

## **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).

### **DOMÍNIO DO ACERVO DIGITAL**

*Domínio e Usuários.* 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 é 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 começo de uso de uma biblioteca digital. Esta biblioteca oferece serviços para submissão, armazenamento, divulgação, disponibilização e preservação de documentos. Seus usuários finais são os 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 na forma digital das novas contribuições. Concomitantemente com estas admissões começou o trabalho de digitalização do acervo ainda exclusivamente impresso.

*Produtores de dados.* Os dados são o resultado da produção intelectual dos pesquisadores e alunos de Pós-Graduação do Instituto e eventualmente por pesquisadores fora da instituição, envolvidos nas áreas do conhecimento mencionadas acima.

### **PROCESSO E INTERFACE DE 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 se o trabalho foi ou deve ser publicado fora do INPE, ou ainda se deve ser publicado pelo INPE.

Caso o trabalho foi publicado fora do INPE, o acordo de submissão consiste apenas em definir o nível de disseminação do trabalho: sem restrição, com restrição ao INPE ou sem permissão de acesso.

Caso o trabalho deve ser publicado fora do INPE, mais precisamente, foi ou deve ser submetido 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. Os mesmos poderão ser cedidos livremente futuramente a qualquer editora. Os trabalhos nesta categoria são chamados de ePrint.

Caso o trabalho for publicado pelo INPE, o acordo de submissão estipula que o produtor está ciente que ao fazer a submissão, ele está transferindo o documento original para o administrador da Biblioteca Digital e que este não será submetido para outro acervo.

*Sessão típica de entrega dos dados.* Dependendo do acordo de submissão, a entrega dos dados pode tomar formas diferentes. No entanto a maioria das entregas são feitas através de formulários on-line personalizados em função do tipo do trabalho a ser submetido: 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, ele deverá ser entregue dentro das normas de editoração definidas pela Instituição. O trabalho submetido será revisado e havendo necessidade de melhoria na sua forma, ele deverá ser re-submetido. Quando o trabalho consiste em mais de um arquivo, o conjunto de arquivos deve ser entregue encapsulado num arquivo no formato .zip. Este arquivo é automaticamente aberto no ato do recebimento do trabalho recriando o conjunto original de arquivos. Caso o trabalho seja um ePrint pode haver também 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, ainda existe uma possibilidade, bem que pouco usada de entrega de PAI já pronto. Neste caso, distingue-se entre os PAIs que contêm o trabalho original daqueles que contêm apenas uma cópia do mesmo.

*Objetos de informação entregues.* Os objetos de informação entregues pelos produtores são todos compostos exclusivamente de objetos digitais nos seguintes formatos: .pdf, .djvu, .doc, .tex, .djuv, .html, .txt, .jpg, .ppt, .prn, .ps, .zip. Na entrega, a informação de representação destes objetos consiste em apenas na presença do tipo de formato no nome dos arquivos submetidos. Todos os objetos de informação entregues consistem em informação de conteúdo, informação descritiva e eventualmente informação de descrição de preservação (IDP). O pacote de submissão de informação (PSI) consiste de duas partes: de um lado, a informação descritiva (geralmente informação de natureza bibliográfica, mais palavras-chaves e resumo) e a IDP (referência dos programas de computadores geradores da informação de conteúdo), e de outro lado, a própria informação de conteúdo. No ato de recebimento do PSI, a IDP é automaticamente acrescentada de outros tipos de informação de preservação (data, local e responsável da submissão, identificador).

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

*Processo de transformação.* Na maioria das submissões não há conversão automática de formato da informação de conteúdo, exceto no caso de ePrint.. Neste caso, todas as submissões são feitas no formato postscript (.prn) e no ato da submissão este formato é convertido automaticamente no formato “portable document format” (.pdf) com a inserção de um identificador no cabeçalho das páginas do documento. No caso dos trabalhos publicados pela Instituição, como teses e dissertações, o documento entregue é guardado no PAI junto com suas conversões nos formatos .pdf, .djvu e .html que são executadas manualmente.

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

*Validação.* No caso de publicação pelo Instituto, quando o trabalho é submetido na sua forma digital à Memória Técnico-Científica, ele já passou 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 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 as normas de edição definidas pelo Instituto.

## **FORMAS INTERNAS**

*Armazenamento.* O armazenamento permanente na biblioteca digital é feito exclusivamente na forma de arquivos no sistema de arquivo disponível no disco rígido dos computadores que hospedem o acervo desta 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 um (ou mais) à informação descritiva. Esta organização segue o padrão de Repositórios Uniformes para uma biblioteca digital (*URLib*). Os nomes dos diretórios criados para hospedar os PAIs obedecem a uma regra de geração que garante sua universalidade dentro do universo de uso da biblioteca digital, mesmo estes sendo gerados em acervos distintos. Esta forma de organizar o sistema de arquivo permite em particular a transferência dos PAIs de um acervo para outro sem necessidade de rever a 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 tiver um caráter sigiloso. Na hora da criação dos PAIs estas informações (metadados) estão capturadas e mantidas no base de dados que pode ser acessada por meio de dois tipos de servidor: um primeiro usando uma comunicação via socket e outro via protocolo http. Os metadados podem ser retornados em vários formatos (ver Acesso).

*Migração.* Até o momento somente foram realizadas migrações de plataforma, por exemplo, de SunOS para Linux ou Windows, de forma a melhorar a razão desempenho/custo. Nestas migrações não ocorre alteração na informação de empacotamento, nas informações de conteúdo e descritiva, e na IDP. Estas migrações são do tipo reprodução e são realizadas usando 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. Esta 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 é atualmente acessível via 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 pode ou não ser visíveis, a não ser pelo administrador da biblioteca, por meio do sistema de busca. Assim grande parte dos PDIs relativos ao software de gerenciamento e manutenção da biblioteca digital (URLibService) são visíveis apenas para o administrador. Caso o PDI seja classificado na categoria de divulgação reservada ele não também não é visível via 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. Usam o recurso de acesso restrito por login/senha novamente os PDIs com divulgação reservada. Os PIDs com divulgação interna ao Instituto possuem restrições para impressão (conforme opções definidas no arquivo PDF). Finalmente os PDIs contendo cópia de publicações feitas fora do INPE podem ter seu acesso restrito à instituição por meio do controle de IPs.

*Ajuda 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 atualizada automaticamente 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 a de Sensoriamento Remoto, têm seu acesso à biblioteca digital personalizado no sentido em que o resultado da busca fica limitada naquelas áreas do conhecimento. O usuário final ainda encontra páginas que ensinam como montar expressões de busca avançadas. Finalmente ele pode também receber via mensagem eletrônica, telefone ou visita apóio dos funcionário do SID que são responsáveis pelo uso da biblioteca digital.

*Formato disponível.* Dependendo do PDI, sua informação de conteúdo pode ser acessado 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 a informação descritiva, ela 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 pode também ser feito via o 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 copiados na íntegra via download.

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