

# **Biblioteca Digital da SBC: Disponibilizando os Artigos do JEMS Através de um Provedor de Dados Compatível com OAI\***

**Diego Fraga Contessa, José Palazzo Moreira de Oliveira**

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)  
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brasil

{contessa,palazzo}@inf.ufrgs.br

***Abstract.** Digital libraries based on Open Archives Initiative and Dublin Core format came to ease the dissemination process for published works and research results. In the other hand, Brazilian Computer Society offers JEMS for the paper submission and evaluation management. This work presents the data provider developed in order to turn available metadata about published papers on JEMS events. This data provider offers a great quantity of metadata to a great quantity of potential “consumers” (the digital libraries). The data provider offers metadata in simple and qualified Dublin Core, and disseminate a significantly greater quantity of metadata.*

***Resumo.** As bibliotecas digitais baseadas na Iniciativa de Arquivos Abertos (OAI) e no formato Dublin Core surgiram para facilitar o processo de disseminação de trabalhos publicados e resultados de pesquisas. Por sua vez, a Sociedade Brasileira de Computação oferece o JEMS para o gerenciamento da submissão e avaliação de artigos de conferências e periódicos. O presente trabalho apresenta o provedor de dados que foi desenvolvido para disponibilizar os metadados dos artigos publicados nos eventos do JEMS, oferecendo uma grande quantidade de metadados a uma grande quantidade de potenciais “consumidores” (as bibliotecas digitais). O provedor de dados oferece metadados nos formatos Dublin Core simples e qualificado, e aumenta significativamente a quantidade dos metadados disponibilizados.*

## **1. Introdução**

Com o aumento do número de usuários e da capacidade da Internet, a mesma tornou-se peça chave na publicação e divulgação dos resultados de pesquisas científicas. Como solução para atender a esse tipo de demanda surgiram as bibliotecas digitais, que facilitam a publicação de trabalhos científicos e fornecem acesso simples às publicações. Porém, com as bibliotecas digitais surgiu a necessidade de garantir a interoperabilidade e a padronização dos metadados, de modo que as informações

---

\* Este trabalho é parcialmente financiado pelo projeto CTInfo, CNPq, DIGITEX, proc. 550845/2005-4 e pelo projeto PRONEX/FAPERGS, proc. 0408993.

armazenadas em um determinado repositório possam ser utilizadas por outras bibliotecas digitais. Estes são os objetivos da *Open Archives Initiative* - (Iniciativa de Arquivos Abertos) - OAI [Lagoze e Sompel 2004], e do *Dublin Core* [Hillmann 2005]. No Brasil, a Sociedade Brasileira de Computação – SBC - oferece aos seus eventos ou àqueles apoiados e atualmente a algumas outras publicações e conferências internacionais o JEMS (*Journal and Event Management System*). O objetivo deste sistema é oferecer o gerenciamento completo dos eventos, e para suportar este gerenciamento o JEMS contém uma base de dados que armazena uma grande quantidade de metadados sobre os autores, artigos e as conferências. Nota-se, portanto, que este sistema é uma fonte em potencial para um repositório compatível com a OAI.

O objetivo deste trabalho é apresentar o sistema desenvolvido para possibilitar a disponibilização dos metadados dos artigos do JEMS, através da implementação de um provedor de dados compatível com o padrão OAI. Além disso, são apresentadas algumas extensões necessárias ao JEMS para permitir o fornecimento de um conjunto mais completo de metadados sobre os artigos publicados.

## **2. *Open Archives Initiative* (OAI)**

O padrão OAI define um protocolo para a disseminação de metadados entre bibliotecas digitais, o OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) [OAI-PMH 2006]. Este protocolo é baseado em XML (o que garante a interoperabilidade necessária ao OAI), e define duas entidades: os provedores de dados e de serviços. Provedores de dados são responsáveis pela obtenção de metadados, armazenamento dos mesmos em um repositório e disponibilização a provedores de serviços, que têm oferecem serviços específicos a uma determinada aplicação. Para realizar a comunicação entre provedores de dados e de serviços, o OAI-PMH baseia-se num esquema de requisições (HTTP) e respostas (XML) para seis verbos (*Identify*, *ListMetadataFormats*, *ListIdentifiers*, *ListSets*, *GetRecord* e *ListRecords*). Para que um provedor de dados seja compatível com a OAI deve oferecer suporte às seis requisições.

Criado para garantir a padronização do formato dos metadados sobre um recurso (no caso, artigos publicados), o *Dublin Core* pode ser codificado tanto em HTML quanto em XML (o qual é usado neste trabalho). O formato define dois “níveis” de representação dos artigos: o simples (quinze elementos básicos) e o qualificado (três elementos adicionais e qualificadores para os elementos). O conjunto dos metadados do *Dublin Core* pode ser encontrado em [Hillmann 2005]. Para que um provedor de dados possa ser considerado compatível com o padrão OAI ele deve ao menos oferecer metadados no formato *Dublin Core* simples. Porém, o padrão permite (e até recomenda) que se ofereça suporte a mais de um formato, podendo ser o *Dublin Core* qualificado ou algum formato “proprietário”, específico para uma determinada aplicação. Tendo em vista o objetivo de disponibilizar metadados sobre os artigos de eventos promovidos pela SBC, pode-se destacar a BDBComp (Biblioteca Digital Brasileira de Computação) [Laender, Gonçalves e Roberto 2005], que foi criada para disponibilizar metadados sobre artigos publicados pela comunidade brasileira de Ciência da Computação. Segundo [Silva 2004], a BDBComp pode receber metadados pela colheita de metadados em páginas de eventos, por um serviço de auto-arquivamento de artigos ou pela colheita de metadados em provedores de dados compatíveis com o OAI.

### 3. Journal and Event Management System (JEMS)

O JEMS é o sistema utilizado para o gerenciamento completo de conferências e periódicos. O sistema é baseado no EDAS [Schulzrinne 2006], e recebeu uma grande quantidade de modificações e melhorias que o tornaram bastante diferente do sistema original. Pelo fato de gerenciar todas as etapas de submissão e avaliação das publicações, o JEMS possui uma completa base de dados com informações sobre os autores e os artigos (título, abstract, formato, datas associadas ao artigo, e palavras-chave, por exemplo), além de uma boa quantidade de dados sobre as conferências (coordenadores do comitê de programa, data de realização e tópicos de interesse, por exemplo). Contudo, o JEMS não possui todos os metadados *Dublin Core*. Assim, se por um lado o JEMS é uma fonte em potencial de metadados sobre artigos, por outro há a necessidade de serem feitas extensões para que ele tenha a possibilidade de fornecer a grande quantidade de metadados que podem ser disponibilizados no *Dublin Core*.

### 4. Disponibilização dos Metadados dos Artigos do JEMS

As extensões propostas ao JEMS alteram significativamente o processo de submissão de artigos. A Figura 1 apresenta o processo antes das extensões.

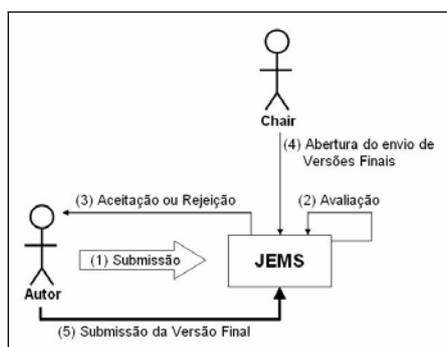
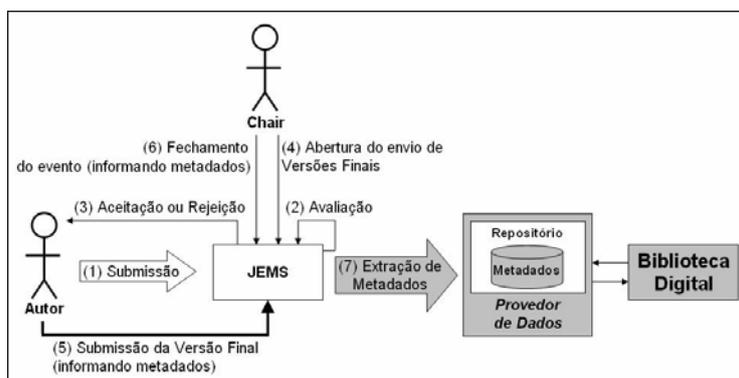


Figura 1. Processo de submissão de artigos no JEMS

A primeira fase é a submissão do artigo, onde o autor se cadastra no sistema e em seguida registra seu artigo. Neste ponto são identificados co-autores e as palavras-chave do artigo são selecionadas dentre as opções anteriormente configuradas pelo *chair*. Em seguida há a avaliação do artigo pelo comitê de programa, e então os autores são notificados a respeito da aceitação ou rejeição de seu artigo. O passo seguinte é a abertura do período de submissões de versões finais, o que é feito pelo *chair* da conferência. Por fim, os autores de artigos aprovados submetem a versão final ao JEMS.

A Figura 2 apresenta o processo de submissão de artigos no JEMS depois das extensões deste sistema de disponibilização de metadados. Até a terceira fase (notificação aos autores) o processo segue o mesmo, sendo modificado a partir da quarta etapa. No momento em que o *chair* abre o período de submissão das versões finais, ele deve definir este novo período em uma interface específica para este fim. Na etapa de submissão da versão final de um artigo o sistema deve solicitar metadados adicionais ao autor, para que ele possa enviar seu arquivo. Neste momento são solicitadas informações sobre contribuições ao trabalho e relacionamentos com outros trabalhos, por exemplo. Também neste ponto o autor deve informar metadados fundamentais como palavras-chave, caso ainda não o tenha feito. Em seguida há o

fechamento da conferência, em que o *chair* deve informar metadados como meio de publicação, local de realização da conferência e informações sobre direitos autorais, por exemplo. Depois disto um módulo de extração de metadados gera um “pacote” formado pelos artigos publicados na conferência e por um arquivo XML que contém dados sobre o evento e os artigos publicados. Este “pacote” será enviado ao servidor do provedor de dados, para que os metadados sejam armazenados no repositório.

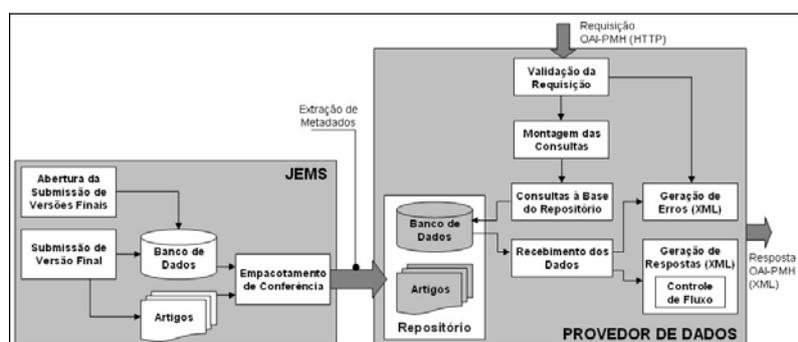


**Figura 2. Processo de submissão de artigos no JEMS com as extensões**

O servidor em que o provedor de dados é executado é o mesmo da Biblioteca Digital da SBC, a qual é um portal que disponibiliza os artigos publicados em conferências promovidas pela SBC. Para armazenar os dados necessários a este portal, foi desenvolvida uma base de dados para as informações dos artigos publicados, com uma reduzida quantidade de dados sobre os artigos. Como há uma correlação entre os dados já armazenados no banco de dados do portal da Biblioteca Digital e os dados que são enviados pelo JEMS ao provedor de dados, foi definida uma extensão do modelo do banco de dados da Biblioteca Digital para que ele pudesse receber todos os metadados que o JEMS passa a disponibilizar. Deste modo, pode-se executar o provedor de dados no servidor da biblioteca digital da SBC, de modo a torná-la compatível com a OAI.

## 5. O Provedor de Dados do JEMS

A Figura 3 apresenta a arquitetura do provedor de dados desenvolvido. A operação do sistema inicia pelo recebimento de uma requisição HTTP (o verbo do OAI-PMH e seus parâmetros). Em seguida, a requisição é validada para garantir a inexistência de erros.



**Figura 3. Arquitetura do provedor de dados implementado**

Se a requisição é válida, o provedor de dados monta a consulta SQL adequada para a obtenção dos metadados do repositório. Em seguida, esta consulta é enviada ao repositório, e os resultados da consulta são retornados ao provedor de dados, que verifica se foram retornados metadados. Caso haja metadados que correspondam à requisição, a resposta XML é gerada (no formato Dublin Core simples ou qualificado, de acordo com a requisição), e o resultado é retornado. Caso haja algum erro nos parâmetros ou a requisição seja inválida, ou não haja metadados que atendam à consulta SQL montada, é gerado um erro (XML) definido pelo OAI-PMH.

O provedor de dados oferece controle de fluxo por *resumption tokens*, atendendo à recomendação do OAI-PMH para listas de tamanho grande. Com isso, uma transferência de metadados pode ser feita por partes, favorecendo casos onde haja falha na transmissão de alguma parte de uma lista de metadados, já que a mesma pode ser continuada do ponto onde parou, facilitando a recuperação. O formato *Dublin Core* simples é disponibilizado para todos os registros do provedor de dados. Além dele, decidiu-se disponibilizar todos os registros também no formato *Dublin Core* qualificado. O provedor de dados implementado não trabalha com o conceito de registros deletados, já que uma vez que um artigo é publicado seus metadados são armazenados no repositório e ele não é mais removido, porque não deixa de ser publicado. Outro ponto importante é a organização hierárquica que se obteve pelo uso do conceito de conjuntos. A cada evento ou periódico corresponde um conjunto, e cada artigo pertence somente a um conjunto. Foram usados vocabulários restritos e específicos conforme recomendado em [Hillmann 2005]. Isto garante que o conteúdo dos metadados tenha também uma padronização. Foram usados os *MIME types* definidos pelo IANA para o metadado *format*, e para o idioma do artigo (metadado *language*), foi usada a RFC 3066, que trata das abreviações dos nomes de idiomas.

## **6. Avaliação Experimental da Disponibilização de Metadados**

Para verificar a compatibilidade do provedor de dados desenvolvido com a OAI, o mesmo foi submetido aos testes do *OAI Repository Explorer* [Suleman 2001]. A verificação do *Repository Explorer* consiste de uma bateria de testes automáticos que inclui testes de parâmetros (corretos e incorretos), verificação de respostas de erro, verificação de respostas XML com metadados, validação correta de verbos e testes do mecanismo de controle de fluxo. O provedor de dados respondeu corretamente a todos os testes feitos pela ferramenta. Logo, pode-se dizer que o provedor de dados é completamente compatível com o padrão OAI. A segunda avaliação foi com relação à quantidade e qualidade dos metadados disponibilizados. Para isso, a base de dados do repositório foi povoada com dados sobre dois eventos que foram retirados do JEMS (a ERBD 2005 e a ERBD 2006), através de uma simulação do processo de submissão de artigos ao JEMS após as extensões. Além destes, a base já continha dados sobre os eventos que já eram disponibilizados pelo portal da Biblioteca Digital da SBC, e que haviam sido inseridos na base manualmente. Para as conferências cujos artigos publicados já estavam na base de dados do repositório a quantidade de metadados disponibilizados pelo provedor de dados é menor, sendo que apenas alguns elementos possuem conteúdo. Para os eventos que foram inseridos no repositório pelo processo de obtenção dos metadados do JEMS após as extensões o conjunto de metadados disponibilizados é significativamente maior. Metadados como datas relacionadas aos

artigos (submissão, aceitação, publicação, entre outras) são agora facilmente obtidos. Outros podem ser disponibilizados a partir da informação por parte dos autores.

## 7. Conclusões

Com o processo de obtenção de metadados através das extensões do JEMS, o sistema de submissões da SBC torna-se mais completo. Assim, ele pode disponibilizar uma maior quantidade de metadados, melhorando a quantidade e qualidade dos metadados disponibilizados. Além disso, a verificação feita usando o *OAI Repository Explorer* demonstrou que o provedor de dados é compatível com a OAI, de modo a aumentar bastante a disseminação dos artigos publicados nos eventos ou periódicos gerenciados pelo JEMS. Qualquer biblioteca digital que tenha um provedor de serviços pode realizar a colheita de metadados no provedor de dados desenvolvido. Uma biblioteca digital que obterá grande benefício pelo uso do provedor de dados do JEMS é a BDBComp. Como trabalho futuro pode-se citar a criação de outros formatos para os metadados disponibilizados, usando trabalhos relacionados como [Warpechowski, Souto e Oliveira 2006] para a disponibilização de metadados de objetos de aprendizagem (LOM). A criação de formatos proprietários para os metadados é outra possibilidade, além do desenvolvimento de provedores de serviços que usem o provedor de dados deste trabalho para agregar funcionalidades a aplicações existentes.

## Referências

- Hillmann, D. (2005) "Using Dublin Core". Disponível em: <<http://dublincore.org/documents/usageguide>>. Acesso em Julho, 2006.
- Laender, A. H. F., Gonçalves, M. A. e Roberto, P. A. (2004) "BDBComp: Building a Digital Library for the Brazilian Computer Science Community". In Proceedings of the Joint Conference on Digital Libraries (JCDL'04) (Tucson, Arizona, USA, June 7-11, 2004). 2004.
- Lagoze, C. e Sompel, H. V. de. (2004) "The Open Archives Initiative: Building a Low-Barrier Interoperability Framework". In Proceedings of the Joint Conference on Digital Libraries (JCDL'04) (Tucson, Arizona, USA, June 7-11, 2004). 2004.
- Schulzrinne, H. (2006) Editor's Assistant v3. Disponível em: <<http://edas.info>>. Acesso em Julho, 2006.
- Silva, L. V. e. (2004) "Um Serviço de Auto-arquivamento de Publicações Científicas Compatível com o Padrão OAI". Dissertação de Mestrado, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil, 2004.
- Suleman, H. (2001) "Enforcing Interoperability with the Open Archives Initiative Repository Explorer". In Proceedings of the Joint Conference on Digital Libraries (JCDL'01) (Roanoke, VA, USA, June 24-28, 2001). 2001.
- The Open Archives Initiative Protocol for Metadata Harvesting (2006). Disponível em: <<http://openarchives.org/openarchivesprotocol.html>>. Acesso em Julho, 2006.
- Warpechowski, M., Souto, M.A.M. e Oliveira, J. Palazzo M. (2006) "Techniques for Metadata Retrieval of Learning Objects", SW-EL International Workshop on Applications of Semantic Web Technologies for E-Learning, with AH '06, Dublin.