

PRESERVAÇÃO DIGITAL DA MEMÓRIA TÉCNICO-CIENTÍFICA DO INPE

G. J. F. BANON, M. L. RIBEIRO e L. C. BANON

Instituto Nacional de Pesquisas Espaciais

CP 515

12 245-970, São José dos Campos, SP, Brasil

banon@dpi.inpe.br, marciana@sid.inpe.br, lise@iconet.com.br

Resumo. Nesta publicação, fazemos um rápido histórico da Memória Técnico-Científica do Instituto Nacional de Pesquisas Espaciais (INPE). Devido a sua importância na área digital, recordamos o problema da identificação do original de uma obra. Mostramos uma solução e mencionamos sua implementação no software *URLibService*. Comentamos também sobre a importância de termos um software que garanta a persistência dos vínculos bibliográficos, ou seja, a robustez frente a possível re-estruturação do acervo ou ainda a transferência de direitos patrimoniais entre detentores de obras intelectuais. Finalmente, fazemos uma breve descrição do estado atual do desenvolvimento do acervo digital desta Memória.

Palavras-chave: biblioteca digital, identificação do original de uma obra, direitos autorais, direito patrimonial, direito moral do autor, preservação digital, vínculo persistente, Memória Técnico-Científica, acervo digital, acervo distribuído, OAI, ePrint.

1. Introdução

A grande inovação da área digital foi a possibilidade de criar com grande facilidade e rapidez, cópias idênticas a um original. Por outro lado, temos também uma grande facilidade em editar e alterar um conteúdo digital. Com o advento da Internet, surgiu uma grande necessidade de saber a proveniência de uma obra digital. Esta necessidade está também ligada a questão dos direitos autorais.

Um primeiro desafio é então, reconhecer dentro de inúmeras cópias (alteradas ou não) qual é o original de uma obra. Resolvendo este problema, será solucionada a questão dos direitos patrimoniais, pois o detentor da obra original é também detentor destes direitos.

Um segundo desafio é a questão da anterioridade. Como qualquer um pode copiar um conteúdo digital, é de grande importância ter um mecanismo seguro de datação das obras digitais. Resolver este problema, soluciona a questão dos direitos morais dos autores.

Encontrar soluções para o primeiro desafio é importante porque contribui na preservação das obras digitais. De fato, além do problema técnico, o problema da preservação digital é também econômico. Ao identificar a obra original, esta passa a ser valorizada, o que ajuda a garantir sua permanência.

A Memória Técnico-Científica do INPE tem muito a ganhar com a tecnologia digital, tanto no plano da preservação como na sua divulgação. No entanto, pelos motivos apresentados acima, é importante fazer a opção certa no momento de recorrer a esta nova tecnologia.

Nesta publicação, faremos um rápido histórico da Memória Técnico-Científica do INPE. Em seguida, apresentaremos as principais características da Biblioteca Digital escolhida para hospedar esta Memória. Finalmente faremos uma descrição sucinta do estado atual do desenvolvimento do acervo digital.

2. Memória Técnico-Científica

A Memória Técnico-Científica do Instituto nasceu em 1963 com as primeiras publicações da então Comissão Nacional de Atividades Espaciais (CNAE).

Foi em 1995 que nasceu o acervo digital da Memória Técnico-Científica com o primeiro armazenamento no formato digital de um trabalho sobre Morfologia Matemática. Desde o início adotou-se o conceito de Repositório

Uniforme para garantir a criação de um espaço digital próprio para cada obra depositado no acervo (ver próxima seção).

Em 1998, com o auxílio da FAPESP, foi adquirida uma servidora SUN Enterprise 250 que desde então é usada como repositório principal para as obras intelectuais produzidas pelo Instituto.

O relatório de projeto submetido no final de 2003 ao CNPq, intitulado “Preservação, enriquecimento e disseminação das Memórias Técnico-Científica e Documental do INPE” [5], contém maiores informações sobre a Memória Técnico-Científica do INPE.

3. Biblioteca Digital URLib

Como comentado na introdução, a preservação digital passa pela identificação do original de uma obra. Para permitir esta identificação é preciso de dois ingredientes: primeiro, que a obra seja depositada num espaço digital dentro de um acervo possuindo um identificador universal e, segundo, que este identificador conste (explicitamente ou não) em algum lugar neste espaço.

Por exemplo, um artigo científico no site de uma revista on-line é o original pois na sua URL (“Uniform Resource Locator”) consta a URL do site e ao abrir o artigo temos no seu cabeçalho, o nome da revista que podemos, em seguida, relacionar com a URL do site.

Neste exemplo, no entanto, existe um grave problema: o identificador do acervo é escolhido como sendo a sua localização na Web. Ou seja, a obra não poderia mudar de site sem correr o risco de perdermos a capacidade de relacionar ainda o nome da revista com a URL do novo site.

O que precisamos então, é de um mecanismo para gravar o identificador universal do acervo junto com a obra (ou como indicado acima, dentro do espaço digital reservado à obra).

O software chamado URLibService implementa precisamente este mecanismo. Este software foi criado para servir na montagem e manutenção de uma biblioteca digital com acervos distribuídos. Nesta biblioteca digital que chamamos de URLib, do Inglês “Uniform Repositories for a Library” [1], os espaços criados para depositar as obras digitais nos acervos locais são chamados de repositórios uniformes ou simplesmente de *repositórios*.

Os repositórios criados num determinado acervo local podem ser copiados sem atropelo para um outro acervo, e cada repositório contém o nome do acervo local que hospeda a obra original. Assim, como indicado acima, mesmo fazendo cópias de um repositório para outros acervos locais, ficamos sabendo qual é o acervo local que possui o original.

Isto é ilustrado na Figura 1 onde temos dois acervos locais: Acervo1 e Acervo2, com dois repositórios em cada um: Rep 1 e Rep 2 no primeiro e Rep 2 e Rep 3 no segundo. Observamos que os dois acervos têm o mesmo repositório: Rep 2. Com a informação que o acervo hospedeiro de Rep 2 é Acervo 1, sabemos que o original pertence ao Acervo 1 e que o Acervo 2 contém apenas uma cópia. Nesta figura, as obras originais são pintadas de cinza.

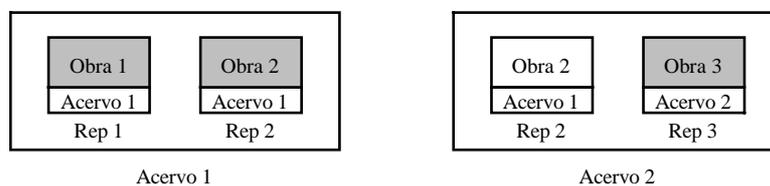


Fig. 1 - Exemplo mostrando a identificação das obras originais dentro de dois acervos: o original da Obra 2 está no Acervo 1.

de cinza.

Com este cuidado, o detentor de um acervo local passa a ser, sem ambiguidade, o detentor dos direitos patrimoniais de todas as obras originais depositadas no mesmo.

Ao acessar um repositório na *URLib*, o *URLibService* exibe uma capa com um cabeçalho indicando se a versão acessada é a oficial, isto é, se o documento aberto é idêntico ao original (Figura 2).

Com o *URLibService* é possível também controlar o grau de reprodutibilidade de um repositório. Está é a

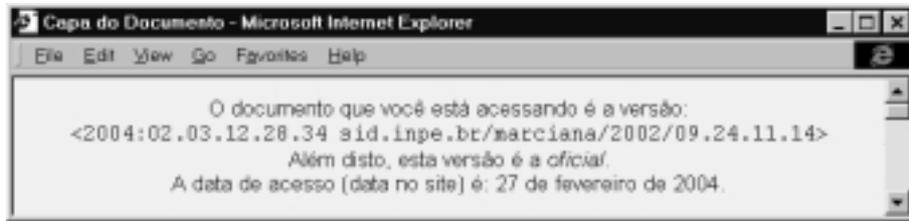


Fig. 2 - Cabeçalho da Capa indicando que a versão acessada é a oficial, isto é, o documento aberto é idêntico ao original.

primeira forma de valorizar os direitos patrimoniais.

A segunda forma é a possibilidade de transferir os direitos patrimoniais de uma obra, entre detentores de dois acervos distintos. Isto é possível através da simples troca de nome do acervo hospedeiro contido no repositório da obra.

Para garantir a integridade do nome do acervo local contido num repositório, o *URLibService* recorre ao uso de senhas, uma senha é criada aleatoriamente para cada repositório, e a de um serviço independente de registro da tripla (nome do repositório, identificador do acervo local, senha).

Finalmente, afim de evitar que a mesma obra intelectual seja armazenada em dois repositórios diferentes, o *URLibService* oferece também um serviço independente de registro da obra junto com os nomes dos seus autores. Através da data do registro, este mecanismo garante a anterioridade e assim os direitos morais dos autores.

Como comentado acima, a valorização e conseqüentemente a preservação de uma obra depende da capacidade que temos de poder trocar as obras de acervos, ou seja de mudar o detentor dos direitos patrimoniais. Isto cria um novo desafio tecnológico, pois uma obra intelectual muitas vezes se insere no contexto do conhecimento universal. Neste caso, ela pode citar outras obras ou ainda ser citada por outras. Como estas citações bibliográficas podem se tornar elementos de navegação entre obras e levando-se em conta que as obras não tem um acervo predefinido, há necessidade de se ter um mecanismo automático de busca das obras citadas. O *URLibService* possui um tal mecanismo, garantindo assim a persistência, além das obras, dos vínculos ("links") entre as mesmas. Detalhes sobre como criar vínculos persistentes estão apresentados em [3] e em [4] encontram-se alguns exemplos destes vínculos.

As características do *URLibService* apresentadas acima e outras listadas em [2] fizeram da *URLib* uma plataforma conveniente para a receber a Memória Técnico-Científica do Instituto.

4. Acervo digital da Memória Técnico-Científica

Como mencionado acima, a montagem do acervo digital da Memória Técnico-Científica começou em 1995 e desde então está se desenvolvendo a partir da plataforma *URLib*.

Atualmente o acervo está distribuído em nove acervos locais hospedados em dois computadores. Por conveniência os acervos locais foram criados para abrigar a Memória por áreas do conhecimento, por exemplo: Sensoriamento Remoto, Meteorologia e Combustão, ou ainda por eventos científicos, por exemplo: Congresso Brasileiro de Sensoriamento Remoto, WORCAPs do curso de Pós-Graduação em Computação Aplicada.

Esta divisão atende a necessidade de repartir melhor as responsabilidades de entrada dos dados e permite também resolver limitações de armazenamento num determinado disco rígido ou partição. Por outro lado, para o usuário final que consulta a Memória (endereço atual: <<http://iris.sid.inpe.br:1905>>), esta divisão em acervo locais é irrelevante, já que através do *URLibService*, todos estes acervos passam a ser integrados.

Figura 3 mostra, de forma simplificada, os componentes deste acervo. Cada componente é interligado aos outros via Internet.

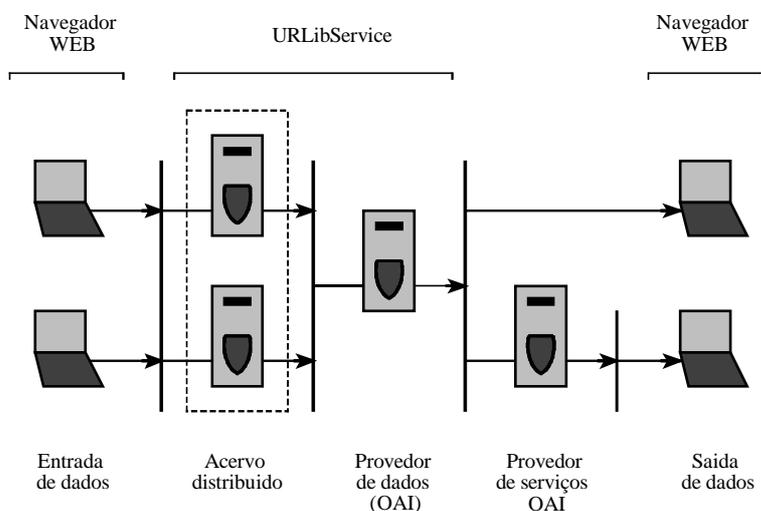


Fig. 3 - Componentes do acervo digital da Memória Técnico-Científica.

Além da consulta aos servidores da *URLib*, é possível consultar parte do acervo, especificamente as teses e dissertações, através do “OAI service providers” já que o *URLibService* disponibiliza também os metadados por meio do protocolo de comunicação da OAI (Open Archives Initiative) [7].

A entrada de dados é feita on-line, através de formulários de submissão, por pessoas especializadas ou pelos próprios autores (no caso de artigos em eventos organizados pelo INPE). Atualmente, o acervo contém mais de 9.000 referências, contemplando toda produção Técnico-Científica do INPE desde 1963 (7.400 foram

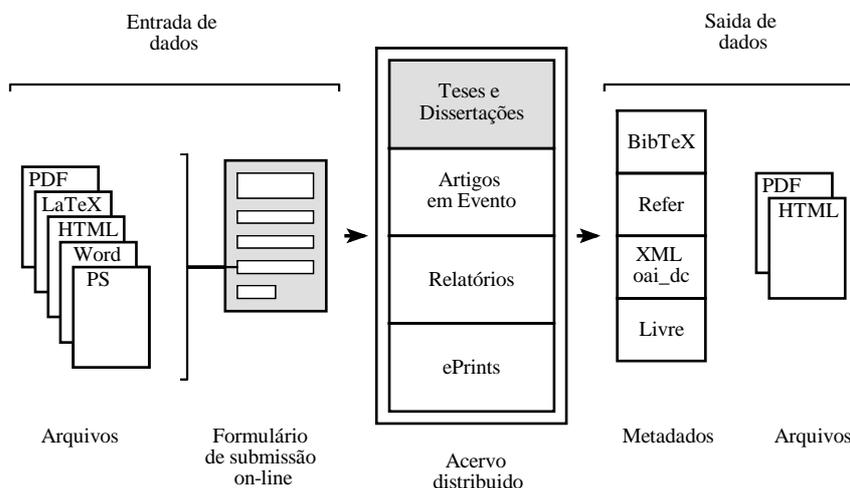


Fig. 4 - Formatos de entrada e saída.

importadas do banco ISIS), sendo mais de 1.200 com o texto completo (280 sendo de teses e dissertações). A Figura 4 mostra os formatos usados na entrada e saída para os arquivos e os metadados.

Visões parciais do acervo podem ser pré-programadas e exibidas no formato HTML através de páginas dinâmicas. Por exemplo, o acompanhamento da produção Técnico-Científica do Instituto para o ano de 2003 pode ser feito através de uma tabela mostrando os indicadores desta produção por divisões [6].

5. Conclusão

Nesta publicação, fizemos um rápido histórico da Memória Técnico-Científica do Instituto.

Recordamos o problema da identificação do original de uma obra digital. Mostramos uma solução e mencionamos sua implementação no software *URLibService*. Comentamos sobre a importância de termos um software que garanta a persistência dos vínculos bibliográficos, ou seja, a robustez frente a possível reestruturação do acervo.

Finalmente, fizemos uma breve descrição do estado atual do desenvolvimento do acervo digital desta Memória. Para o ano de 2004, a biblioteca digital está oferecendo um novo serviço de submissão on-line, armazenamento e disponibilização de ePrints para os pesquisadores do Instituto. Esta iniciativa deve resgatar a memória de muitas obras produzidas pela Instituição, especialmente as obras publicadas como artigo em eventos que são geralmente difíceis de serem reunidas na Memória da Instituição. Esta publicação é inclusive o primeiro ePrint desta Memória.

6. References

As referências mencionando a *URLib*, podem ser encontradas a partir de qualquer site rodando o *URLibService*, por exemplo: <<http://iris.sid.inpe.br:1905>>. Basta introduzir o identificador (nome do repositório) no campo de busca.

- [1] Banon, G. J. F. *Uniform repositories for a digital library - URLib*. 1998. Disponível na biblioteca digital *URLib*: <dpi.inpe.br/banon/1997/08.14.12.32>. Acesso em: 25 fev. 2004.
- [2] Banon, G. J. F.; Banon, L. C. *O que é a URLib?* Disponível na biblioteca digital *URLib*: <iconet.com.br/banon/2001/05.25.16.44>. Acesso em: 25 fev. 2004.
- [3] Banon, G. J. F.; Banon, L. C. *Creating Persistent Citation Links* Disponível na biblioteca digital *URLib*: <iconet.com.br/banon/2001/07.27.00.56>. Acesso em: 25 fev. 2004.
- [4] Banon, G. J. F. *Hiperdocumentos versus URLib*. Disponível na biblioteca digital *URLib*: <dpi.inpe.br/banon/2002/10.10.08.39>. Acesso em: 25 fev. 2004.
- [5] Banon, G. J. F.; Ribeiro, M. L.; Silva, M. D. *Preservação, enriquecimento e disseminação das Memórias Técnico-Científica e Documental do INPE*. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2003. Disponível na biblioteca digital *URLib*: <sid.inpe.br/marciana/2003/10.21.16.25>. Acesso em: 25 fev. 2004.
- [6] Instituto Nacional de Pesquisas Espaciais *Tabela com os indicadores Físicos e Operacionais IPUB e IGPUB: ano de 2003*. Disponível na biblioteca digital *URLib*: <sid.inpe.br/marciana/2003/08.21.16.18>. Acesso em: 28 fev. 2004.
- [7] Open Archives Initiative. *Open Archives Initiative*. Disponível em: <<http://www.openarchives.org>>. Acesso em: 25 fev. 2004.