

Biblioteca Digital: a experiência do Impa

Maria Cecilia Pragana Chataignier

Bacharel em Arquitetura – UFRJ. Especialização em Análise de Sistemas – UERJ. Trabalha no Impa como bolsista do CNPq, na informatização da biblioteca.
E-mail: cecilia@impa.br

Margareth Prevot da Silva

Bacharel em Informática pela UFRJ. Mestre em Informática pela PUC-Rio, na área de Banco de Dados. Trabalha no Impa como bolsista pelo CNPq, na implantação do servidor de *preprints*.
E-mail: margot@impa.br

Resumo

Relato da experiência do Impa na informatização de sua biblioteca, utilizando o software Horizon, e na construção de um servidor de preprints (dissertações de mestrado, teses de doutorado e artigos ainda não publicados) através da participação no projeto internacional Math-Net.

Palavras-chave

Biblioteca digital; Horizon; Impa; Automação; Preprint; Metadados; XML; Dublin core; Math-net; MPRESS; Harvest.

Digital Library: Impa's experience

Abstract

Report on Impa's experience in its Digital Library Project, using the software Horizon, and in the construction of a Preprint Server (Dissertations, Thesis and non-published articles) through a cooperation with the international Math-Net project.

Keywords

Digital library; Horizon; Impa; Automation; Preprint; Metadados; XML; Dublin core; Math-net; MPRESS; Harvest.

INTRODUÇÃO

As bibliotecas são, com freqüência, descritas como o coração das universidades e institutos de pesquisas. Um bom acervo bibliográfico, em geral, reflete o nível da pesquisa que o instituto ou universidade desenvolve. A eficiência com a qual esse acervo é acessado e utilizado pelos leitores é de grande importância para o crescimento científico e tecnológico desses institutos e universidades e conseqüentemente do nosso país.

A Biblioteca do Instituto de Matemática Pura e Aplicada (Impa) possui notável acervo de cerca de 30 mil volumes em livros e 32 mil volumes em periódicos, correspondentes a mais de 600 títulos em assinaturas correntes, em matemática e áreas mais afins, tendo sido escolhida como biblioteca-base na área de matemática pura e aplicada*.

Além desse acervo, a biblioteca possui um volume considerável de dissertações de mestrado e teses de doutorado (tanto publicações do próprio instituto, como de outras instituições), apostilas e atas de colóquios, bem como outras pré-publicações.

PROJETO DE INFORMATIZAÇÃO DA BIBLIOTECA

O projeto de Informatização da Biblioteca do Impa tem como principal objetivo disponibilizar o acervo da biblioteca via Internet, não só para a comunidade do Impa, como também para as comunidades acadêmicas nacional e internacional. Esse projeto teve início por volta de 1994, a partir de um trabalho em que foram analisados vários sistemas para bibliotecas disponíveis no mercado¹.

O Impa optou, como solução, pela compra do *software* Horizon, desenvolvido pela Ameritech Library Services, hoje Epixtech, Inc. (<http://www.epixtech.com>), amplamente utilizado como solução de automação de bibliotecas nos EUA e em outros países.

O Horizon é um sistema em ambiente gráfico com arquitetura cliente/servidor que utiliza um banco de dados relacional SQL. O formato de catalogação utilizado pelo Horizon é o padrão internacional MARC. A arquitetura

* Comut – Programa de Comutação Bibliográfica

desse sistema consiste em um servidor de banco de dados, destinado ao catálogo, em plataforma Windows NT; um servidor web em plataforma UNIX; e clientes que servem aos administradores do sistema e bibliotecários, também em Windows NT, como mostra a figura 1.

O sistema pode ser acessado pelos usuários, via Internet, através de um *browser* qualquer (Netscape, Internet Explorer etc.) que se comunica com o banco de dados por meio do protocolo Z39.50 da interface web.

O sistema foi implementado em módulos independentes, com funções específicas, mas que interagem. Os módulos mais importantes – como Administração, Catalogação, Consulta ao Catálogo, Circulação e Periódicos – são instalados nas máquinas dos administradores do sistema e dos bibliotecários, com permissões de uso específicas para cada função ou funcionário.

A atual interface web (webPAC) permite apenas a Consulta ao Catálogo. Em breve, uma nova interface (iPAC) permitirá, além da consulta, a reserva e/ou renovação de material bibliográfico, via Internet, de forma segura, substituindo as anteriores.

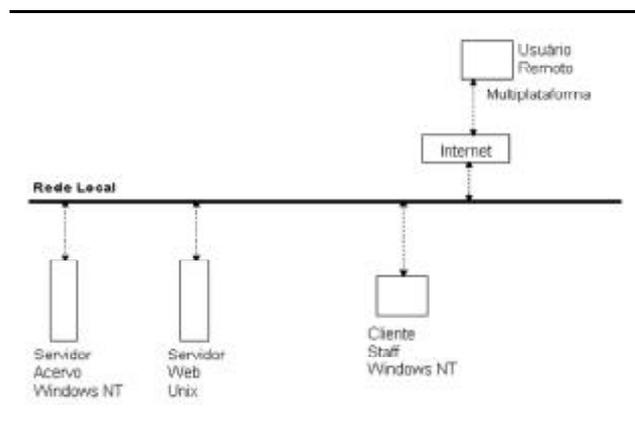
A informatização da biblioteca está sendo feita em etapas, sendo que grande parte delas – mas não todas – já foram concluídas. Após a aquisição do *software*, em 1996, foi montada uma equipe com três analistas de sistemas.

No início de 1997, a Ameritech realizou o primeiro curso, no Impa, para a equipe técnica e bibliotecários, sobre o formato MARC. Foram definidos os parâmetros iniciais básicos de instalação, e o sistema foi instalado, provisoriamente, com uma base de dados qualquer, para fins de treinamento.

A etapa seguinte compreendeu a conversão do catálogo (parte em dBASE) para o formato MARC, adotado internacionalmente como padrão pelos sistemas de automação de bibliotecas.

Considerando o grande volume e a especificidade do material bibliográfico a ser catalogado no padrão MARC e sendo relevante a alta qualidade, além da precisão dessa catalogação, ficou determinado que essa tarefa seria terceirizada. Assim sendo, foram identificadas as empresas que ofereciam esse serviço e a proposta escolhida foi a da OCLC Online Computer Library Center, Inc. (<http://www.oclc.org>), por se tratar de uma das mais respeitadas instituições atuando nesta área.

FIGURA 1
Diagrama da arquitetura do sistema Horizon



Lotes de informações bibliográficas, em formatos diversos (dBASE, Excel), foram gerados e enviados periodicamente à OCLC, via FTP. Os lotes foram processados e convertidos, através de vários processos, sendo que alguns ainda retornaram para confirmação e voltaram para reprocessamento. Todo esse processo durou cerca de um ano.

Finalizado o trabalho de conversão, os arquivos convertidos foram enviados à Ameritech para a geração da carga inicial do banco de dados.

Nessa época (julho/99), a Ameritech realizou o segundo treinamento no Impa, desta vez enfocando os módulos do sistema Horizon, propriamente dito. O sistema, com todos os seus módulos, foi instalado para os bibliotecários e equipe técnica, já com a base de dados do Impa. Foi também instalada, configurada e customizada, a interface web, para consulta ao catálogo, via Internet. Esse acesso ficou restrito ao domínio Impa, a princípio.

A etapa seguinte, completada em 2001, consistiu na etiquetagem dos livros, com códigos de barras, e geração de seus respectivos vínculos com os registros bibliográficos correspondentes. Aproveitou-se a oportunidade para fazer uma conferência e acrescentar informações locais. No final dessa etapa, o módulo de Consulta ao Catálogo foi aberto sem restrições, a toda a comunidade científica, disponibilizando a totalidade de seu acervo, via Internet.

Para a catalogação do material adquirido após a conversão, o Impa optou por se associar à OCLC e utilizar um de seus *softwares* de catalogação *on-line/off-line* – o CatME (OCLC Cataloging Micro Enhancer).

O módulo de Circulação foi implantado no início de 2000, com algumas limitações, mas já permitindo que o controle de empréstimos fosse automatizado. Em seguida, foram emitidos crachás com códigos de barras para os usuários e gerados os respectivos vínculos com seus cadastros.

O processo de etiquetagem com códigos de barras, para os periódicos encadernados, foi iniciado recentemente e será feito de forma semelhante ao procedimento com livros, com conferências e inclusão de informações locais.

Ainda em relação aos periódicos, a etapa seguinte será a de gerar as informações sobre a periodicidade de cada um, para que o sistema possa controlar, automaticamente, o recebimento dos fascículos e as falhas das coleções, emitindo avisos e relatórios.

A catalogação e etiquetagem de outros materiais, como CDs, *softwares*, vídeos, filmes etc., ficarão para o final, por se tratar de casos específicos e de menor volume.

OUTROS SERVIÇOS

Ainda como parte do processo de informatização da biblioteca, comprou-se o *software* Ariel e um *scanner* de mesa, que serão implantados em breve, para agilização dos serviços do Programa de Comutação Bibliográfica – COMUT (<http://www.ibict.br/comut>). A Biblioteca do Impa faz parte do programa, como biblioteca-base, na área de matemática pura e aplicada, enviando cópias de documentos a outras instituições, em atendimento a pedidos.

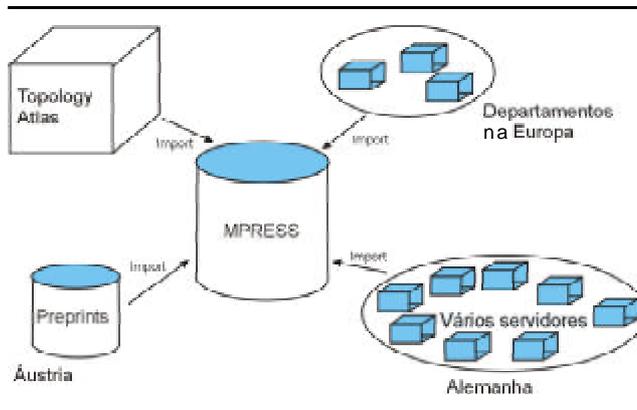
A biblioteca mantém, também, uma página na Internet (<http://www.impa.br/Biblioteca>), com informações de interesse para a comunidade de matemática, assim como acesso a portais, como o Periódicos – CAPES, bases de dados, como o MathSciNet, o MATH Database e o web of Science, além de periódicos assinados pela instituição (alguns desses acessos são restritos ao domínio Impa).

SERVIDOR DE PREPRINTS

Uma biblioteca de pesquisa, como a do Impa, é composta, basicamente, de livros, periódicos e *preprints* (dissertações de mestrado, teses de doutorado e artigos ainda não publicados). No caso de *preprints*, é importante que o conteúdo integral, e não apenas uma referência ao mesmo, esteja disponível à comunidade acadêmica.

A solução adotada para disponibilizar o seu acervo de *preprints* foi participar do projeto Math-Net (<http://>

FIGURA 2
Estrutura do MPRESS



www.math-net.org) – iniciativa internacional de criação de uma rede distribuída de repositórios de *preprints* de matemática na Internet, acessáveis por um serviço de busca e navegação.

PROJETO MATH-NET

O projeto Math-Net é uma iniciativa de instituições alemãs, com o apoio do Deutsches Forschungsnetz (DFN) e do Deutsche Telekom, de criar uma rede de informações sobre matemática na Internet. O objetivo do projeto é estabelecer uma infra-estrutura técnica e organizacional para que informações eletrônicas de matemática possam ser disponibilizadas a baixo custo, com eficiência e de fácil uso pelos usuários em geral. A ênfase são informações sobre publicações, coleções de *softwares* e dados, atividades de ensino e pesquisa, bem como informações organizacionais e bibliográficas².

Um dos serviços oferecidos pelo projeto Math-Net é disponibilizar na Internet uma coleção distribuída e indexada de repositórios de *preprints* de matemática, seguindo um padrão predeterminado. Este serviço é chamado de MPRESS – Mathematical PREprint Server System (<http://MathNet.preprints.org>)³. O MPRESS começou com o apoio do EMS (European Mathematical Society) e, recentemente, o CEIC (Committee on Electronic Information and Communication), estabelecido pelo IMU (International Mathematical Union), criou um subcomitê para aumentar o suporte ao projeto. Atualmente, várias instituições em diversos países (Brasil*, Alemanha, França, Itália, EUA, Canadá e outros) fazem parte do MPRESS. Uma relação dos integrantes pode ser obtida no endereço <http://MathNet.preprints.org/msc/Mstatistics.html>.

* Atualmente, apenas o Impa participa do Math-Net.

Os *preprints* são armazenados utilizando o conceito de metadados – dados descrevendo os dados⁴. O padrão utilizado pelo projeto Math-Net é o Dublin Core Metadata (<http://purl.oclc.org/dc>). Os metadados são armazenados no *header* do documento html utilizando tags META.

Estes metadados poderiam ser produzidos manualmente. Mas, para que se tenha uma garantia de que o padrão Dublin Core está sendo produzido corretamente e para que os autores não necessitem de conhecimento prévio deste padrão, o projeto oferece o editor MMM – Mathematics Metadata Markup, onde, através de um formulário eletrônico, o autor publica o documento html do *preprint* com os metadados gerados automaticamente.

Um repositório de dados necessita de alguma forma de indexação para facilitar a busca por informações específicas. No caso do MPRESS, os *preprints* são indexados por meio do *software* Harvest^{5, 6}, que é livre de encargos para instituições de ensino.

O *software* é formado por três componentes: *gatherer*, *broker* e *replicator*. O *gatherer* acessa URLs definidas em um arquivo de configuração e, recursivamente, coleta dados dos documentos existentes nestas URLs. A recursão é importante porque evita que o usuário precise incluir uma URL para cada *preprint*. A profundidade da recursão é configurável. O *gatherer* armazena as informações sobre o documento de forma resumida. A forma com que os dados são resumidos depende do conteúdo dos documentos do arquivo. Arquivos *postscript* são resumidos como um conjunto de palavras não estruturadas. Já os arquivos html com os metadados são resumidos com alta qualidade, já que eles contêm metadados. O *broker*, ou interface com o usuário, indexa os dados dos *gatherers* e de outros *brokers* remotos, armazenando-os em um base de dados, e oferece uma interface de busca ao usuário. E o *replicator* copia a base de dados do *broker* de forma customizada, possibilitando cópias atualizadas do banco de dados em diversos lugares.

MATH-NET NO IMPA

A partir de acordos iniciados em 1998 e troca de informações entre técnicos da Universidade de Osnabrück (<http://www.mathematik.uni-osnabrueck.de>) e do Impa, o projeto para instalação de um servidor de *preprints* no Impa foi iniciado. Em julho de 1999, a pesquisadora Judith Plümer fez uma visita ao Impa e apresentou o projeto Math-Net no 22º Colóquio Brasileiro de Matemática⁷ (<http://w3.impa.br/~coloquio/CBM22/Atividades/horario.html>). Houve grande interesse de outras

FIGURA 3

Header HTML com TAGS descrevendo os metadados

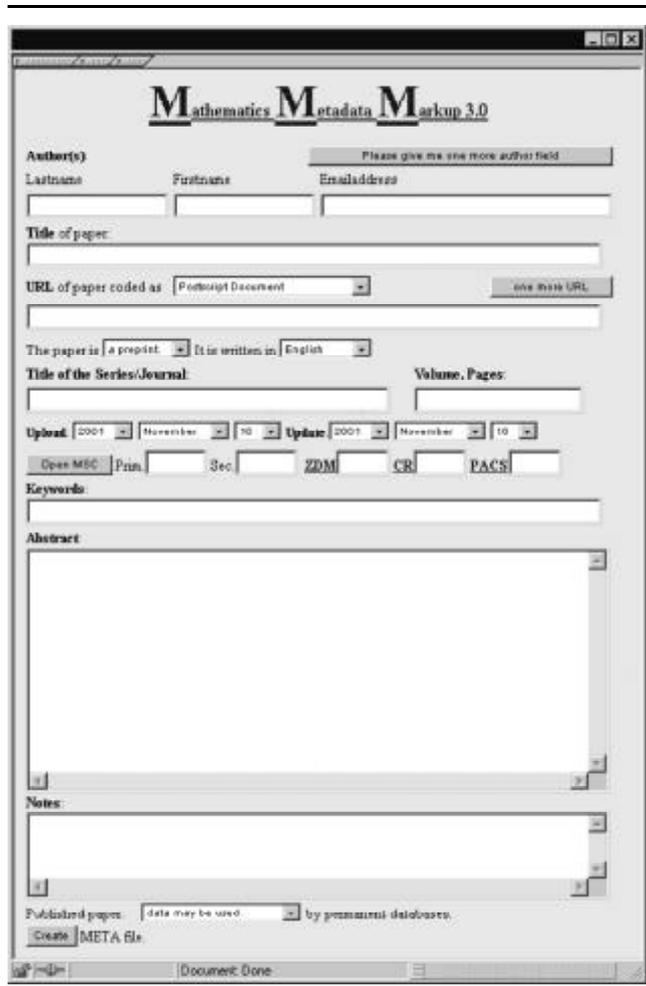
```
<META NAME = "DC.Vreator.Personalname"
      CONTENT = "Erwin Mustermann">

<META NAME = "DC.Title"
      CONTENT = "Theorems on Potatoes">

<META NAME = "DC.Identifier"
      CONTENT = "http://www.here/ToP/ ">
```

FIGURA 4

Editor MMM



instituições brasileiras pelo projeto. Em 2000, um servidor foi instalado no Impa para testes e adaptações dos *softwares*, sendo disponibilizado na Internet em 2001 (<http://www.preprint.impa.br>). Nesse período, foram observados e resolvidos pequenos problemas e propostas algumas novas implementações que irão contribuir para a melhor *performance* do sistema.

No projeto Math-Net, a versão integral dos *preprints* pode estar localizada em qualquer servidor na Internet, e o documento com os metadados contém a informação de sua localização. Uma das adaptações do projeto feita no Impa foi oferecer a opção de armazenar os arquivos com a versão integral em um servidor preestabelecido. Isto é importante, pois há casos em que autores não têm como e onde disponibilizar seus *preprints* em um servidor web. O arquivo é enviado junto com as informações no momento da inclusão via formulário eletrônico.

Para que a qualidade do conteúdo dos *preprints* seja garantida, outra adaptação importante foi estabelecer um processo de validação dos *preprints* a serem incluídos. A metodologia escolhida foi a validação, via correio eletrônico, de um pesquisador responsável pelo *preprint*, que pode ou não ser o próprio autor. Após a validação do responsável, o administrador do servidor de *preprints* utiliza um módulo de aprovação, desenvolvido no Impa, para efetivamente publicar o *preprint* na Internet.

Com relação ao problema de *copyright*, o próprio autor tem controle sobre o *preprint*, podendo a qualquer momento retirar a versão texto completo do *preprint*, conservando os seus metadados.

MATH-NET BRASIL

Dentro do contexto do projeto MathNet, é importante uma grande abrangência nacional. Para tal, é necessário criar estímulos para que outras instituições brasileiras participem do projeto, para que futuramente possamos ter uma rede nacional de servidores de *preprints* integrada ao Math-Net e com servidor e endereço próprios. Como forma de incentivo à participação da comunidade acadêmica brasileira de matemática no projeto Math-Net, o Impa irá disponibilizar um servidor único (<http://www.externalpreprints.impa.br>) específico para as instituições ou departamentos onde o custo de uma infraestrutura própria de máquina, *software* e suporte técnico não se aplique, como, por exemplo, institutos de pequeno porte. A quantidade de *preprints* por instituição será limitada, evitando que institutos com uma grande produção de *preprints* deixem de optar pela instalação de

FIGURA 5
Sistema Harvest

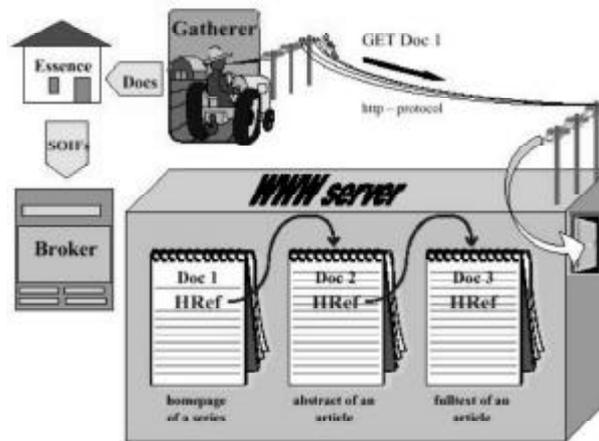


FIGURA 6
Campo no formulário MMM para o arquivo versão integral do preprint

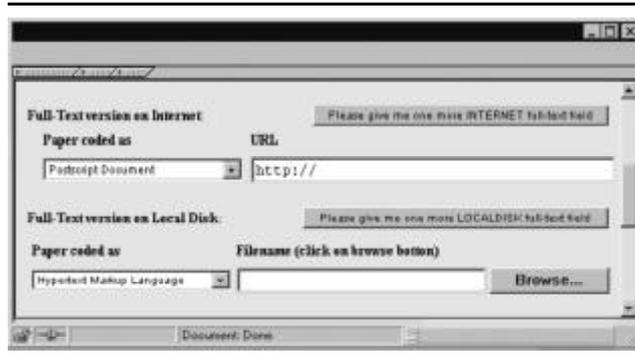
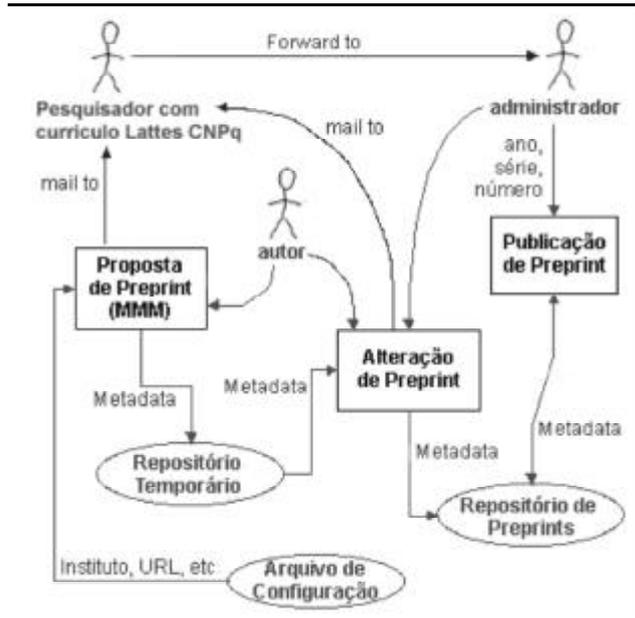


FIGURA 7
Processo de validação

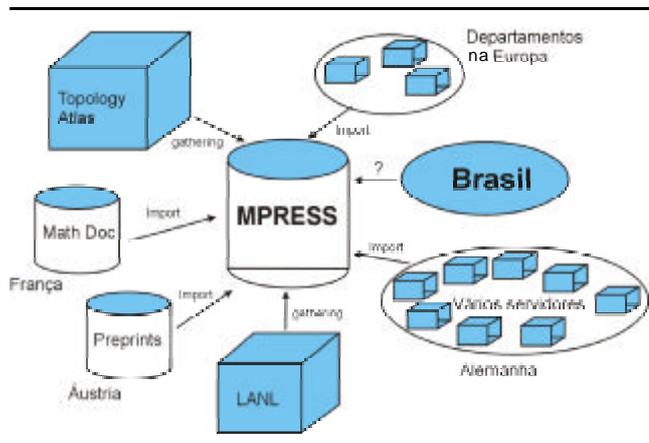


um servidor próprio. Os responsáveis que irão validar estes *preprints* deverão ser pesquisadores brasileiros, na área de matemática, cadastrados na base de dados da plataforma Lattes do CNPq.

Para os institutos que decidirem por um servidor de *preprints* em suas próprias instalações, o *software* será adaptado de forma que os dados específicos de uma determinada instituição estejam em arquivos de configuração, fáceis de serem alterados, não exigindo conhecimentos específicos de linguagem de programação.

Os que quiserem receber informações sobre o projeto e também sobre a produção de *preprints* do Impa devem visitar a página <http://www.preprint.impa.br> para inscrever-se na lista de divulgação do projeto: mathnet-l@impa.br.

FIGURA 8
Brasil no MPRESS



REFERÊNCIAS BIBLIOGRÁFICAS

1. PÉREZ-ALCÁZAR, J. J. *Projeto de automação da biblioteca do IMPA*. [S. l. : s. n.], 1994.
2. DALITZ, Wolfgang; GRÖTSCHEL, Martin; LÜGGER, Joachim. *Information Services for Mathematics in the Internet (Math-Net)*. In: IMACS WORLD CONGRESS ON SCIENTIFIC COMPUTATION: modelling and applied mathematics.15. *Proceedings*... [S. l. : s. n., 2001?].
3. SYDOW, A. (Ed.). *Artificial Intelligence and Computer Science*, v. 4, p. 773-778, 1997.
4. PLÜMER, Judith. *Metadata in mathematics*. In: ICM 98 BERLIN, 24., 1998. [S. l. : s. n.], 1998.
5. GRÖTSCHEL, Martin; LÜGGER, Joachim. *Scientific Information Systems and Metadata*. [S. l. : s. n.], 1998.
6. PLÜMER, Judith; SCHWÄNZL, Roland. *Harvesting mathematics*. *Euromath Bulletin*, v. 2, n. 1, 1996.
7. HARDY, Darren R.; SCHWARTZ, Michael F.; WESSELS, Duane. *Harvest user's manual*. Version 1.4 patchlevel 2. [S. l. : s. n.], 1996.
8. PLÜMER, Judith. *MathNet: a network for mathematicians*. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 22., 1999. [S. l. : s. n.], 1999.

Artigo recebido em 19/11/2001.