

## **Biblioteca Digital da Memória Técnico-Científica do INPE**

Este documento foi elaborado por Gerald Jean Francis Banon em dezembro de 2005 para ser incluído no Anexo A intitulado: Exemplos de arquivos existentes, do Projeto 08:001.06-059 da ABNT intitulado: Sistemas espaciais de dados e informações – Modelo de referência para sistema aberto de arquivamento de informação (SAAI). A segunda versão inclui as melhorias de português sugeridas por Katia P Thomaz.

### **DOMÍNIO**

#### *Domínio e Consumidores*

A Memória Técnico-Científica do Instituto Nacional de Pesquisas Espaciais (INPE) é o acervo institucional que contém toda produção técnico-científica do INPE desde da criação da então Comissão Nacional de Atividades Espaciais (CNAE) em 1963.

Este acervo está sob a responsabilidade do Serviço de Documentação e Informação (SID) e do Conselho de Editoração. A partir de 1995 este acervo entrou na fase digital com o início de uso de uma biblioteca digital. Esta biblioteca oferece serviços para submissão, armazenamento, divulgação, acesso e preservação de documentos. Seus usuários finais são pesquisadores e alunos de Pós-Graduação do Instituto e da comunidade científica em geral, envolvidos em Ciência Espacial e da Atmosfera, Mecânica Espacial e Controle, Meteorologia e Sensoriamento Remoto e áreas correlatas.

Atualmente o acervo da Memória Técnico-Científica comporta 16316 registros, sendo que 4750 deles contêm o texto completo de forma digital. A partir de 2001 as submissões tornaram-se remotas facilitando as admissões de novas contribuições na forma digital. No mesmo período, iniciou-se o trabalho de digitalização do acervo que ainda se encontrava exclusivamente na forma impressa.

#### *Produtores de dados*

Os dados originam-se da produção intelectual dos pesquisadores e alunos de Pós-Graduação do Instituto e, eventualmente, de pesquisadores externos à instituição, envolvidos nas áreas do conhecimento mencionadas acima.

### **ADMISSÃO**

#### *Acordo de submissão*

Pelas normas internas do Instituto, todo trabalho intelectual produzido dentro da instituição deve ser registrado, armazenado, publicado e disseminado na Biblioteca Digital da Memória Técnico-Científica. Existem três acordos de submissão, dependendo

do estágio e local de publicação do documento. Para os trabalhos produzidos pelo INPE e publicados fora do INPE, o acordo de submissão consiste apenas da definição do nível de disseminação do trabalho, qual seja: sem restrição, com restrição ao INPE ou sem permissão de acesso. Para os trabalhos que tenham sido ou serão submetidos fora do INPE, o acordo de submissão estipula que, ao submeter qualquer versão do seu trabalho, os autores não estão cedendo seus direitos patrimoniais ao INPE. Esses direitos poderão, no futuro, ser livremente cedidos a qualquer editora. Os trabalhos classificados nesta categoria são chamados de ePrint. Para os trabalhos a serem publicados pelo INPE, o acordo de submissão estipula que o produtor está ciente que ao fazer a submissão, está transferindo o documento original para o administrador da Biblioteca Digital e que o mesmo não será submetido a outro acervo.

#### *Sessão típica de entrega dos dados*

Dependendo do acordo de submissão, a entrega dos dados pode tomar formas diferentes. A maioria das entregas são feitas através de formulários *on-line* personalizados em função do tipo do trabalho a ser submetido, qual seja: artigo em revista, artigo em conferência, livro, livro editado, capítulo de livro, relatório, tese e dissertação, programa de computador, ePrint, material audiovisual para apoio a apresentação em eventos. Em todos os formulários devem ser preenchidos pelo menos o nome dos autores e o título do trabalho. Caso o trabalho seja considerado uma publicação do INPE, deverá ser entregue dentro das normas de editoração definidas pela Instituição. O trabalho submetido será revisado e, havendo necessidade de melhoria na sua forma, deverá ser submetido novamente. Caso o trabalho contemple mais de um arquivo, o conjunto de arquivos deve ser entregue encapsulado em um arquivo no formato *.zip*. Esse arquivo é automaticamente aberto no ato do recebimento do trabalho, recriando-se o conjunto original de arquivos. Caso o trabalho seja um ePrint poderá também haver atualizações na forma de submissões de novas versões do mesmo trabalho.

Todos os processos de entrega mencionados acima são seguidos da criação de um pacote de arquivamento de informação (PAI). No entanto, há possibilidade, apesar de pouco usada, de entrega de PAI já formatado. Nesse caso, distingue-se entre os PAIs que contêm o trabalho original daqueles que contêm apenas uma cópia do mesmo.

#### *Objetos de informação*

Todos os objetos de informação entregues pelos produtores são compostos exclusivamente de objetos digitais nos formatos: *.pdf*, *.djvu*, *.doc*, *.tex*, *.djvu*, *.html*, *.txt*, *.jpg*, *.ppt*, *.prn*, *.ps*, *.zip*. No ato da entrega, a informação de representação desses objetos consiste em apenas na presença do tipo de formato no nome dos arquivos submetidos. Todos os objetos de informação entregues contemplam a informação de conteúdo, informação descritiva e eventualmente informação de descrição de preservação (IDP). O pacote de submissão de informação (PSI) compõe-se de duas partes. A primeira parte contém informação descritiva (geralmente informação de natureza bibliográfica, acrescida do resumo e das palavras chaves) e IDP (referência aos programas aplicativos geradores da informação de conteúdo). A segunda parte contém a própria informação de conteúdo. No ato de recebimento do PSI, a IDP é automaticamente ampliada com outros

tipos de informação de preservação, quais sejam: data, local e responsável da submissão, identificador.

#### *Series*

Nos casos de submissão de artigos em eventos organizados pelo Instituto, os mesmos podem ser reunidos em uma serie de arquivamento de informação (SAI) correspondente ao conceito de Anais. Isto acontece no momento da criação do sumário e do índice por autor.

#### *Transformação*

Na maioria das submissões não há conversão automática de formato da informação de conteúdo, exceto no caso de ePrint.. Nesse caso, todas as submissões devem ser feitas no formato *postscript* (.prn) e no ato do recebimento, esse formato é automaticamente convertido para o formato *portable document format* (.pdf) com a inserção de um identificador no cabeçalho das páginas do documento. Para os trabalhos publicados pela Instituição, como teses e dissertações, o documento entregue é guardado no PAI juntamente com suas conversões para os formatos .pdf, .djvu e .html que são executadas voluntariamente.

A informação descritiva geralmente é adquirida via formulário. Até 2003, essa informação também era alternativamente adquirida a partir de arquivos textos no formato REVISA do ISIS e transformado no formato Refer.

#### *Validação*

Os trabalhos publicados pelo Instituto, submetidos à Memória Técnico-Científica na forma digital, passam por uma revisão de conteúdo e de forma. Por exemplo, no caso de teses e dissertações, o trabalho já foi objeto de uma avaliação por parte da banca de doutores reunidos na ocasião da sua defesa. Após sua submissão à biblioteca digital, tanto a informação de conteúdo como a informação descritiva é ainda revisada pelo SID para verificação da sua conformidade com as normas de edição definidas pelo Instituto.

## **FORMATOS INTERNOS**

#### *Armazenamento*

O armazenamento permanente na biblioteca digital é feito exclusivamente na forma de arquivos dedados no sistema de arquivo disponível no disco rígido dos computadores que hospedem o acervo da biblioteca. O sistema de arquivo é organizado de forma que cada PAI ocupa dois ou mais diretórios. Um diretório é dedicado à informação de conteúdo e os demais à informação descritiva. Essa organização segue o padrão de Repositórios Uniformes para uma biblioteca digital (*URLib*). Os nomes de diretórios, criados para hospedar os PAIs, obedecem a uma regra de geração que garante sua universalidade dentro da biblioteca digital, mesmo que gerados em acervos distintos. Essa forma de organizar do sistema de arquivo permite em particular a transferência dos PAIs de um acervo para outro sem necessidade de revisão da informação de empacotamento.

O armazenamento da informação descritiva é feito no formato Refer Extendido em um arquivo tipo texto. O armazenamento da IDP é feito em um conjunto de arquivos tipo texto ou binário caso a informação tenha caráter sigiloso. No momento da criação dos PAIs, essas informações (metadados) são capturadas e mantidas num banco de dados que pode ser acessado por meio de dois tipos de servidor. O primeiro servidor usa uma comunicação via *socket* e outro servidor use uma comunicação via protocolo http. Os metadados podem ser recuperados em diversos formatos (ver Acesso).

### *Migração*

Até o momento, foram somente realizadas migrações de plataforma, por exemplo, de SunOS para Linux ou Windows, de forma a melhorar a razão desempenho/custo. Nessas migrações não ocorreram alterações na informação de empacotamento, na informação de conteúdo e descritiva, e na IDP. Foram reproduções realizadas através do uso de ferramentas que duplicam as estruturas de diretórios de um sistema de arquivo para outro.

### *Segurança*

Somente parte do PAI é disseminado, geralmente a parte em formato .pdf, .djvu ou .html. Essa disseminação pode ser livre ou com restrições de acesso e de busca.

Todos os PAIs são periodicamente (três vezes por semana) copiados em fitas localizadas em um prédio distinto do prédio hospedando o acervo digital *on-line*.

## **ACESSO**

### *Sistema de busca*

O sistema de busca da biblioteca digital da Memória Técnico-Científica do INPE encontra-se disponível na Internet no endereço: <http://iris.sid.inpe.br:1905>. A recuperação dos pacotes de disseminação de informação (PDI) é feita a partir de expressão de busca composta por palavras-chave cuja ocorrência é procurada dentro da informação descritiva e da IDP. A expressão de busca pode, alternativamente, consistir em expressão booleana e/ou ainda conter o nome dos campos onde as ocorrências de palavras devem ser procuradas.

### *Segurança*

Existem vários níveis de segurança no acesso aos PDIs. Em primeiro lugar, os PDIs podem ou não estar visíveis para o sistema de busca, com exceção administrador da biblioteca. Assim, grande parte dos PDIs relativos ao *software* de gerenciamento e manutenção da biblioteca digital (URLibService) estão visíveis apenas para o administrador. Caso o PDI seja classificado na categoria de divulgação reservada, ele também não estará visível para o sistema de busca. Em segundo lugar, o acesso a informação de conteúdo pode ou não ser restrito a determinados IPs ou a determinadas pessoas devidamente registradas, por meio de login/senha. Novamente, os PDIs com divulgação reservada usam o recurso de acesso restrito por login/senha. Os PDIs com divulgação interna ao Instituto possuem restrições para impressão (conforme opções

definidas no arquivo PDF). Finalmente, os PDIs que contêm cópia de publicações feitas fora do INPE podem ter seu acesso restrito à instituição por meio do controle de IPs.

#### *Apoio ao consumidor*

Além do sistema de busca *on-line*, o usuário final pode consultar na página da Memória Técnico-Científica a lista das últimas aquisições automaticamente atualizada todos os dias no primeiro clique do dia. No caso de artigos em eventos organizados pelo INPE (eg., SBSR, WORCAP), o usuário final pode também navegar dentro dos anais *on-line* por meio de sumários e índices por autor. Algumas áreas do conhecimento, como o sensoriamento remoto, têm acesso personalizado à biblioteca digital para garantir que o resultado da busca fique limitado àquelas áreas de conhecimento. O usuário final ainda encontra páginas que ensinam como montar expressões de busca avançadas. Finalmente, o usuário também pode receber apoio dos funcionários do SID responsáveis pelo uso da biblioteca digital, via mensagem eletrônica, telefone ou visita.

#### *Mídias e formatos*

Dependendo do PDI, sua informação de conteúdo pode ser acessada via o protocolo http em mais de um formato. Assim no caso de teses e dissertações, o acesso pode ser feito nos formatos .html, .jpg, .pdf e .djvu. Quanto à informação descritiva, pode ser acessada em vários formatos: tabela HTML, BibTeX, Refer, XML, xrefer, oai\_dc e mtd-br no caso de teses e dissertações. O acesso também pode ser feito via protocolo de coleta de metadados da Open Archives Initiative (OAI-PMH).

Mediante pagamento, o SID pode fornecer cópias impressas ou CD-ROM da informação de conteúdo de um determinado PDI. Finalmente alguns PAIs podem ser integralmente copiados via *download*.

#### *Transformação*

O processo de disseminação da informação descritiva em vários formatos é o resultado da transformação desta informação do formato usado internamente para o formato desejado. Essa transformação ocorre no ato do pedido de acesso. Não existe transformação similar para a informação de conteúdo.