



Ministério da
Ciência e Tecnologia



dpi.inpe.br/banon-pc3@80/2009/11.10.13.03

AUDITORIA INTERNA SOBRE O REPOSITÓRIO DIGITAL DA MEMÓRIA CIENTÍFICA DO INPE

Gerald Jean Francis Banon

URL do documento original:
<<http://urlib.net/J8LNKAN8PW/36CT2G2>>

INPE
São José dos Campos
2010

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):

Presidente:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr^a Regina Célia dos Santos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr. Horácio Hideki Yanasse - Centro de Tecnologias Especiais (CTE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Deicy Farabello - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Vivéca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)



Ministério da
Ciência e Tecnologia



dpi.inpe.br/banon-pc3@80/2009/11.10.13.03

AUDITORIA INTERNA SOBRE O REPOSITÓRIO DIGITAL DA MEMÓRIA CIENTÍFICA DO INPE

Gerald Jean Francis Banon

URL do documento original:
<<http://urlib.net/J8LNKAN8PW/36CT2G2>>

INPE
São José dos Campos
2010

Dados Internacionais de Catalogação na Publicação (CIP)

Banon, Gerald Jean Francis.

B227 Auditoria interna sobre o repositório digital da Memória Científica do INPE / Gerald Jean Francis Banon. – São José dos Campos : INPE, 2010.

40 p. ; (dpi.inpe.br/banon-pc3@80/2009/11.10.13.03)

1. Auditoria. 2. Memória científica. 3. Repositório digital.
4. Arquivo digital. 5. Biblioteca digital. I. Título.

CDU 021.61:657.6

AGRADECIMENTOS

Agradecemos aqui o Serviço de Informação e Documentação (SID) do INPE, e em particular à Marciana Leite Ribeiro, Silvia Castro Marcelino, Simone Angélica Del-Ducca Barbedo, Viveca Sant'Ana Lemos e Yolanda Ribeiro da Silva Souza, pelas sugestões de melhoria à primeira versão deste texto.

RESUMO

Neste relatório, são apresentados os primeiros elementos de uma auditoria interna, realizada para avaliar a confiabilidade do Repositório Digital que hospede a Memória Científica do INPE. A avaliação é conduzida seguindo as recomendações contidas no documento intitulado *Audit and certification of trustworthy digital repositories* preparado pelo *Consultative Committee for Space Data Systems*.

INTERNAL AUDIT ON INPE SCIENTIFIC MEMORY DIGITAL REPOSITORY

ABSTRACT

In this report, the first elements of an internal audit, made to evaluate the trustworthiness of the Digital Repository which is hosting the INPE Scientific Memory, are presented. The evaluation is conducted following the recommendation contained in the document entitled “Audit and certification of trustworthy digital repositories” prepared by the “Consultative Committee for Space Data Systems”.

LISTA DE TABELAS

	<u>Pág.</u>
5.1 Matriz Processos × Responsabilidades	40
5.2 Fatores de riscos	42

SUMÁRIO

	<u>Pág.</u>
1 Introdução	1
2 Terminologia	3
3 Infraestrutura organizacional	5
3.1 Governança e viabilidade organizacional	5
3.2 Estrutura organizacional e quadro de funcionários	7
3.3 Responsabilidade regulamentar e quadro da política de preservação	9
3.4 Sustentabilidade financeira	12
3.5 Contratos, licenças, e responsabilidades	12
4 Gerenciamento de objetos digitais	17
4.1 Admissão: aquisição de conteúdo	17
4.2 Admissão: criação de PAI	19
4.3 Plano de preservação	23
4.4 Preservação de PAI	25
4.5 Gerenciamento de informação	29
4.6 Gerenciamento de acesso	31
5 Gerenciamento de riscos de infraestrutura e segurança	35
5.1 Gerenciamento de riscos de infraestrutura técnica	35
5.2 Gerenciamento de riscos de segurança	41
REFERÊNCIAS BIBLIOGRÁFICAS	45

1 Introdução

A Memória Científica do INPE é o Repositório de toda produção intelectual da Instituição. Com o evento da Internet, essa produção pode se tornar ao alcance do público via Web desde que ela seja convertida e organizada na forma digital, o que vem acontecendo ao longo dos anos. A questão é verificar se isto está acontecendo de forma a garantir a manutenção do acesso por longo prazo.

Desde o início, em 1995, o Repositório digital do INPE vem sendo hospedado na plataforma *URLib*. O objetivo desta auditoria é confrontar as escolhas feitas até o momento, com as recomendações contidas no documento CCSDS 652.0-R-1, intitulado: *Audit and certification of trustworthy digital repositories* (CCSDS, 2009), e preparado pelo *Consultative Committee for Space Data Systems*. Isto, afim da Comunidade Alvo do Repositório ser esclarecida sobre as condições atuais de preservação da Memória Científica do INPE.

Nessa auditoria interna, procuramos responder a todos itens constantes do documento CCSDS 652.0-R-1. A numeração de cada subseção correspondente a esses itens foram reproduzidos aqui, traduzindo para o Português seu enunciado. As subseções que não requerem comprovações foram omitidas.

Na próxima seção, a terminologia usada na área dos Repositórios Digitais é relembrada. As três últimas seções seguem o roteiro das Seções 3 a 5 do documento CCSDS 652.0-R-1 sobre a infraestrutura organizacional, o gerenciamento de objetos digitais e, finalmente, o gerenciamento de riscos.

2 Terminologia

Neste documento, ao se referir ao conceito de Repositório Digital, ou simplesmente Repositório, é feito uso da terminologia definida no documento CCSDS 650.0-B-1, intitulado: *Reference Model for an Open Archival Information System (OAIS)* (CCSDS, 2002), e preparado também pelo *Consultative Committee for Space Data Systems*. Em outros trabalhos, um Repositório Digital é ainda denominado Arquivo Digital ou Biblioteca Digital.

As definições abaixo são adaptadas da norma ABNT intitulada: “Sistemas espaciais de dados e informações – Modelo de referência para sistema aberto de arquivamento de informação (SAAI)” (ABNT, 2007), baseada na recomendação CCSDS 650.0-B-1.

Acesso: Entidade do **Repositório** que contém serviços e processos necessários para possibilitar a visualização dos conteúdos de informação arquivada e o uso dos serviços relacionados pelo **Consumidor**.

Admissão: Entidade do **Repositório** que contém serviços e processos necessários para aceitar **Pacotes de Submissão de Informação** dos **Produtores**, preparar **Pacotes de Arquivamento de Informação** para arquivamento e assegurar que **Pacotes de Arquivamento de Informação** e sua **Informação Descritiva** de apoio sejam inseridos dentro do **Repositório**.

Comunidade Alvo: Grupo identificado de potenciais **Consumidores** capazes de compreender determinado conjunto de informações. A **Comunidade Alvo** pode ser composta por múltiplas comunidades de usuários.

Consumidor: Papel desempenhado por aquelas pessoas ou sistemas-cliente que interagem com os serviços do **Repositório** para encontrar informação preservada de interesse e para acessar essa informação de forma detalhada.

Informação de Conteúdo: Conjunto de informações alvo original da preservação – usa-se aqui também os termos “documento” e “texto completo”.

Informação de Descrição de Preservação - IDP: Informação necessária para a adequada preservação da **Informação de Conteúdo** – inclui o que se chama aqui de “metadados”.

Informação Descritiva: Conjunto de informações fornecido à entidade gerenciamento de dados para apoiar a pesquisa, a solicitação e a recuperação dos conteúdos

de informação do **Repositório** pelo **Consumidor**.

Informação de Fixidez: Informação que documenta mecanismos de autenticação e fornece chaves de autenticação para garantir que o objeto de **Informação de Conteúdo** não tenha sido alterado de forma não documentada.

Informação de Representação: Informação que mapeia um objeto composto por um conjunto de cadeias de *bits* em conceitos mais compreensíveis.

Objeto digital: Objeto composto por um conjunto de cadeias de bits.

Pacote de Arquivamento de Informação - PAI: Pacote de informação preservado pelo **Repositório**, composto por **Informação de Conteúdo** e sua **Informação de Descrição de Preservação**.

Pacote de submissão de informação - PSI: Pacote de informação entregue pelo **Produtor** ao **Repositório** usado na formação de um ou mais **Pacote de Arquivamento de Informação**.

Produtor: Papel desempenhado por aquelas pessoas ou sistemas-cliente que fornecem informação a ser preservada.

Repositório: Organização que pretende preservar informação para acesso e uso por determinada **Comunidade Alvo**.

3 Infraestrutura organizacional

3.1 Governança e viabilidade organizacional

3.1.1 O Repositório deve ter uma declaração de missão que reflete um compromisso com a preservação, retenção por longo prazo, gerenciamento, e acesso à informação digital.

Considerando que a produção científica do INPE é parte da documentação governamental, a missão do Repositório da Memória Científica da Instituição é regida pelo Parágrafo 2º do Artigo 216º da Constituição da Republica Federativa do Brasil de 1988:

Cabem à administração pública, na forma da lei, a gestão da documentação governamental e as providências para franquear sua consulta a quantos dela necessitem (BRASIL, 1988).

O Projeto de Lei 41, de 2010, deve regulamentar o acesso a informações previsto no § 2º do art. 216, e revogar a Lei nº 8.159 mencionada em seguida.

Como parte do Poder Público, o INPE tem o dever de proteger sua produção científica conforme o Artigo 1º da Lei nº 8.159, de 8 de janeiro de 1991 que dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências:

É dever do Poder Público a gestão documental e a de proteção especial a documentos de arquivos, como instrumento de apoio à administração, à cultura, ao desenvolvimento científico e como elementos de prova e informação (BRASIL, 1991).

Considerando o Parágrafo 1º do Artigo 1º do Decreto nº 4.915, de 12 de dezembro de 2003:

Para os fins deste Decreto, consideram-se documentos de arquivo aqueles produzidos e recebidos por órgãos e entidades da administração pública federal, em decorrência do exercício de funções e atividades específicas, qualquer que seja o suporte da informação ou a natureza dos documentos (BRASIL, 2003).

a produção científica do INPE consiste em documentos de arquivo e como tal sua gestão fica sob a responsabilidade do Sistema de Gestão de Documentos de Arquivo (SIGA) instituído por esse decreto.

Considerando que a produção de Teses e Dissertações é parte importante da produção científica do Instituto, o Repositório da Memória Científica deve atender o

Artigo 1 da Portaria nº 13/2006 da CAPES que institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos:

Para fins do acompanhamento e avaliação destinados à renovação periódica do reconhecimento, os programas de mestrado e doutorado deverão instalar e manter [...] arquivos digitais, acessíveis ao público por meio da Internet, para divulgação das dissertações e teses de final de curso (CAPES, 2006).

3.1.2 O repositório deve ter um Plano Estratégico de Preservação que define a abordagem que o repositório tomará no apoio por longo prazo à sua missão.

O Plano Estratégico de Preservação da Memória Científica ainda tem que ser detalhado. Ele se insere nas: *Estratégias de aprimoramento e ampliação da participação do SID nas atividades de PD&I do INPE* (RIBEIRO et al., 2009). Ele deverá ser revisado em 2011.

3.1.2.1 O Repositório deve ter um plano de sucessão apropriado e formal, planos de contingência, e/ou acordo de depósito existentes caso o Repositório deixar de operar ou a instituição que o rege ou financia, alterar substancialmente seu objetivo.

Um parceiro natural no plano de sucessão seria o Arquivo Nacional, desde que pelo *Artigo 18 da Lei nº 8.159, de 8 de janeiro de 1991* que dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências:

Compete ao Arquivo Nacional a gestão e o recolhimento dos documentos produzidos e recebidos pelo Poder Executivo Federal, bem como preservar e facultar o acesso aos documentos sob sua guarda, e acompanhar e implementar a política nacional de arquivos (BRASIL, 1991).

No entanto, não existe um plano formal entre o INPE e o Arquivo Nacional. Este plano poderia ser elaborado dentro do quadro do projeto *International Research on Permanent Authentic Records in Electronic Systems* (InterPARES).

3.1.2.2 O repositório deve monitorar o ambiente organizacional para determinar quando executar o seu plano de sucessão formal, planos de contingência e / ou acordo de depósito.

Não existe procedimentos formais para monitorar a viabilidade do Repositório. No entanto, essa auditoria interna é um primeiro passo nesse sentido.

3.1.3 O repositório deve ter uma Política de Acervo ou outro documento que especifica o tipo de informação que irá preservar, manter, gerenciar e prover acesso à.

Os tipos de documentos alvos da preservação pelo Repositório são apresentados na Seção 2.1 do Manual para elaboração, formatação e submissão de teses, dissertações e outros trabalhos do INPE (INPE, 2010) e na Apêndice A.1 da nota técnica: Aspectos da Memória Científica do Instituto Nacional de Pesquisas Espaciais (INPE) (RIBEIRO, 2007). Parte desses tipos aparecem também relacionados na Tabela dos formulários de submissão de trabalhos integrando a Memória Científica do INPE (BANON, 2006).

Não foram ainda definidas prioridades sobre quais são as outras Informações de Conteúdo (e.g., notícias, *clippings*, documentos de programas e atividades institucionais) que o INPE deveria preservar na forma digital. Ainda precisaria, também, dimensionar os recursos necessários para tornar essa preservação uma realidade.

Os documentos que ingressam o Repositório Digital podem ser divididos em duas grandes categorias: os que nascem digitais e os que nasceram na forma impresso. Os documentos dessa última categoria requer o trabalho adicional de digitalização para ingressar o Repositório. A Instituição tem dada prioridade à digitalização das Teses e Dissertações, no entanto esse processo está sendo muito moroso por uma falta de recursos.

3.2 Estrutura organizacional e quadro de funcionários

3.2.1.1 O Repositório deve ter identificado e estabelecido as obrigações que ele precisa cumprir.

O SID está empenhado no mapeamento dos processos envolvidos nos serviços de informação e documentação, os quais incluem os serviços da Memória Científica (BARBEDO et al., 2009).

Com relação à Memória Científica, destacam-se os seguintes processos (entre parêntese são definidos o respectivo responsável e eventualmente a respectiva competência: AS = Analista de Sistema; P = Programador; AdmRd = Administrador de Rede; AdmSt = Administrador de Sistema):

- configuração de admissão (configurar procedimentos de admissão e atualização (INPE - AdmSt));

- manutenção de tabelas (manter tabelas para o preenchimento automático de campos (INPE - AdmSt));
- importação Lattes (importar registros da Plataforma Lattes (INPE - AdmSt));
- auto-arquivamento (inserir o texto completo (autores));
- ePrint (submeter e autalizar trabalhos ainda não publicados (autores));
- anais de Evento (submeter trabalhos em evento organizado pelo INPE (autores) e publicar anais (INPE - AdmSt));
- Tese e Dissertação (editar e publicar T&Ds (autores/SID));
- MAN, NTC, PRP, PUD, RPQ (editar e publicar outras publicações (autores/SID));
- revisão de admissão (revisar metadados e formatação das admissões - (SID));
- digitalização (digitalizar o material impresso - (SID));
- provedor OAI (acompanhar o processo de exportação dos registros via o protocolo OAI (INPE - AdmSt));
- configuração de acesso (configurar procedimentos de acesso (INPE - AdmSt));
- geração de tabelas e relatórios (criar e gerar tabelas, relatórios e indicadores dinâmicos (INPE - AdmSt));
- copia de segurança (programar ciclos de cópia de segurança (*backup*) (INPE - AdmRd));
- manutenção do resolvedor `urlib.net` (terceiro);
- manutenção do aglomerado:
 - definir necessidades computacionais (INPE - AS);
 - instalar/manter redes e computadores (INPE - AdmRd);
 - migrar conteúdo entre acervos (INPE - AdmSt);
 - programar o acesso aos acervos distribuídos após remanejamento (INPE - AdmSt).

3.2.1.2 O Repositório deve ter o número adequado de funcionários para realizar todas as funções e serviços.

A equipe técnica encarregada do Repositório é constituída por 4 servidores e 3 estagiários INPE, 1 estagiário SELPER, 3 colaboradores dos Centros Regionais contribuindo em tempo parcial, mais 2 administradores de rede e 1 pesquisador em tempo parcial também.

3.2.1.3 O Repositório deve ter um programa de desenvolvimento profissional ativo existente que resulta em funcionários com habilidades e oportunidades de desenvolvimento de competências.

Existe um plano anual de capacitação da equipe técnica (ver relatórios de avaliação das atividades do SID mencionados no Item 3.3.3). Por outro lado, o preparo sobre questões operacionais vem acontecendo por meio de troca de experiência no dia-a-dia.

Para os alunos de Pós-Graduação do INPE, existe um programa de treinamento sobre os diversos aspectos da editoração por meio de **seminários de orientação para pesquisa e publicação digital**. (LEMOS, 2009).

3.3 Responsabilidade regulamentar e quadro da política de preservação

3.3.1 O Repositório deve definir sua Comunidade Alvo e base(s) de conhecimento associada(s) e terá estas definições adequadamente acessível.

A Comunidade Alvo da Memória Científica é definida na seção intitulada “Domínio” do ePrint: **Biblioteca Digital da Memória Técnico-Científica do INPE** (BANON, 2005).

A Comunidade Alvo é suposta usar os navegadores: Mozilla Firefox, Internet Explorer e/ou Safari com *plugin* do Adobe Acrobat Reader para abertura dos arquivos no formato PDF (*Portable Document Format*).

3.3.2 O Repositório deve ter Políticas de Preservação existentes para garantir que seu Plano Estratégico de Preservação será cumprido.

A política de preservação do Repositório é definida por meio de resoluções da direção do INPE:

RE/DIR-204-v01 – Política de Editoração e Preservação da Produção Intelectual do INPE (de 20/07/2007 - revogada em 10/02/2010).

RE/DIR-204-v02 – Política de Editoração e Preservação da Produção Intelectual do INPE (de 10/02/2010).

Essa política define quais são as responsabilidades do Repositório frente à Comunidade Alvo no que diz respeito à Produção Intelectual (PI) do INPE.

As principais responsabilidades do Repositório são:

- R1: Registrar a PI (coletar metadados de cada trabalho);
- R2: Depositar a PI (depositar o texto completo de cada trabalho registrado);
- R3: Editar a PI (compilar as componentes de um trabalho e criar o pdf);
- R4: Publicar a PI (autorizar a publicação de um trabalho);
- R5: Encontrar a PI (fornecer meio de buscar um o mais trabalhos);
- R6: Disseminar a PI (disseminar com ou sem restrições os trabalhos);
- R7: Preservar a PI (preservar os originais submetidos para publicação e a navegação);
- R8: Verificar a PI (verificar a completude e exatidão dos trabalhos originais);
- R9: Mensurar a PI (gerar relatórios, tabelas e indicadores).

3.3.2.1 O Repositório deve ter mecanismos de revisão, atualização e desenvolvimento contínuo de suas Políticas de Preservação enquanto o Repositório cresce e enquanto a tecnologia e a prática da Comunidade evoluem.

As revisões da Política de Preservação são propostas à Direção do Instituto pelo Conselho de Editoração e Preservação da Produção Intelectual do INPE:

RE/DIR-202 – Conselho de Editoração e Preservação da Produção Intelectual do INPE.

Em fevereiro de 2010, foi feita uma primeira revisão da Política de Preservação (ver Item 3.3.2).

3.3.3 O Repositório deve ter uma história documentada das alterações das suas operações, procedimentos, *software* e *hardware*.

Ainda não existe procedimentos sistemáticos para guardar o histórico da evolução do Repositório da Memória Científica. No entanto, existem relatórios anuais sobre as atividades do SID onde é possível encontrar marcos da evolução do Repositório para os anos: 2005 (INPE/SID, 2005), 2006 (INPE/SID, 2006), 2007 (INPE/SID, 2007), 2008 (INPE/SID, 2008) e 2009 (INPE/SID, 2009).

Existem também alguns artigos e materiais audiovisuais descrevendo o Repositório numa determinada época.

Em 2004:

Preservação digital da Memória Técnico-Científica do INPE (ePrint) (BANON et al., 2004a);

Preservação digital da Memória Técnico-Científica do INPE (material audiovisual)

(BANON et al., 2004b).

Em 2005:

Biblioteca digital da Memória Técnico-Científica do INPE (material audiovisual) (BANON; BANON, 2005).

Em 2006:

Reflexões sobre o resguardo da memória científica do INPE (artigo) (RIBEIRO, 2006a);

Reflexões sobre o resguardo da memória científica do INPE (material audiovisual) (RIBEIRO, 2006b).

Em 2007:

Aspectos da memória científica do Instituto Nacional de Pesquisas Espaciais (INPE) (nota técnica) (RIBEIRO, 2007).

Em 2008:

Repositório digital do INPE (material audiovisual) (BANON; RIBEIRO, 2008).

3.3.4 O Repositório deve se comprometer com a transparência e a responsabilidade em todas as ações de apoio à operação e gerenciamento do Repositório que afetam a preservação do conteúdo digital ao longo do tempo.

O presente documento constitui uma primeira referência sobre o gerenciamento do Repositório em termo de preservação do conteúdo digital. Ele é acessível sem restrição apontando para: <http://urlib.net/J8LNKAN8PW/36CT2G2>.

3.3.5 O Repositório deve definir, coletar, controlar e fornecer adequadamente suas medidas de integridade da informação.

As atuais medidas de integridade são: a somatória dos tamanhos dos arquivos, o número de arquivos que compõe o texto completo e a data de última atualização. No futuro, poderia-se considerar como mais uma possível medida de integridade a assinatura MD5 (*Message-Digest algorithm 5*) de cada Informação de Conteúdo (RIVEST, 1992).

Caso a integridade de um registro for questionada, é possível, consultando a série histórica deste registro, perceber eventuais alterações (ver Item 4.2.9).

3.3.6 O Repositório deve comprometer-se a uma programação regular de auto-avaliação e certificação externa.

Não existe no momento, auditorias externas regulares do conteúdo do Repositório como um todo. Existem apenas validações de cada nova admissão e mecanismos para detectar duplicações. Não existem verbas alocadas para a realização periódica de certificações.

3.4 Sustentabilidade financeira

3.4.1 O Repositório deve ter procedimentos de planejamento de negócios a curto e longo prazo existentes para manter o repositório ao longo do tempo.

No momento, existe apenas previsões anuais para aquisição de máquinas (computadores) com objetivo de garantir a renovação do suporte computacional e o acompanhamento do crescimento do acervo.

3.4.2 O Repositório deve ter práticas financeiras e de procedimentos que são transparentes, em conformidade com as normas e práticas contábilísticas relevantes, e são auditadas por terceiros, em conformidade com os requisitos legais locais.

O Repositório está sob a responsabilidade do Serviço de Informação e Documentação (SID). As atividades do SID recebem auditorias periódicas.

3.4.3 O Repositório deve ter um compromisso permanente para analisar e relatar sobre os riscos, benefícios, investimentos e despesas (incluindo ativos, licenças e passivos).

Os relatórios anuais de atividades do SID levantam eventuais problemas decorrentes da manutenção e crescimento do Repositório.

3.5 Contratos, licenças, e responsabilidades

3.5.1 O Repositório deve ter e manter acordos ou contratos de depósito adequados para os materiais digitais que ele gerencia, preserva e/ou para os quais fornece acesso.

O Repositório hospeda dois tipos de documentos: documentos originais e cópias de documentos originais. Os documentos originais são os documentos publicados

pelo INPE como ePrints, Teses e Dissertações, Relatórios de Pesquisa, Publicações Didáticas, Notas Técnicas e Manuais. Cópias de originais são documentos publicados por terceiros como Artigos em Revista. Os Artigos em Eventos são originais quando fazem parte de anais publicados pelo INPE, e cópias de originais no caso contrário.

Ao submeter eletronicamente um documento original o autor aceita transferir os direitos patrimoniais para o INPE, isto é autoriza o INPE em publicar o documento. O autor dá seu acordo ao INPE no ato da submissão por meio de um clique de aceite das condições sobre direitos autorais, que libera a funcionalidade de submissão (ver o texto sobre direitos autorais [no caso de T&D](#)).

O texto sobre direitos autorais apresentado ao autor constitui o acordo de submissão/depósito. No ato da submissão, este acordo é automaticamente gravado numa pasta chamada de **agreement** ao lado da pasta **doc** contendo o documento submetido, preservando assim os termos do acordo na eventualidade de uma mudança destes termos em acordos futuros.

Caso uma T&D seja submetida pelo SID (caso das T&Ds antigas que estão sendo digitalizadas pelo SID), seguindo as [recomendações](#) do Conselho de Editoração e Preservação da Produção Intelectual do INPE, a publicação acontecerá somente após o autor ter dado sua autorização por meio de uma [carta](#).

Atualmente, pela [RE/DIR-204-v02](#) (Política de Editoração e Preservação da Produção Intelectual do INPE) de 10/02/2010, este cuidado estende-se a todas T&Ds publicadas pelo INPE. Neste caso, as cópias digitais das cartas de autorização dos autores estão armazenadas nas respectivas pastas **agreement**.

O acordo de submissão para ePrints é definido por meio de uma [política editorial](#) própria.

Quando o documento hospedado no Repositório é uma cópia de um original publicado por terceiro, o acesso ao documento é restrito ao INPE, a não ser que o INPE tenha conseguido uma autorização do detentor dos direitos patrimoniais para disponibilizar o documento sem restrições (ver, por exemplo, o texto da autorização do CCSDS clicando no link vermelho **não é original**, ao abrir o documento: <http://urlib.net/rep/8JMKD3MGP8W/35LNQS8>).

3.5.1.1 O Repositório deve ter contratos ou acordos de depósito, que especificam e transferem todos os direitos de preservação necessários e os direitos transferidos devem ser documentados.

No momento, não está explícito os direitos do Repositório vinculados à preservação dos documentos eletrônicos. No futuro, será necessário mencionar no texto sobre direitos autorais que o INPE está autorizado pelo autor em alterar a codificação do documento original em todas as situações, onde isto é necessário por motivo de preservação do acesso ao conteúdo do documento.

3.5.1.2 O repositório deve ter especificado todos os aspectos relevantes de aquisição, manutenção, acesso e remoção em acordos escritos com os depositantes e outras partes interessadas.

Os vários aspectos, em particular, os de Admissão e Acesso, estão definidos nas respectivas sessões do ePrint: Biblioteca Digital da Memória Técnico-Científica do INPE (BANON, 2005).

Ver também Itens 3.5.1 e 4.6.1.

3.5.1.3 O Repositório deve ter políticas escritas, que indicam quando ele aceita a responsabilidade de preservação de conteúdo de cada conjunto de objetos de dados submetidos.

No presente momento, o Repositório assume todas as responsabilidades na preservação dos documentos submetidos conforme estipulado no texto sobre direitos autorais, a não ser que o documento seja uma cópia de um original publicado por terceiro. No entanto, não existem detalhes especificando a natureza da preservação (e.g., preservação do conteúdo, da formatação).

De modo geral, é aconselhado ao autor manter uma cópia do original em seu poder, imaginando a pior situação em que o Repositório ou seus sucessores ficassem impossibilitados de manter o original apesar de terem assumido a responsabilidade em preservá-lo.

No futuro, esta responsabilidade poderia ser explícita no ato de confirmação da admissão.

3.5.1.4 O Repositório deve ter políticas existentes, que tratam de responsabilidade e de questões de propriedade/direitos.

Ver Itens 3.5.1 e 3.5.1.1.

Por outro lado, futuras incrementações das políticas existentes deveriam ser submetidas à aprovação de especialistas em assunto de leis.

3.5.2 O Repositório deve controlar e gerenciar os direitos de propriedade intelectual e restrições ao uso de conteúdo do Repositório como requerido pelo acordo contrato ou licença de depósito.

A validação das admissões pelo SID deveria incluir a verificação do cumprimento dos direitos autorais que decorrem do reaproveitamento do conteúdo produzido por terceiros.

4 Gerenciamento de objetos digitais

4.1 Admissão: aquisição de conteúdo

4.1.1 O Repositório deverá identificar a Informação de Conteúdo e as propriedades da informação que o repositório irá preservar.

De um modo geral, os documentos inseridos no Repositório estão no formato PDF, Word e PowerPoint da Microsoft, e HTML.

Ainda não foi discutida a questão da preservação da formatação dos documentos publicados pelo INPE. No entanto, atualmente existe uma providência quanto aos documentos editados em \LaTeX . Neste caso, o arquivo `.cls` é preservado junto com os arquivos `.tex` usados na compilação do documento.

Ainda não foi discutido se todos os registros devem ser tratados da mesma forma em termo de preservação.

Hoje, estão sendo preservados, quando existem, os identificadores DOI[®] e Número de série INPE dos documentos inseridos no Repositório.

4.1.1.1 O Repositório deve ter um procedimento(s) para identificar as propriedades de informação que irá preservar.

No presente momento, não existem procedimentos para identificar as propriedades da Informação de Conteúdo que deveriam ser preservadas em relação a cada documento inserido no Repositório. No entanto é possível identificar o tipo de arquivo que contém a Informação de Conteúdo.

Para determinar quais documentos e quais propriedades devem ser preservadas, seria necessário criar um novo campo, por exemplo `preservationrequirement`.

4.1.1.2 O Repositório deve ter um registro das Informações de Conteúdo e das propriedades da informação que irá preservar.

No momento, o Repositório não dispõe de mecanismos para acompanhar o requisito de preservação. Para isto, poderia-se definir uma modalidade de uso do atual campo `lineage`.

4.1.2 O Repositório deve especificar claramente a informação que precisa ser associada a Informação de Conteúdo específico no momento do seu

depósito.

As informações necessárias na admissão de um documento definem o próprio formulário de submissão.

4.1.3 O Repositório deve ter características adequadas que permitam o reconhecimento e a análise do PISs.

No momento, o Repositório não possui mecanismos para confirmar o tipo dos arquivos submetidos. Ele apenas reconhece o conteúdo de um arquivo com base no seu tipo.

4.1.4 O Repositório deve ter mecanismos para verificar adequadamente a identidade do Produtor de qualquer material.

No momento, por razão de simplicidade no processo de admissão, não estão sendo usados mecanismos para o reconhecimento do produtor do material. Assim, o produtor não precisa se cadastrar antes de submeter seu material. No entanto, caso seja necessário (por exemplo, previamente à decisão de remover o registro que resultou de uma submissão indevida), é possível interagir com o produtor, já que seu endereço de e-mail é guardado.

4.1.5 O Repositório deve ter um processo de admissão que verifica a completude e exatidão de cada PIS.

No momento, o Repositório apresenta, para o produtor, formulários de submissão personalizados de acordo com cada tipo de documento. Os campos dos formulários podem ser com preenchimento obrigatório ou não, com menu para seleções de opções de preenchimento ou de preenchimento livre.

Para evitar duplicidade, durante a submissão o Repositório verifica a existência ou não de um registro do mesmo tipo previamente criado com os mesmos dados (autores, ano e título).

4.1.6 O Repositório deve possuir um controle suficiente sobre os Objetos Digitais para preservá-los.

A qualquer momento, é possível criar listagens relacionando os documentos submetidos com informações como: a somatória dos tamanhos dos arquivos, o número de arquivos que compõe o texto completo e a data de última atualização (ver Item 3.3.5). Desta forma, é possível conferir a integridade das informações submetidas.

Atualmente, com a esperança de contribuir para a preservação do conteúdo e da formatação dos originais, existem recomendações para que os autores forneçam seus arquivos fontes junto com o arquivo PDF.

4.1.7 O Repositório deve fornecer ao Produtor/depositante as respostas adequadas em pontos acordados durante os processos de Admissão.

No final de cada submissão, o Produtor/depositante é informado sobre o sucesso da admissão e pode consultar imediatamente o Repositório para conferir a integridade das informações submetidas.

4.1.8 O Repositório deve ter registros contemporâneos de ações e processos administrativos que sejam relevantes para a aquisição de conteúdos.

Dados relevantes ao processo de admissão estão armanezados como parte dos metadados associados ao documento submetido.

4.2 Admissão: criação de PAI

4.2.1.1 O Repositório deve ser capaz de identificar qual a definição se aplica a qual PAI.

O Repositório usa essencialmente uma única definição de Pacote de Arquivamento de Informação (PAI).

No entanto, pode-se dividir os PAIs em duas categorias. Na primeira, os PAIs possuem um arquivo alvo, sendo definido como o arquivo que deve ser aberto ao acessar o documento (a Informação de Conteúdo). Na segunda categoria, os PAIs não possuem um arquivo alvo, neste caso ao acessar o documento a lista dos arquivos que o compõe é exibida na forma de uma página HTML, a partir da qual cada arquivo pode ser acessado separadamente.

A presença ou não do arquivo alvo num PAI define sem ambiguidade sua categoria.

4.2.1.2 O Repositório deve ter uma definição de cada PAI que é adequada para a preservação por longo prazo, permitindo a identificação e análise de todos os componentes necessários a esse PAI.

Uma breve definição do PAI é dada na subseção “Armazenamento” (na seção “Formatos internos”) do ePrint: [Biblioteca Digital da Memória Técnico-Científica do INPE \(BANON, 2005\)](#).

A exibição dos arquivos que compõe o documento (Informação de Conteúdo) fica por conta do navegador Web usado pelo usuário final (Consumidor) e dos respectivos aplicativos instalados como *plugin* do navegador. No entanto, antes da exibição, a abertura e leitura dos arquivos, assim como o direcionamento dos seus conteúdos para o canal de comunicação apropriado, fica por conta dos servidores envolvidos no acesso à informação.

4.2.2 O Repositório deve ter uma descrição de como os PAIs são construídos a partir dos PSIs.

No Repositório, os PAIs começam a existir logo após as submissões. Eles podem ser alterados por meio de atualizações sucessivas.

No momento, a descrição da construção dos PAIs a partir dos pacotes de submissão de informação (PSIs) não existe de forma explícita, mas sim de forma implícita por meio da própria codificação do processo de admissão.

No caso da admissão de documentos publicados pelo INPE, a construção dos PAIs é feita em pelo menos duas etapas como descrito na Seção 2.9 do documento:

[Manual de instruções operacionais para publicação e disponibilização de trabalhos na biblioteca digital do INPE](#)

4.2.3.1 O Repositório deve seguir os procedimentos documentados caso um SIP não esteja incorporado à um PAI ou esteja descartado e deve indicar as razões pela quais o PSI não foi incorporado ou foi descartado.

Um primeiro teste automático de validação dos PSIs ocorre no momento da submissão. Caso um PSI passa por este teste, o Repositório imediatamente o transforma em um PAI. Enquanto o PAI não foi gerado, o próximo PSI fica numa fila de espera (primeiro entrou – primeiro saiu).

4.2.4.1 O Repositório deve identificar cada PAI dentro do Repositório.

O Repositório identifica cada PAI por meio de identificadores chamados de IBIs (*Internet Based Identifier*) criados com base na Internet. Uma primeira versão do identificador é criado seguindo as regras descritas na Seção 3 do documento:

[What is URLib? \(BANON; BANON, 2008\).](#)

Existem também duas outras versões alternativas do IBI.

Assim, por exemplo, o artigo em evento intitulado: “Um Estudo Experimental do Efeito do Número de Grashof em Chamas Difusivas” recebeu, quando depositado no Repositório, os três identificadores:

sid.inpe.br/mtc-m17@80/2007/04.09.16.57
8JMKD3MGPBW/TM6792
KUqcbtHDZ/PGQKK

O primeiro e o segundo são insensíveis a maiúsculas e minúsculas. O segundo e o terceiro são opacos.

Para localizar/abrir este documento pode-se usar qualquer uma das três URLs:

<http://urlib.net/sid.inpe.br/mtc-m17@80/2007/04.09.16.57>
<http://urlib.net/8JMKD3MGPBW/TM6792>
<http://urlib.net/KUqcbtHDZ/PGQKK>

Diferentemente do DOI[®] (*Digital Object Identifier*) (IDF, 2006) o uso do IBI não está atrelado a um registro explícito do responsável pela identificação. No Item 4.4.1.2 é dado um esboço de como seriam preservados os direitos autorais.

4.2.4.2 O Repositório deve ter um sistema confiável de serviços de vinculação/resolução a fim de encontrar o objeto identificado com unicidade, independentemente da sua localização física.

Por meio do resolvedor `urlib.net`, o Repositório dispõe de um sistema de resolução dos identificadores que permite a localização dos PAIs. Esse sistema aceita os três tipos de identificações como exemplificado no item anterior.

O atual proprietário do nome de domínio `urlib.net` é Gerald Jean Francis Banon. Com a [autorização do INPE](#), o nome `urlib.net` aponta atualmente para a servidora `hermes.dpi.inpe.br` hospedada na Divisão de Processamento de Imagens do Instituto. O *software* que implementa o resolvedor é o aplicativo `URLibService`¹ licenciado por Gerald Jean Francis Banon.

4.2.5 O Repositório deve ter acesso às ferramentas e recursos necessários para fornecer a Informação de Representação oficial para todos os Objectos Digitais que contém.

No momento, os tipos de formatos de arquivos armazenados no Repositório não são descritos localmente, nem sua descrição identificada num outro Repositório confiável,

¹Registro INPI: n° 00072253.

por meio de uma Informação de Representação pertinente a cada um.

4.2.6 O Repositório deve ter processos documentados de aquisição da Informação de Descrição de Preservação (IDP) para sua Informação de Conteúdo associada e adquirir a IDP em conformidade com os processos documentados.

Dentre os diferentes tipos de Informação de Descrição de Preservação (IDP), o Repositório mantém a Informação de Fixidez por meio de três informações: a somatória dos tamanhos dos arquivos, o número de arquivos que compõe o texto completo e a data de última atualização (ver Item 3.3.5). O Repositório mantém também a Informação de Referência por meio de um identificador associado a cada Informação de Conteúdo. O Repositório ainda mantém alguns elementos da Informação de Proveniência como: quando e quais foram os atores que participaram do processo de submissão e atualização das Informações de Conteúdo e Descritivas. Finalmente, o Repositório mantém na forma de Informação de Contexto, a existência de edições anteriores e posteriores de uma obra, toda vez que pertinente.

4.2.7 O Repositório deve garantir que a Informação de Conteúdo dos PAIs é compreensível para a Comunidade Alvo no momento da criação do PAI.

Cada Informação de Conteúdo submetido no Repositório é validado pela Memória Científica/SID por meio de um processo manual. Uma das etapas da validação consiste em verificar se informação depositada pode ser aberto por meio de pelo menos um navegador. Uma vez a validação terminada, a Memória Científica/SID encerra o processo de atualização do PAI.

4.2.8 O Repositório deve verificar cada PAI em termo de sua completude e exatidão no momento em que ele está criado.

No caso da admissão de documentos publicados pelo INPE, uma outra etapa da validação pelo SID consiste em verificar se a Informação de Conteúdo submetida ao Repositório apresenta-se na forma de arquivos-fontes usados na geração do PDF e se esses arquivos efetivamente servem na geração do PDF.

4.2.9 O Repositório deve fornecer um mecanismo independente para verificar a integridade do acervo/conteúdo do Repositório.

Existem levantamentos mensais de todos os registros contendo um texto completo.

Para cada registro é informado o nome do repositório, o site contendo o repositório, a somatória dos tamanhos dos arquivos, o número de arquivos que compõe o texto completo, a data de última atualização, o tipo de referência, o ano de publicação, o título e o nome dos autores. Desta forma, é possível montar uma série histórica de cada registro.

Estes levantamentos são acessíveis sem restrição apontando para: <http://urlib.net/8JMKD3MGPCW/36GLR3P>.

Cada levantamento é obtido por meio da URL:

```
http://bibdigital.sid.inpe.br/col/dpi.inpe.br/banon-pc2@1905/2006/02.16.12.09/doc/script.cgi?query=size+*&fieldlist=repository+site+size+numberoffiles+lastupdate+referencetype+year+title+author+programmer+editor+reporter+cartographer+director&progressiveoutput=yes
```

Esses levantamentos podem ser realizados por qualquer usuário credenciado, tornando-se assim independente o mecanismo de verificação da integridade do acervo/conteúdo do Repositório.

4.2.10 O Repositório deve ter registros contemporâneos de ações e processos administrativos que sejam relevantes para a criação de PAI.

A Informação de Proveniência, parte da Informação de Descrição de Preservação (IDP) descrita no Item 4.2.6, é registrada na forma de metadados nos respectivos PAIs. No entanto, essa informação não menciona o tipo de ações executadas. Existe apenas, para o supervisor do processo de publicação, a opção de registrar essas ações manualmente no campo **notes** dos metadados. Esses procedimentos foram adotados com base na experiência vividas de vários anos.

4.3 Plano de preservação

4.3.1 O Repositório deve ter estratégias de preservação documentadas relevantes para suas posses.

No momento, não existe uma estratégia de preservação documentada.

Os principais riscos de preservação são a obsolescência dos equipamentos, dos *softwares* e dos formatos.

Para lidar com a obsolescência dos equipamentos, conta-se com os mecanismos de

migração de arquivos do equipamento velho para o novo, apostando que os arquivos exportados do equipamento velho poderão ser importados corretamente no equipamento novo. Atualmente, as migrações são feitas usando os programas `zip/unzip`.

O Repositório está hospedado na plataforma *URLib* gerenciado pelo aplicativo *URLibService*² escrito na linguagem interpretada Tcl/Tk. Para lidar com a obsolescência das versões do interpretador, conta-se com a existência de novas versões.

Para lidar com a obsolescência de formato de arquivos, ver Item 4.3.3.1.

4.3.2 O Repositório deve ter mecanismos existentes para acompanhar a seu ambiente de preservação.

No momento, o Repositório não tem mecanismo para detectar quando um formato está se tornando obsoleto.

Por outro lado, os usuários finais (Consumidores) têm atualizados por conta própria seus softwares para abertura dos documentos disponibilizados pelo Repositório. O Repositório não tem testadas sistematicamente as novas versões dos navegadores para alertar sobre o aparecimento de disfunções como ocorreu com a versão 3.6.2 do Mozilla Firefox em Português (Brasil) para Windows.

4.3.2.1 O Repositório deve ter mecanismos existentes de vigilância e notificação de quando a Informação de Representação é inadequada para entendimento pela Comunidade Alvo dos dados possuídos.

Caso seja necessário identificar os PAIs contendo arquivos com formato obsoleto, seria possível procurá-los por meio dos campos `update` e `targetfile`. O primeiro permite identificar uma data na qual o formato era operacional, o segundo permite identificar o tipo de formato.

4.3.3 O Repositório deve ter mecanismos para alterar os seus planos de preservação como resultado das suas atividades de acompanhamento.

O Repositório ainda não tem rotina para ajustar seu plano de preservação.

4.3.3.1 O Repositório deve ter mecanismos para a criação, identificação ou coleta de qualquer Informação de Representação extra necessária.

Para lidar com a obsolescência de formato de arquivos, procura-se arquivar junto com

²Registro INPI: nº 00072253.

um arquivo no formato PDF, os arquivos-fontes que foram usados para sua geração (arquivos MSWord ou L^AT_EX), contando-se com a existência de novas versões de *software* gerador de PDF (ver também Item 4.4.1.1). No entanto, falta ainda registrar a versão da codificação dos arquivos fontes.

4.3.4 O Repositório deve apresentar provas da eficácia das suas actividades de preservação.

O Repositório não dispõe de mecanismos automáticos para verificar o sucesso na abertura dos arquivos armazenados. No entanto, o Repositório, poderia disponibilizar um aviso para os usuários finais (Consumidores) indicando que, em caso de dificuldade ao abrir um arquivo, ele deveria informar do problema ao administrador do Repositório.

4.4 Preservação de PAI

4.4.1 O Repositório deve ter as especificações de como os PAIs são armazenados até o nível de bit.

No Repositório, os PAIs consistem em um conjunto de arquivos que estão dispostos em pastas no próprio sistema de arquivos (ver Seção 4.2.1.2). Quando necessário (por exemplo em caso de migração), os arquivos que compõe um PAI são empacotados na forma de um arquivo ZIP. A Informação de Representação não especifica como é armazenada a Informação de Conteúdo ao nível dos bits, ela consiste apenas na presença do tipo de formato no nome dos arquivos. Quando esse tipo é omitido, o arquivo é considerado como sendo do tipo `.txt`.

Quanto à Informação de Descrição de Preservação (IDP), ela pode ser guardada em um arquivo tipo binário por questão de privacidade. Nesse caso, seu armazenamento e recuperação envolvem, respectivamente as funções `Tcl Shift` e `UnShift` definidas a seguir:

```
proc Shift {rep var} {
    regexp {.$} $rep shift
    binary scan [list $rep $var] c* varC
    set data {}
    set space 0
    set i 0
    foreach number $varC {
        if $space {
```

```

        lappend data [expr $number + $shift - $i - 70]
        incr i
    } else {
        lappend data [expr $number - 81]
        if {$number == 32} {set space 1}
    }
}
set data [concat $data $shift]
return [binary format c* $data]
}

proc UnShift {data} {
    binary scan $data c* dataC
    set shift [lindex $dataC end]
    set dataC [lreplace $dataC end end]
    set data {}
    set space 0
    set i 0
    foreach number $dataC {
        if $space {
            lappend data [expr $number + 70 + $i - $shift]
            incr i
        } else {
            lappend data [expr $number + 81]
            if {$number == -49} {set space 1}
        }
    }
}
return [binary format c* $data]
}

```

Os comandos Tcl usados nestas funções estão descritas no [ActiveTcl User Guide](#) (ACTIVESTATE, 2010).

A abertura dos arquivos no formato PDF dependem essencialmente do programa Adobe Acrobat Reader usado pela Comunidade Alvo. No entanto, o Repositório não dispõe, no momento, da Informação de Representação descrevendo o formato PDF.

4.4.1.1 O Repositório deve preservar a Informação de Conteúdo dos PAIs.

Todos os documentos inseridos no Repositório e aceitos para publicação devem ser preservados por longo prazo e, por isto, o correspondente PAI não pode ser removido. A remoção de um PAI somente pode ocorrer antes de seu conteúdo ser aceito para publicação e em caso de duplicação. Não existe ainda política de preservação a respeito de um PAI, cujo conteúdo não foi aceito para publicação.

Ainda não foram definidos procedimentos no caso de necessitar de uma ação de transformação dos PAIs com objetivo de preservar a Informação de Conteúdo. Procedimentos estes, que poderiam garantir que as propriedades da Informação de Conteúdo continuam preservadas. Não existem, por exemplo, procedimentos sistemáticos de armazenamento de versões de um mesmo documento, a não ser, e por outros motivos, no caso dos ePrints ou no caso da criação de uma nova edição de um documento.

No entanto, existem medidas preventivas de preservação da Informação de Conteúdo quando este é produzido a partir do \LaTeX ou MSWORD. Nesse caso, existem recomendações para que os arquivos-fontes sejam submetidos junto com o arquivo PDF (cf., Seção 2.9 do [Manual de instruções operacionais para publicação e disponibilização de trabalhos na biblioteca digital do INPE](#)). No PAI, os arquivos fontes que constituem o documento (Informação de Conteúdo), estão guardados na pasta `source` (invisível na Web) e o PDF na pasta `doc` (visível na Web), sendo estas duas pastas no mesmo nível na hierarquia de pastas dentro o PAI.

No passado, foram tomadas outras medidas preventivas, como a de apresentar a mesma Informação de Conteúdo em vários formatos (HTML, DJVU e PDF). No entanto, depois do PDF se tornar em 2008 uma norma ISO (cf., [Long live ISO 32000-1 - The PDF standard \(KING, 2009\)](#)), fazendo com que a propriedade do padrão PDF passasse da Adobe para a ISO, ele adquiriu mais confiabilidade e estabilidade, e por isto, as novas Informações de Conteúdo passaram apenas a ser apresentadas no formato PDF.

4.4.1.2 O Repositório deve monitorar ativamente a integridade dos PAIs.

No momento, não existe um sistema independente automático para testar a integridade de cada documento (Informação de Conteúdo).

Um sistema desta natureza poderia fazer uso da assinatura MD5 da Informação de Conteúdo. Caso ele guardasse as assinaturas MD5 de todas as versões de um documento, junto com as datas de recebimento das assinaturas, ele poderia servir

também para quase³ comprovar a anterioridade de um documento em relação a um outro. Caso ele guardasse a identificação de um documento junto com a identificação dos seus proprietários e as respectivas datas de aquisição, ele poderia servir também para comprovar quem é o proprietário do original de um documento. Finalmente, caso ele guardasse todas as identificações dos sucessivos proprietários do original, ele poderia servir para comprovar quem é o autor do documento (ver em (BANON; BANON, 2009) esboço de um sistema com algumas dessas propriedades).

Atualmente, existe apenas um mecanismo com base na data de modificação dos arquivos, para alertar sobre uma modificação do conteúdo de um documento quando esta é realizada fora de padrões pre-definidos. Essa alerta é acessível explicitamente para a Comunidade Alvo ao consultar o resultado de uma busca. No caso de uma modificação incontrolada, os títulos dos registros normalmente azuis se tornam vermelhos.

Existe também, um mecanismo automático para gravar as datas de alterações ocorridas no campo `year` usado para indicar o ano de publicação dos documentos.

4.4.2 O Repositório deve ter registros contemporâneos de ações e processos administrativos que sejam relevantes para o armazenamento e preservação do PAIs.

Não existem ainda relatórios específicos sobre armazenamento e preservação dos PAIs. No entanto, existem alguns relatórios sobre a natureza dos PAIs. Estes relatórios podem ser gerados automaticamente como exemplificado no documento intitulado: “Breve relatório com dados oriundos da Memória Científica do INPE” (BANON, 2009), que apresenta os números de inserções e publicações por anos, assim como os números de PAIs por categorias, ou ainda estatísticas de acesso ao Repositório por anos.

Para cada PAI, existe um levantamento automático (chamado também de histórico) das datas de todas as alterações ocorridas junto com os nomes dos usuários responsáveis pelas alterações. No entanto, não é guardado de forma automática a natureza das alterações. Existe apenas a possibilidade para o supervisor do processo de publicação, de gravar no campo `notes` uma descrição das ações tomadas (ver Item 4.2.10). Estas informações podem ser comprovadas pela Comunidade Alvo consultando a página intitulada *Referência Completa* associada a cada PAI.

³A assinatura MD5 não é um identificador.

4.4.2.1 O Repositório deve ter procedimentos para todas as ações tomadas sobre os PAIs.

Existem procedimentos relativos ao processo de publicação e em particular, às ações a serem tomadas sobre os PAIs, na Seção 2.9 do documento:

[Manual de instruções operacionais para publicação e disponibilização de trabalhos na biblioteca digital do INPE](#)

No entanto, como mencionado no Item 4.4.1.1, ainda não foram definidos procedimentos no caso de necessitar de uma ação de transformação dos PAIs com objetivo de preservar a Informação de Conteúdo.

4.4.2.2 O Repositório deve ser capaz de demonstrar que todas as ações tomadas sobre os PAIs foram compatíveis com a especificação dessas ações.

O resultado das ações de preservação descritas no item anterior pode ser comprovado pela Comunidade Alvo ao consultar, na página intitulada *Referência Completa* associada a cada PAI, o campo `Conteúdo da Pasta source` para verificar a existência na pasta `source` dos arquivos-fontes, e verificando a ausência do campo `username` para comprovar que o PAI não pode ser mais modificado.

4.5 Gerenciamento de informação

4.5.1 O Repositório deve especificar os requisitos mínimos de informação que permitam à Comunidade Alvo de descobrir e identificar material de interesse.

Para encontrar os documentos inseridos no Repositório, basta digitar palavras-chave no campo de busca “Biblioteca Digital” na página da Biblioteca On-Line no endereço: <http://www.inpe.br/biblioteca>, ou no campo de busca na página da Memória Científica do INPE no endereço: <http://bibdigital.sid.inpe.br>, e clicar “OK”.

Detalhes sobre procedimentos para realizar buscas avançadas podem ser encontrados ao lado do botão “OK” na página da Memória mencionada acima. Outros detalhes encontram-se na seção intitulada “Acesso” do ePrint: [Biblioteca Digital da Memória Técnico-Científica do INPE \(BANON, 2005\)](#).

4.5.2 O Repositório deve capturar ou criar um mínimo de Informações Descritivas e assegurar que ela está associada com o PAI.

Toda Informação Descritiva, fornecida pelo produtor, é guardada sistematicamente no PAI. Outros elementos da Informação Descritiva são criados pelo Repositório, como por exemplo, o número de arquivos contendo a Informação de Conteúdo e o espaço ocupado por esses arquivos. A associação da Informação Descritiva com o PAI é mantida por meio do registro, na própria Informação Descritiva, do identificador do PAI.

4.5.3 O Repositório deve manter ligação bi-direcional entre cada PAI e sua Informação Descritiva.

Por meio dos identificadores do Item 4.2.4.1, é possível acessar tanto o texto completo quanto seus metadados. Por exemplo, o vínculo:

<http://urlib.net/LK47B6W/E6H5HH>

permite acessar o texto completo do documento intitulado “What is URLib?”, e o vínculo:

<http://urlib.net/LK47B6W/E6H5HH??>

permite acessar seus metadados. Por sua vez, a partir desses metadados é possível acessar o texto completo (clikando no vínculo no campo **Repositório**). Nesse caso, abre-se uma nova janela, cujo conteúdo é o resultado da resolução do vínculo:

<http://urlib.net/rep/LK47B6W/E6H5HH>

nela, é possível tanto visualizar o texto completo quanto acessar novamente os metadados clicando no vínculo **Metadados** na barra de menu na parte superior da janela.

4.5.3.1 O Repositório deve manter as associações entre seus PAIs e suas Informações Descritivas ao longo o tempo.

Como mencionado no Item 4.5.2, a associação da Informação Descritiva com o PAI é mantida por meio do registro, na própria Informação Descritiva, do identificador do PAI. Na prática, esse registro é guardado no arquivo `doc/@metadatada.refer` do repositório de metadados, e é replicado no arquivo `service/reference` desse mesmo repositório. Além disso, ele é também inserido no arquivo `col/dpi.inpe.br/banon/1998/08.02.08.56/auxdoc/.referenceTable.tcl`. Estas duplicidades são usadas por restabelecer a relação entre o PAI e a sua Informação Descritiva, quando um ou dois desses registros estão danificados.

Caso os três registros estejam perdidos, ao tentar restabelecer a relação, uma mensagem de error é gravada no arquivo `@errorLog`. Então, a relação entre o PAI e a sua Informação Descritiva pode ainda ser restabelecida manualmente usando o fato de o nome da pasta que contém a Informação Descritiva segue em geral imediatamente o nome da pasta que contém a Informação de Conteúdo do ponto de visto da ordem alfanumérica.

4.6 Gerenciamento de acesso

4.6.1 O Repositório deve cumprir as Políticas de Acesso.

A política de Acesso do INPE segue o Decreto no 4.553, de 27 de dezembro de 2002 (BRASIL, 2002) que deve ser substituído pelo texto do Projeto de Lei 41, de 2010.

De um modo geral, a política de acesso à Informação de Conteúdo consiste em dar um acesso livre a todos os trabalhos editados pelo INPE a não ser que o autor obtenha do INPE que seja de acesso restrito. Nesse caso, o acesso é feito por meio de login/senha. Quando os trabalhos são editados fora do INPE (caso das publicações em revistas, por exemplo), seu acesso é restrito ao INPE por meio do controle de IPs, a não ser que o INPE obtenha da editora a autorização de acesso livre.

Outros detalhes encontram-se na seção intitulada “Acesso” do ePrint: *Biblioteca Digital da Memória Técnico-Científica do INPE* (BANON, 2005).

Para caracterizar melhor as condições de acesso ao seu acervo, o Repositório definiu 3 indicadores.

Sejam:

- T o número total de obras registradas no Repositório;
- TC o número total de obras registradas com texto completo;
- TCL o número total de obras registradas com texto completo de acesso livre.

O Indicador de Armazenamento Digital (IAD) é, por definição, a percentagem:

$$IAD = 100TC/T.$$

Um IAD de 100% significa que todas as obras registradas no Repositório estão na forma digital.

O Indicador de Acesso Livre (IAL) é, por definição, a percentagem:

$$\text{IAL} = 100\text{TCL}/\text{TC}.$$

Um IAL de 100% significa que todas as obras na forma digital estão disponíveis livremente na Web.

O Indicador de Presença na Web (IPW) é, por definição, a percentagem:

$$\text{IPW} = 100\text{TCL}/\text{T}.$$

Um IAL de 100% significa que todas as obras registradas no Repositório estão disponíveis livremente na Web.

O Repositório mantém também o registro do número de acessos por dia à cada documento depositado em seu acervo. Com base nesse registro, a Comunidade Alvo pode consultar estatísticas de acesso à cada documento. Por exemplo, a URL: <http://urlib.net/statistics.cgi/sid.inpe.br/bibdigital@80/2006/11.11.23.17> retorna a estatística de acesso à página de busca no Repositório, cujo acesso é feito principalmente a partir da página principal do site da Biblioteca On-Line.

O Repositório não está fazendo ainda um levantamento sistemático do perfil dos usuários, por exemplo distinguindo entre os usuários internos ao INPE, dos externos, e nessa última categoria, entre os robots e os outros.

4.6.1.1 O Repositório deve registrar e passar em revista todas as falhas de gerenciamento e anomalias de acesso.

Existem, atualmente, dois tipos importantes de falha de acesso.

No primeiro tipo, a falha é de um acervo local. Nesse caso, a falha é detectada e informada no momento do retorno a uma solicitação de acesso ou de busca. No caso de um acesso, é retornada uma mensagem alertando que o documento não foi encontrado porque uma conexão com um acervo local não pode ser estabelecida (no entanto, não há informação sobre quais são os acervos locais inacessíveis naquele momento). No caso de uma busca, o retorno contém o número de acervos locais que atenderam a solicitação. Há uma falha quando este número não é igual ao número total de acervos locais pre-programados para atender às solicitações de busca (pedindo a exibição do código fonte da página de retorno, é possível encontrar a lista dos acervos locais inacessíveis naquele momento).

No segundo tipo, a falha é do resolvedor `urlib.net`. Nesse caso, o acesso aos do-

cumentos via URLs persistentes deixa de funcionar e a mensagem padrão “Servidor não encontrado” é exibida pelo navegador.

4.6.2 O Repositório deve seguir as políticas e procedimentos que permitem a difusão de Objectos Digitais que são rastreáveis até os originais, com provas que sustenta sua autenticidade.

Na fase de disseminação, as informações retornadas ao navegador são resultados de buscas, ou páginas dinâmicas de metadados ou finalmente cópias de arquivos contendo os Objetos de Informação dos PAIs. Essas informações são exibidas pelos navegadores Web e seus respectivos aplicativos instalados como *plugin*.

Nesta fase, a garantia de autenticidade decorre das propriedades do protocolo HTTP e dos aplicativos de exibição.

4.6.2.1 O Repositório deve registrar e agir de acordo com relatórios de problemas a respeito de erros em dados ou respostas dos usuários.

Os problemas encontrados pela Comunidade Alvo podem ser encaminhados ao administrador da Biblioteca Digital via e-mail (marciana@sid.inpe.br).

5 Gerenciamento de riscos de infraestrutura e segurança

5.1 Gerenciamento de riscos de infraestrutura técnica

5.1.1 O Repositório deve identificar e gerenciar os riscos de suas operações e objetivos de preservação dependentes da infra-estrutura do sistema.

Como mencionado no Item 4.3.1, o Repositório está hospedado na plataforma *URLib* que por sua vez pode ser hospedado em ambiente Windows ou Unix. O aplicativo de gerenciamento da plataforma, o *URLibService*¹, trabalha com acervos distribuídos, rompendo assim a barreira das limitações de espaço e velocidade das máquinas disponíveis. Atualmente o Repositório está hospedado em um aglomerado de 12 máquinas do tipo PC comum. O crescimento do acervo é possível, por exemplo, apenas integrando mais uma máquina no aglomerado, ou ainda substituindo uma máquina obsoleta por uma nova com maior capacidade e melhor desempenho.

As principais componentes da plataforma são o interpretador Tel/Tk, o servidor Apache e os programas ZIP/UNZIP. A troca de uma versão destes componentes por outra necessita apenas da execução da seqüência `unpost/post`, isto é, uma interrupção do serviço por alguns segundos.

O *URLibService* possui ferramentas para exportar e/ou migrar os PAIs de um acervo local para outro. Existem dois tipos de transferência entre acervos, dependendo se ela se acompanha ou não de uma transferência de direitos patrimoniais. No entanto, essas transferências são somente possíveis entre acervos no padrão da *URLib*.

Dentro de certos limites, o Repositório é capaz de recriar arquivos perdidos a partir de dois ciclos de cópias de segurança (*backups*) (ver Item 5.1.1.2).

5.1.1.1 O Repositório deve empregar tecnologia de vigilância ou outra tecnologia monitorando sistemas de notificação.

Ainda não foi elaborado nenhum relatório avaliando os riscos de obsolescência de cada componente da plataforma *URLib*.

5.1.1.1.1 O Repositório deve ter tecnologias adequadas de *hardware* para os serviços que presta aos suas Comunidades Alvos.

Os equipamentos necessários para o funcionamento do Repositório consistem apenas

¹Registro INPI: n° 00072253.

em um aglomerado de microcomputadores (atualmente são 13) onde estão armazenados os Pacotes de Arquivamento de Informação (PAIs).

Quanto à Comunidade Alvo, cada Produtor ou Consumidor, necessita como equipamento apenas de um microcomputador.

Todas essas máquinas estão interligadas entre si via Internet.

5.1.1.1.2 O Repositório deve ter procedimentos existente para monitorar e receber notificações quando modificações de *hardware* são necessárias.

O Repositório ainda não têm relatórios dinâmicos de fácil acesso para monitorar os espaços em disco ocupados e ainda disponíveis.

5.1.1.1.3 O Repositório deve ter procedimentos para avaliar quando mudanças no *hardware* atual são necessárias.

Com a plataforma *URLib* o problema de limite de capacidade de armazenamento dos microcomputadores em uso pode ser solucionado com relativa facilidade acrescentando um novo microcomputador ao aglomerado. Não há necessidade de recorrer necessariamente a novas tecnologias para superar os limite de capacidade. Por outro lado, novas técnicas podem ser usadas para reduzir o tamanho do aglomerado e assim simplificar a manutenção.

5.1.1.1.4 O Repositório deve ter procedimentos, compromisso e financiamento para substituição de *hardware* quando a avaliação indica a necessidade de fazê-lo.

A renovação e aquisição de novos microcomputadores é feita por meio de requisição de compra. O prazo de entrega dos equipamentos pode demorar vários meses, por isto é importante antecipar suas necessidades.

No futuro próximo as trocas de equipamentos poderiam decorrer da aplicação do Plano Diretor de Informática do INPE.

5.1.1.1.5 O Repositório deve ter tecnologias de *software* apropriadas para os serviços que presta a suas Comunidades Alvos.

A princípio, a plataforma *URLib* não depende nem do número, nem do tamanho dos registros que são inseridos nela, assim como não depende do formato dos dados. O

mecanismo de transferência de dados entre o Repositório e a Comunidade Alvo tem como base o protocolo HTTP que está sendo usado tanto nos scripts Tcl via o pacote `http`, quanto pelo servidor Web Apache. O uso de novos formatos de dados é sempre possível desde que o Consumidor possua no seu navegador os *plugins* apropriados.

5.1.1.1.6 O Repositório deve ter procedimentos existentes para monitorar e receber notificações quando alterações de *software* são necessárias.

O Repositório não tem procedimentos regulares para monitorar e receber avisos sobre necessidades de modificação de *software*. Existe apenas um acompanhamento aleatório sobre a existência de novas versões *software* e os relatórios de erros (arquivos: `@errorLog`).

5.1.1.1.7 O Repositório deve ter procedimentos para avaliar quando mudanças no *software* atual são necessárias.

O Repositório não têm procedimentos predefinidos para avaliar quando uma nova versão do *software* ou uma nova tecnologia de *software* são necessárias. Seriam necessárias reunir conhecimentos sobre várias tecnologias de *software* e realizar estudos comparativos. Por exemplo, seria necessário um estudo comparativo entre as plataformas `URLib` e `DSpace`. Até o momento, as adequações às necessidades da Comunidade Alvo foram realizadas por meio de melhorias do aplicativo `URLibService`².

5.1.1.1.8 O Repositório deve ter procedimentos, compromisso e financiamento para substituir o *software* quando uma avaliação indica a necessidade de fazê-lo.

Até o momento, todas as componentes de *software* parte da plataforma `URLib` e seus aprimoramentos são gratuitos. Como não houve questionamento sobre uma possível substituição, não houve também estudos sobre como financiar troca para *software* pagos.

5.1.1.2 O Repositório deve ter o *hardware* e suporte de *software* adequados para a funcionalidade de *backup* suficiente para preservar o conteúdo do Repositório e controlar suas funções.

O Repositório possui dois ciclos de cópias de segurança (*backups*). Com o primeiro ciclo, é possível recriar um arquivo desde que sua perda ocorreu depois das duas horas da manhã e a sua tentativa de recriação ocorre antes das próximas duas horas

²Registro INPI: n° 00072253.

da manhã. Com o segundo ciclo, é possível recriar um arquivo desde que sua perda ocorreu depois da décima primeira segunda-feira (quarta-feira ou sexta-feira, resp.) passada às vinte duas horas e dez minutos e a sua recriação ocorre antes das próximas vinte duas horas e dez minutos da segunda-feira (quarta-feira ou sexta-feira, resp.).

O primeiro ciclo é feito na forma de um anel de sincronização usando os aplicativos abertos `rsync` (<http://www.samba.org/rsync>) e `ssh`. Cada máquina do anel tem duas partições chamadas `dados1` e `dados2`. Uma vez por dia, às duas da manhã, a partição `dados1` de uma máquina é copiada na partição `dados2` da máquina seguinte do anel. Atualmente, apenas 9 das 12 máquinas fazem parte deste anel.

O segundo ciclo é feito com a utilização do software livre AMANDA (*Advanced Maryland Automatic Network Disk Archiver* – <http://www.amanda.org>). AMANDA é um software cliente/servidor, que pode ser utilizado para fazer cópia de segurança dos arquivos de várias máquinas conectadas em rede, numa máquina central, em discos e/ou fitas magnéticas.

O AMANDA é programado para ser executado automaticamente todas as segundas, quartas e sextas, a partir das vinte duas horas e dez minutos, e as cópia de segurança das partições `dados1` são feitas em fitas. Nove das máquinas do aglomerado são situadas num prédio diferente do prédio onde são gravadas e guardadas as fitas.

São utilizadas 33 fitas DLT de 160 GBytes de capacidade, de forma circular, ou seja, na trigésima quarta ocorrência de backup, a primeira fita é reutilizada. O software AMANDA se encarrega de gerenciar os backups, de modo a garantir que nas 33 fitas sempre há um backup completo dos sistemas de arquivo.

Ao fazer cópias de segurança das partições `dados1` de todas as máquinas do aglomerado, todo o conteúdo do Repositório fica preservado em fitas.

No entanto, não existe procedimento para alertar sistematicamente da perda de um arquivo. Apenas existem levantamentos mensais de todos os registros contendo um texto completo (ver Item 4.2.9). Esses levantamentos podem confirmar a perda de um arquivo.

5.1.1.3 O Repositório deve ter mecanismos eficazes para detectar a corrupção ou perda de bits.

Não existe no momento, como parte da Informação de Descrição de Preservação (IDP), e mais precisamente, como parte da Informação de Fixidade, a assinatura

MD5 da Informação de Conteúdo.

Alguns arquivos que contêm cópia de parte dos metadados codificados na linguagem Tcl, tem seus conteúdos indiretamente verificados quando interpretado. Nesse momento, o interpretador Tcl é capaz de detectar a existência de dados corrompidos.

5.1.1.3.1 O Repositório deve registrar e comunicar à sua administração todos os incidentes de corrupção de dados ou perda, e deverão ser tomadas medidas de reparação/substituição de dados corrompidos ou perdidos.

Não existe no momento, um registro sistemático de todas as perdas ocorridas e das medidas tomadas para recriá-las.

Durante os últimos anos, as razões de perdas de dados eram analisadas, e soluções implementadas afim de evitar novas ocorrências dos mesmos tipos de perda. Desta forma, o sistema de gerenciamento dos dados se tornou cada vez mais robusto.

No momento de queda de energia, pode ocorrer corrupção de dados, sobre tudo em arquivos grandes como aqueles que contêm cópia de parte dos metadados. Quando ocorre corrupção nesse tipo de arquivos, existem procedimentos automáticos para restaurar seus conteúdos.

5.1.1.4 O Repositório deve ter um processo de registro e reagir à disponibilidade de novas atualizações de segurança com base numa avaliação de risco-benefício.

Na plataforma *URLib* não existe até o momento atualizações automáticas. Todas as atualizações tem que ser efetivadas manualmente. Uma atualização, antes de ser efetivadas de forma generalizada, é testada por um certo tempo em poucos acervos locais.

5.1.1.5 O Repositório deve definir os processos para mídia de armazenamento e/ou alteração de hardware (e.g., renovação, migração).

Nesta auditoria, o Repositório sobe análise é considerado utilizar apenas, como meio de armazenamento, os discos rígidos dos microcomputadores. Nesse Repositório não é feito uso de CD.

5.1.1.6 O Repositório deve ter identificados e documentados os processos críticos que afetam sua capacidade de cumprir com suas responsabilidades obrigatórias.

Os principais processos são apresentados em detalhe no Item 3.2.1.1.

As principais responsabilidades do Repositório são (ver Item 3.3.2):

- R1: Registrar a PI (coletar metadados de cada trabalho);
- R2: Depositar a PI (depositar o texto completo de cada trabalho registrado);
- R3: Editar a PI (compilar as componentes de um trabalho e criar o pdf);
- R4: Publicar a PI (autorizar a publicação de um trabalho);
- R5: Encontrar a PI (fornecer meio de buscar um o mais trabalhos);
- R6: Disseminar a PI (disseminar com ou sem restrições os trabalhos);
- R7: Preservar a PI (preservar os originais submetidos para publicação e a navegação);
- R8: Verificar a PI (verificar a completude e exatidão dos trabalhos originais);
- R9: Mensurar a PI (gerar relatórios, tabelas e indicadores).

A Tabela 5.1 mostra como as responsabilidades do Repositório estão sendo cumpridas.

Tabela 5.1 - Matriz Processos × Responsabilidades

Processo	R1	R2	R3	R4	R5	R6	R7	R8	R9
Configuração de admissão	×	×							
Manutenção de tabelas	×				×				×
Importação Lattes	×				×				×
Auto-arquivamento	×	×			×	×			×
ePrint	×	×	×		×	×			×
Anais de Evento	×	×	×	×	×	×			×
Tese e Dissertação	×	×	×	×	×	×			×
MAN, NTC, PRP, PUD, RPQ	×	×	×	×	×	×			×
Revisão de admissão					×			×	×
Digitalização						×	×		
Provedor OAI					×	×			
Configuração de acesso						×			
Geração de tabelas e relatórios					×			×	×
Cópia de segurança							×		
Manutenção do resolvedor							×		
Manutenção do aglomerado							×		

Todos os processos mencionados na Tabela 5.1, não foram ainda documentados de forma exaustiva.

5.1.1.6.1 O Repositório deve ter um processo documentado de gestão de mudanças que identifica alterações em processos críticos que potencialmente afetam a capacidade do Repositório de cumprir com suas responsabilidades obrigatórias.

Por não serem ainda totalmente consolidados, os processos do Item 3.2.1.1, não possuem registros de modificações.

5.1.1.6.2 O Repositório deve ter um processo para testar e avaliar o efeito de mudanças em processos críticos do Repositório.

Não existe documentação sobre as modificações realizadas nos processos do Item 3.2.1.1, nem procedimentos preestabelecidos de teste dessas modificações. No entanto, todas as modificações realizadas até hoje passaram por testes adhoc mas não necessariamente exaustivos.

5.1.2 O Repositório deve gerenciar o número e localização de cópias de todos os objectos digitais.

O Repositório dispõe de mecanismos para criar cópias de PAIs em acervos locais distintos. Ao acessar o documento (Informação de Conteúdo) de um PAI, o sistema retorna de preferência (caso ela seja encontrada) a versão original do documento, caso contrário, o sistema retorna a primeira cópia encontrada. No entanto, ao abrir o documento que seja uma cópia, existe um mecanismo para verificar se o original não foi modificado depois que a cópia foi feita. Nesse caso, o sistema exibe em fundo azul o alerta **é original** para informar que o documento é igual ao original.

Exemplo: o documento no PAI dpi.inpe.br/banon/1998/07.28.15.31 armazenado no site mtc-m12.sid.inpe.br é atualmente uma cópia. Ao abrir essa cópia com a URL: <http://mtc-m12.sid.inpe.br/rep/dpi.inpe.br/banon/1998/07.28.15.31>, o alerta em fundo azul **é original** é exibida, indicando que o documento é igual ao original.

5.1.2.1 O Repositório deve ter mecanismos existentes para garantir que quaisquer/múltiplas cópias de Objetos Digitais estejam sincronizadas.

Não há no momento um sistema automático de sincronização das cópias. No entanto, se uma cópia é diferente do original, ao abrir essa cópia o alerta em fundo azul **é original** é omitido.

5.2 Gerenciamento de riscos de segurança

5.2.1 O Repositório deve manter uma análise sistemática dos fatores de riscos de segurança associados aos dados, sistemas, pessoal e instalação física.

Os fatores de riscos dos principais componentes do Repositório são apresentados na Tabela 5.2.

Tabela 5.2 - Fatores de riscos

Tipo	C. prazo	M. prazo	Observação
computadores	30%	10%	hoje, não há computador sobressalente
<i>software</i>	10%	10%	depende de <i>software</i> livre
plataforma <i>URLib</i>	10%	40%	somente usada no INPE
rede	10%	10%	Internet e rede do INPE
incêndio	20%	20%	33 fitas e 3 máq. estão no mesmo prédio
digitalização	50%	30%	falta de recursos
gerenciamento	30%	30%	falta uma auditoria independente
pessoal	10%	50%	falta um administrador de sistema
renovação da equipe	20%	60%	não tem contratação planejada
expertisa da equipe	10%	30%	não tem treinamento operacional planejado
ataques externos	20%	40%	não tem avaliação independente
serviços terceiros	10%	20%	resolvedor <code>urlib.net</code>

Caso os altos fatores de risco se mantiverem, o Repositório poderá entrar em colapso.

Uma solução para baixar significativamente alguns riscos a médio prazo seria a criação de uma Organização (Fundação) sem fins lucrativos reunindo membros interessados em preservar a navegação, isto é, o funcionamento dos vínculos persistentes. Esta Fundação recolheria fundos para pagar as despesas do registro do nome de domínio `urlib.net`, da manutenção de uma servidora resolvidora de URLs e do *software* `URLibService`³ que implementa esse resolvidor. Este *software* seria licenciado para certos usos e dentro de certas condições a serem definidas.

5.2.2 O Repositório deve ter implementado controles para tratar adequadamente cada um dos riscos de segurança definidos.

O INPE tem interesse em repassar sua experiência em construção e manutenção de repositório digital para outras instituições, em particular do MCT. A ampliação do uso da plataforma *URLib* criaria melhores condições para essa plataforma se sustentar.

O INPE tem que identificar um Repositório alternativo (por exemplo, MCT, Arquivo Nacional, ...) onde seria armazenada uma cópia de segurança sincronizada do seus

³Registro INPI: n° 00072253.

acervos locais.

O INPE teria que contratar uma auditoria externa sobre seu Repositório.

O INPE tem que identificar um funcionário da área de Tecnologia da Informação para cuidar da administração técnica do Repositório. Sem ele, o INPE está perdendo oportunidades de preservar sua memória (por exemplo: preservação dos anais em eventos organizados pelo INPE).

O INPE tem que contratar novos funcionários para, recompor e consolidar a atual equipe do SID encarregada do Repositório, pois a atual equipe não tem pessoal sobrando em caso de necessidade de compensar a ausência de um dos seus membros e deve, à meio prazo, perder membros em consequência de aposentadorias.

O INPE deve se articular para organizar cursos de treinamento sobre o uso da plataforma *URLib* para a equipe do SID encarregada do Repositório.

O INPE deverá treinar o novo administrador técnico do Repositório quando este for contratado.

5.2.3 Os funcionários do Repositório devem ter funções, responsabilidades e autorizações delineadas relacionadas à implementação de mudanças no sistema.

No momento, qualquer usuário (cadastrado ou não) pode criar um registro. Todo usuário que criou pelo menos um registro passa a ser automaticamente registrado. Cada registro tem um único responsável tanto de parte dos metadados quanto do texto completo. No momento da criação de um registro, o responsável é o usuário que o criou. Este pode transferir sua responsabilidade de atualização para um outro usuário cadastrado. Dependendo do seu nível de responsabilidade um usuário poderá o não atualizar certos metadados.

No futuro, pode ser necessário, em caso de abuso, restringir o processo de cadastramento ao INPE. Então, os usuários fora do INPE poderão submeter novos documentos somente se forem previamente cadastrados.

Existem no Repositório vários atores, cada um com seu nível de responsabilidade:

- o administrador do sistema tem o poder de trocar o responsável pela atualização de cada registro (inclusive metadatos), alterar o conteúdo de qualquer campo de qualquer registro e remover registros;

- o administrador de ePrint tem o poder de atualizar cada ePrint finalizado pelos seus autores;
- os usuários com poder de finalizar o processo de atualização de um determinado tipo de registro;
- os usuários com poder (ou mandato) para autorizar a publicação de um trabalho (preenchimento do campo `year`) e escolher o nível de restrição de acesso;
- os usuários com poder para preencher o campo `abstract` das T&Ds;
- os usuários com o privilégio de poder ser escolhido para atualizar um determinado registro;
- os usuários com poder para preencher os campos: `ano`, `nextedition` e o `stageofalternatepublication` (inclusive com a opção “publicado”) dos ePrints;
- os usuários com poder para preencher os campos: `notes`, `url` (inclusive com opção para download) e `dissemination` (no caso das T&Ds);
- os usuários com acesso aos computadores autorizados para o preenchimento do campo `notes` das T&Ds publicadas pelo INPE.

5.2.4 O Repositório deve ter acertos escritos adequados para as catástrofes e plano(s) de recuperação, incluindo pelo menos um *backup* externo de todas as informações preservadas, juntamente com uma cópia externa do(s) plano(s) de recuperação.

O Repositório precisaria identificar um Repositório alternativo (por exemplo: MCT ou Arquivo Nacional) onde seria armazenada uma cópia de segurança sincronizada do seus acervos locais (ver Item 3.1.2.1).

REFERÊNCIAS BIBLIOGRÁFICAS

- ACTIVESTATE. **ActiveTcl User Guide**. 2010. Disponível em:
<http://docs.activestate.com/activetcl/8.4/tcl/tcl_contents.htm>.
Acesso em: 23 abr. 2010. 26
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). **NBR 15472:2007**: sistemas espaciais de dados e informações – modelo de referência para sistema aberto de arquivamento de informação (SAAI). Rio de Janeiro, mar. 2007. 95 p. 3
- BANON, G. J. F. **Biblioteca Digital da Memória Técnico-Científica do INPE**. São José dos Campos: INPE, 2005. Disponível em:
<<http://urlib.net/J8LNKAN8PWT2T/QNQFSH>>. Acesso em: 23 abr. 2010. 9, 14, 19, 29, 31
- _____. **Tabela dos formulários de submissão de trabalhos integrando a Memória Técnico-Científica do INPE**. 2006. Depositado no acervo da URLib. Disponível em:
<<http://urlib.net/iconet.com.br/banon/2006/11.22.22.25>>. Acesso em: 23 abr. 2010. 7
- _____. **Breve relatório com dados oriundos da Memória Científica do INPE**. São José dos Campos: INPE, 2009. Disponível em:
<<http://urlib.net/8JMKD3MGP8W/34483PP>>. Acesso em: 24 abr. 2010. 28
- BANON, G. J. F.; BANON, L. C. **Biblioteca digital da Memória Técnico-Científica do INPE**. São José dos Campos: INPE, 2005. Transparências preparadas para o Dia do Bibliotecário - São José dos Campos, 2005. Disponível em:
<<http://urlib.net/dpi.inpe.br/banon/2005/03.15.13.44>>. Acesso em: 23 abr. 2010. 11
- _____. **O que é a URLib?** 2008. Depositado no acervo da URLib. Disponível em:
<<http://urlib.net/iconet.com.br/banon/2001/05.25.16.44>>. Acesso em: 24 abr. 2010. 20
- _____. **Identificador global com base na Internet: Protocolo de comunicação com o resolvedor urlib.net**. 2009. Depositado no acervo da URLib. Disponível em: <<http://urlib.net/NENDTJMTKW/335L8GH>>. Acesso em: 01 maio 2010. 28

BANON, G. J. F.; RIBEIRO, M. L. **Repositório digital do INPE**. São José dos Campos: INPE, 2008. Transparências preparadas para o I Workshop sobre Acesso Livre nas Unidades do MCT - Rio de Janeiro, 2008. Disponível em: <<http://urlib.net/LK47B6W/34CEBGE>>. Acesso em: 26 abr. 2010. 11

BANON, G. J. F.; RIBEIRO, M. L.; BANON, L. C. **Preservação digital da Memória Técnico-Científica do INPE**. São José dos Campos: INPE, 2004a. Disponível em: <<http://urlib.net/dpi.inpe.br/lise/2004/03.02.15.20>>. Acesso em: 23 abr. 2010. 10

_____. _____. São José dos Campos: INPE, 2004b. Transparências preparadas para o II Simpósio Internacional de Biblioteca Digital - Campinas, 2004. Disponível em: <<http://urlib.net/dpi.inpe.br/banon/2004/05.13.19.36>>. Acesso em: 23 abr. 2010. 11

BARBEDO, S. A. D.-D.; RIBEIRO, M. L.; MARCELINO, S. C. Mapeamento dos processos no serviço de informação e documentação do inpe. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTACAO, 23. **Anais...** 2009. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m18@80/2009/07.16.18.41>>. Acesso em: 23 abr. 2010. 7

BRASIL. Constituição da Republica Federativa do Brasil de 1988. **Diário Oficial da União**, Brasília, DF, Secão I, ANO CXXXVI, n. 191-A, 5 out. 1988. Disponível em: <<http://www.jusbrasil.com.br/diarios/3640291/dou-secao-1-05-10-1988-pg-145/pdfView>>. Acesso em: 27 de maio de 2010. 5

_____. Lei nº 8.159, de 8 de janeiro de 1991. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. **Diário Oficial da União**, Brasília, DF, Secão I, p. 455, 9 jan. 1991. Disponível em: <<http://www.jusbrasil.com.br/diarios/940758/dou-secao-1-09-01-1991-pg-3/pdfView>>. Acesso em: 28 de maio de 2010. 5, 6

_____. Decreto nº 4.553, de 27 de dezembro de 2002. Dispõe sobre a salvaguarda de dados, informações, documentos e materiais sigilosos de interesse da segurança da sociedade e do Estado, no âmbito da Administração Pública Federal e dá outras providências. **Diário Oficial da União**, Brasília, DF, Secão I, n. 251, p. 6, 30 dez. 2002. Disponível em: <<http://www.jusbrasil.com.br/diarios/828241/dou-secao-1-30-12-2002-pg-6/pdfView>>. Acesso em: 28 de maio de 2010. 31

_____. Decreto nº 4.915, de 12 de dezembro de 2003. Dispõe sobre o Sistema de Gestão de Documentos de Arquivo - SIGA, da Administração Pública Federal e dá outras providências. **Diário Oficial da União**, Brasília, DF, Secão I, n. 243, p. 2, 15 dez. 2003. Disponível em: <<http://www.jusbrasil.com.br/diarios/801953/dou-secao-1-15-12-2003-pg-2/pdfView>>. Acesso em: 28 de maio de 2010. 5

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS).

CCSDS 650.0-B-1: reference model for an open archival information system (OAIS). [S.l.], 2002. 148 p. 3

_____. **CCSDS 652.0-R-1**: audit and certification of trustworthy digital repositories. [S.l.], 2009. 1

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). Portaria nº 13/2006. Institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos. 2006. Disponível em: <http://www.abmes.org.br/_download/Associados/Legislacao/2006/Portaria/Port_CAPES_13_20060215.doc>. Acesso em: 8 de abril de 2010. 6

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Manual para elaboração, formatação e submissão de teses**. São José dos Campos: INPE, 2010. xv + 97 p. (INPE-13269-MAN/45-V.1). Disponível em: <<http://urlib.net/8JMKD3MGP8W/PGU542>>. Acesso em: 08 abr. 2010. 7

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. SERVIÇO DE INFORMAÇÃO E DOCUMENTAÇÃO (INPE. SID). **Relatório de avaliação das atividades do SID 2005**. São José dos Campos: INPE, 2005. 48 p.

Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m17@80/2008/05.12.17.28>>. Acesso em: 23 abr. 2010. 10

_____. **Relatório de avaliação das atividades do SID 2006**. São José dos Campos: INPE, 2006. 48 p. Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m17@80/2008/05.12.17.29>>. Acesso em: 23 abr. 2010. 10

_____. **Relatório de avaliação das atividades do SID 2007**. São José dos Campos: INPE, 2007. 83 p. Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m17@80/2008/03.31.13.38>>. Acesso em: 23 abr. 2010. 10

_____. **Relatório de avaliação das atividades do SID 2008**. São José dos Campos: INPE, 2008. 102 p. Disponível em:
<<http://urlib.net/8JMKD3MGPBW/352D2AP>>. Acesso em: 23 abr. 2010. 10

_____. **Relatório de avaliação das atividades do SID 2009**. São José dos Campos: INPE, 2009. 72 p. Disponível em:
<<http://urlib.net/8JMKD3MGPBW/37CJS5P>>. Acesso em: 23 abr. 2010. 10

INTERNATIONAL DOI FOUNDATION (IDF). **The DOI Handbook**. 4.4.1. ed. Oxford, Oct. 2006. 173 p. Disponível em: <<http://dx.doi.org/10.1000/186>>. 21

KING, J. C. Long live iso 32000-1 - the pdf standard. **ISO Focus**, v. 6, n. 4, 2009. Disponível em: <http://www.iso.org/iso/p.24__main_focus.pdf>. Acesso em: 24 abr. 2010. 27

LEMOS, V. S. **Folheto Orientação para Pesquisa e Publicação Digital no INPE**. São José dos Campos: INPE, 2009. Disponível em:
<<http://urlib.net/dpi.inpe.br/banon/2009/04.08.18.50>>. Acesso em: 23 abr. 2010. 9

RIBEIRO, M. L. Reflexões sobre o resguardo da Memória Científica do INPE. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 14. (SNBU), 2006, Salvador, BA. **Anais...** Salvador, 2006a. p. 15. Disponível em:
<<http://urlib.net/sid.inpe.br/mtc-m16@80/2006/10.31.16.16>>. Acesso em: 23 abr. 2010. 11

_____. **Reflexões sobre o resguardo da memória científica do INPE**. São José dos Campos: INPE, 2006b. Transparências preparadas para o 14. Seminário Nacional de Bibliotecas Universitárias - Salvador, 2006. Disponível em:
<<http://urlib.net/sid.inpe.br/ePrint@1905/2006/10.31.15.52>>. Acesso em: 23 abr. 2010. 11

_____. **Aspectos da Memória Científica do Instituto Nacional de Pesquisas Espaciais (INPE)**. São José dos Campos: INPE, 2007. 57 p. INPE-11278-NTC/364. Disponível em:
<<http://urlib.net/sid.inpe.br/mtc-m17@80/2007/07.10.18.17>>. Acesso em: 23 abr. 2010. 7, 11

RIBEIRO, M. L.; MARCELINO, S. C.; BARBEDO, S. A. D.-D. Estratégias de aprimoramento e ampliação da participação do SID nas atividades de P,D&I do

INPE. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTACAO, 23. **Anais...** 2009. Disponível em: <<http://urlib.net/8JMKD3MGP8W/35LTCTE>>. Acesso em: 08 abr. 2010. 6

RIVEST, R. **The MD5 Message-Digest Algorithm**. 1992. RFC 1321. Disponível em: <<http://tools.ietf.org/html/rfc1321>>. Acesso em: 23 abr. 2010. 11