Institute of Science, Bangalore, India 66 records (c) http://eprints.iisc.ernet.in/
AMS Campus 80 records (c) http://amscampus.cib.unibo.it/	Institute of Mathematics and Statistics of University of São Paulo http://eprints.ime.usp.br/
AMS Miscellanea 9 records (c) http://amsmisc.cib.unibo.it/	Instructional Technology ePrint Server 13 records (c) http://preprint.usu.edu/
ANU: Australian National University 898 records (c) http://eprints.anu.edu.au/	LADARK iPrints, a free self-archiving service for Latin Americanist scholars (The Johns Hopkins University) 5 records (c) http://eprints.ladark.dkc.jhu.edu/
Archive Lyon 2 8 records (c) http://eprints.univ-lyon2.fr:8050/	LU:research (Lund University) 2143 records (c) http://eprints.lub.lu.se/
Archive of European Integration 298 records (s) http://aei.pitt.edu/	Mathematical Sciences Eprint Archive (The Johns Hopkins University) 11 records (c) http://eprints.mts.jhu.edu/
CaltechBOOK: Books by Caltech Authors 1 records (c) http://caltechbook.library.caltech.edu/	NDLTD Papers http://ndltdpapers.dlib.vt.edu:9090/
CaltechCDSTR: Caltech Control and Dynamical Systems Technical Reports 21 records (c) http://caltechcdstr.library.caltech.edu/	NUI Maynooth Eprint Archive 61 records (c) http://eprints.may.ie/

<p>CaltechCSTR: Caltech Computer Science Technical Reports 410 records (c) http://caltechcstr.library.caltech.edu/</p>	<p>Nottingham eTheses http://etheses.nottingham.ac.uk/</p>
<p>CaltechEERL: Caltech Earthquake Engineering Research Laboratory Technical Reports 303 records (c) http://caltecheerl.library.caltech.edu/</p>	<p>Organic Eprints 356 records (c) http://www.orgprints.org/</p>
<p>CaltechGALCITFM: Aeronautical Laboratories (Fluid Mechanics) Technical Reports 6 records (c) http://caltechgalcitfm.library.caltech.edu/</p>	<p>Papyrus - Institutional Eprints Archive of Université de Montréal 22 records (c) http://papyrus.bib.umontreal.ca/</p>
<p>CaltechLIB: Caltech Library System Papers and Publications 21 records (c) http://caltechlib.library.caltech.edu/</p>	<p>PhilSci Archive 483 records (s) http://philsci-archive.pitt.edu/</p>
<p>CaltechOH: Caltech Oral Histories 37 records (c) http://oralhistories.library.caltech.edu/</p>	<p>PhysiologieAnimale 9 records (c) http://phy043.tours.inra.fr:8080/</p>
<p>CaltechPARADISE: Caltech Parallel and Distributed Systems Group 54 records (c) http://caltechparadise.library.caltech.edu/</p>	<p>Psychology 720 records (c) http://psycprints.ecs.soton.ac.uk/</p>
<p>CogPrints Cognitive Science Eprint Archive 1713 records (c) http://cogprints.ecs.soton.ac.uk/</p>	<p>Sammelpunkt. Elektronisch archivierte Theorie 413 records (c) http://sammelpunkt.philo.at:8080/</p>
<p>DList - Digital Library of Information Science and Technology 79 records (c) http://dlist.sir.arizona.edu/</p>	<p>StrathPrints: Academic writing produced by members of Strathclyde University 0 records (c) http://eprints.cdlr.strath.ac.uk/</p>
<p>DemoPrints 123 records (c) http://demoprints.eprints.org/</p>	<p>UFPR Eprints. Universidade Federal do Paraná, Brasil http://eprints.c3sl.ufpr.br/</p>
<p>Digitale Publikationen der Ludwig-Maximilians-Universität München 667 records (c) http://edoc.ub.uni-muenchen.de/</p>	<p>University of Bath (Pilot Service) 4 records (c) http://eprints.bath.ac.uk/</p>
<p>Drexel University Electronic Theses and Dissertations 78 records (c) http://thesis.library.drexel.edu/</p>	<p>University of Melbourne ePrints Repository 169 records (s) http://eprints.unimelb.edu.au/</p>

<p>E-LIS: E-Prints in Library and Informations Science 123 records (c) http://eprints.rclis.org/</p>	<p>University of Queensland ePrint Archive 393 records (c) http://eprint.uq.edu.au/</p>
<p>ENS-LSH (Human Sciences Archive) 56 records (c) http://eprints.ens-lsh.fr/</p>	<p>University of Southampton: Department of Electronics and Computer Science 7178 records (c) http://eprints.ecs.soton.ac.uk/</p>
<p>EPSILON EPrints2 Dissertation Test Archive 143 records (c) http://diss-epsilon.slu.se/</p>	<p>VT CS Technical Reports 586 records (s) http://eprints.cs.vt.edu:8000/</p>
<p>ETD Individuals 15 records (s) http://etdindividuals.dlib.vt.edu:9090/</p>	<p>Wisdom Archive: Faculty of Mathematics and Computer Science, Weizmann Institute. 234 records (s) http://wisdomarchive.wisdom.weizmann.ac.il:81/</p>
<p>EXYSTENC: Complex Systems Network of Excellence http://www.complexityscience.net/index.html</p>	<p>cav2001: Fourth International Symposium on Cavitation. Hosted by Caltech, June, 2001 145 records (c) http://cav2001.library.caltech.edu/</p>
<p>HyNIC_NJIT http://hynic.njit.edu/</p>	<p>ePrints.FRI (Faculty of Computer and Information Science (FRI), University of Ljubljana, Slovenia) [INITIAL DEPLOYMENT]. 58 records (c) http://eprints.fri.uni-lj.si/</p>

Tabela 6 - Repositórios que implementam Eprints 2.

Repositórios EPrints 1

@rchiveSIC : Science de l'information et de la Communication 135 records (s) http://archivesic.ccsd.cnrs.fr/	IJN : Institut Jean Nicod - Paris 237 records (s) http://jeannicod.ccsd.cnrs.fr/
Arquivos Abertos - IBICT 20 records (s) http://www.sbg.ibict.br/	Iowa State University Computer Science Technical Reports http://archives.cs.iastate.edu/
Behavioral and Brain Sciences 160 records (c) http://www.bbsonline.org/	Nottingham ePrints: University of Nottingham - 46 records (c) http://www-db.library.nottingham.ac.uk/ep1
CCSD: these-EN-ligne - 579 records (s) http://theses-EN-ligne.in2p3.fr/	PASTEL - 114 records (c) http://pastel.paristech.org/
CSIRO Exploration and Mining Archive http://www.ned.dem.csiro.au/eprints/	RIACS - Research Institute for Advanced Computer Science - Eprints Archive - 65 records (c) http://eprints.riacs.edu/
Digital Library of the Commons 769 records (c) http://dlc.dlib.indiana.edu/	Sissa Multimedia Database http://mmdb.sissa.it/
Formations Media Studies Archive 21 records (s) http://formations2.ulst.ac.uk/	Theoretical and Applied Linguistics (Edinburgh) - 90 records (s) http://archive.ling.ed.ac.uk/
History and Theory of Psychology 63 records (c) http://htpprints.yorku.ca/	Università degli studi di Firenze 107 records (c) http://biblio.unifi.it/
HofPrints 41 records (c) http://hofprints.hofstra.edu/	ePrints @ Glasgow 36 records (c) http://eprints.lib.gla.ac.uk/
Hofstra University EPrint Archive 41 records (c) http://hofprints.hofstra.edu/	

Tabela 7 - Repositórios que implementam Eprints 1.

ANEXO III - UFPR EPRINTS: TESTE DE CONFORMIDADE OAI

Open Archives Initiative

Repository Explorer

<http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai>

O *Repository Explorer* [20] é uma ferramenta desenvolvida por Hussein Suleman (*Virginia Tech*) para auxiliar nos testes de conformidade do protocolo da Iniciativa *Open Archives*, o OAI-PMH. Antes de se solicitar o registro à Iniciativa, o "candidato" a provedor de dados deve obter êxito na execução desta ferramenta, e, conseqüentemente, comprovar sua conformidade ao protocolo.

Data da realização do teste: 04/06/2003.

Endereço fornecido : <http://eprints.c3sl.ufpr.br/perl/oai2>

Resposta do *Repository Explorer*:

Open Archives Initiative :: Protocol for Metadata Harvesting v2.0
RE Protocol Tester 1.45a :: Virginia Tech DLRL :: March 2003

(1) Testing : Identify

URL : <http://eprints.c3sl.ufpr.br/perl/oai2?verb=Identify>

Test Result : OK

--- [Repository Name = UFPR EPrints]

--- [Protocol Version = 2.0]

--- [Base URL = <http://eprints.c3sl.ufpr.br/perl/oai2>]

--- [Admin Email = <mailto:admin@eprints.c3sl.ufpr.br>]

--- [Granularity = YYYY-MM-DD]

(2) Testing : Identify (illegal_parameter)

URL : <http://eprints.c3sl.ufpr.br/perl/oai2?verb=Identify&test=test>

Test Result : OK

(3) Testing : ListMetadataFormats

URL : <http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListMetadataFormats>

Test Result : OK

--- [Only oai_dc supported]

(4) Testing : ListSets

URL : <http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListSets>

Test Result : OK

--- [Sample Set Spec = 7374617475733D696E7072657373]

(5) Testing : ListIdentifiers (oai_dc)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc

Test Result : OK

--- [Sample Identifier = oai:GenericEPrints.OAI2:21]

(6) Testing : ListIdentifiers (oai_dc, from/until)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2000-01-01&until=2000-01-01

Test Result : OK

(7) Testing : ListIdentifiers (oai_dc, set, from/until)

URL :

http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&set=7374617475733D696E7072657373&from=2000-01-01&until=2000-01-01

Test Result : OK

(8) Testing : ListIdentifiers (oai_dc, illegal_set, illegal_from/until)

URL :

http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&set=really_wrong_set&from=some_random_date&until=some_random_date

Test Result : OK

(9) Testing : ListIdentifiers (oai_dc, from granularity != until granularity)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2001-01-01&until=2002-01-01T00:00:00Z

Test Result : OK

(10) Testing : ListIdentifiers (oai_dc, from > until)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2000-01-01&until=1999-01-01

Test Result : OK

(11) Testing : ListIdentifiers ()

URL : <http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers>

Test Result : OK

(12) Testing : ListIdentifiers (illegal_mdp)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=illegal_mdp

Test Result : OK

(13) Testing : ListIdentifiers (mdp, mdp)

URL :

http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&metadataPrefix=oai_dc

Test Result : OK

(14) Testing : ListIdentifiers (oai_dc, from YYYY-MM-DD)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2001-01-01

Test Result : OK

(15) Testing : ListIdentifiers (oai_dc, from YYYY-MM-DDThh:mm:ssZ)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2001-01-01T00:00:00Z

Test Result : OK

(16) Testing : ListIdentifiers (oai_dc, from YYYY)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2001

Test Result : OK

(17) Testing : ListMetadataFormats (identifier)

URL :

<http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListMetadataFormats&identifier=oai:GenericEPrints.OAI2:21>

Test Result : OK

--- [Only oai_dc supported]

(18) Testing : ListMetadataFormats (illegal_id)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListMetadataFormats&identifier=really_wrong_id

Test Result : OK

(19) Testing : GetRecord (identifier, oai_dc)

URL :

http://eprints.c3sl.ufpr.br/perl/oai2?verb=GetRecord&identifier=oai:GenericEPrints.OAI2:21&metadataPrefix=oai_dc

Test Result : OK

(20) Testing : GetRecord (identifier)

URL : <http://eprints.c3sl.ufpr.br/perl/oai2?verb=GetRecord&identifier=oai:GenericEPrints.OAI2:21>

Test Result : OK

(21) Testing : GetRecord (identifier, illegal_mdp)

URL :

http://eprints.c3sl.ufpr.br/perl/oai2?verb=GetRecord&identifier=oai:GenericEPrints.OAI2:21&metadataPrefix=really_wrong_mdp

Test Result : OK

(22) Testing : GetRecord (oai_dc)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=GetRecord&metadataPrefix=oai_dc

Test Result : OK

(23) Testing : GetRecord (illegal_id, oai_dc)

URL :

http://eprints.c3sl.ufpr.br/perl/oai2?verb=GetRecord&identifier=really_wrong_id&metadataPrefix=oai_dc

Test Result : OK

(24) Testing : ListRecords (oai_dc, from/until)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListRecords&metadataPrefix=oai_dc&from=2000-01-01&until=2000-01-01

Test Result : OK

(25) Testing : ListRecords (oai_dc, illegal_set, illegal_from/until)

URL :

http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListRecords&metadataPrefix=oai_dc&set=really_wrong_set&from=some_random_date&until=some_random_date

Test Result : OK

(26) Testing : ListRecords

URL : <http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListRecords>

Test Result : OK

(27) Testing : ListRecords (oai_dc, from granularity != until granularity)

URL : http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListRecords&metadataPrefix=oai_dc&from=2001-01-01&until=2002-01-01T00:00:00Z

Test Result : OK

(28) Testing : ListIdentifiers (oai_dc, set)

URL :
http://eprints.c3sl.ufpr.br/perl/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc&set=7374617475733D696E7072657373
Test Result : OK

(29) Testing : IllegalVerb
URL : <http://eprints.c3sl.ufpr.br/perl/oai2?verb=IllegalVerb>
Test Result : OK

--- Total Errors : 0

i.

Congratulations !

Your archive satisfied all the tests we performed. You may now add it to the list of archives on the front page if you wish by filling in the following additional details:

Base URL of Archive : <http://eprints.c3sl.ufpr.br/perl/oai2>

ANEXO IV - UFPR EPRINTS: REGISTRO OAI

O fato de constar na lista de repositórios *Eprints* (Anexo IV) não significa que o *Harvester* esteja coletando os metadados de um repositório. Para que isto aconteça, é preciso estar registrado na Iniciativa *Open Archives*.

Logo após o resultado satisfatório do **UFPR Eprints** no teste de conformidade com o protocolo, foi solicitado o seu registro no site da OAI [20] na madrugada de 05/06/2003. Após algumas horas, recebemos a mensagem a seguir, confirmando nossa compatibilidade com o protocolo e informando nossa aceitação na Iniciativa *Open Archives*:

De: "Nobody" nobody@nsdljib.nsdlib.cornell.edu
 Para: admin@fantomas.inf.ufpr.br
 Enviada em: quinta-feira, 5 de junho de 2003 12:35
 Assunto: OAI Validation Request

We received your request to register this site:

Base URL: <http://eprints.c3sl.ufpr.br/perl/oi2>

.....
 Using Xerces-J 2.2.1 to validate verb responses:
 Validating the Identify response...OK
 Validating the ListSets response...OK
 Validating the ListIdentifiers response...OK
 Validating the ListMetadataFormats response...OK
 Validating the GetRecord response...OK
 Validating the ListRecords response...OK

Number of bad verb responses: 0
 Number of verbs responses that validated: 6
 Total number of errors: 0

.....
 Your repository validated (at Protocol Version 2.0)
 according to our OAI protocol conformance tester.
 Your repository has been registered in the OAI database of
 conforming repositories.
 The OAI identifier GenericEPrints.OAI2 has been reserved for your use.

.....
 Total number of exceptions missed: 0
 Total error count: 0
 Your site is robust as well as compliant

Note: only spot checks have been made to check for UTF-8 correctness. To make sure your entire site is clean, or to fix it so that it works, you might use Simeon Warner's utf8conditioner, found on the OAI tools page or downloadable from <http://www.cs.cornell.edu/people/simeon/software/utf8conditioner/>.

.....

We will poll all registered sites from time to time to check for continuing conformance. In the meantime, you might want to add the OAI logo to your site: "<http://www.openarchives.org/images/OA100.gif>" Thank you for registering your site with the OAI

Desde esta data, o *Harvester* tem coletado os metadados de nosso repositório, o que pode ser facilmente comprovado através de uma pesquisa ao site OAIster [8].

ANEXO V - RECOMENDAÇÕES PARA USO DO DUBLIN CORE NO EPRINTS

Sendo o Conjunto de Elementos Dublin Core o padrão recomendado pelo Protocolo da Iniciativa Arquivos Abertos (*Open Archives Initiative Protocol for Metadata Harvesting - OAI-PMH*), a Universidade de Bath (Reino Unido) publicou um artigo recentemente (12/03/2003), com recomendações de uso do *Dublin Core* para descrever os documentos armazenados nos repositórios *Eprints* [28]. Trata-se de recomendações de preenchimento dos elementos de metadados adotados como padrão pela OAI e, conseqüentemente, pelo *Eprints*. Algumas destas recomendações estão direcionadas à realidade da origem do *Eprints* (Inglaterra) e, ao serem utilizadas como base para criação de normas e políticas no uso do *Eprints*, certamente devem ser adaptadas à nossa realidade.

Segundo o artigo, os repositórios *Eprints* devem implementar um conjunto mínimo de dez elementos *Dublin Core*, de acordo com o formato do registro 'oai-dc' da Iniciativa *Open Archives*.

Os elementos são:

1. dc:title
2. dc:creator
3. dc:subject
4. dc:description
5. dc:publisher
6. dc:date
7. dc:type
8. dc:identifier
9. dc:language
10. dc:relation

Recomendações de uso dos Elementos *Dublin Core* no *Eprints*:

1. **dc:title(*)**

(*) indica que o elemento é parte do conjunto mínimo de dez elementos.

Título do Eprint.

Preservar o título original, colocando substantivos próprios em letras maiúsculas. Os subtítulos devem ser separados do título por dois pontos (:).

Ex.: `<dc:title> Initial sequencing and analysis of the human genome</dc:title>`

`<dc:title> The new nationalism and the old history: perspectives the West German Historikerstreit</dc:title>`

Se necessário, repetir este elemento para múltiplos títulos.

2. **dc:creator(*)**

Autor do eprint.

Nomes próprios devem ser listados primeiramente pelo último sobrenome, seguidos pelo primeiro nome e demais sobrenomes, ou pela(s) inicial(ais) abreviadas com ponto. Separar o último sobrenome do primeiro nome e demais sobrenomes por uma vírgula. Os títulos (Prof., Dr., etc) devem preceder o primeiro nome, enquanto o sufixo Jr. deve seguir o último sobrenome. Quando em dúvida, colocar o nome como aparece, sem invertê-lo.

Exemplos:

Evans, R. J.

Walker Jr., Dr. John

No caso de organizações onde há claramente uma hierarquia presente, listar esta hierarquia da mais alta para a mais baixa, separadas por pontos. Se não há uma hierarquia ou não é possível identificar sua estrutura, relacionar os nomes como aparecem no *eprint*.

Exemplos:

International Human Genome Sequencing Consortium.

Loughborough University. Department of Computer Science.

Se necessário, repetir este elemento para múltiplos autores.

3. dc:subject(*)

O tópico do *eprint*.

Em geral, escolher as palavras mais significantes e únicas para palavras-chaves, evitando aquelas muito genéricas para a descrição do *eprint*. Múltiplos termos podem ser separados por ponto-e-vírgula (;). Subtópicos podem ser adicionados, separados por dois traços (—) e os espaços podem ser omitidos.

Exemplos:

Polar oceanography; boundary current; water masses; halocline

World War, 1939-1945—Germany

Germany—History—1933-1945

Hitler, Adolf, 1889-1945

4. dc:description(*)

Um resumo do conteúdo do *eprint*, tipicamente na forma de um “*abstract*”.

A descrição pode incluir mas não é limitada a: resumo (*abstract*), sumário, referência a uma representação gráfica do conteúdo ou um texto livro sobre o conteúdo.

5. dc:publisher(*)

A entidade publicadora do *eprint*, seja a instituição do autor ou outra de caráter comercial.

No caso de organizações onde há claramente uma hierarquia presente, listar esta hierarquia da mais alta para a mais baixa, separadas por pontos. Se não

há uma hierarquia ou não é possível identificar sua estrutura, relacionar os nomes como aparecem no eprint.

Exemplos:

Loughborough University. Department of Computer Science.

University of Reading. Rural History Centre.

Nomes próprios devem ser listados primeiramente pelo último sobrenome, seguidos pelo primeiro nome e demais sobrenomes, ou pela(s) inicial(ais) abreviada(s) com ponto. Separar o último sobrenome do primeiro nome e demais sobrenomes por uma vírgula. Os títulos (Prof., Dr., etc) devem preceder o primeiro nome, enquanto o sufixo Jr. deve seguir o último sobrenome. Quando em dúvida, colocar o nome como aparece, sem invertê-lo.

Exemplos:

Sulston, John E.

Evans, R. J.

Walker Jr., Dr. John

6. dc:contributor

Nome daqueles que contribuíram para o *eprint* (diferentes dos autores primários). Podem ser supervisores, editores, técnicos, etc.

Nomes próprios devem ser listados primeiramente pelo último sobrenome, seguidos pelo primeiro nome e demais sobrenomes, ou pela(s) inicial(ais) abreviadas com ponto. Separar o último sobrenome do primeiro nome e demais sobrenomes por uma vírgula. Os títulos (Prof., Dr., etc) devem preceder o primeiro nome, enquanto o sufixo Jr. deve seguir o último sobrenome. Quando em dúvida, colocar o nome como aparece, sem invertê-lo.

Exemplos:

Evans, R. J.

Walker Jr., Dr. John

No caso de organizações onde há claramente uma hierarquia presente, listar esta hierarquia da mais alta para a mais baixa, separadas por pontos. Se não há uma hierarquia ou não é possível identificar sua estrutura, relacionar os nomes como aparecem no eprint.

Exemplos:

International Human Genome Sequencing Consortium.

Loughborough University. Department of Computer Science.

Incluir organizações neste elemento somente para indicar a contribuição de uma organização, não para indicar a afiliação de uma pessoa.

7. dc:date(*)

Data da última modificação do *eprint* e/ou data de sua inclusão no repositório (*archive*).

O formato da data (YYYY-MM-DD) deve estar de acordo com 'W3C encoding rules for dates and times' [29]

Exemplos:

2003-03-19

2001

2003-01

8. dc:type(*)

O tipo do *eprint*. É recomendado que o valor deste elemento seja selecionado a partir da lista:

- *Book*
- *BookChapter*
- *ConferenceProceedings*
- *ConferencePaper*
- *ConferencePoster*

- *InCollection*
- *Technical Report*
- *OnlineJournalArticle*
- *JournalArticle*
- *NewsArticle*
- *Other*
- *Preprint*
- *Thesis*

Se necessário, repetir este elemento para múltiplos tipos.

Se necessário, repetir este elemento para indicar o status de arbitragem do *eprint*, usando um dos seguintes valores:

- *PeerReviewed*
- *NonPeerReviewed*

9. dc:format

O tipo de mídia do *eprint*.

É recomendado que seja selecionado um termo da lista disponível em *IANA Internet Assigned Numbers Authority registered media types (MIME)* [30].

10. dc:identifier(*)

A URL ou citação bibliográfica do *eprint*.

Exemplos

<http://eprints.bath.ac.uk/archive/00000003/>

HEERY, R. Information Gateways: collaboration on content. Online Information Review (2000), 24 (1), 40-45.

11. dc:source

A URL, título ou citação bibliográfica de um recurso de onde o *eprint* é derivado. Em geral, este elemento não deve ser usado.

12. dc:language(*)

A linguagem na qual o *eprint* está escrito.

Usar os códigos definidos em RFC 3066 [31]

Se necessário, repetir este elemento para indicar múltiplas linguagens.

13. dc:relation(*)

A URL de cada formato disponível do *eprint*. Se necessário, repetir este elemento para indicar múltiplos formatos.

<http://eprints.bath.ac.uk/archive/00000003/01/1097.pdf>

<http://eprints.bath.ac.uk/archive/00000003/01/1097.html>

14. dc:coverage

A localização geográfica ou período temporal sobre o qual versa o *eprint*. É recomendado selecionar o valor de um vocabulário controlado como o *Getty Thesaurus of Geographic Names* [32].

Se necessário, repetir este elemento para indicar múltiplos locais ou períodos.

15. dc:rights

Contém informações sobre direitos autorais do *eprint*.

Exemplos:

University of Bath.

Andrew Smith, 2003.

<http://creativecommons.org/licenses/by-nd-nc/1.0>

