



Recomendações para a produção de
PLANOS DE PRESERVAÇÃO DIGITAL

Lisboa, 25 de Março de 2008

Ficha técnica MIP

Título: Recomendações para a produção de Planos de Preservação Digital

Autor: Francisco Barbedo – coordenador

Autor: Glória Santos

Autor: Luís Corujo

Autor: Mário Sant’Ana

Id.

Classificação: 260.01.02/DSIPE 25 PPD

Descritores: Preservação Digital, Planos de Preservação Digital, Documentos de Arquivo Electrónico,

Colaborador: Câmara Municipal de Lisboa

Colaborador: INFARMED I.P

Colaborador: Instituto de Segurança Social, I.P.

Data/Hora: 2008-03-25

Formato de dados: Texto, PDF

Estatuto de utilização: Acesso público

Relação: versão – 1.0

© DGARQ, 2008

0	Introdução	4
0.1	A necessidade de um Plano de Preservação Digital (PPD)	4
0.2	Estado da situação.....	5
0.3	Preservação digital	5
0.4	O Plano de Preservação Digital.....	6
0.5	Objectivos e destinatários deste documento	7
0.6	Estrutura e conteúdo do documento	7
1	Elaboração de um Plano de Preservação Digital.....	10
1.1	Pré-requisitos	10
1.2	Metodologia para a elaboração do PPD	13
1.2.1	Caracterização dos sistemas de informação e identificação dos requisitos dos documentos de arquivo electrónicos.....	15
1.2.2	Planeamento da estratégia de preservação	17
2	Implementação do Plano de Preservação Digital	23
2.1	Teste da solução proposta.....	23
2.2	Produção de cópias de segurança.....	23
2.3	Aplicação da Estratégia de Preservação.....	23
2.4	Medidas de monitorização e auditorias internas.....	24
2.5	Afectação de recursos.....	24
3	Ameaças	25
3.1	Obsolescência de formatos e software.....	25
3.2	Obsolescência do hardware e suportes.....	26
3.3	Elementos que afectam os suportes	27
4	Caracterização de formatos e aplicações.....	28
4.1	Formatos	28
4.1.1	Identificação de formatos alternativos normalizados.....	32
4.2	Aplicações.....	33
5	Sistemas de armazenamento.....	35
5.1	Atributos desejáveis.....	35
5.2	Arquitecturas Possíveis.....	36
5.2.1	Critérios para avaliação de soluções de armazenamento	38
5.3	Cópias de segurança.....	39
5.3.1	Gestão do repositório de dados.....	41
6	Metainformação	43
7	Estratégias de Preservação mais comuns.....	48
7.1	Preservação de tecnologia.....	48
7.2	Emulação.....	48
7.3	Monitorização de suportes e formatos	49
7.4	Encapsulamento.....	49
7.5	Transposição de Formatos e Suportes (Migração e transferência de suporte)	49
8	Conclusão.....	52
	Anexos.....	53
A	Lista de requisitos para a prática de preservação digital.....	54
B	Questionário	58
C	Exemplo de um ficheiro PREMIS contendo MI técnica NISO Z39.87.....	72
D	Recursos Aconselhados.....	75
E	Glossário.....	83

0 Introdução

Conceito – Ao longo deste documento serão utilizados os termos “documento de arquivo electrónico” (DAE) e “objecto digital” (OD). Estes termos representam duas realidades relacionadas. O DAE é a entidade lógica que possui conteúdo, contexto e estrutura de forma a ter um significado específico. O OD é a componente física do DAE, normalmente equivalente a ficheiros. Um DAE é sempre composto por, pelo menos, um OD, podendo ser composto por vários OD. Por exemplo, um relatório em formato <.doc> pode ser composto por um ficheiro de texto, um ficheiro de imagem e um ficheiro tabular, estando todos integrados para a representação completa do documento. A mesma situação se aplica a documentos estruturados como bases de dados (simples ou de suporte a sistemas de informação complexos) ou ainda documentos multimédia.

Em termos práticos apenas se consegue preservar um documento de arquivo electrónico se forem realizadas acções sobre as suas componentes, ou seja, os OD.

0.1 A necessidade de um Plano de Preservação Digital (PPD)

As tecnologias de informação são, actualmente, o principal suporte para a produção e armazenamento de informação. As actividades organizacionais são, em graus de menor ou maior intensidade, mediadas tecnologicamente. Informação de diversos tipos, seja ela de apoio à decisão, operacional, probatória, ou qualquer outra, é produzida e mantida electronicamente estando portanto dependente de um sistema intermediário composto pelo *software* e *hardware* que contribuíram para a sua criação e, naturalmente, indispensáveis para recuperar e utilizar essa informação. A rápida taxa de obsolescência tecnológica, inerente à indústria informática, levanta problemas críticos de preservação de informação operacionalmente indispensável à organização. No entanto, a tarefa de preservação digital está longe de ser simples ou isenta de custos. A experiência constatada no terreno, pela DGARQ, identificou cenários prevaletentes em que a informação é produzida com carácter de utilização imediata sem serem consideradas necessidades operacionais sobre essa mesma informação a médio ou longo prazo. O resultado desta atitude resulta na perda, muitas vezes irreversível, de informação com consequências mais ou menos dramáticas para a instituição que a perdeu.

A produção do presente documento visa dar uma linha de orientação para as organizações que produzam e dependam, em maior ou menor percentagem, de informação criada e mantida electronicamente. Essa orientação servirá para tomar medidas que possam garantir as condições materiais mínimas para preservar informação digital, durante o período pelo qual a organização dela necessite.

Não é considerada, neste documento, a circunstância em que um produtor deseje criar um arquivo digital, orientado para a preservação de objectos digitais (OD) de conservação permanente e acessível a utilizadores externos à organização. Neste caso muitos outros requisitos serão necessários bem como recursos humanos e materiais indispensáveis para gerir esse tipo de função. Neste caso, a metodologia aconselhada é a de proceder a transferências desses objectos para um arquivo digital especializado nessa função. Actualmente a DGARQ encontra-se a desenvolver um sistema desta natureza – o RODA (Repositório de Objectos Digitais Autênticos) - que se prevê estar apto a entrar em exploração a partir de Janeiro de 2009.

0.2 Estado da situação

As novas tecnologias digitais permitem novas possibilidades que, trazendo consigo benefícios e problemas, tornam os objectos delas resultantes complicados de gerir, particularmente se considerarmos períodos prolongados de tempo. Alguns exemplos:

- O mesmo documento ou formato de documento pode ser representado em diferentes suportes para diferentes finalidades, obrigando à distinção entre formato e suporte;
- Diferentes suportes podem ser utilizados em diferentes fases de criação/captura, armazenamento, acesso, distribuição e uso do documento, podendo os documentos ser codificados de maneiras completamente diferentes em cada uma destas fases;
- A separação entre formato e suporte e o tratamento separado de cada uma das fases do documento, permite uma estruturação mais livre dos documentos sem constrangimentos inerentes ao suporte papel;
- Os conteúdos originalmente digitais podem representar material de leitura, som, vídeo, fotografias, artes gráficas, radiografias, jogos, registos, dados e outros;
- A qualidade da cópia digital permanece, em tudo, idêntica, independentemente do número de cópias sucessivas realizadas;
- A transposição para papel não é aplicável por ocasionar perda de informação significativa que comprometeria decisivamente a sua utilidade e inteligibilidade¹.

O mercado informático não oferece soluções orientadas para a preservação digital, embora haja bastantes aplicações vocacionadas para gestão de documentos electrónicos. Tais ferramentas permitem apenas gerir eficazmente funcionalidades de gestão de documentos, tais como produção, circulação ou armazenamento, não sendo previstas capacidades conducentes à preservação dos documentos ao longo de períodos de tempo prolongados.

No entanto as organizações necessitam, sob o ponto de vista operacional, da informação produzida digitalmente durante períodos mais ou menos prolongados, consoante a natureza da informação e a perenidade do seu valor administrativo, legal ou de qualquer outra natureza.

Neste cenário a preservação que, em ambiente papel, é normalmente confiada a instituições especializadas, torna-se um problema premente das organizações produtoras. É fundamental que estas sejam capazes de gerir a sua informação digital de forma a mantê-la utilizável e garantir os propósitos operacionais da organização.

O presente documento pretende contribuir para esse propósito.

0.3 Preservação digital

A preservação digital, considerada no contexto deste documento, assume três vectores:

1. Conjunto de actividades desenvolvidas com o fim de aumentar a vida útil dos DAE, salvaguardando a utilização operacional e protegendo-os das falhas de suportes, perda física e obsolescência tecnológica;
2. Conjunto de actividades que promovem a acessibilidade continuada aos conteúdos;

¹ Exemplos típicos desta situação são por exemplo, bases de dados ou documentos multimédia

3. Conjunto de actividades que assistem na preservação do conteúdo intelectual, forma, estilo, aparência e funcionalidade.

Assim sendo, para maximizar as perspectivas da preservação de OD a longo prazo, devem ser aplicadas medidas de preservação logo que possível, ou seja, preferencialmente enquanto os OD são/estão ainda acessíveis e ainda na fase de produção.

Quanto mais rápida a instituição integrar a preservação digital no seu plano de acção (e orçamento!), definir e implementar soluções apropriadas, maior a probabilidade de sucesso na preservação dos OD.

Pretende-se pois que a organização possua um documento estratégico que determine, à partida, quais os procedimentos a realizar para evitar obsolescência tecnológica e a consequente probabilidade de perder informação.

Os organismos são encorajados a serem proactivos na prossecução das suas estratégias de preservação digital e na definição e implementação dessas medidas, antes que a sua informação digital se torne obsoleta e consequentemente inutilizável. Na realidade as medidas de preservação são normalmente tomadas reactivamente – quando o são de todo – como resposta a situações emergentes de constatação de perda de informação necessária, quando tais medidas deveriam ser planeadas e implementadas como parte de uma solução de preservação digital a longo prazo.

0.4 O Plano de Preservação Digital

Um Plano de Preservação Digital:

1. É um documento estratégico que
2. contém políticas e procedimentos orientados
3. para a constituição de uma estrutura técnica e organizacional que
4. permita preservar de forma continuada documentos de arquivo electrónicos (DAE) através de
5. acções realizadas sobre os objectos digitais (OD) que os compõem.

Destina-se a garantir que os DAE sejam conservados de forma legível e acessível, mantendo simultaneamente as suas propriedades de autenticidade e integridade durante tanto tempo quanto a organização deles necessitar.

O Plano de Preservação Digital permitirá identificar quais as funcionalidades que devem ser implementadas e a forma de as implementar, para manter a integridade e usabilidade dos documentos de arquivo electrónicos ao longo do tempo.

O desenvolvimento de um Plano de Preservação Digital e a selecção das estratégias apropriadas, deve ser o resultado de um esforço de colaboração entre as unidades orgânicas referentes ao arquivo (gestão documental) e à informática (tecnologias da informação), com a participação de todas as unidades orgânicas afectadas pelo processo ou que produzam informação electrónica.

Trata-se de um processo directamente ligado a práticas de gestão documental, como adiante será explicado.

AVISO – A preservação digital implica custos significativos que devem ser analisados e previstos pela Organização através do reforço orçamental, considerado adequado à situação diagnosticada aquando da elaboração do Plano de Preservação Digital.

0.5 Objectivos e destinatários deste documento

Estas recomendações para a elaboração de planos de preservação digital destinam-se a ser utilizadas por qualquer organismo que:

- Produza documentação/informação em formato electrónico (especificamente documentação de cariz arquivístico);
- Mantenha em formato electrónico essa informação;
- Tenha reconhecidas necessidades de utilização operacional dessa informação que excedam sete anos de longevidade;
- Pretenda produzir um documento de planeamento que lhe permita determinar procedimentos de forma a reunir as condições materiais necessárias a preservar de facto essa informação.

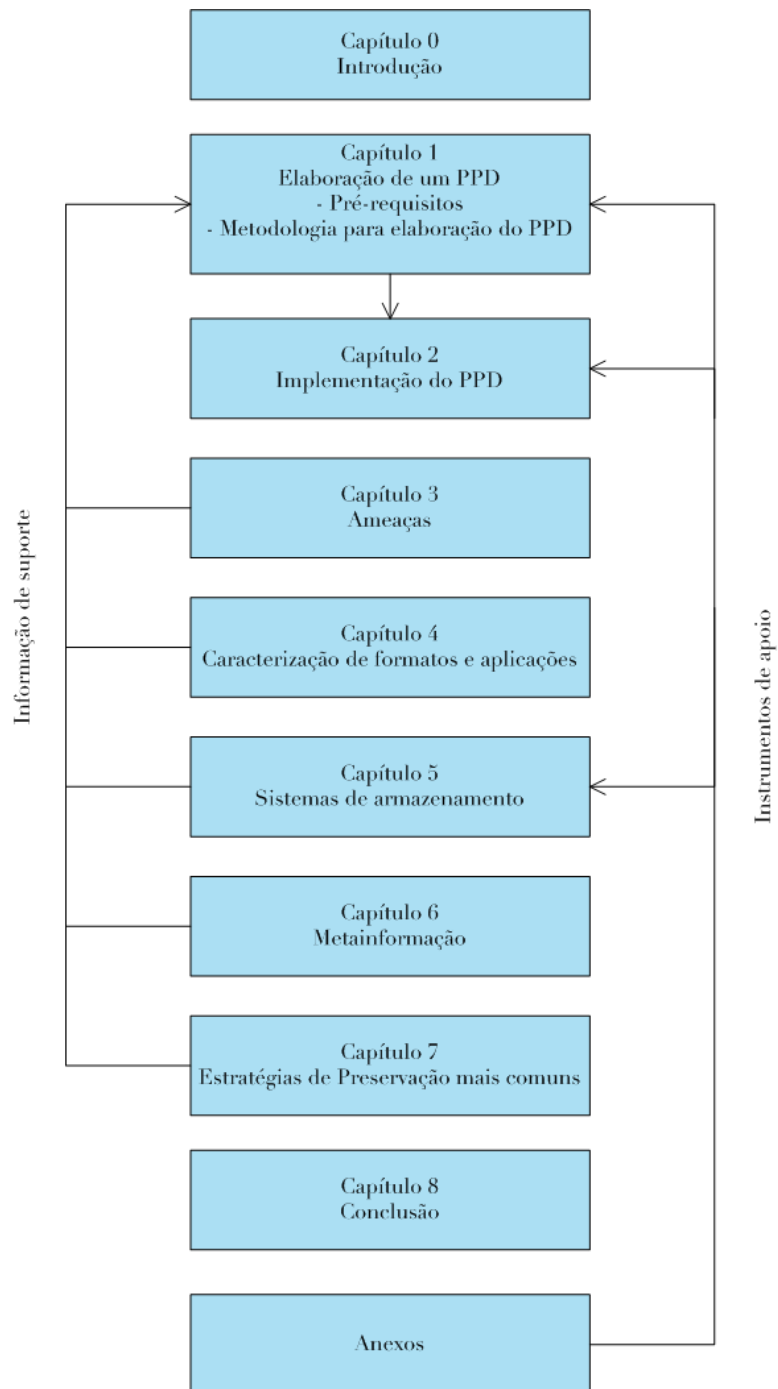
0.6 Estrutura e conteúdo do documento

Este documento é composto por nove capítulos, conforme o esquema que abaixo se apresenta. Salienta-se que, no início de cada capítulo, foi introduzido um resumo do conteúdo do mesmo.

A Introdução (Cap. 0) e a Conclusão (Cap. 8) destinam-se a contextualizar e evidenciar as problemáticas relativas a elaboração de Planos de Preservação Digital.

Entre os Cap. 1 e 2 existe uma relação cronológica sequencial, enquanto que os Cap. 3 a 7 contêm informação de suporte ao Cap. 1.

Finalmente, os Anexos contêm instrumentos de apoio aos Cap. 1, 2 e 4.



(página intencionalmente deixada em branco)

1 Elaboração de um Plano de Preservação Digital

Este capítulo encontra-se dividido em dois grandes pontos:

O primeiro destina-se a apresentar e explicar os pré-requisitos considerados como basilares para uma organização que pretenda proceder à elaboração de um PPD, nomeadamente, o plano de classificação e a tabela de selecção.

O segundo desenvolve as etapas e actividades da metodologia para a elaboração do PPD:

- A avaliação do desempenho do sistema de arquivo e dos documentos electrónicos que o integram, através do levantamento e análise dessa mesma informação. Com vista a agilizar a recolha de dados, são fornecidos em anexo os instrumentos necessários.
- Planeamento da estratégia de preservação, a qual tem como finalidade a escolha da melhor solução de preservação para o organismo e a consequente produção do PPD. Referem-se factores a considerar ao nível da definição de estratégias de preservação e formatos de preservação, bem como da escolha de aplicações informáticas, soluções de armazenamento e metainformação.

1.1 Pré-requisitos

Consideram-se como instrumentos estruturantes de base à elaboração de um PPD:

- O plano de classificação e
- A tabela de selecção.

A classificação destina-se a organizar a informação de acordo com critérios pré-determinados (funcionais, orgânicos, etc.) de forma a compartimentá-la para sobre ela realizar de forma eficiente algum tipo de trabalho. A existência de um sistema de classificação é essencial para a organização da informação.

O **plano de classificação** (um tipo de sistema classificativo) permite otimizar a aplicação de um PPD, na medida em que se apresenta como uma ferramenta de gestão de documentos de arquivo que visa, nomeadamente:

Plano de classificação:

Elemento estruturante do sistema de arquivo de qualquer organismo, na medida em que se apresenta como um conjunto de regras claramente definidas que promovem a organização dos documentos de arquivo.

Fonte: Manual para a gestão de documentos, p. 2-5.

- Perceber as relações entre os documentos;
- Assegurar a continuidade da nomenclatura dos documentos ao longo do tempo;
- Facilitar a recuperação dos documentos relativos a uma função ou a uma actividade;
- Definir os níveis de protecção e de acesso a diferentes conjuntos documentais.

O plano de classificação deve ser funcional, reflectir as actividades da organização da qual deriva e resultar da análise das suas funções/actividades. Constitui a base

Para informações adicionais sobre sistemas de classificação, consultar:

- SIADÉ caderno 2: Recomendações para a Gestão de Documentos de Arquivo Electrónicos (MOREQ)
- Manual para gestão de documentos
- Norma Portuguesa NP4438 – partes 1 e 2

crítica para a organização e recuperação dos documentos.

O sistema de **classificação funcional** tem por base as funções atribuídas a uma instituição para dar cumprimento aos fins para os quais foi criada, às actividades levadas a cabo no desempenho de cada função e às acções que as concretizam. Assim, a classificação deverá reflectir as funções/actividades da organização.

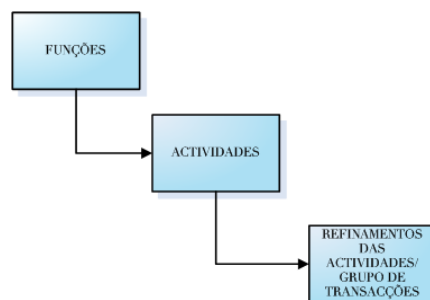
Considerando que o sistema de classificação funcional não está dependente das alterações orgânicas que ocorrem nas instituições, consegue-se uma vantajosa estabilidade organizativa dos documentos produzidos.

De acordo com a NP 4438-2², os sistemas de classificação permitem:

- Organizar, descrever e relacionar entre si os seus documentos;
- Relacionar e partilhar documentos de áreas diferentes, dentro da organização ou com entidades externas;
- Melhorar o acesso, a recuperação, a utilização e a disseminação dos seus documentos.

A estrutura de um sistema de classificação funcional é hierárquica, sendo que:

- O primeiro nível representa as funções,
- O segundo as actividades que constituem a função
- O terceiro e subsequentes níveis são refinamentos das actividades ou grupo de transacções que decorrem no âmbito de cada actividade.



A **tabela de selecção** é uma ferramenta crucial na elaboração de um PPD, porque permite determinar os adequados prazos de retenção e acções de eliminação ou de transferência de documentos. Desta forma torna-se possível circunscrever os DAE a serem abrangidos pelo PPD, i.e., aqueles cujo prazo de conservação, segundo a referida tabela, seja superior a sete anos.

Tabela de selecção:

Regista o resultado do processo de avaliação documental e fixa, para os conjuntos de documentos identificados (séries), os respectivos prazos de conservação administrativa e destino final.

Salienta-se que qualquer tabela de selecção que careça de aprovação da DGARQ, ao abrigo do Decreto-Lei 447/88, de 10 de Dezembro, e que inclua documentos de arquivo electrónicos implica a elaboração de um PPD desde que se verifiquem cumulativamente as seguintes circunstâncias:

1. O prazo de conservação previsto seja superior a sete anos,.

² INSTITUTO PORTUGUÊS DA QUALIDADE - NP 4438-2, *Informação e documentação. Gestão de documentos de arquivo. Parte 2: Recomendações de aplicação*. Lisboa: Instituto Português da Qualidade, 2005, p.17

2. O destino final seja de conservação permanente, ou conservação permanente parcial

Considerando que as actividades relativas à preservação de OD são dispendiosas, pois implicam a mobilização de recursos materiais e humanos especializados, assim como a afectação de tempo para as concretizar, a rigorosa aplicação da tabela de selecção assume óbvias vantagens organizativas e orçamentais.

A criticidade de um documento sob o ponto de vista probatório e a sua longevidade, (i.e., o tempo de vida operacional considerado como necessário pela organização para a manutenção do documento no seu sistema de arquivo), são variáveis independentes que não se alteram de forma directamente proporcional.

Exemplo:

Uma factura tem um elevado nível de criticidade ou relevância operacional que dura apenas um curto período de tempo.

Um contrato, no entanto, pode ter um grau de criticidade probatória menos elevado mas bastante mais prolongado no tempo (pelo menos durante o período de vigência da transacção veiculada no contrato).

Em termos de preservação digital este aspecto adquire especial relevância. De facto se tiver sido avaliada a informação produzida na organização sob o ponto de vista de relevância operacional e longevidade necessária, podemos facilmente determinar sobre que classes de informação ou sobre que sistemas de informação deveremos prever acções de preservação digital.

Exemplo 1:

Uma base de dados contém informação sobre execução orçamental. Essa informação de acordo com a Lei tem um prazo de conservação que atinge os dez anos, podendo a partir daí ser eliminada. Neste caso não se justifica investir tempo e recursos para preservar essa informação pois de qualquer forma poderá ser eliminada após 10 anos de vigência.

Exemplo 2

Um sistema de *web conference* através do qual se processam reuniões à distância envolvendo diversos serviços deslocalizados de uma organização. Neste contexto as gravações realizadas equivalem às **actas** e por esse motivo deverão ser objecto de conservação permanente. Neste caso é necessário prever medidas de preservação digital que permitam a manutenção desta informação.

É muito pouco provável que, no que toca a documentos electrónicos, haja uma longevidade única, isto é, um único e idêntico período temporal durante o qual é útil guardar a informação. Assim, apenas interessa planear e empreender acções de preservação sobre as classes ou sistemas de informação cuja necessidade de utilização operacional exceda um determinado período de tempo, a partir do qual os formatos em que a informação foi produzida se começam a tornar obsoletos. Chamaremos a essa “barreira” temporal o **horizonte de obsolescência** de um determinado OD.

O horizonte de obsolescência é estimado em sete anos, pelo que toda a informação digital cujo prazo de vida operacional não exceda este período de tempo não deverá ser abrangida pelo PPD.







Note-se que este período não é rígido reportando-se ao prazo de retrocompatibilidade normalmente assegurado pelas empresas vendedoras, pelo que pode ser alterado consoante a realidade específica de cada organização. Por exemplo, os sistemas de informação que suportam a actividade contabilística contêm informação cujo prazo de vida útil esperado é de 10 anos pelo que, apesar de ultrapassar o prazo sugerido, pode não ser necessário considerar estes sistemas para serem objecto de acções de preservação.

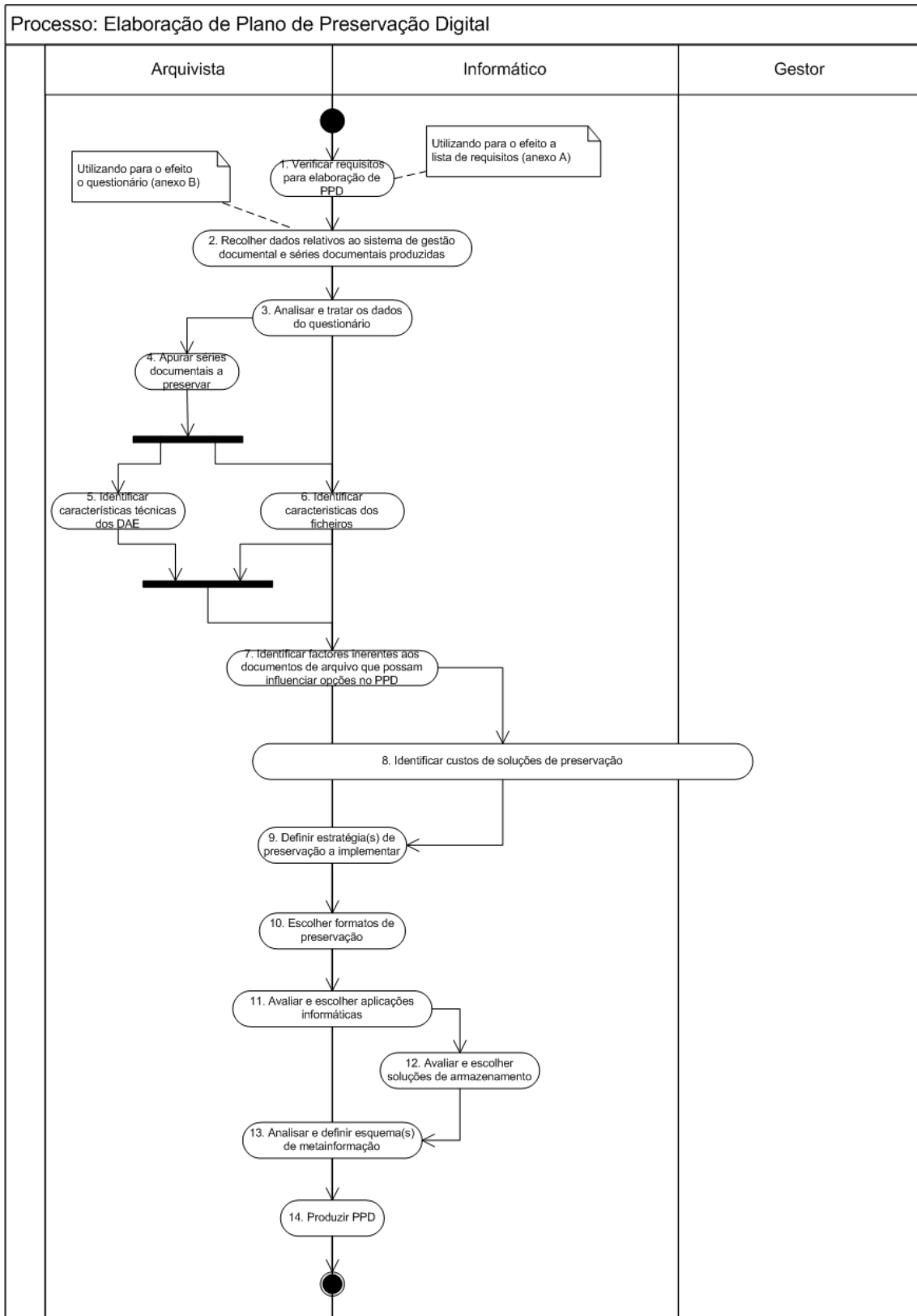
Salienta-se que os dois requisitos acima descritos são parte integrante da “Lista de requisitos para a prática de preservação digital” (Anexo A).

1.2 Metodologia para a elaboração do PPD

Apresenta-se, de seguida, a descrição gráfica do processo de elaboração de um PPD, sob a forma de diagrama de actividades, contemplando os actores intervenientes e todas as actividades que o compõem. De referir que a sequência apresentada para as actividades 9 a 13 é apenas uma sugestão, podendo a sua ordem ser alterada.

Legenda:

#	Símbolo	Descrição
1		Actividade.
2		Fluxo ou sentido do processo (diagrama de actividades).
3		Nota: indica observações gerais que completam o sentido de elementos.
4		Início do processo.
5		Fim do processo (pode haver mais que um).
6		Transição (união): indica a junção de múltiplas actividades.



1.2.1 Caracterização dos sistemas de informação e identificação dos requisitos dos documentos de arquivo electrónicos

Nesta etapa pretende-se caracterizar e avaliar o desempenho do sistema de arquivo e dos documentos electrónicos que o integram.

Documentos de Arquivo Electrónicos:

São documentos de arquivo produzidos, transmitidos e conservada em forma electrónica.

Os sistemas de arquivo produzem ou armazenam informação, transaccional ou analítica, directamente relacionada com as actividades desempenhadas pela organização. Neste sentido o seu desempenho terá impacto nos processos a utilizar para preservar essa informação.

Um sistema de informação adquire e gere informação(ões) necessária(s) para:

- A condução das actividades da organização;
- A percepção das necessidades e expectativas dos clientes internos e externos;
- Solidificar posições num ambiente competitivo;
- Apoiar as actividades dos actores organizacionais;
- Adaptar a organização à envolvente (diminuir a incerteza e a complexidade).

O desenvolvimento de um PPD implica a análise da informação de arquivo electrónica, produzida pela organização, e que obedece ao seguinte conjunto de propriedades:

- Ser produzida na rede organizacional;
- Ser considerada como propriedade institucional;
- Ter um prazo de conservação superior a sete anos, de acordo com a respectiva tabela de selecção (horizonte de obsolescência);
- Possuir reconhecida importância (vital) de acordo com tabela de selecção, análise de risco e criticidade de sistemas para o organismo.

Isto significa que nem todos os documentos electrónicos produzidos na organização deverão ser objecto de análise para a elaboração do PPD. Na organização existem, certamente, centenas de documentos produzidos a título individual, como documentos pessoais de trabalho, que não foram comunicados e que nessa qualidade não têm representatividade institucional. Da mesma forma há, muitas vezes, pequenas aplicações desenvolvidas a título pessoal pelos colaboradores para os auxiliar nas suas actividades e que podem não ter expressão institucional, ou seja, não constituírem para todos os efeitos documentos de arquivo.

O conhecimento do sistema é uma fase essencial no processo, pois permite aferir a situação real no que respeita à existência de condições viáveis para assegurar a preservação dos DAE, registando os pontos fortes e fracos da organização no que respeita a preservação digital e indicando quais as áreas a desenvolver para implementar o PPD.

Para tal deverá ser realizado o levantamento e análise dessa mesma informação, tendo por base o questionário que se apresenta no **anexo B**, o qual contempla duas vertentes:

A) Análise e caracterização do sistema de gestão documental e suas funcionalidades. Nesta fase pretende-se apreender as características do sistema em vigor na organização relativamente a um conjunto de funcionalidades relevantes para a organização e para a segurança da informação:

- Controlo e segurança;
- Armazenamento;

- Funções administrativas;
- Encriptação;
- Desempenho e escalabilidade;
- Contratação de serviços externos;
- Conservação a longo prazo;
- Metainformação.

B) Análise das séries documentais ou actividades suportadas pelas mesmas (nos seguintes casos: inexistência de Plano de Classificação, de Tabela de Selecção ou qualquer outra identificação das séries documentais). Trata-se de particularizar dentro do sistema, caracterizado na parte anterior, as séries documentais existentes em que sejam produzidos documentos de arquivo electrónicos e avaliar quais os elementos do sistema que se lhes aplicam.

Conceito:

É necessário, ainda, explicar o que se entende, neste contexto, por série documental e qual a sua relação com OD.

Uma série documental (independentemente do suporte) é um conjunto de documentos que obedecem a idêntica competência funcional e tipológica. Neste contexto pode tratar-se de um:

- Sistema de informação que suporte uma determinada actividade;
- Conjunto de sistemas de informação que suporte a mesma função ou actividade, ou ainda
- Conjunto de documentos electrónicos não estruturados (por exemplo documentos de texto e mensagens de correio electrónico que façam parte de um processo administrativo).

Alguns conceitos arquivísticos que nos são habituais e raramente suscitam dúvidas podem tornar-se confusos quando passamos do papel para o digital.

Por exemplo, um sistema de bases de dados transaccional produz informação de arquivo, na medida em que os dados contidos estão vinculados a transacções específicas. Qual a identificação arquivística que corresponde a esta entidade informacional? Uma série? Um documento composto? Ditó de outro modo, podemos interrogar-nos sobre a equivalência de um sistema de informação relativamente a entidades conceptuais arquivísticas tradicionais.

No entanto pode-se apontar semelhanças nos dois universos. Em digital uma base de dados corresponde ao nível de séries quando se refere a dados de carácter organizacional, funcional ou administrativo. Os ficheiros que a constituem são descritos ao nível do documento que pode ser composto ou simples.

Tomando como exemplo o SRH (Sistema de gestão de Recursos Humanos) teríamos:

Série: Sistema de gestão de Recursos Humanos

- ⇒ Documento composto1: act_admg.sql
- ⇒ Documento composto 2: pr-m002-02288.txt
- ⇒ Etc.

Neste caso o sistema de informação como um todo consiste na série documental. Este sistema é composto fisicamente por unidades físicas – ficheiros que equivalem individual ou de forma agregada a documentos simples ou compostos. (No caso apresentado, cada ficheiro equivale a um documento)

Repare-se que um ficheiro **pode não equivaler a** um documento. A cardinalidade pode variar da seguinte forma:

1 (documento) → 1..n (ficheiros)

Para o efeito de preservação digital é essencial identificar estes dois níveis de existência, embora apenas nos casos em que se justifique a preocupação em preservar estes objectos digitais (ver p.12). A relação entre estes dois níveis apresenta uma cardinalidade 1:n com uma associação do tipo agregação. Ou seja, as partes (documentos/ficheiros) fazem parte do todo (série/base de dados) e aquelas podem existir mesmo se o elemento agregador desaparecer. Com efeito os ficheiros podem ser reutilizados noutra base de dados.

Salienta-se ainda que para efeitos de preservação a unidade atómica é o ficheiro mesmo que este isoladamente não corresponda a uma unidade de descrição “documento”.

A observação de diversos sistemas organizacionais permite afirmar que nem sempre todas as características de um sistema de informação se aplicam a todos os OD produzidos. Efectivamente, a decisão desta escolha depende, muitas vezes, de factores como:

- A criticidade atribuída à informação;
- A irrelevância, por vezes injustificada, atribuída a outro tipo de DAE (por exemplo sítios *web*);
- O desconhecimento de que determinados OD são de facto produzidos na organização ou onde estes, efectivamente, se encontram.

A parte B do inquérito incide sobre as seguintes questões:

- Identificação do produtor;
- Referenciação da série;
- Controlo e segurança;
- Avaliação;
- Assinaturas digitais;

A análise e tratamento dos dados do questionário permitirá obter uma descrição das características do sistema, bem como a informação necessária sobre os documentos de arquivo electrónicos que se pretendem preservar no âmbito do PPD.

1.2.2 Planeamento da estratégia de preservação

O planeamento da estratégia de preservação tem como finalidade a escolha da melhor solução de preservação para o organismo e a consequente produção do documento “Plano de Preservação Digital”, por forma a assegurar a integridade e funcionalidade continuadas dos

documentos de arquivo electrónicos e dos objectos digitais constituintes. Esse planeamento deve ser feito com base na informação reunida anteriormente e comporta:

- Os procedimentos que devem ser realizados pelo organismo sobre as diversas classes de informação identificadas (ver Cap. 7);
- A definição de formatos para preservação para cada uma das classes de informação caracterizadas (ver Cap. 4.1);
- A seriação fundamentada de soluções de software alternativas para um determinado formato (ver Cap. 4.2);
- A selecção e utilização de soluções de armazenamento adequadas a preservação digital (ver Cap. 5);
- Os esquemas de metainformação que devem ser utilizados (ver Cap. 6).

No planeamento da estratégia de preservação as instituições devem ter em conta, os seguintes factores:

Os processos de preservação requerem a utilização de normas reconhecidas, nomeadamente de normas para gestão de documentos (Ex: NP 4438).

- Custos de implementação e custos de manutenção para as medidas de preservação;
- Complexidade técnica da abordagem escolhida e a capacidade da organização para suportar essa abordagem ao longo do tempo (técnica e financeiramente);
- Compatibilidade com o *software* e *hardware* existentes;
- Impacto nos processos de negócio (ex.: se a abordagem escolhida obriga a alterações das práticas de trabalho);
- Eficácia e robustez da abordagem na protecção da integridade, acessibilidade e funcionalidade dos documentos de arquivo electrónicos ao longo do tempo.

Naturalmente a intensidade das medidas de preservação a empreender devem estar directamente relacionadas com o grau de criticidade atribuída aos objectos digitais, assim como a análise de risco que deve ser efectuada.

1.2.2.1 Definição de estratégias de preservação

A definição da estratégia de preservação implica a identificação, análise, caracterização e adequação de procedimentos para a salvaguarda de informação digital, de acordo com a informação reunida e as decisões tomadas anteriormente. Estas vão definir a especificidade da estratégia (ou conjunto de estratégias) de preservação que o organismo irá assumir e materializar no PPD.

Os elementos a identificar e recolher para a definição da estratégia de preservação são os seguintes:

- As características dos ficheiros (que descrevem a experiência visual e contextual dos utilizadores ao lidar com os OD) ao nível de:

- **Forma (aparência);**
- **Conteúdo;**
- **Estrutura;**
- **Comportamento.**

Ficheiro de base de dados

Forma: dependente do formulário ou relatório;

Conteúdo: dados colocados via formulário;

Estrutura: tabular;

Comportamento: actualização de data e hora (caso esteja assim definido nas características).

Deve-se, ainda, ter em conta a discriminação das características específicas que um determinado formato possa ter. Por exemplo a profundidade de cor, resolução de imagem, tipos de interactividade, possibilidade de utilização de macros, e metainformação embebida;

- As características técnicas dos documentos de arquivo electrónicos (que incluem o contexto, o suporte de armazenamento, relações existentes com outros documentos de arquivo, e a metainformação que lhes está associada ou é atribuída pela organização);

Contexto: Justificação de faltas;

Suporte de armazenamento: servidor;

Relações existentes com outros documentos de arquivo: inclui informação contida na base de dados do telepono; faz parte do processo individual do funcionário.

Metainformação: Assunto, Data, Colaborador, Data da Falta, Justificação, Despacho.

- Os factores inerentes aos documentos de arquivo electrónicos que irão influenciar opções a incluir no PPD por exemplo:

- A usabilidade;
- A complexidade;
- A escalabilidade;

Usabilidade: significa que um documento é utilizável ou seja, que pode ser localizado, recuperado, apresentado e interpretado.

Escalabilidade: característica de um sistema ou equipamento que pode crescer em escala, isto é, que possibilita incrementos de capacidade ou funcionalidades acompanhando as necessidades dos utilizadores (aumento da capacidade de armazenamento do sistema através da substituição ou instalação de novos discos rígidos; versões mais actuais do formato PDF que permitem a inclusão de informação de dados CAD).

Complexidade: significa que um documento integra diversas componentes digitais (por exemplo as páginas *Web*).

- Os custos, tanto com recursos humanos como materiais, têm uma importância elevada na escolha das soluções de preservação. É importante, nesta fase, empreender uma análise custo/benefício que compare os diversos cenários de preservação digital adequados à informação que é identificada e avaliada em função das seguintes variáveis em parte já referidas neste documento:

- Criticidade dos OD, i.e., a sua importância para o desempenho das actividades da organização;
- Longevidade operacional, ou seja, qual o tempo de vida operacional atribuído ao objecto digital;
- Risco aceitável pela organização, ou seja, a avaliação pela organização do grau de risco aceitável de perda ou obsolescência dos OD;
- Nível de formalização dos DAE produzidos (ignorar DAE produzidos a título individual, por ex.).

Tal análise deve ter em conta os custos correspondentes, de forma a obter uma solução valorizada, de acordo com as variáveis referidas. Aconselha-se ainda a criação e adopção de regras e procedimentos, de forma a ajudar a normalizar a aplicação das técnicas escolhidas por todos os sistemas da instituição.

Uma abordagem mais específica das estratégias encontra-se no Cap. 7.

1.2.2.2 Definição de formatos de preservação

A escolha do formato de preservação deve ser feita antes da criação do OD que se pretende preservar evitando-se, deste modo, custos adicionais na sua conversão para o formato final. A utilização de

Pretende-se que um dos resultados do PPD seja a produção de uma tabela que relacione a cada um dos formatos utilizado na Organização, um ou mais formatos de preservação definidos para o efeito.

formatos normalizados e bem documentados de aplicações com especificações abertas permitem garantir uma continuidade de utilização mais estável e prolongada, quando comparados com os formatos e as aplicações proprietários (ver Cap. 3.1).

Apresenta-se uma lista de categorias de formatos no Cap. 4.1.

1.2.2.3 Escolha de aplicações informáticas

A escolha do sistema aplicacional da organização deverá ter em conta, como um dos critérios de escolha, as vantagens que a aplicação oferece para a estabilização e normalização da informação produzida através dela.

Dependendo da informação recolhida no âmbito do capítulo anterior (caracterização dos sistemas de informação e identificação dos requisitos dos documentos de arquivo electrónico).

- Poderá ser necessário proceder à aquisição de *software*, caso o existente não dê resposta às medidas previstas no PPD;
- Poderá ser necessário proceder à adaptação, alteração ou substituição de *software* utilizado no Organismo, de acordo com o indicado no PPD;
- Deverão constar, no PPD, orientações que sirvam de base a futuras aquisições de *software*.

Em todos estes casos há que ter em consideração o conjunto de estratégias de preservação e os formatos definidos anteriormente, bem como a verificação da existência, no mercado, de soluções compatíveis com essas estratégias e formatos. Caso estas não existam ou não sejam

satisfatórias, a opção passa por se investir no desenvolvimento de soluções específicas, ou empreender acções de “congelamento” dos OD produzidos nesse formato até ao aparecimento de uma solução viável.

Neste último caso aconselha-se a utilização, sempre que possível, de aplicações informáticas com especificações abertas, não proprietárias e bem documentadas, pelas razões indicadas no Cap. 3.1. Estas devem permitir executar todas as operações necessárias definidas no âmbito do PPD, ser robustas, fáceis de utilizar e conter uma interface intuitiva.

Identificamos as tipologias de aplicações de software quanto à sua origem no Cap. 4.2.

1.2.2.4 Escolha de soluções de armazenamento

O sistema de armazenamento utilizado é de importância crítica para a preservação digital. É com efeito imprescindível que este obedeça a características que assegurem a persistência e segurança dos objectos digitais a preservar. É ainda preciso notar que o grande volume de informação digital normalmente produzida exige elevada capacidade de armazenamento acompanhada dos imprescindíveis mecanismos de segurança (entre os quais o backup se conta). A escolha da solução de armazenamento deve considerar os recursos e as infra-estruturas de apoio existentes, de forma a avaliar:

- A sua adequação a necessidades de preservação digital;
- Quais as necessidades de aquisição de sistemas de armazenamento adequados (no caso da resposta ao quesito anterior ser negativa);
- Os custos com necessidades de armazenamento de informação a médio/longo prazo.

Uma opção que se venha a revelar menos adequada, poderá ter como resultado indesejável obrigar a uma inflexão de todo o planeamento elaborado. Uma análise mais aprofundada sobre as soluções de armazenamento aconselhadas encontra-se no Cap. 5.

1.2.2.5 Escolha de Metainformação

A metainformação tem também um papel primordial na preservação digital, visto que será esta que permitirá uma melhor gestão dos:

- Objectos Digitais;
- Sistemas de informação (por exemplo registo de correspondência, Sistemas de informação geográficos);
- Utilizadores;
- Funcionalidades a disponibilizar.

Por exemplo o sistema de registo de documentos utiliza metainformação, os serviços de Tecnologias da Informação e Comunicação (TIC), utilizam-na igualmente quando guardam rotinas de auditoria que registam o comportamento dos sistemas e os acessos aos mesmos.

A metainformação é normalmente utilizada dentro da organização – seja de forma implícita ou explícita – com diversos propósitos e por diferentes áreas funcionais. Basicamente pretende-se referenciar ou descrever outra informação de forma a ser mais fácil a sua posterior recuperação ou ainda garantir a evidência de uma determinada transacção (veja-se por exemplo a metainformação embebida no cabeçalho de imagens ou nos documentos de texto). Na maior parte dos casos observados não existe, todavia, uma formalização organizacional

(num documento único por exemplo) de todos os esquemas de metainformação utilizados e de quais os respectivos elementos. Propõe-se assim a criação de um documento deste tipo, de forma a oficializar e disseminar a utilização de metainformação e evitar redundâncias e repetições muitas vezes verificadas.

Tal documento deve adequar-se igualmente à documentação analógica e à documentação digital. A vantagem desta prática reside em saber exactamente que metainformação é utilizada para cada classe de informação.

A utilização de metainformação específica de preservação digital é ainda uma das componentes que deve ser integrada nesse documento.

Os processos de preservação devem estar totalmente documentados e a documentação guardada para apoio a futuros esforços de preservação. Qualquer cópia ou reformatação de dados para migração ou conversão deve ser documentada na metainformação de gestão documental.

2 Implementação do Plano de Preservação Digital

Neste capítulo procede-se à abordagem da implementação do Plano de Preservação Digital, mencionando algumas recomendações a tomar aos seguintes níveis: teste da solução proposta; produção de cópias de segurança; aplicação da estratégia de preservação; medidas de monitorização e auditorias internas; e afectação de recursos.

São apresentadas, de seguida, algumas recomendações de acções a tomar na fase de implementação do PPD. Estas não pretendem ser rígidas, visto que a implementação do plano de preservação digital, ou seja, a aplicação prática no terreno das medidas nele preconizadas, deverá ser programada pela própria instituição. Aconselha-se a utilização de bibliografia ligada à temática de Gestão de Projecto³, como apoio a esta fase.

A DGARQ, como entidade coordenadora do sistema nacional de arquivos, é neutra quanto ao método escolhido pela organização no que concerne à implementação do PPD. À DGARQ importa apenas, em sede de auditoria, o resultado dessa mesma implementação. Nesse âmbito, o anexo A estabelece uma lista de requisitos considerados obrigatórios para a prática de preservação digital, a qual servirá de base a uma eventual auditoria. Este facto não inviabiliza que a DGARQ forneça apoio técnico para a construção de uma estratégia de preservação digital quando uma organização o solicite.

2.1 Teste da solução proposta

Antes da implementação do PPD, os procedimentos técnicos e soluções decididas a partir do processo de avaliação descrito em 1.2.2 devem ser sujeitos a testes. Estes devem ser levados a cabo pelos colaboradores do organismo sobre duplicados dos documentos de arquivo electrónicos de forma a salvaguardar os originais de quaisquer resultados indesejáveis, sempre possíveis em fases de teste.

2.2 Produção de cópias de segurança

Antes da intervenção de implementação devem ser feitas cópias de segurança de todos os OD seleccionados para preservação. A integridade destas cópias deve ser verificada antes da sua remoção para uma área de armazenamento segura. Estes duplicados não devem ser sujeitos a qualquer tratamento de preservação, devendo funcionar como matrizes caso os procedimentos preconizados para a estratégia de salvaguarda sejam mal sucedidos.

2.3 Aplicação da Estratégia de Preservação

Após a finalização desta fase e caso os resultados tenham sido satisfatórios, o tratamento de preservação deve ser estendido a todos os OD seleccionados para tal.

³ Por exemplo. ANSI/PMI - A Guide to the Project Management Body of Knowledge: PMOBOK Guide, 3ª ed.. Pennsylvania: Project Management Institute. 2004 .ISBN 193069945X

2.4 Medidas de monitorização e auditorias internas

A eficácia da aplicação do plano de preservação digital deverá ser regularmente monitorizada internamente considerando o conjunto de tópicos que compõem o PPD descritos em 1. Desta forma será possível identificar eventuais disfunções e empreender acções correctivas.

Tópicos a considerar na Monitorização:

- Os procedimentos de preservação a realizar pelo organismo (ver Cap. 7);
- A definição de formatos para preservação; (ver ponto 3.1)
- A listagem de soluções de software alternativas para um determinado formato (ver ponto 3.2);
- A selecção e utilização de soluções de armazenamento; (ver ponto 5)
- Esquemas de metainformação a utilizar (ver ponto 6)

Depois de sujeitos a acções de preservação, os OD deverão ser reavaliados de forma a confirmar a permanência das suas funcionalidades de acordo com os padrões pré-definidos.

A integridade de toda a metainformação relevante associada aos OD deve ser verificada após a aplicação da estratégia de preservação. A metainformação deve também ser actualizada, registando as ocorrências verificadas durante o processo de preservação de forma a constituir um histórico essencial para provar a integridade do sistema ou a verificação e comprovação de eventuais disfunções.

O que monitorizar nos documentos de arquivo electrónicos preservados:

- Integridade;
- Funcionalidade;
- Estrutura;
- Conteúdo;
- Metainformação associada.

Por exemplo.

- ⇒ Após a migração de um sistema de bases de dados produzido originalmente em ORACLE, para um formato em MySQL ou DBML, é necessário verificar se os dados migrados mantêm a sua qualidade e se as relações existentes entre tabelas permanecem explícitas.
- ⇒ Após a migração de documentos produzidos em Word para formato PDF é necessário verificar se não houve desformatação, se as notas de pé de página não foram apagadas, se os cabeçalhos permanecem, etc.
- ⇒ Após a migração de documentos de áudio e vídeo para formato MPEG é necessário verificar se não houve dessincronização entre as imagens e o som com, se não existem fenómenos de arrastamento de *pixels*, etc.
- ⇒ Após migração de publicações Multimédia Interactivas é necessário verificar se as hiperligações se mantêm correctas, se é possível o acesso ao conteúdo dos ficheiros, se todos os ficheiros permanecem presentes, etc.

2.5. Afectação de recursos

É essencial prever antecipadamente os recursos humanos, materiais, bem como os elementos de gestão basilares (cronograma), considerados como necessários para assegurar o esforço de implementação do plano.

3 Ameaças

Este capítulo destina-se a explicitar o conjunto de ameaças que pesam sobre os OD, abordando a problemática da obsolescência dos formatos, software, hardware e suportes, assim como elementos que afectam os suportes.

A especificidade dos OD implica que estes estejam sujeitos às seguintes ameaças:

■ **Obsolescência tecnológica:**

Ocorre ao nível de formatos e *software*, mas também do próprio *hardware* e suportes utilizados.

■ **Danos físicos:**

Estes podem ser acidentais ou propositados, ocorrem principalmente ao nível dos materiais que compõem o *hardware* e os suportes, mas também podem ocorrer nos ficheiros (muitas vezes como consequência dos danos de *hardware* e/ou dos suportes).

3.1 Obsolescência de formatos e software

A obsolescência de formatos e *software* pode ocorrer quando:

- Um formato de ficheiro é ultrapassado por outro formato ou por uma versão mais recente que comporta mais complexidade;
- Um dado formato não vinga ou as empresas não criam *software* compatível;
- Um dado formato falha, estagna, ou já não é compatível com os sistemas actuais;
- O *software* que suporta o formato falha comercialmente ou é adquirido por um concorrente que o retira do mercado.

No que diz respeito ao *software*, este pode dividir-se em três grupos, de acordo com o exposto no Cap. 4.2:

- Proprietário com especificações fechadas;
- Proprietário com especificações abertas;
- Não proprietário com especificações abertas.

Para escolher o melhor *software* no âmbito do planeamento da estratégia de preservação é necessário avaliar se este:

- Tem grande adopção;
- É retrocompatível;
- Tem suporte de metainformação (como XML);
- Dispõe de várias funcionalidades, sem ser muito complexo;
- Permite a utilização de vários formatos;
- Inclui funcionalidades de verificação de erros;

- Dispõe de um ciclo de actualizações razoável.

3.2 Obsolescência do *hardware* e suportes

No âmbito do *hardware* e dos suportes, a rápida obsolescência é uma constante, devido aos avanços tecnológicos a nível de:

- Velocidade do processador;
- Densidade dos *chips* de memória;
- Capacidade dos periféricos de armazenamento;
- Velocidade de processamento de imagem;
- Velocidade de transmissão.

Tais alterações levam à rápida substituição dos suportes, não só devido à maior rapidez, capacidade e produtividade disponíveis, mas também devido ao cada vez maior número de funções oferecidas.

Elementos como:

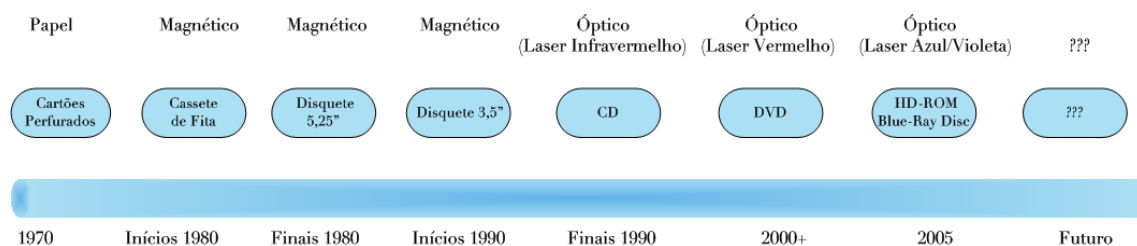
- A diminuição do tamanho físico;
- O aumento da capacidade e diminuição do custo por unidade de armazenamento;
- As tendências em termos de fragilidade, estabilidade, segurança e duração antes da obsolescência e
- Outras melhorias,

As normas ISO/IEC 17799, ISO/IEC 27000 a ISO/IEC 27008 fornecem informações, códigos de práticas, e requisitos para auditoria e certificação de técnicas de segurança para sistemas de gestão de informação. De referir, no entanto, que a análise destas normas não está no âmbito deste documento e que não cabe à DGARQ a certificação destas técnicas.

Levam a novas e melhores aplicações de *software*, que aumentam a probabilidade obsolescência do equipamento anterior.

Assim os dispositivos necessários para leitura do suporte de armazenamento podem deixar de ser produzidos. Os fabricantes de todas estas tecnologias competem, emergem, unem-se, ou desaparecem, tornando cada vez mais difícil a manutenção dos conteúdos digitais ao longo do tempo.

Cronologia: evolução de suportes



3.3 Elementos que afectam os suportes

Os principais elementos que afectam os suportes e, directamente, os OD que se pretendem preservar, são:

- A instabilidade do material;
- As condições ambientais de armazenamento (temperatura, humidade relativa e exposição à luz);
- O manuseamento e utilização (riscos em suportes ópticos, dedadas, borrões, sujidade, poeiras, água e humidade, solventes orgânicos, marcadores de escrita, etiquetas adesivas);
- Desastres naturais (fogo, inundações, abalos de terra);
- Falhas de infra-estrutura (canalização, electricidade, sistema de controlo climatérico);
- Manutenção inadequada e mau funcionamento do *hardware*;
- Erro humano, sabotagem (roubo e vandalismo);
- Campos magnéticos, raios – X e microondas;
- Desgaste derivado da utilização.

4 Caracterização de formatos e aplicações

Este capítulo pretende caracterizar o conjunto de formatos e aplicações existentes, sendo definidas:

- Os formatos, as suas características e propostas para estratégias de preservação;
- As diferentes categorias de aplicações e suas especificações.

São igualmente apresentados conselhos relativos formatos e aplicações a utilizar.

4.1 Formatos

Qualquer tentativa de categorização/compartimentação de tipos de documentos electrónicos é necessariamente artificial, devido ao facto de muitos conterem, em si, elementos de outras categorias. Por exemplo, são raros os documentos electrónicos de texto estruturado que não incluam actualmente tabelas ou imagens; são raros os formatos de registos de vídeo que não incluam também som e/ou voz. Visto que, para efeitos de definição de formatos, é necessário definir tais categorias, optou-se pela taxionomia dos tipos de documentos de arquivo electrónicos⁴ baseada na definida pelos Arquivos Nacionais do Reino Unido⁵.

Na tabela abaixo apresentada, inclui-se uma coluna relativa à categoria de formato utilizada, uma segunda coluna que descreve as principais características dessa categoria e uma terceira coluna onde se indica a estratégia de preservação recomendada.

Categoria	Características	Estratégia de Preservação Recomendada	Informação Adicional
1. Dados Tabulares (<i>Data Set</i>)	Dados alfanuméricos existente numa vasta variedade de aplicações de processamento de dados; Dados geridos em ficheiro linear (<i>flatfile</i>), em rede, hierárquica; Bases de dados relacionais e orientadas a objectos.	Migração para novas plataformas. Conversão de dados e estrutura para formatos normalizados. Actualmente é impossível migrar as componentes comportamentais relacionadas com motores de gestão de bases de dados, normalmente proprietários.	O Arquivo Nacional dos Países Baixos produziu uma proposta para preservação de: Folhas de cálculo < www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility-permanence-spreadsh-en.pdf >; Bases de Dados < www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility-permanence-databases-en.pdf >.
2. Texto estruturado/ documentos do Office	Dados alfanuméricos; Dados de marcação (<i>markup</i>); Etiquetas para outros tipos de dados (imagens vectoriais e	Migração para novas plataformas. Conversão para formatos normalizados. Refrescamento de suportes.	O Arquivo Nacional dos Países Baixos produziu propostas para preservação de: Documentos de texto < www.digitaleduurzaamheid.nl/bibliotheek/

⁴ In STEPHENS, David O. – *Digital Preservation in the United Kingdom* in THE INFORMATION MANAGEMENT, vol. 34, n.º. 4, Outubro 2000

⁵ *The National Archives*, <http://www.nationalarchives.gov.uk/>

Categoria	Características	Estratégia de Preservação Recomendada	Informação Adicional
	<p>mapas de bits (<i>raster</i>) existentes em processadores de texto/programas para escritório e digitalização de documentos/sistemas aplicativos de gestão de documentos electrónicos, bases de dados relacionais, e outros ambientes de tipos de documentos de aplicações específicas.</p>		<p>docs/volatility-permanence-textdocs-en.pdf>;</p> <p>Mensagens de Correio Electrónico <www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility-permanence-email-en.pdf>.</p>
<p>3. Dados de Desenho [design data]</p>	<p>Imagens vectoriais e de mapa de bits e dados alfanuméricos em sistemas de CAD e conversão orientada a objectos para bases de dados normalizadas.</p>	<p>Retrocompatibilidade, interoperabilidade, e formatos com emulação de tecnologia como estratégia a curto prazo para dados do produto (<i>product data</i>) em sistemas obsoletos.</p> <p>Refreshamento de suportes.</p>	
<p>4. Apresentações</p>	<p>Dados alfanuméricos em imagens empresariais, <i>clipart</i> [coleção de imagens], vídeo, e multimédia de treino/ensino.</p> <p>Diagramas para a gestão de informações, de conhecimento e de capital intelectual; para a compreensão e solução de problemas.</p>	<p>Retrocompatibilidade.</p> <p>Conversão para formatos normalizados.</p> <p>Refreshamento de suportes.</p>	
<p>5. Imagens</p>	<p>Imagens de mapas de bit e dados alfanuméricos existentes no software de captura/edição de imagens, bases de dados orientadas a objectos, relacionais ou simples de livrarias de fotos, colecções de digitais de belas-artes, colecções de imagens</p>	<p>Retrocompatibilidade.</p> <p>Conversão para formatos normalizados.</p> <p>Refreshamento de suportes.</p>	

Categoria	Características	Estratégia de Preservação Recomendada	Informação Adicional
	médicas.		
6. Documentos de som e voz	Dados de som em processamento de voz, bases de dados relacionais ou simples (<i>flat</i>) de registos áudio em colecções de musica e repositórios similares.	Retrocompatibilidade. Conversão para formatos normalizados. Refrescamento de suportes (Esta é uma área especializada que necessita de trabalho adicional por especialistas de preservação digital).	
7. Documentos de Vídeo	Vídeo Digital, ecrã total (<i>fullscreen</i>)/ vídeo (em movimento) em fotogramas de vídeo digital armazenadas em ficheiros de mapas de bit e som armazenadas em bases de dados relacionais ou simples (<i>flat</i>) em colecções de musica e repositórios similares	Retrocompatibilidade. Conversão para formatos normalizados. Refrescamento de suportes. A tecnologia de preservação é necessária em situações a curto prazo quando os dados estão fechados em sistemas proprietários.	
8. Dados geográfico/cartográficos	Imagens vectoriais e de mapa de bits e dados alfanuméricos em sistemas de informação geográfica (SIG) e <i>software</i> de cartografia. Dados armazenados orientada em bases de dados relacionais orientadas a objectos.	Retrocompatibilidade. Conversão para formatos normalizados. Refrescamento de suportes. A tecnologia de preservação é necessária em situações a curto prazo quando os dados estão fechados em sistemas proprietários.	
9. Publicações Multimédia Interactivas	Dados de som e vídeo, imagens em movimento, imagens vectoriais e de mapa de bits e dados alfanuméricos armazenados em <i>software</i> de criação e edição em ambientes de publicação electrónica.	A tecnologia de preservação é necessária em situações a curto prazo quando os dados estão fechados em sistemas proprietários até que existam normas aceites e amplamente utilizadas.	O Arquivo Nacional dos Países Baixos aborda a problemática da preservação de sítios Web em < www.digitaleduurzaamheid.nl/index.cfm?paginaakeuze=374 >.

Categoria	Características	Estratégia de Preservação Recomendada	Informação Adicional
10. Dados Científicos e Financeiros	Dados específicos de carácter científico e financeiro.	A tecnologia de preservação é necessária em situações a curto prazo quando os dados estão fechados em sistemas proprietários até que existam normas aceites e amplamente utilizadas.	

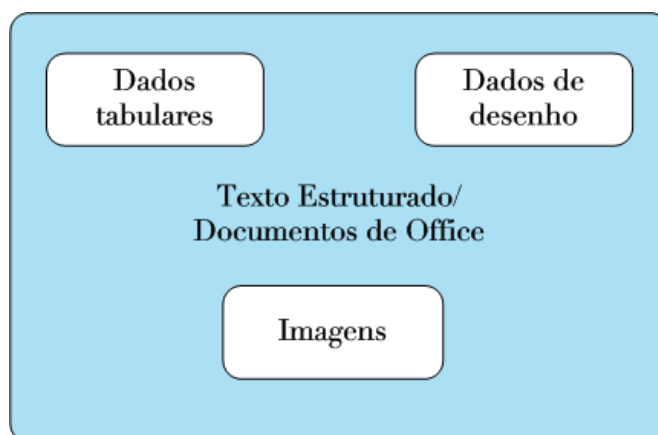
Apresentam-se, a seguir, dois exemplos que ilustram relações possíveis entre categorias que podem coexistir num mesmo formato.

Importa referir que qualquer esquema que tente representar relações entre categorias não está, nem pode estar nunca, encerrado, uma vez que a listagem de categorias existentes e as suas características distintivas nunca estarão definitivamente estabilizadas. As necessidades nesta área nunca estarão, por definição, satisfeitas assim como as respostas a essas necessidades podem ser imprevisíveis em termos de possibilidades de combinação.

Assim é natural que, no futuro, surjam novos formatos que, à partida, não sejam passíveis de categorizar na lista actual. Tal remete para a perenidade deste documento e para a necessidade constante da sua actualização.

■ **Exemplo 1: Documento de texto**

Dados Estruturados/Documentos de Office: um documento de texto pode conter, para além do texto, Dados Tabulares (tabelas), Imagens e ainda Dados de Desenho.



Exemplo 2: *Web-conference*

Dentro das Publicações Multimédia Interactivas podemos incluir a *Web-conference*, a qual pode conter Registos de som e voz, Registos de vídeo, Apresentações, Imagens, Texto Estruturado.



4.1.1 Identificação de formatos alternativos normalizados

A criação de documentos de arquivo electrónicos, de acordo com as necessidades operativas do organismo, obriga à utilização de uma panóplia de formatos decorrentes das aplicações de *software* utilizado na sua produção. Muitas vezes isso traduz-se na produção de formatos proprietários e não normalizados, o que dificulta a sua preservação a longo prazo. Nestes casos é fundamental prever quais os formatos normalizados a utilizar no âmbito da preservação digital.

No caso das categorias Texto estruturado/documentos do Office e Imagens a escolha é facilitada pela existência de formatos normalizados muito utilizados e que mantêm o conteúdo informacional e a estrutura dos documentos (mais concretamente o PDF e o TIFF). Outras categorias, como *Data Set*, *Design Data*, Dados geográficos/cartográficos e Publicações Multimédia Interactivas, ainda não dispõem de formatos normalizados que garantam a preservação de todas as características dos OD.

Recomenda-se a não utilização de formatos proprietários para armazenamento e preservação digital a longo prazo, pelas razões indicadas nos Cap. 3.1, devendo ser necessário identificar, à partida, quais os formatos que, por serem normalizados e/ou não proprietários e de fácil preservação a longo prazo, se poderão utilizar alternativamente. Tal escolha deve assentar em primeiro lugar na categoria de formatos em que o DAE se insere.

O projecto RODA definiu para as taxonomias de OD com que opera actualmente, os seguintes formatos de preservação:

- Texto estruturado: PDF/A ([0Hwww.pdfa.org](http://www.pdfa.org))
- Imagens fixas bidimensionais: TIFF ([1Hhttp://partners.adobe.com/public/developer/tiff](http://partners.adobe.com/public/developer/tiff))
- Bases de dados relacionais: DBML, um formato

É preciso salientar que não existem formatos normalizados para cada uma das categorias de formatos identificados. Tal comporta riscos para a preservação desses OD que não podem ser evitados. No caso em que não haja formato alternativo e o DAE se encontra inactivo há que

manter o formato original sujeito a monitorização regular até ser desenvolvido um formato normalizado aplicável ao formato original em questão.

Deve ser produzida uma tabela que permita definir quais os formatos de preservação que o organismo pretende utilizar.

4.2 Aplicações

As aplicações informáticas ou *software* são, conceptualmente, uma sequência de instruções a serem seguidas e/ou executadas, na manipulação, redireccionamento ou modificação de dados ou informação, e que permitem ao utilizador fazer uma ou mais tarefas específicas. Do ponto de vista técnico, integra igualmente, não só o programa de computador propriamente dito, mas também manuais, especificações, planos de teste, etc.

As aplicações fazem parte do sistema intermediário que permite a produção, manipulação, gestão e preservação dos OD.

As aplicações que produzem os documentos podem dividir-se em três tipos:

■ Proprietário com especificações fechadas:

A cópia, redistribuição ou modificação são proibidas pelo seu criador ou distribuidor. Tais restrições são conseguidas através de meios legais e/ou técnicos. Os meios legais podem incluir o licenciamento de *software*, *copyright* (direitos de autor), e uso de patentes, enquanto que os meios técnicos são caracterizados pela venda ao utilizador de binários executáveis legíveis pelo *hardware*, ocultando ou encriptando o código fonte legível pelos humanos.

■ Proprietário com especificações abertas:

Vários produtores de *software* definem e utilizam especificações ou padrões para garantir a interoperabilidade entre aplicações e plataformas. Estes padrões devem estar disponíveis para livre acesso e implementação, sem quaisquer *royalties* e outras taxas e sem discriminação de uso.

■ Não proprietário com especificações abertas:

Software livre, segundo a definição criada pela Free Software Foundation, é qualquer programa de computador que pode ser usado, copiado, estudado, modificado e redistribuído sem nenhuma restrição. Este tipo de *software* opõe-se ao conceito de *software* proprietário, mas não à comercialização (*software* comercial).

A forma usual de distribuição de *software* livre é anexar a este uma licença de *software* livre, e tornar o código fonte do programa disponível. Para tal é muito usual a utilização de especificações abertas (normas) para manter a interoperabilidade de todas as distribuições originárias do mesmo código fonte base.

O *software* livre permite as seguintes modalidades:

- **Gratuito (*freeware*):** não implica no pagamento de licenças de uso ou *royalties*;
- ***Shareware*:** pode ser copiado, distribuído e utilizado experimentalmente por um determinado período, após o qual se deve pagar uma taxa.

Aconselha-se a utilização, sempre que possível, de aplicações e formatos com especificações abertas, não proprietários e bem documentados, pelas razões indicadas no Cap. 3.1.

Esta prática é objecto da Resolução de Conselho de Ministros 21/2002, de 31 de Janeiro.

Para mais informações sobre a promoção da utilização de *software* livre na Administração Pública, aceda a:
<http://www.softwarelivre.citiap.gov.pt/>

5 Sistemas de armazenamento

Este capítulo destina-se a abordar os sistemas de armazenamento de acordo com um conjunto de critérios que devem obedecer. Em seguida é feita uma análise comparativa das arquitecturas de armazenamento (NAS, SAN e CAS), sendo definidos critérios de avaliação das mesmas.

Finalmente é dada relevância à problemática das cópias de segurança no que concerne à criação de réplicas, modelos de armazenamento, tipos de gestão de repositórios de dados.

5.1 Atributos desejáveis

A solução de armazenamento deve ser distinta das réplicas de segurança (também designados de ficheiros de salvaguarda ou *backups*), abordadas no Cap.5.3. Assim, um sistema de armazenamento consiste no dispositivo ou conjunto de dispositivos utilizados para armazenar primariamente os dados e a informação produzida (dispositivos primários).

O *backup* destina-se a salvaguardar os dados contidos nesses dispositivos primários. Consiste portanto numa réplica desses dados, a qual poderá ser efectuada seguindo diferentes métodos oferecendo a garantia de, em caso de perda ou corrupção dos dados primários contidos no sistema de armazenamento, estes poderem ser recuperados através das cópias de *backup* e o sistema reposto na situação original antes da ruptura verificada.

Um sistema de armazenamento é uma estrutura de complexidade variável, mas não se pode confundir com um suporte de armazenamento individualizado. Por exemplo um DVD, ou um conjunto de DVD considerados individualmente sem um sistema de organização mecânica ou electrónica, não constitui um sistema de armazenamento na acepção deste documento, embora contenha informação. Mais à frente abordar-se-á os suportes considerados de forma descontextualizada do sistema de armazenamento, pelo facto de, em muitas organizações, ainda se recorrer a esse processo para armazenar informação.

É normal um sistema de armazenamento incluir um sistema de cópia de segurança. Deverá incluir réplicas, geográfica e administrativamente independentes. Desta forma a solução de armazenamento deverá ter as seguintes características:

- Estar disponível em linha;
- Apresentar baixa latência.

Latência (em inglês *lag*): refere-se ao atrasos que se podem verificar na comunicação entre computadores (Internet, por exemplo) e comunicações via satélite. Em computação, *lag* refere-se ao tempo que um pacote de dados leva a completar um percurso de ida e volta entre um computador local e o seu destino.

A combinação destes atributos permite auditorias fáceis e frequentes e, ainda, reparações rápidas. Deve suportar a introdução progressiva de novos componentes heterogéneos (ex. discos rígidos de diferentes marcas), permitindo assim a escalabilidade e a existência de um plano de sucessão facilmente implementável.

O sistema de armazenamento deve:

- Ter um desempenho adequado às necessidades específicas de cada organização, de forma a evitar “engarrafamentos” (bloqueios e atrasos de reposta);
- Estar em conformidade com normas *de jure*;

- Ser independente da plataforma aplicacional utilizada (sistema operativo e vendedor de *hardware*).

Preferencialmente o custo inicial e de manutenção deverá ser mínimo, tal como o esforço necessário para a sua administração.

5.2 Arquitecturas Possíveis

As arquitecturas de armazenamento a longo prazo, actualmente disponíveis são as seguintes:

- Network Attached Storage (NAS)
- Storage Area Network (SAN)
- Content-Addressable Storage (CAS).

Estes cenários aplicam-se a organizações com uma forte produção de DAE que requerem, portanto, elevada capacidade de armazenamento e de resposta. No caso da organização produzir DAE em pequena escala é possível recorrer a sistemas de armazenamento com desempenho inferior que podem, inclusivamente, incluir suportes ópticos e dispositivos de armazenamento conexos.

Por questões que se prendem com a gestão e segurança, aconselha-se a utilização de um sistema de armazenamento, caso o volume de documentos de arquivo electrónicos utilizados pelo organismo for superior a 10Gb.

As **NAS** são dispositivos dedicados à partilha de ficheiros entre diferentes clientes de rede. Permitem adicionar armazenamento na rede sem ser necessário desligar o servidor. Podem estar ligadas em qualquer parte da rede, não integrando, porém, o servidor que faz a gestão do processamento dos dados.

É uma solução barata, fácil de instalar, administrar e escalar, e baseia-se em protocolos abertos e bastante disseminados. No entanto a ligação entre servidor aplicacional e armazenamento é lenta; apresenta limitações no que respeita ao tamanho máximo de um bastidor (actualmente 16 terabytes); precisa de utilizar sistemas de virtualização de armazenamento (*storage virtualization*) para suportar múltiplos sistemas de ficheiros com o objectivo de guardar os dados.

As **SAN** são redes projectadas para agrupar dispositivos de armazenamentos de computador. Existem duas variações de SAN:

1. Uma rede na qual o propósito principal é a transferência de dados entre computadores e dispositivos de armazenamento. Neste caso a SAN consiste numa infra-estrutura de comunicação que dota conexões físicas com uma camada de gestão, com o objectivo de organizar dispositivos de armazenamento, conexões e computadores, tornando a transferência de dados mais robusta e segura;
2. Um sistema de armazenamento formado por dispositivos de armazenamento, computadores e/ou aplicações, e todo um controlo via *software*, comunicando-se através de uma rede de computadores.

As SAN caracterizam-se por:

- Apresentar um bom desempenho;
- Ser normalmente mais dispendiosas;
- Ter maior exigência de administração (gestão de unidades lógicas e de sistemas de ficheiros);
- Ter reduzida interoperabilidade (o que funciona numa SAN pode não funcionar noutra);
- O suporte aos servidores estar dependente do sistema operativo (com as consequentes problemas de limitações de drivers de controlos).

As CAS são mecanismos para armazenamento de informação, cuja recuperação se baseia no seu conteúdo e não na sua localização de armazenamento. As CAS caracterizam-se por disponibilizarem um acesso simples, sem as limitações normais de um sistema de ficheiros ou protocolo de partilha.

São fáceis de instalar e administrar e têm um elevado grau de escalabilidade. No entanto são dispendiosas, pelo menos na fase inicial de aquisição, e a sua interface de programação (API) é restrita e pouco disseminada.

Apresenta-se, abaixo, um quadro comparativo destes sistemas de armazenamento:

Quadro comparativo de sistemas de armazenamento

	NAS	SAN	CAS
Heterogeneidade	Boa	Má	Nenhuma
Escalabilidade	Adicionando mais NAS	<i>Upgrade</i>	Até 24 Petabytes
Desempenho	Baixo	Muito Alto	Alto
Segue normas <i>standard</i>	Sim	“ <i>Flavours</i> ” da norma standard de base	API restrita
Independência	Completa	Parcial. Só alguns sistemas operativos são suportados	Sim. Desde que implemente a API
Custo	Baixo	Alto	Muito Alto (inicial)
Administração	Depende da virtualização	Difícil	Muito Fácil

5.2.1 Critérios para avaliação de soluções de armazenamento

São seguidamente referidas as variáveis que deverão ser consideradas para avaliar a solução de armazenamento mais adequada às necessidades de preservação identificadas. Sem detrimento de outros critérios que a organização pretenda considerar os que são a seguir enunciados são básicos para assegurar a qualidade da preservação digital a longo prazo.

Estas variáveis estão organizadas nos seguintes grupos: custo, escalabilidade, interoperabilidade, segurança de dados e facilidade de programação. Compete à organização definir a ponderação a atribuir a cada uma das variáveis, de forma a permitir essa avaliação.

Para cada uma destas variáveis deve ser atribuída uma ponderação parcial de acordo com os critérios que a organização queira privilegiar. O conjunto de variáveis por grupo dará a pontuação final desse grupo. A média ponderada dos quais dará uma classificação final ao sistema de armazenamento avaliado.

1. Custo

- Custo de instalação
- Manutenção e suporte/ano
- Prazo da garantia
- Custo da solução para 1 TB
- Custo da solução para 10 TB
- Custo da solução para 100 TB
- Custo de cada nó (1 nó suporta vários discos)
- Custo de cada unidade de controlo

2. Escalabilidade

- Número máximo de nós
- Número máximo de nós por cada unidade de controlo
- Capacidade máxima de cada nó (número de discos)
- Limite máximo em TB de cada nó

3. Interoperabilidade

- Suporta nós de marcas concorrentes
- Suporta discos de marcas concorrentes
- Disponibiliza interfaces de programação
- Suporta conectividade NFS
- Suporta conectividade SMB

4. Segurança de dados

- Suporta RAID

- Suporta *backups*
- Suporta replicação remota

5. Facilidade de programação

- *Storage Virtualization*
- API de controlo do número de réplicas
- API de controlo dos processos de auditoria

5.3 Cópias de segurança

A capacidade de fazer réplicas de igual qualidade (digital para digital) significa que é possível, e recomendável, armazenar cópias de OD num ambiente diferente das armazenadas para acesso frequente. Estas cópias terão a mesma qualidade da cópia ou original arquivado, e estes só serão necessários para inspecções, para fazer novas cópias, ou para efectuar a migração para novos formatos.

As cópias digitais de arquivo devem também ser guardadas em locais diferentes das cópias de acesso frequente, a fim de prevenir perdas de informação por desastres, roubo ou mau manuseamento.

Assim, aconselha-se a utilização de uma solução fora de linha, que limite o acesso apenas quando a informação armazenada estiver corrompida.

Cópia arquivada – uso limitado (<i>backup</i> /réplica)	Cópia de acesso frequente – uso múltiplo (armazenamento)
1. Armazenar num ambiente controlado e separado, se possível, da cópia de acesso frequente; 2. Armazenar em condições ambientais recomendadas para armazenamento de arquivo.	1. Armazenar para fácil acesso; 2. Verificar, visualmente, se o disco foi alvo de algum dano após a utilização; 3. Armazenar em condições ambientais similares às condições de utilização.

O armazenamento de *backups* deve seguir um dos seguintes modelos:

■ **Não estruturado**

Este modelo pode ser constituído apenas por um conjunto de suportes ópticos graváveis, com a informação mínima sobre o que foi replicado e quando. Trata-se do modelo mais fácil de implementar mas, provavelmente, o que piores resultados obterá em termos de recuperação de alto nível.

■ **Total+Incremental**

Neste caso pretende-se armazenar várias réplicas dos OD originais da seguinte forma:

1. Em primeiro lugar é feita uma cópia de segurança de todos os OD;
2. Posteriormente poderá ser feita uma cópia de segurança incremental;
3. Finalmente, quando se proceder ao restauro do sistema na totalidade e até determinada data, é necessária a localização da cópia de segurança total, bem como do ponto (data) de restauro pretendido.

Este modelo oferece um alto nível de segurança, em que qualquer OD pode ser restaurado e pode ser utilizado com suportes removíveis, tais como tapes e suportes ópticos. O aspecto menos positivo é a existência de longas séries de incrementações e necessidade de espaço.

■ **Total + Diferencial**

Difere do modelo anterior no sentido em que, após a cópia de segurança total, cada cópia de segurança parcial deve capturar todos os OD criados ou modificados após aquela, mesmo que já tenham sido incluídos na cópia de segurança parcial anterior. A vantagem é que o restauro apenas obriga à recuperação da cópia de segurança total e a última cópia de segurança parcial.

■ **Réplica + Incremental invertido (Sincronização)**

Este modelo é similar ao repositório Total + Incremental, diferindo no facto de fornecer uma réplica que reflecte o estado do sistema, aquando da última cópia de segurança, bem como uma lista das incrementações invertidas.

Um dos benefícios reside no facto de requer uma cópia de segurança total. Cada cópia de segurança incremental é imediatamente aplicada à réplica, e os OD a substituir são movidos para uma incrementação invertida, i.e., é feita uma sincronização. Este modelo não deve utilizar suportes móveis, visto que cada cópia de segurança necessita de ser efectuada por comparação com o *mirror*.

■ **Protecção de dados contínua**

Neste modelo são registadas todas as alterações do sistema, em vez de calendarizar cópias de segurança periódicas. Esta operação é geralmente realizada através da gravação das diferenças a nível de bytes ou blocos de dados, e não das diferenças a nível de ficheiros. Este modelo difere da simples replicação de discos, uma vez que permite retornar ao estado anterior do registo e, assim, restaurar a anterior imagem de dados.

Unidades de suporte para *backups* mais aconselhadas:

Tapes (magnéticas): são um suporte de acesso sequencial, pelo que, mesmo que os tempos de acesso sejam baixos, a velocidade de escrita ou leitura contínua de dados é rápida. São, desde há muito tempo, o suporte mais utilizado para armazenamento, cópias de segurança, arquivo e transmissão.

Existe uma variedade de formatos, muitos dos quais proprietários ou específicos de alguns nichos de mercado, como é o caso das *mainframes* ou de uma marca de computadores específica.

Discos rígidos: podem ser ligados localmente através de interfaces *SCSI*, *USB* ou *Firewire*, ou através de tecnologias de maior distância, tais como *Ethernet* (Rede), *SCSI* ou Canais de Fibra. As suas principais vantagens são: tempos de acesso baixos, disponibilidade, capacidade e facilidade de uso. A relação capacidade/preço dos discos rígidos tem vindo a melhorar ao longo dos anos.

Cópia de Segurança Remota: são feitas através Internet para um local remoto e permitem proteger os dados contra alguns dos maiores perigos, como o fogo, inundações, sismos ou detonação nuclear. À medida que as ligações de Internet de banda larga se tornam cada vez mais disseminadas, os serviços de cópia de segurança remota são cada vez mais utilizados. Um dos aspectos menos positivos desta solução, prende-se com o facto da velocidade da ligação de Internet ser normalmente mais baixa do que a velocidade de transmissão dos sistemas de armazenamento de dados locais. Este facto pode ser problemático quando se lida com grandes volumes de dados. Adicionalmente, é necessário ter em conta os perigos associados à subcontratação (*outsourcing*) de terceiros para armazenarem dados considerados sensíveis ou pessoais.

As unidades de suporte para *backups* menos aconselhadas são:

Discos Ópticos: podem ser utilizados como suporte de cópias de segurança. Uma das vantagens do CD e do DVD é poderem ser restaurados em qualquer máquina com um leitor. Muitos formatos de disco óptico são *WORM*, o que faz com que sejam tradicionalmente utilizados para arquivo, visto os dados não poderem ser alterados. Outros formatos regraváveis podem também ser utilizados, como os *CDRW* ou *DVD-RAM*.

Os novos discos *HD-DVD* e *BluRay* aumentaram de forma significativa a quantidade de dados que podem ser armazenados num único disco óptico. No entanto é preciso ter em consideração a inexistência de estudos sobre a real perdurabilidade de suportes ópticos.

Disquetes: muito utilizadas durante a década de 80 e inícios de 90 e associadas às cópias de segurança. Hoje em dia, a capacidade das disquetes tornaram esta solução obsoleta.

Dispositivos de memória: também conhecidos por memórias *flash*, *Pen-drives*, cartões de memória, cartões digitais, etc., estes suportes são relativamente caros em comparação com a sua baixa capacidade, mas são bastante portáteis e fáceis de utilizar.

5.3.1 Gestão do repositório de dados

Independentemente do modelo de repositório ou suporte utilizado para as cópias de segurança, é necessário estabelecer um equilíbrio entre as questões de acessibilidade, segurança e custos.

Em linha (*Online*)

Normalmente o tipo de armazenamento de dados mais acessível, que pode iniciar o restauro em milissegundos. Um exemplo disso é um disco rígido interno ou um conjunto de discos rígidos (ligados a uma *SAN*). Este tipo de armazenamento é muito conveniente e rápido, mas relativamente caro. Por outro lado, sofre de alguma vulnerabilidade, pois os dados podem ser apagados ou alterados, seja por acidente ou por vírus que elimine dados.

Perto de Linha (*Near-line*)

Normalmente trata-se de um tipo de armazenamento de dados menos acessível, mas mais barato, do que o armazenamento em linha, sendo útil para o armazenamento de dados de réplicas (cópias de segurança). Um exemplo disso é um conjunto de tapes com tempos de restauro que vão desde segundos até alguns minutos. Normalmente está associado a um aparelho mecânico que move as unidades de suporte do local de armazenamento para o dispositivo de leitura e/ou gravação de dados.

Fora de Linha (*Off-line*)

Similar ao perto de linha, embora necessite de intervenção humana para disponibilizar as unidades de suporte. Isto pode incluir simplesmente o armazenamento das tapes de cópias de segurança nas estantes. O tempo de acesso deste suporte pode ser muito demorado.

Cofre-Forte (*Off-site vault*)

Com vista à protecção contra desastres naturais, algumas instituições preferem enviar os suportes de armazenamento das cópias de segurança para um cofre-forte fora de linha. O cofre pode ser simplesmente o gabinete do Administrador do sistema ou, algo tão sofisticado como, um *bunker* de alta segurança com sistemas anti-desastre e controlo climatérico, que permitem o armazenamento de suportes de cópias de segurança.

Unidade de Cópias de Segurança/Centros de Recuperação de Dados (*Backup site, Disaster Recovery Center*)

No caso de um desastre, os dados nos suportes de cópia de segurança não serão suficientes para a recuperação da informação. Também são necessários sistemas informáticos e redes bem configuradas para os dados poderem ser restaurados. Algumas organizações têm os seus próprios centros de recuperação de dados equipados para este tipo de cenário. De notar que, sendo estas unidades de recuperação de dados um investimento tão dispendioso, raramente se utilizam as cópias de segurança como método de transferência de dados para a unidade de recuperação de dados. O procedimento mais usual é a replicação de disco (*disk mirroring*) remota, que permite manter os dados no centro de recuperação de dados o mais actualizado possível.

6 Metainformação

Este capítulo aborda a problemática da necessidade de metainformação (MI) para uma correcta e viável preservação digital. Para tal refere-se a utilização de várias camadas de MI, como forma de a organizar, indicando-se, para cada camada, exemplos de esquemas de MI passíveis de serem utilizados.

É proposta uma lista de elementos de MI recomendados para a camada de MI de preservação. Finalmente, alerta-se para a utilização de uma linguagem cuja sintaxe permita ser legível e pesquisável pelo ser humano e pelas máquinas.

Com vista à sistematização de informação, disponibiliza-se um instrumento de apoio (anexo E) relativo à definição de esquemas de MI.

A Metainformação – ou metadados – é transversal a todos os elementos (identificados nos Cap. 4 e 5) do PPD.

Na aceção mais básica, a MI é informação sobre informação, ou mais especificamente dados estruturados sobre informação capturada no sistema de arquivo.

A metainformação descreve atributos do documento de arquivo electrónico dando-lhe significado, contexto e organização, permitindo a produção, gestão e utilização de documentos de arquivo ao longo do tempo, assim como nos, e através dos, domínios em que são produzidos.

Note-se que existem muitos esquemas de metainformação – normalizados ou não – destinados a descrever aspectos temáticos ou funcionalmente especializados. Por exemplo existe metainformação específica para dados estatísticos ou para sistemas espaciais como os SIG.

A metainformação pode ainda existir implicitamente embebida nos objectos digitais sendo automaticamente produzida pelas soluções aplicacionais que lhes deram origem

No âmbito da documentação de arquivo electrónica, a MI consiste num conjunto de elementos (atributos) associados a cada OD no momento da sua integração no arquivo e ao longo da sua gestão.

A metainformação pode ser associada não apenas a documentos individuais mas também a pastas, utilizadores, plano de classificação e qualquer outra entidade de um Sistema de Gestão Documental/Arquivo.

A importância da MI neste âmbito prende-se com a dependência dos documentos electrónicos relativamente ao sistema intermediário, sendo então necessária para que os OD fiquem contextualizados em termos de produção, gestão e preservação. A metainformação associada aos OD permite:

- A referência única e persistente de cada OD;
- A localização e recuperação expedita de cada OD;
- A criação do histórico de cada OD;
- O registo das transformações realizadas sobre cada OD;

Embora de uma forma geral os documentos electrónicos contenham alguma metainformação fornecida pelas aplicações que os produzem, esta é insuficiente e/ou não se encontra estruturada (organizada). Para ultrapassar esse facto dispomos de esquemas de metainformação que designam os elementos que têm de ser associados aos documentos electrónicos. Importa clarificar que não existem esquemas de MI obrigatórios, pese embora a sua utilização se inclua no conjunto de boas práticas para a preservação digital.

Visto que a MI pode ser repartida por áreas, é aconselhável a utilização de diferentes esquemas no âmbito da preservação digital, que devem ser aplicados por diversas camadas:

■ **Metainformação Descritiva ou de Identificação:**

O objectivo é a pesquisa, recuperação e identificação. Pode incluir elementos como o título, o assunto, autor e palavras-chave. O esquema EAD (*Encoded Archival Description*) é um bom exemplo, por ser um esquema muito completo e que respeita as normas de descrição arquivística;

O sítio Web da Library of Congress inclui a página oficial do EAD: <URL:http://www.loc.gov/ead>.

O RODA utiliza um modelo de dados baseado no EAD, designado por EADPART, abordado em <URL: http://roda.iantt.pt/pt/system/files/roda-relatorioFinal.pdf>

■ **MI Administrativa:**

Fornecer informação para apoio à gestão do documento de arquivo electrónico, tal como quando e como foi criado, tipo de ficheiro e outra informação técnica, e quem tem privilégios de acesso;

■ **MI Estrutural:**

Relaciona hierarquicamente os diferentes OD que fazem parte de um mesmo documento. O esquema METS (*Metadata Encoding & Transmission Standard*) permite agrupar metainformação descritiva, administrativa e estrutural sobre objectos guardados num repositório digital.

O sítio Web da Library of Congress inclui a página oficial do METS: <URL: http://www.loc.gov/standards/mets>.

O esquema METS foi escolhido no âmbito do RODA, por permitir a codificação da MI relativa ao OD, nomeadamente a sua estrutura hierárquica, o seu nome e localização.

■ **MI Técnica:**

Descreve as características técnicas dos ficheiros e dos seus formatos, pelo que é a mais específica, por depender do formato do OD. Um exemplo deste tipo de MI é o esquema baseado na norma ANSI/NISO Z39.87 - *Metadata for Images in XML schema*, define um conjunto normalizado de elementos de metainformação para imagens digitais;

O sítio Web da Library of Congress inclui a página oficial da ANSI/NISO Z39.87: <URL:http://www.loc.gov/standards/mix>

O RODA utiliza o esquema baseado na norma ANSI/NISO Z39.87 no caso das imagens digitais.

■ **MI de preservação:**

Contém informação necessária para arquivar e preservar o OD. O esquema PREMIS (*PREservation Metadata Implementation Strategies*), produto da parceria entre a OCLC (*Online Computer Library Center*) e a RLG (*Research Libraries Group*) é o mais conhecido,

O sítio Web da Library of Congress inclui a página oficial do PREMIS: <URL:http://www.loc.gov/standards/premis>

O RODA definiu o PREMIS como o esquema MI de preservação a ser utilizado no âmbito do projecto.

Existem, adicionalmente, algumas subcamadas de tipos de metainformação, que são aqui identificadas em separado:

- MI de gestão de controlo e direitos: que tem a ver com os direitos de propriedade intelectual;
- MI de contexto de produção;
- MI de controlo e gestão.

O organismo deve fazer uma descrição de todos os esquemas, modelos ou elementos de MI utilizados ou que lhes interesse utilizar, considerados suficientes. Para tal deve produzir uma tabela que contenha essa informação.

É preciso ter em conta que, no contexto da preservação digital, existem elementos de MI que devem ser considerados obrigatórios. Para tal, é avançada, no quadro abaixo, uma proposta que especifica os elementos recomendados, retirados do esquema *PREMIS*⁶.

A primeira coluna indica o sujeito do conjunto dos atributos; a coluna ID apresenta o número do atributo de acordo com a ordenação existente no dicionário de dados do *PREMIS*; a coluna atributo identifica as unidades semânticas; a coluna Descrição estabelece o âmbito dos atributos.

⁶ O Esquema PREMIS está disponível em <http://www.oclc.org/research/projects/pmwg/>

	Id.	Atributo (<i>semantic unit</i>)	Descrição
Objecto	1	• <i>{objectIdentifier}</i>	Uma designação para identificar de forma unívoca um objecto num arquivo digital.
	1.1	• <i>objectIdentifierType</i>	
	1.2	• <i>objectIdentifierValue</i>	O valor do identificador do objecto.
	2	• <i>preservationLevel</i>	Um valor a indicar o conjunto de funcionalidades de preservação a aplicar ao objecto. Pode ser adequado consoante aquilo que se pretende que o OD assegure.
	3	• <i>objectCategory</i>	A categoria de objecto a que a MI é aplicável. A profundidade de MI aplicável, de acordo com o elemento 2.
	4	• <i>{objectCharacteristics}</i>	As propriedades técnicas de um ficheiro ou <i>bitstream</i>
	4.1	• <i>compositionLevel</i>	Indicação sobre se o OD está sujeito a uma ou mais operações de descodificação ou decomposição. No caso do OD estar comprimido ou agregado a outros OD e ser necessário isolá-lo.
	4.4	• <i>{format}</i>	Organização de um ficheiro ou <i>bitstream</i> em que se entende por formato a organização da informação digital de acordo com condições pré-definidas. Registrar o formato para cada ficheiro que compõem o OD.
	4.4.1.1	• <i>formatName</i>	Designação do formato do ficheiro ou <i>bitstream</i> .
	7	• <i>{storage}</i>	Informação sobre como e onde o OD é armazenado no Arquivo Digital.
	7.1.1	• <i>contentLocationType</i>	Esquema de utilização utilizado (os meios para aceder à localização).
	7.1.2	• <i>contentLocationValue</i>	A referência à localização do conteúdo.
7.2	• <i>storageMedium</i>	O suporte físico em que o objecto está armazenado.	
Evento	14	• <i>{eventIdentifier}</i>	Uma designação para identificar de forma única um evento decorrido dentro do Arquivo Digital.
	14.1	• <i>eventIdentifierType</i>	Designação para o domínio em que o Identificador do evento é único (ex. sistema de armazenamento, integração, monitorização, etc. ou um identificador único sequencial para todo o Arquivo Digital).
	14.2	• <i>eventIdentifierValue</i>	Valor do identificador.
	15	• <i>eventType</i>	Categorização da natureza do evento (de acordo com as funcionalidades asseguradas pelo Arquivo Digital e que devem constar de uma lista de autoridade).
	16	• <i>eventDateTime</i>	A data ou intervalo de datas em que o evento decorreu.
Actor	21	• <i>{agentIdentifier}</i>	A designação usada para identificar de forma única um actor dentro de um sistema de arquivo digital.
	21.1	• <i>agentIdentifierType</i>	Designação do domínio em que o id do actor é único.
	21.2	• <i>agentIdentifierValue</i>	Valor do id do actor.
Direitos	28	• <i>{permissionGranted}</i>	As acções permitidas pelo depositante ao AD.
	28.1	• <i>act</i>	Acção que o AD pode empreender.
	28.3.1	• <i>startDate</i>	Data de início do acordo.
	28.3.2	• <i>endDate</i>	Data em que termina o acordo.

A metainformação estruturada deve basear-se numa sintaxe que facilite a pesquisa e a identificação por parte tanto do utilizador como da máquina. A sintaxe mais utilizada na actualidade é o XML (eXtensible Markup Language), desenvolvida pelo W3C (World Wide

O Arquivo Nacional da Holanda produziu um *Digital Preservation Testbed White Paper* sobre o XML
www.digitaleduurzaamheid.nl/bibliotheek/docs/white-paper_xml-en.pdf

Web Consortium), uma forma estendida de HTML que permite a definição localizada de etiquetas e a troca facilitada de informação. Actualmente, existem cada vez mais ferramentas para XML, e esta linguagem tem assumido um papel cada vez mais crucial na troca de uma variedade de dados na Web.

Apresentamos um exemplo da utilização do XML no **Anexo C**.

7 Estratégias de Preservação mais comuns

Este capítulo identifica as estratégias de preservação mais utilizadas, a saber, preservação de tecnologia, emulação, monitorização de suportes e formatos, encapsulamento e a migração e transferência de suportes. Neste âmbito, são referidas as características, bem como os aspectos positivos e negativos de cada uma destas estratégias.

A selecção da(s) estratégia(s) de preservação apropriada(s) deve ser, como já foi referido, o resultado de um esforço de colaboração entre as unidades orgânicas referentes ao arquivo e à informática, com a participação de todas as unidades orgânicas afectadas pelo processo ou que produzam documentos de arquivo electrónicos.

Apresenta-se, seguidamente, o conjunto das estratégias de preservação mais comumente utilizadas, elencando características, vantagens e desvantagens de cada uma delas. As estratégias apresentadas podem ser utilizadas a curto, médio e longo prazo, cabendo à entidade produtora definir a sua calendarização e utilização no seu PPD

7.1 Preservação de tecnologia

A preservação de tecnologia implica a conservação e manutenção de todo o *hardware* e *software* necessários à correcta apresentação dos OD. Neste âmbito, o foco da preservação concentra-se, não no objecto conceptual, mas sim na preservação do objecto digital na sua forma original.

A preservação de tecnologia apresenta algumas desvantagens, nomeadamente:

- O facto de qualquer plataforma tecnológica acabar por se tornar obsoleta;
- A existência de dificuldades na gestão do espaço físico;
- A manutenção e custos de operação;
- O facto do acesso à informação estar limitado a alguns locais físicos do globo e com condicionalismos acrescidos ao nível da reutilização da informação.

7.2 Emulação

A emulação corresponde à utilização de um *software* - o emulador - capaz de reproduzir o comportamento de uma plataforma de *hardware* e/ou *software*, numa outra plataforma que, à partida, seria incompatível.

As principais vantagens da emulação incluem o facto de preservar, de forma fiel, as características e as funcionalidades

Tipos de emulação possíveis:

Emulação de aplicações informáticas (ex: *Lotus 123*);
Emulação de sistemas operativos (ex: emulador de MS Windows para Linux);
Emulação de plataformas de hardware (ex: emulador de ZX Spectrum para PC).

Jeff Rothenberg, defensor da emulação, propõe um modelo teórico passível de emular plataformas actuais em computadores futuros, através da conservação do objecto digital juntamente com o software necessário à sua execução/apresentação e na criação de uma especificação escrita numa linguagem independente da plataforma de hardware que suporta a execução desse software.
www.clir.org/pubs/reports/rothenberg/pub77.pdf

do objecto digital original e, embora se centre na preservação do objecto lógico no seu formato original, não sofre de alguns dos problemas da estratégia de preservação de tecnologia (Ex.: envelhecimento do *hardware*).

No entanto, o emulador pode tornar-se obsoleto e a sua utilização pressupõe que os utilizadores do futuro sejam capazes de operar adequadamente aplicações e sistemas operativos há muito desaparecidos (Ex.: num futuro próximo será difícil conceber que os utilizadores estejam aptos a enfrentar as particularidades do sistema operativo MS-DOS).

T. Hendley defende que a emulação apenas deveria ser utilizada em situações em que haja interesse em preservar o ambiente tecnológico original ou quando não é possível converter os OD para formatos actuais.

David Bearman considera que a emulação pretende conservar o objecto errado, i.e., as funcionalidades dos sistemas de informação em vez dos DAE, pois são estes últimos que têm valor evidencial, através do seu conteúdo, estrutura e contexto.

O Arquivo Nacional dos Países Baixos produziu um *Digital Preservation Testbed White Paper* sobre a Emulação <www.digitaleduurzaamheid.nl/bibliotheek/docs/white_paper_emulatie_EN.pdf>

7.3 Monitorização de suportes e formatos

A monitorização prevê processos de verificação automática, manual e semi-automática dos OD, sendo as duas últimas opções mais realistas por razões de custos.

As preocupações da monitorização prendem-se com as seguintes questões:

- Tempo de vida estimado dos suportes;
- Tempo médio de prevalência de uma versão de aplicação informática (3 anos);
- Retrocompatibilidade assegurada pelos fabricantes (em média 3 versões anteriores).

7.4 Encapsulamento

O encapsulamento consiste em preservar, juntamente com o objecto digital, toda a informação necessária e suficiente para permitir o futuro desenvolvimento de conversores, visualizadores ou emuladores (por exemplo, a descrição formal e detalhada do formato do objecto preservado).

O encapsulamento está orientado a objectos que apenas serão acedidos num futuro longínquo. Permite adiar a responsabilidade de preservação, e o desenvolvimento futuro de visualizadores, migradores ou emuladores.

No entanto, é preciso estar ciente que objectos complexos possuem especificações complexas e que uma especificação incompleta poderá ter um efeito desastroso para a preservação do OD.

7.5 Transposição de Formatos e Suportes (Migração e transferência de suporte)

A transposição de formatos e suportes refere-se à transferência de documentos contidos num determinado suporte ou formato para outro suporte ou formato mais actualizado. É o processo responsável pela reorganização dos elementos de informação que constituem um OD. Centra-se, sobretudo, na preservação do seu conteúdo intelectual, ou seja, na preservação do objecto conceptual e na verificação frequente da integridade dos suportes físicos.

O principal objectivo é evitar a obsolescência tecnológica, mantendo os OD compatíveis com tecnologias actuais, de forma a permitir a sua interpretação sem necessidade de recorrer a artefactos menos convencionais.

A transposição prevê processos de:

- Refrescamento a nível de suportes (exemplo: de Disquete para CD-R ou de CD-R para DVD-R), nos casos em que existam e sejam utilizados estes suportes (cf. Cap. 5).
- Processos de migração entre formatos (exemplo: de Word 97 para Word 2000);
- Transposição conjunta de formatos e suportes.

O Arquivo Nacional dos Países Baixos produziu um *Digital Preservation Testbed White Paper* sobre a Migração <www.digitaleduurzaamheid.nl/bibliotheek/docs/Migration.pdf>

Esta estratégia apresenta, no entanto, algumas desvantagens como sejam a:

- Probabilidade de algumas propriedades dos OD não serem correctamente transferidas para o formato de destino adoptado;
- Existência de incompatibilidades entre os formatos de origem e destino;
- Inadequação dos conversores;
- Obsolescência de formatos.

Adicionalmente, durante a transposição de documentos para novos formatos há informação que se perde ao nível da estrutura, da metainformação e, por vezes, do conteúdo.

Face a estas situações de perda de informação, deve recorrer-se a estratégias nas quais se incluem documentar exaustivamente o processo de migração

(metainformação) para que fique claro o que se perdeu, quando e como se perdeu. Desta forma tenta-se garantir que o que se perde não compromete a autenticidade e fidedignidade do documento. Este aspecto de confirmação da perda de informação em objectos migrados foi já referida no capítulo 2.

Esta questão é particularmente evidente em documentos compostos e complexos como, por exemplo:

- Bases de dados (muitas tabelas com ligações entre si e circulação constante de informação);
- Documentos multimédia (páginas *web* em que as actualizações e modificações são frequentes).

Apesar das desvantagens apresentadas, **esta é de longe a estratégia de preservação mais aplicada até à data e a única que tem vindo a dar provas da sua eficácia.**

Existem algumas variantes no caso específico da transferência de formatos, a saber:

■ Actualização de versões:

Caracteriza-se pela actualização dos formatos utilizando *software* retrocompatível (por exemplo, a utilização do Word 6 para actualizar um documento criado com o Word 5);

■ **Conversão para formatos concorrentes:**

Permite ultrapassar o risco de descontinuidade de formatos, convertendo OD para formatos análogos, independentemente da aplicação utilizada na sua criação (por exemplo, converter um documento Word para PDF);

■ **Normalização:**

Corresponde à migração para um número reduzido de formatos compatíveis, o que poderá evitar futuras complicações a nível de direitos de autor ou pagamento de *royalties*. Promove, também, a interoperabilidade entre sistemas distintos.

O facto de serem utilizados formatos abertos e independentes da plataforma permite que diferentes configurações de *hardware* e *software* sejam capazes de os interpretar. Existem vários tipos de normas:

● Normas *de facto*:

Quando o produto está firmemente estabelecido no mercado, a compatibilidade das outras aplicações é medida referenciando esse produto e as alterações de compatibilidade dependem do proprietário (ex: Word);

● Especificações públicas:

Quando algumas empresas líderes de mercado criam um consórcio para definir uma norma de ligação que possibilita o desenvolvimento de produtos compatíveis;

● Normas *de jure*

Quando organismos oficiais promovem o consenso acerca de uma especificação, que depois se torna uma norma oficial (ex: ISO, NP, etc.)

O tempo de intervalo entre transferências de formato está dependente da evolução da tecnologia, bem como do facto de usarmos ou não formatos normalizados, podendo variar entre os 7 e os 10 anos.

8 Conclusão

A preservação digital é uma função que não pode ser evitada. Cada vez mais as organizações dependem, a médio e longo prazo, de informação produzida e mantida electronicamente. Tal dependência deve-se a motivos que se prendem com a operação da organização, salvaguarda dos seus interesses, incrementação da transparência administrativa e credibilidade da instituição relativamente aos seus *stakeholders*, entre os quais se contam o Governo e o cidadão.

A organização pode ter a opção de não preservar a sua informação digital. Julgamos, no entanto, ser esse risco inaceitável atendendo às circunstâncias referidas. Embora tal implique custos que podem ser significativos, compete à organização ponderar seriamente o que tem a perder caso decida nada fazer.

Este documento é um primeiro passo para guiar as organizações na tarefa de preservação digital. A DGARQ, enquanto órgão responsável pela coordenação e execução da política arquivística nacional, posiciona-se como parte interessada nesse processo, disponibilizando o seu conhecimento na área para auxiliar as organizações a construir e implementar processos de preservação, sem se pretender substituir de forma alguma às responsabilidades que toda e qualquer organização deve manter relativamente ao seu património arquivístico, independentemente do suporte em que este se encontra.

Anexos

A Lista de requisitos para a prática de preservação digital

Indicar na coluna *Comentário* a seguinte informação:
 Se o requisito for considerado cumprido colocar “S”
 Se o requisito for considerado não cumprido colocar “N”

#	Designação	Obrigatoriedade	Comentário
1	O organismo dispõe de:		
1.1	Plano de Classificação	Aconselhável	
1.2	Tabela de Selecção	Aconselhável	
2	Foi elaborada uma análise e caracterização da informação electrónica produzida pela organização no âmbito:		
2.1	do sistema de gestão documental	Obrigatório	
2.2	das séries documentais	Obrigatório	
3	A análise anterior permitiu:		
3.1	definir classes de informação a serem preservadas	Obrigatório	
3.2	identificar as características dos ficheiros por cada classe de informação definida	Obrigatório	
3.3	identificar as características técnicas dos Documentos de Arquivo Electrónico	Obrigatório	
3.4	identificar os seguintes factores inerentes aos Documentos de Arquivo:		
3.4.1	Usabilidade	Obrigatório	
3.4.2	Complexidade	Obrigatório	
3.4.3	Escalabilidade	Obrigatório	
4	Foram elaborados e analisados estudos no intuito de identificar os custos de soluções de Preservação na fase de desenvolvimento e implementação em termos de:		
4.1	<i>software</i>	Obrigatório	

#	Designação	Obrigatoriedade	Comentário
4.2	<i>hardware</i>	Obrigatório	
4.3	preparação e adaptação de infraestruturas	Obrigatório	
4.4	recursos humanos	Obrigatório	
5	Foram analisados e identificados os custos de funcionamento, manutenção e actualização em termos de:		
5.1	parque informático	Obrigatório	
5.2	infraestrutura	Obrigatório	
5.3	formação	Obrigatório	
5.4	recursos humanos	Obrigatório	
6	Foi produzida uma avaliação e escolha de:		
6.1	formatos normalizados de preservação a longo prazo, para cada formato de ficheiro a preservar?	Obrigatório	
6.2	aplicações informáticas:		
6.2.1	com capacidade de incremento de utilizadores simultâneos	Obrigatório	
6.2.2	com especificações abertas	Obrigatório	
6.2.3	que permitam a exportação para formatos normalizados	Obrigatório	
6.3	soluções armazenamento:		
6.3.1	que permitam o escalamento da capacidade de armazenamento	Obrigatório	
6.3.2	que incluam sistemas de <i>backup</i>	Obrigatório	
6.3.3	que incluam sistema de rotinas de auditoria e monitorização	Obrigatório	
7	Analisaram-se e adaptaram-se os esquemas de metainformação:		
7.1	já utilizados no organismo, com elementos de metainformação:		
7.1.1	Descritiva	Obrigatório	

#	Designação	Obrigatoriedade	Comentário
7.1.2	Técnica	Obrigatório	
7.1.3	Estrutural	Obrigatório	
7.1.4	Preservação	Aconselhável	
7.2	a serem implementados e que contenham elementos de metainformação:		
7.2.1	Descritiva	Obrigatório se aplicável	
7.2.2	Técnica	Obrigatório se aplicável	
7.2.3	Estrutural	Obrigatório se aplicável	
7.2.4	Preservação	Obrigatório se aplicável	
8	Foram afectados os recursos definidos para implementação	Obrigatório se aplicável	
9	Foi definido um cronograma para execução da implementação	Aconselhável	
10	Foram definidas medidas de auditoria internas a nível de:		
10.1	procedimentos de preservação	Obrigatório	
10.2	formatos para preservação	Obrigatório	
10.3	soluções de software alternativas para um determinado formato	Obrigatório	
10.4	soluções de armazenamento	Obrigatório	
10.5	esquemas de metainformação	Obrigatório	
11	Foram definidas medidas de monitorização interna dos documentos de arquivo electrónicos preservados, em termos de:		
11.1	integridade	Obrigatório	
11.2	funcionalidade	Obrigatório	
11.3	estrutura	Obrigatório	
11.4	conteúdo	Obrigatório	

#	Designação	Obrigatoriedade	Comentário
11.5	metainformação associada	Obrigatório	
12	Foi produzido um documento que defina o PPD de acordo com as soluções propostas anteriormente	Obrigatório	
13	A solução proposta foi testada preventivamente sobre cópias de documentos de arquivo electrónicos, para despistar falhas	Aconselhável	
14	Foram produzidas cópias de segurança dos documentos de arquivo electrónicos a preservar?	Aconselhável	
15	A estratégia de preservação digital foi implementada de acordo com o definido no PPD desenvolvido pelo organismo?	Obrigatório	

B Questionário

Instruções para o preenchimento do questionário

Parte I - Sistema de gestão documental

Esta primeira parte respeita aos recursos de manutenção e apoio ao sistema de gestão documental.

1. Controlo e Segurança

a) Segurança do Sistema:

Refere-se às funcionalidades do sistema que asseguram a protecção dos dados.

O objectivo é identificar as medidas de protecção de dados utilizadas pelo sistema. Estas podem consistir em *Firewalls* (dispositivo ou aplicação de controlo de tráfego de dados de e para a rede), sistemas de detecção de intrusos, e outros a especificar pelo Organismo. Um sistema pode utilizar um ou mais medidas de segurança.

a) Rotinas de Auditoria:

Permitem o registo de procedimentos efectuados a nível do sistema.

Pretende-se saber:

- Quais os elementos que estão sujeitos a estas rotinas de auditoria;
- Se essas rotinas dão origem a relatórios;
- Que tipo de informação consta nesses relatórios;
- Se esses relatórios são protegidos contra alterações ou eliminações não autorizadas;
- Quem tem acesso a esses relatórios.

2. Armazenamento

a) Sistemas de Informação

Inclui questões relativas ao armazenamento de OD originais.

O objectivo é identificar onde estão instalados os repositórios de dados que se pretendem preservar a longo prazo, quais as arquitecturas de armazenamento utilizados para o efeito (ver Cap. 5.2 para informações mais pormenorizadas) e onde se localizam. Esta localização poderá referir-se à sala, cofre ou empresa com quem se estabeleceu um protocolo de *outsourcing*.

Pretende-se, igualmente, saber que outros dispositivos portáteis são também utilizados fora do local, qual a infra-estrutura de acesso aos repositórios de dados (ver os Cap. 5.3.1 para informações mais pormenorizadas), e se é possível a sincronização de dados com outro repositório.

Finalmente, é solicitada a identificação dos formatos lógicos dos dados dos diferentes recursos da série e categorias utilizados (que estão identificados no Cap. 4.1 e listados no anexo C), e as respectivas aplicações informáticas (de acordo com o Cap. 4.2), a dimensão, o espaço ocupado e a taxa de crescimento de OD em unidades binárias, medidas em *terabytes*.

b) Backups

Estas questões referem-se ao armazenamento das réplicas ou cópias de segurança produzidas a partir dos OD originais.

O objectivo é identificar onde estão instalados os repositórios das réplicas dos dados que se pretendem preservar a longo prazo, quais as arquiteturas de armazenamento utilizados para o efeito (ver Cap. 5.3 para informações mais pormenorizadas) e onde se localizam. Esta localização poderá referir-se à sala, cofre ou empresa com quem se estabeleceu um protocolo de *outsourcing*.

Finalmente, é pedida a dimensão, em unidades binárias medidas em *terabytes*.

3. Funções Administrativas

Inclui questões respeitantes aos recursos de manutenção e apoio ao sistema.

a) Administração Geral

Relativa à gestão de parâmetros, salvaguarda e reconstituição do sistema, bem como à administração de utilizadores.

Pretende-se a identificação das UO que administram (normalmente a UO ligada à informática) e das que detêm a propriedade dos sistemas de informação.

4. Outras funções

a) Encriptação

O objectivo é saber se a encriptação é utilizada. Em caso afirmativo pretende-se identificar qual(ais) o(s) sistema(s) de encriptação utilizado(s) (ex: MD5, SSH, etc.), e em que tipos de documentos são utilizados.

5 Requisitos não funcionais

a) Desempenho e escalabilidade

Inclui questões relacionadas com a capacidade de satisfazer a gama de requisitos para o qual o sistema foi concebido.

A **questão 28** reporta-se às capacidades do sistema de gestão documental, actuais e possíveis no futuro, relativamente ao número de utilizadores possíveis em simultâneo; à velocidade de retorno de pesquisa e de carregamento para visualização (medidas em segundos); e ainda à capacidade de armazenamento (medida em *terabytes*).

b) Contratação de serviços externos

Reflecte a opção de utilizar prestadores de serviços externos ao organismo para gestão do sistema.

Pretende-se saber se a subcontratação de serviços é utilizada para a administração e alojamento dos sistemas de informação. Adicionalmente, pede-se a descrição do tipo e duração desse(s) contrato(s).

c) **Conservação a longo prazo** (entendida em termos de preservação digital como um período superior a sete anos).

O objectivo é identificar:

- Quais as normas e legislação (Ex. art.º 17º de RCM 5/90, de 28 de Fevereiro) relativas a requisitos ambientais observados;
- Quais as condições ambientais específicas para armazenamento a longo prazo, caso existam;
- Se a monitorização dos sistemas e suportes de armazenamento é feita regularmente.

6 Metainformação

Pretende-se obter informação sobre a utilização de esquemas ou modelos de MI. Em caso afirmativo, solicita-se que se faça a identificação dos mesmos, referindo qual(ais) o(s) sistema(s) de informação em que são utilizados. Caso contrário, deverá ser produzida uma lista com os elementos de MI utilizados em cada sistema de informação.

Parte II - Série/Actividade suportada

Esta parte pretende caracterizar cada uma das séries documentais que contêm OD. Caso a Organização não utilize qualquer plano de classificação ou não possua tabela de selecção, deverá orientar estas questões no âmbito das actividades suportadas pelos OD.

1. Identificação da organização produtora

Campo destinado à identificação da entidade produtora da documentação, objecto de avaliação.

O objectivo é fornecer informação sobre o sector de actividade a que pertence o produtor da série, a sua identificação formal e unívoca e o nome do Organismo, unidade orgânica ou pessoa individual (desde que não coincidente com o anterior) responsável pela compilação e transmissão dos documentos da série.

Os sectores de actividade possíveis (mas não exclusivos) são: Sector Privado (SP); Sector Público Empresarial (SPE); Administração Central (AC); Administração Local (AL); Administração Regional (RA); Segurança Social (SS); Governo (GOV); Tribunais (TRB); Presidência da República (PR); Assembleia da República (AR); Organizações Não Governamentais (ONG); Instituições de Solidariedade Social (ISS); Organismos Comunitários (OC); Organizações Inter-Governamentais (OIG)

2. Referenciação da série – Pretende fornecer informação identificativa da série ou da actividade suportada pela documentação em causa.

a) Número de Referência – Número sequencial exclusivo que identifica a série ou actividade suportada. No caso de não existir tabela de selecção passar ao grupo 3.

b) Código de classificação – Código exclusivo da série, que permite a sua localização intelectual, tal como se apresentam no plano de classificação.

c) Identificação da série documental/actividade - Pretende-se obter informação descritiva sobre a série documental, nomeadamente ao nível de títulos, actividade suportada e assunto.

d) **Série relacionadas** – São séries que contêm documentos e/ou informação que diz igualmente respeito ou outra(s) série(s).

Preencher no caso da série em análise conter informação síntese (o conteúdo informativo existente numa ou em mais séries documentais é sintetizado ou resumido numa outra série), duplicada (uma série documental contém, no todo ou em parte, o mesmo conteúdo informativo que outras séries) ou complementar (o conteúdo informativo de uma série está relacionado com o de outra série, acrescentando-lhe informação adicional).

3. Controlo e Segurança

a) Acesso

O objectivo é saber se existem restrições de acesso à informação da série/actividade suportada.

b) Salvaguarda e Recuperação

Pretende-se conhecer a existência de cópias de segurança e relacionar os tipos de cópias de segurança com o intervalo de tempo entre cada uma. As cópias podem ser integrais (é produzida sobre a totalidade dos OD da série/actividade suportada), parciais (produzidas sobre parte da série, normalmente os OD mais recentes) e distribuídas (são produzidas réplicas que são armazenadas em localizações diferente). Os períodos de intervalo podem ser diários, semanais, mensais, etc.

c) Categorias de Segurança

Visa a identificação da categoria de segurança mais alta existente na série documental/actividade relacionada.

4 Avaliação

a) Tabelas de Selecção

O objectivo é identificar o tempo de retenção definido para os recursos da série e a acção a tomar, findo o prazo de conservação.

A eliminação global significa que:

- A série tem um valor secundário nulo;
- A informação contida na série foi objecto de publicação, com um número de depósito legal e no arquivo definitivo se conserve, pelo menos, um exemplar da publicação;
- Os dados essenciais da série estão incluídos noutra série de conservação permanente no mesmo organismo ou num organismo distinto.

A Conservação permanente global significa que:

- A série tem valor secundário relevante e se verifica que o conjunto da informação não está publicado nem é recuperável noutras séries de conservação permanente.

A Conservação permanente parcial significa que:

- A série é muito extensa, com um conteúdo informativo significativamente homogéneo e repetitivo e cujo interesse informativo não é tão grande que justifique a sua conservação permanente global. Contudo também se entenda que a eliminação da totalidade da série é uma opção demasiado radical fazendo sentido conservar uma amostra desta;
- A série é constituída por OD cujo interesse informativo é muito desigual.

A Revisão significa que:

- Não há certeza sobre o destino a dar à série, pelo que tal decisão será adiada para quando terminar o prazo de conservação.

b) Análise de risco – é necessária, devido à possibilidade de perda ou corrupção dos dados, resultados de danos físicos.

Pretende-se saber a percentagem de tolerância da organização para a perda dos OD que compõem a série em causa, e se esta contém documentos vitais, com vista a identificar se se trata de uma série candidata a preservação digital.

5 Outras Funções

a) Assinaturas Digitais

O objectivo é obter dados referentes à validação duma assinatura electrónica ou digital anexa aos documentos da série ou certificado digital.

Parte I

Sistema de gestão documental

1. Controlo e Segurança

a) Segurança do Sistema

1. Que medidas de segurança são usados pelo sistema para proteger os dados?
Firewall <input type="checkbox"/> Sistemas de detecção de intrusos <input type="checkbox"/> Outros <input type="checkbox"/> Quais? _____

b) Rotinas de Auditoria

2. São efectuadas rotinas de auditoria sobre:
Sistemas de informação <input type="checkbox"/> Dados <input type="checkbox"/> Documentos electrónicos <input type="checkbox"/> Utilizadores <input type="checkbox"/>
3. São produzidos relatórios de auditoria a partir das rotinas?
Sim <input type="checkbox"/> Não <input type="checkbox"/> Alguns <input type="checkbox"/>
4. Que informação contém os relatórios de auditoria sobre os dados?
A nível do sistema
A nível dos documentos
A nível dos dados
A nível de utilizadores
5. Os relatórios de auditoria são protegidos?
Sim <input type="checkbox"/> Não <input type="checkbox"/>

6. Quem tem direitos de acesso e administrativos aos relatórios?
Administradores de sistema <input type="checkbox"/> Arquivistas/Gestores de Informação <input type="checkbox"/> Informáticos <input type="checkbox"/> Outros <input type="checkbox"/>
7. Durante quanto tempo são guardados os relatórios de auditoria?

2. Armazenamento

a) Sistemas de Informação

8. Onde estão instalados ou depositados os sistemas de informação?
Servidores <input type="checkbox"/> Sistema de armazenamento <input type="checkbox"/>
9. Indique o tipo:
10. Suporte físico dos recursos:
CAS <input type="checkbox"/> NAS <input type="checkbox"/> SAN <input type="checkbox"/> Discos ópticos <input type="checkbox"/> Tapes <input type="checkbox"/> Discos rígidos <input type="checkbox"/> Outros: _____
11. Localização topográfica dos suportes utilizados:
12. Existem dados armazenados em aparelhos que são normalmente utilizados fora do local, como portáteis ou PDA? Quais os aparelhos utilizados?
Computadores Portáteis <input type="checkbox"/> Discos Externos <input type="checkbox"/> Pen-drives <input type="checkbox"/> PDA <input type="checkbox"/> Outros <input type="checkbox"/> _____

13. Infra-estrutura para acesso:		
Em linha <input type="checkbox"/>		
Near line <input type="checkbox"/>		
Fora de linha <input type="checkbox"/>		
14. Os dados estão sincronizados com outro repositório? No mesmo edifício ou externo?		
15. Formatos e categorias de formatos de dados:		
16. Aplicações utilizadas no âmbito desses formatos		
Nome do Formato	Nome da Aplicação	Proprietário (S/N) <input type="checkbox"/> Especificações Abertas (S/N) <input type="checkbox"/>
17. Dimensão:		
18. Taxa de crescimento de OD:		

b) Backups

18. Onde estão instalados ou depositados os <i>backups</i> ?	
Servidores <input type="checkbox"/>	Sistema de armazenamento <input type="checkbox"/>
19. Indique o tipo:	
20. Suporte físico dos recursos:	
CAS <input type="checkbox"/>	
NAS <input type="checkbox"/>	
SAN <input type="checkbox"/>	
Discos ópticos <input type="checkbox"/>	

Tapes <input type="checkbox"/>
Discos rígidos <input type="checkbox"/>
Outros: _____
21. Localização topográfica dos suportes utilizados no <i>backup</i> ?
22. Dimensão:

3. Funções Administrativas

a) Administração Geral

23. Indique a Unidade orgânica/funcional que administra os sistemas de informação:
24. Indique a unidade orgânica/funcional proprietária dos sistemas:

4. Outras funções

a) Encriptação

25. É utilizada encriptação?
Sim <input type="checkbox"/> Não <input type="checkbox"/>
26. Se sim, indique quais os sistemas de encriptação que utiliza:
27. Se sim a que documentos:
Documentos de texto [PDF, WORD] <input type="checkbox"/>
Imagens <input type="checkbox"/>
Mensagens de correio electrónico <input type="checkbox"/>
Outros <input type="checkbox"/> Quais? _____

5 Requisitos não funcionais

a) Desempenho e escalabilidade

28. Capacidades actuais e possibilidade de expansão em termos de:
Número de utilizadores simultâneos
Velocidade de pesquisa simples/complexa
Velocidade de recuperação e visualização
Capacidade de armazenamento

b) Contratação de serviços externos

29. A administração dos sistemas de informação é subcontratada?
Sim (no todo ou em parte) <input type="checkbox"/> Não <input type="checkbox"/>
30. O alojamento dos sistemas de informação é subcontratada?
Sim (no todo ou em parte) <input type="checkbox"/> Não <input type="checkbox"/>
31. Tipo de contrato:
Período temporal: Anos <input type="checkbox"/> Horas <input type="checkbox"/> Qual? _____

c) Conservação a longo prazo

32. São observados requisitos ambientais previstos em normativos e legislação?
Sim <input type="checkbox"/> Não <input type="checkbox"/>
33. Se sim, indique as normas e a legislação:
34. Há condições ambientais para suportar longos períodos de armazenamento?
Sim <input type="checkbox"/> Não <input type="checkbox"/>

35. Se sim, indique quais as condições:
36. É feita regularmente monitorização dos suportes/sistemas de armazenamento?
Sim <input type="checkbox"/> Não <input type="checkbox"/>

6 Metainformação

37 Utilizam algum esquema/modelo de metainformação?
Sim <input type="checkbox"/> Não <input type="checkbox"/>
38. No caso de ser(em) utilizado(s) esquema(s)/modelo(s), indique quais:
39. No caso de não ser utilizado um esquema/modelo, indique que elementos de Metainformação são utilizados:

40. Mapeamento dos sistemas de informação com os respectivos esquemas/modelos de metainformação utilizados.	
Sistema de Informação	Esquema/modelo de Metainformação

Parte II

Série/Actividade suportada

1. Identificação da organização produtora

1. Sector de Produção
2. Identificação do Produtor
3. Designação do Produtor

2. Referenciação da série

a) Número de Referência

4. Número de Referência (No caso de existir tabela de selecção passar ao grupo 3)

b) Código de classificação

5. Código de Classificação

c) Identificação da série documental/actividade

6. Título da série documental
7. Actividade suportada
8. Âmbito e conteúdo da série

d) Série relacionadas

9. Título da Série/Subsérie relacionada	Possui informação
1.	Síntese <input type="checkbox"/> Complementar <input type="checkbox"/> Duplicada <input type="checkbox"/>
2.	Síntese <input type="checkbox"/> Complementar <input type="checkbox"/> Duplicada <input type="checkbox"/>

3. Controlo e Segurança

a) Acesso

10. O acesso à informação é:
<p>Livre para todos os funcionários <input type="checkbox"/></p> <p>Livre para todos os funcionários com algumas restrições <input type="checkbox"/></p> <p>Restrito a todos os funcionários <input type="checkbox"/></p> <p>Restrito à maior parte dos funcionários com algumas excepções <input type="checkbox"/></p>

b) Salvaguarda e Recuperação

11. São feitas cópias de segurança?	
Sim <input type="checkbox"/> Não <input type="checkbox"/>	
12. Que tipo de cópias de segurança/Período de intervalo	
Tipo de cópias de segurança	Período de intervalo entre cópias de segurança (ex. diário, semanal, mensal, etc.)
Integrais <input type="checkbox"/>	
Parciais <input type="checkbox"/>	
Distribuídas <input type="checkbox"/>	

c) Categorias de Segurança

13. Tem classificação de segurança?
Sim <input type="checkbox"/> Não <input type="checkbox"/>

14. Qual?
Muito secreto <input type="checkbox"/> Secreto <input type="checkbox"/> Confidencial <input type="checkbox"/>

4 Avaliação

a) Tabelas de Selecção

15. Prazo de Conservação	
16. Destino Final	
Eliminação global <input type="checkbox"/> ; Conservação permanente global <input type="checkbox"/>	Conservação permanente parcial <input type="checkbox"/> ; Revisão <input type="checkbox"/>

b) Análise de risco

17. Quais as tolerâncias para perda ou corrupção de dados?				
0 – 5% <input type="checkbox"/>	6 – 10% <input type="checkbox"/>	11 – 20% <input type="checkbox"/>	21 – 50 % <input type="checkbox"/>	51 - 100% <input type="checkbox"/>
18. Quais os documentos da série considerados vitais?				

5 Outras Funções

a) Assinaturas Digitais

19. Contêm documentos que utilizam a autenticação de assinatura electrónica
Sim <input type="checkbox"/> Não <input type="checkbox"/>

C Exemplo de um ficheiro PREMIS contendo MI técnica NISO Z39.87

```
<?xml version="1.0" encoding="UTF-8"?>
<premis xmlns="http://www.loc.gov/standards/premis/v1"
  xmlns:mix="http://www.loc.gov/mix/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.loc.gov/standards/premis/v1
    http://www.loc.gov/standards/premis/v1/PREMIS-v1-1.xsd
    http://www.loc.gov/mix/http://www.loc.gov/mix/mix.xsd">
  <object>
    <objectIdentifier>
      <objectIdentifierType>Custom</objectIdentifierType>
      <objectIdentifierValue>R2006.01.0001</objectIdentifierValue>
    </objectIdentifier>
    <preservationLevel>full</preservationLevel>
    <objectCategory>representation</objectCategory>
    <relationship>
      <relationshipType>structural</relationshipType>
      <relationshipSubType>hasroot</relationshipSubType>
      <relatedObjectIdentification>
        <relatedObjectIdentifierType>Custom</relatedObjectIdentifierType>
        <relatedObjectIdentifierValue>F2006.01.0001</relatedObjectIdentifierValue>
        <relatedObjectSequence>0</relatedObjectSequence>
      </relatedObjectIdentification>
    </relationship>
    <relationship>
      <relationshipType>structural</relationshipType>
      <relationshipSubType>haspart</relationshipSubType>
      <relatedObjectIdentification>
        <relatedObjectIdentifierType>Custom</relatedObjectIdentifierType>
        <relatedObjectIdentifierValue>F2006.01.0001.0000001</relatedObjectIdentifierValue>
        <relatedObjectSequence>1</relatedObjectSequence>
      </relatedObjectIdentification>
    </relationship>
    <linkingIntellectualEntityIdentifier>
      <linkingIntellectualEntityIdentifierType>Custom</linkingIntellectualEntityIdentifierType>
      <linkingIntellectualEntityIdentifierValue>1</linkingIntellectualEntityIdentifierValue>
    </linkingIntellectualEntityIdentifier>
  </object>
  <object>
    <objectIdentifier>
      <objectIdentifierType>Custom</objectIdentifierType>
      <objectIdentifierValue>F2006.01.0001</objectIdentifierValue>
    </objectIdentifier>
    <preservationLevel>full</preservationLevel>
    <objectCategory>File</objectCategory>
    <objectCharacteristics>
      <compositionLevel>1</compositionLevel>
      <size>3156</size>
      <format>
        <formatDesignation>
          <formatName>image/xml</formatName>
        </formatDesignation>
        <formatRegistry>
          <formatRegistryName>JHOVE</formatRegistryName>
          <formatRegistryKey>XML1.0</formatRegistryKey>
        </formatRegistry>
        <formatRegistry>
          <formatRegistryName>MIME</formatRegistryName>
          <formatRegistryKey>text/xml</formatRegistryKey>
        </formatRegistry>
      </format>
    </objectCharacteristics>
    <creatingApplication>
      <creatingApplicationName>MakeDVDImage.pl</creatingApplicationName>
      <dateCreatedByApplication>2006-06-09T15:34:18Z</dateCreatedByApplication>
    </creatingApplication>
    <storage>
      <contentLocation>
        <contentLocationType>URI</contentLocationType>
        <contentLocationValue>R2006.01.0001/F2006.01.0001.mets</contentLocationValue>
      </contentLocation>
      <storageMedium>DVD</storageMedium>
    </storage>
  </object>
</premis>
```



```

</object>
<object>
  <objectIdentifier>
    <objectIdentifierType>Custom</objectIdentifierType>
    <objectIdentifierValue>F2006.01.0001.0000001</objectIdentifierValue>
  </objectIdentifier>
  <preservationLevel>full</preservationLevel>
  <objectCategory>File</objectCategory>
  <objectCharacteristics>
    <compositionLevel>0</compositionLevel>
    <fixity>
      <messageDigestAlgorithm>MD5</messageDigestAlgorithm>
      <messageDigest>5d5679af155d9752fab9a42e6c724f0a</messageDigest>
    </fixity>
    <size>482190</size>
    <format>
      <formatDesignation>
        <formatName>image/tiff</formatName>
      </formatDesignation>
      <formatRegistry>
        <formatRegistryName>JHOVE</formatRegistryName>
        <formatRegistryKey>TIFF5.0</formatRegistryKey>
      </formatRegistry>
      <formatRegistry>
        <formatRegistryName>MIME</formatRegistryName>
        <formatRegistryKey>image/tiff</formatRegistryKey>
      </formatRegistry>
    </format>
    <significantProperties>
      <mix:mix>
        <mix:BasicImageParameters>
          <mix:Format>
            <mix:MIMEType>image/tiff</mix:MIMEType>
            <mix:ByteOrder>little-endian</mix:ByteOrder>
            <mix:Compression>
              <mix:CompressionScheme>1</mix:CompressionScheme>
            </mix:Compression>
            <mix:PhotometricInterpretation>
              <mix:ColorSpace>0</mix:ColorSpace>
            </mix:PhotometricInterpretation>
            <mix:Segments>
              <mix:StripOffsets>8</mix:StripOffsets>
              <mix:RowsPerStrip>2339</mix:RowsPerStrip>
              <mix:StripByteCounts>481834</mix:StripByteCounts>
            </mix:Segments>
            <mix:PlanarConfiguration>1</mix:PlanarConfiguration>
          </mix:Format>
          <mix:File>
            <mix:Orientation>1</mix:Orientation>
          </mix:File>
        </mix:BasicImageParameters>
        <mix:ImageCreation>
          <mix:ScanningSystemCapture>
            <mix:ScanningSystemSoftware>
              <mix:ScanningSoftware>KofaxImageProcessingPlatformv1.10.018;KIPPTIFFStorageFilterv110.018</mix:ScanningSoftware>
            </mix:ScanningSystemSoftware>
          </mix:ScanningSystemCapture>
          <mix:DateTimeCreated>1992-05-09T15:10:57</mix:DateTimeCreated>
        </mix:ImageCreation>
        <mix:ImagingPerformanceAssessment>
          <mix:SpatialMetrics>
            <mix:SamplingFrequencyUnit>2</mix:SamplingFrequencyUnit>
            <mix:XSamplingFrequency>200</mix:XSamplingFrequency>
            <mix:YSamplingFrequency>200</mix:YSamplingFrequency>
            <mix:ImageWidth>1648</mix:ImageWidth>
            <mix:ImageLength>2339</mix:ImageLength>
          </mix:SpatialMetrics>
          <mix:Energetics>
            <mix:BitsPerSample>1</mix:BitsPerSample>
            <mix:SamplesPerPixel>1</mix:SamplesPerPixel>
          </mix:Energetics>
        </mix:ImagingPerformanceAssessment>
      </mix:mix>
    </significantProperties>
  </objectCharacteristics>

```

```
<creatingApplication>
  <creatingApplicationName>tiffcp(libtiff-tools)</creatingApplicationName>
  <creatingApplicationVersion>3.7.3-1ubuntu1.1</creatingApplicationVersion>
  <dateCreatedByApplication>2006-06-09T15:34:15Z</dateCreatedByApplication>
</creatingApplication>
<originalName>3</originalName>
<storage>
  <contentLocation>
    <contentLocationType>URI</contentLocationType>
    <contentLocationValue>R2006.01.0001/F2006.01.0001.0000001.tiff
  </contentLocationValue>
  </contentLocation>
  <storageMedium>DVD</storageMedium>
</storage>
</object>
</premis>
```

D Recursos Aconselhados

ALVES, Ivone [et. al.] – *Dicionário de terminologia arquivística*. Lisboa: IBL, 1993.

ADOBE - *Tiff developer information site*. 2002 URL: [Consult. 10 Dez. 2007] <<http://partners.adobe.com/public/developer/tiff>>.

ANDERSSON, Ulf - *Short Version of the Sesam Report: Philosophy and Rules Concerning Electronic Archives and Authenticity*. Proceedings of the DLM-Forum on Electronic Records. Brussels, 18-20 December 1996.

ASSOCIAÇÃO PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO -
Glossário da Sociedade da Informação – Versão 2007

BARBEDO, Francisco *et al.* – *RODA: Relatório Final*, Março 2007, Direcção Geral de Arquivos e Universidade do Minho, [Consult. 10 Dez. 2007] URL:<<http://roda.iannt.pt/pt/system/files/roda-relatorioFinal.pdf>>;

BARBEDO, Francisco *et al.* – *RODA: Repositório de Objectos Digitais Autênticos*, in CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 9, Ponta Delgada, Portugal, 2007 – “Bibliotecas e arquivos : informação para a cidadania, o desenvolvimento e a inovação: actas” [CD-ROM]. [S.l.] : Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas, 2007. ISBN 978-972-9067-37-2.

BARBEDO, Francisco – *RODA - De projecto a programa estratégico para a Preservação Digital*, in *Boletim Arquivos Nacionais*, n°19, Janeiro-Março 2007, Lisboa. ISSN1645-5460;

BARBEDO, Francisco; GOMES, Eugénia; HENRIQUES, Cecília – *SIADE caderno 1: Contexto de Suporte* [Em linha]. Lisboa: Instituto dos Arquivos Nacionais; Instituto de Informática, 2000 [Consult. 10 Dez. 2007] Disponível em: <http://www.iannt.pt/downloads/SIADE_Caderno1.pdf>.ISBN 972-8107-59-5

BARBEDO, Francisco; GOMES, Eugénia; HENRIQUES, Cecília; LAGOAS, José – *SIADE caderno 2: Recomendações para a Gestão de Documentos de Arquivo Electrónicos (MOREQ)*. Lisboa: Instituto dos Arquivos Nacionais; Instituto de Informática, 2000 [Consult. 10 Dez. 2007] Disponível em: <URL:http://www.iannt.pt/downloads/SIADE_Caderno2.pdf>. ISBN 972-8107-59-5

BEARMAN, David - *Electronic Evidence. Strategies for Managing Records in Contemporary Organizations*. Pittsburgh: Archives & Museum Informatics. 1994. ISBN: 1-885626-08-8

BEARMAN, David - *Reality and Chimeras in the Preservation of Electronic Records*, in *D-Lib Magazine*, Abril 1999, disponível em: <www.dlib.org/dlib/april99/bearman/04bearman.html>

BEARMAN, David - *Toward a Reference Model for Business Acceptable Communication*, 1994, disponível em: <www.lis.pitt.edu/~nhprc/prog6-5.html>

BESSER, Howard; TRANT Jennifer - *Introduction to Imaging: Issues in Constructing an Image Database*. Santa Monica, CA: Getty Art History Information Program, 1995. ISBN-13: 978-0892363612

CONSELHO INTERNACIONAL DE ARQUIVOS – *Documentos de Arquivo Electrónicos: Manual para arquivistas*. (Estudos, nº 16) trad e adapt. IAN/TT. 2005. [Em linha]. Lisboa: IAN/TT, 2005. [Referência de 8 de Maio de 2006]. [Consult. 10 Dez. 2007] Disponível na internet em: <http://www.ica.org/sites/default/files/ICAEstudo16_PT_4.pdf>.

CONSELHO INTERNACIONAL DE ARQUIVOS - *ISAD(G): Norma geral internacional de descrição arquivística: adoptada pelo Comité de Normas de Descrição, Estocolmo: Suécia, 19-22 de Setembro de 1999*. Trad. Grupo de Trabalho para a Normalização da Descrição em Arquivo. 2ª ed. [Em linha]. Lisboa: IAN/TT, 2004. [Consult. 10 Dez. 2007].Disponível na internet em: <<http://www.iannt.pt/downloads/ISADG.pdf>>.

CONSELHO INTERNACIONAL DE ARQUIVOS - *ISAAR (CPF): Norma internacional de registo de autoridade arquivística para pessoas colectivas, pessoas singulares e famílias*. Trad. Grupo de Trabalho para a Normalização da Descrição em Arquivo. 2ª ed. [Em linha].Lisboa: IAN/TT, 2004. [Consult. 10 Dez. 2007]. Disponível na internet em: <<http://www.iannt.pt/downloads/ISAAR2.pdf>>

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS, *Reference Model for an Open Archival Information System (OAIS) - Blue Book*. Washington: National Aeronautics and Space Administration, 2002.

COOK, Terry - *Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era*. Archives and Manuscripts 22 (Novembro 1994): p.300-328.

DECRETO-LEI N° 447/88. de 10 de Dezembro – Regula a Regula a pré-arquivagem de documentação.

DICIONÁRIO PRÁTICO DE INFORMÁTICA. Alfragide: Editora McGraw-Hill de Portugal, 2000.

DUFF, Wendy - *Will Metadata Replace Archival Description: A Commentary*. Archivaria 39 (Março 1995): 33-38.

FERREIRA, Miguel - *Introdução à preservação digital – Conceitos, estratégias e actuais consensos*. Guimarães: Escola de Engenharia da Universidade do Minho, 2006 ISBN-13: 978-972-8692-30-8 [em linha] [Consult. 10 Dez. 2007] Disponível na Internet <URL:<https://repositorium.sdum.uminho.pt/handle/1822/5820>>

HENDLEY, T. - *Comparison of Methods & Costs of Digital Preservation*, British Library Research and Innovation Center, West Yorkshire 106, 1998.

HENRIQUES, M., LIBREOTTO, G., RAMALHO, J., e HENRIQUES, P. - *Bidirectional conversion between xml documents and relational data bases*. International conference on CSCW in design, 2002. Disponível na Internet <URL: <http://hdl.handle.net/1822/601>>

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/IEC 17799: Information technology -- Security techniques -- Code of practice for information security management. Genebra: ISO, 2005.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/IEC 27001: Information technology -- Security techniques -- Information security management systems – Requirements. Genebra: ISO, 2005.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/IEC 27002: Information technology -- Security techniques -- Code of practice for information security management. Genebra: ISO, 2005.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/IEC 27006: Information technology -- Security techniques -- Requirements for bodies providing audit and certification of information security management systems. Genebra: ISO, 2007.

INSTITUTO PORTUGUÊS DA QUALIDADE - NP 4438-1, *Informação e documentação. Gestão de documentos de arquivo. Parte 1: Princípios directores*. Lisboa: Instituto Português da Qualidade, 2005.

INSTITUTO PORTUGUÊS DA QUALIDADE - NP 4438-2, *Informação e documentação. Gestão de documentos de arquivo. Parte 2: Recomendações de aplicação*. Lisboa: Instituto Português da Qualidade, 2005.

INSTITUTO DOS ARQUIVOS NACIONAIS/TORRE DO TOMBO - *Manual para a gestão de documentos*. Coord. Madalena Garcia e Maria João Pires de Lima; elab. Cecília Henriques, Francisco Barbedo e Luís Montalvão. 2ª ed. Lisboa: IAN/TT, 2004.

LAVOIE, B. F. - *The open archival information system reference model: Introductory guide*. Digital Preservation Coalition. 2004

MACNEIL, Heather - *Metadata Strategies and Archival Description: Comparing Apples to Oranges*. *Archivaria* 39 (Março 1995): 22-32.

MORELLI, Jeffrey - *Defining Electronic Records: a Terminology Problem... or Something More*. *Electronic Information Resources and Historians: European Perspectives*, Eds. Seamus Ross, and Edward Higgs, 83-91. St Katharinen: Scripta Mercaturae Verlag, 1993.

NISO Z39.87-2006, *Data Dictionary – Technical Metadata for Digital Still Images*. Bethesda: National Information Standards Institute; 2006

OCLC e RLG - *PREMIS: Data Dictionary for Preservation Metadata.*, 2005

PARER, Dagmar; PARROTT, Keith - *Management Practices in the Electronic Records Environment*. *Archives and Manuscripts* 22, no. 1 (Maio 1994):106-122.

RLG EAD Advisory Group - *RLG Best Practice Guidelines for Encoded Archival Description*, 2002

ROTHENBERG, J. - *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation A Report to the Council on Library & Information Resources (CLIR)*, 1999, ISBN - 1-887334-63-7 (disponível em www.clir.org/pubs/reports/rothenberg/contents.html ou www.clir.org/pubs/reports/rothenberg/pub77.pdf)

ROTHENBERG, Jeff - *Ensuring the Longevity of Digital Documents*. Scientific American. (Janeiro 1995), 42-47.

SARAMAGO, M. - *Metadados para a preservação digital e aplicação do modelo OAIS*. In VIII Congresso da BAD. 2004.

STEPHENS, David O. – *Digital Preservation in the United Kingdom* in THE INFORMATION MANAGEMENT, vol. 34, n.º. 4, Outubro 2000

STEPHENS, David O.; WALLACE, Roderick C. – *Electronic Records Retention: Fourteen Basic Principles* in THE INFORMATION MANAGEMENT, vol. 34, n.º. 4, Outubro 2000

WALLACE, David - *Managing the Present: Metadata as Archival Description*. Archivaria 39 (Março 1995): 11-21.

WALLACE, David - *Metadata and the Archival Management of Electronic Records: A Review*. Archivaria 36 (1993): 87-110.

Endereços electrónicos

CASPAR Project - <http://www.casparpreserves.eu/caspar-project>

Conselho Internacional de Arquivos - <http://www.ica.org/>

Conservation on Line (vários links) - <http://palimpsest.stanford.edu/>

DELOS - www.delos.info/

DGARQ - <http://www.dgarq.gov.pt>

Digitale Duurzaamheid - <http://www.digitaleduurzaamheid.nl/home.cfm>

DLM-Forum - http://www.europa.eu.int/historical_archives/dlm_forum

Electronic recordkeeping resources (Gestão de documentos electrónicos) -
<http://www.kshs.org/archives/ermlinks/ermlinks.htm>

Erpanet - <http://www.erpanet.org/>

Estudos de caso (Universidade de Indiana) - <http://www.indiana.edu/~libarche/index.html>

Interpares - www.interpares.org/

Jeff Rothenberg's Digital Longevity Papers - <http://www.panix.com/~jeffr/Prof/digilong.html>

Joint Interoperability Test Command Records Management Application (RMA) -
<http://jite.fhu.disa.mil/recmgt/>

MIP – http://www.dgarq.gov.pt/downloads/MIIP_v1.0a.doc

Model Requirements for the Management of Electronic Records (versão original) -
<http://www.cornwell.co.uk/moreq>

National Archives of Australia - <http://www.naa.gov.au/>

National Archives and Records Administration (USA) - <http://www.nara.gov/>

New South Wales ELECTRONIC RECORDKEEPING WEB SITE GUIDE -
<http://www.records.nsw.gov.au/>

PDF/A Competence Center - The pdf/a competence center webpage, 2007,
URL:<<http://www.pdfa.org>>.

PLANETS - www.planets-project.eu/

Projecto Interpares - <http://www.interpares.org/index.htm>

Projecto Duranti - <http://www.interpares.org/UBCProject/index.htm>

Public Record Office - <http://www.pro.gov.uk/>

Records Continuum Research Group (Meta-informação) - <http://rcrg.dstc.edu.au/>

Research Libraries Group (preservação de documentos electrónicos) - <http://www.rlg.org/>

RLG - *RLG EAD report card*. 2002, - <<http://www.rlg.org/eadreport-card>>.

RODA - <http://roda.iantt.pt/>

SOCIETY OF AMERICAN ARCHIVISTS - *Society of American Archivists home page*. 2003 -
<<http://www.archivists.org>>

SOCIETY OF AMERICAN ARCHIVISTS - *EAD tools survey*. 2006, URL:
<<http://www.archivists.org/saagroups/ead>>.

THE LIBRARY OF CONGRESS. *NISO metadata for images in XML schema official web site*.
2004. - <<http://www.loc.gov/standards/mix>>

THE LIBRARY OF CONGRESS - *Página oficial do EAD* versão de 2002. -
<<http://www.loc.gov/ead/>>.

THE LIBRARY OF CONGRESS. *Página oficial do METS*. 2006. -
<<http://www.loc.gov/standards/mets/>>.

THE LIBRARY OF CONGRESS - *Página oficial do PREMIS*. 2006, -
<<http://www.loc.gov/standards/premis/>>.

THE LIBRARY OF CONGRESS - *Sustainability of Digital Formats_Planning for Library of Congress Collection: TIFF rev6.0*, Março 2007. -
<<http://www.digitalpreservation.gov/formats/fdd/fdd000022.shtml>>

E Glossário

API - *Application Programming Interface* (ou **Interface de Programação de Aplicativos**) é um conjunto de rotinas e normas/regras estabelecidos por um software para utilização das suas funcionalidades por programas aplicativos - isto é: programas que não se querem envolver em detalhes da implementação do software, mas apenas utilizar os seus serviços. De modo geral, a API é composta por uma série de funções acessíveis somente por programação, e que permitem utilizar características do software menos evidentes ao utilizador tradicional.

(fonte: Wikipédia Consult. Dez. 2007)

Aplicação – Utilização de técnicas informáticas para resolver um problema de aplicação, que conduz normalmente ao desenvolvimento ou à utilização de programas de aplicação como, por exemplo, folhas de cálculo ou processadores de texto.

Nota: O termo “aplicação informática” é muitas vezes tomado no sentido de “programa de aplicação”.

(fonte: Glossário da sociedade da informação - Versão 2007)

Programa concebido como ferramenta auxiliar para o desempenho de uma determinada tarefa, como processamento de texto, a contabilidade ou a gestão de inventários

(fonte: Dicionário prático de informática - McGraw Hill)

Administrador - Função responsável pela execução rotineira da política de gestão de documentos de arquivos numa organização

(fonte: MoReq)

Arquivo Digital – Consiste numa organização de pessoas e sistemas, que aceitou a responsabilidade de preservar informação e torná-la disponível para uma dada comunidade.

(fonte: OAIS adapt.)

Autenticidade – Um documento de arquivo autêntico é aquele do qual se pode provar a) ser aquilo que pretende ser, b) ter sido produzido ou enviado pelo alegado produtor ou remetente, e c) ter sido produzido ou enviado no alegado momento de produção ou envio. Para assegurar a autenticidade dos documentos de arquivo, as organizações devem implementar e documentar políticas e procedimentos de controlo da produção, recepção, transmissão, manutenção e destino dos mesmos, assegurando que quem produz documentos está devidamente autorizado e identificado e que os documentos estão protegidos contra qualquer acção não autorizada (adicionar, apagar ou alterar informação, uso não autorizado ou imposição de restrições de acesso indevidas).

(fonte: NP 4438-1, p. 14.)

Backup – Cópia de um programa, disco ou dados, destinada a ser arquivada ou a salvaguardar ficheiros valiosos, caso os originais sejam destruídos

(fonte Dicionário prático de informática - McGraw Hill)

Classificação: identificação sistemática e organização de actividades e/ou documentos de arquivo em classes, de acordo com conversões logicamente estruturadas, métodos e regras representados num sistema de classificação.

(fonte: NP 4438-1, p. 9)

Código fonte – conjunto de palavras escritas de forma ordenada, contendo instruções numa das linguagens de programação existentes no mercado, de maneira lógica. Entende-se como o código base de um programa de computador, implementando o núcleo de funcionalidade do programa.

Complexidade - quando um documento integra diversas componentes digitais (por exemplo as paginas Web)

Conservação a longo prazo: Acto de manter a informação numa forma independente e acessível a longo prazo

(fonte: OAIS adapt.)

Consumidor – papel das pessoas ou sistemas clientes, que interagem com serviços do arquivo digital, com o fim de pesquisar informação preservada de interesse e acederem a essa informação de forma detalhada. Isto poderá incluir pessoas ou sistemas de arquivos digitais internos ou de outros arquivos digitais.

Cópias de Segurança - ver *Backup*

Cópia de segurança incremental – Cópias de seguranças de dados armazenados e/ou alterados posteriormente à cópia de segurança inicial.

Documento de arquivo electrónico (DAE) – Documento de arquivo produzido, transmitido e conservado, em forma electrónica

Documento Electrónico (DE) – Documento produzido ou recebido, e fixo para armazenamento e manipulação num sistema informático, e que requer a utilização do sistema para ser inteligível pelos humanos

(fonte Interpairs 2, glossary adapt.)

Documentos estruturados – contêm informação de forma a ser interpretada primariamente por aplicações informáticas (Ex: registos de sistemas contabilísticos, registos de dados de controlo de tráfego aéreo, etc.)

Documentos não estruturados – apresentam a informação de forma a ser inteligível primariamente por humanos (Ex: cartas, memorandos, mensagens de correio electrónico, imagens, fotocópias, imagens digitalizadas, registos áudio e/ou vídeo.

Documento Vital – que asseguram a continuidade do negócio em caso de acidente ou catástrofe

(fonte: NP 4438-2 p.21)

Documento de que a instituição depende em absoluto para a prossecução das suas funções. Ou seja, caso se verificasse um acidente que destruísse esses documentos, o organismo ver-se-ia impedido de prosseguir com eficácia a sua actividade. Em ambiente electrónico, toda a documentação técnica relativa às características do sistema de informação instalado e ao seu desempenho, assim como toda a metainformação descritiva na tramitação documental, terá de ser considerada vital, pois caso se verifique quebra do sistema ela terá papel fundamental para que se proceda ao processo de recuperação.

(fonte: Manual de Gestão de Documentos)

Em linha (Online) – Normalmente o tipo de armazenamento de dados mais acessível, que pode iniciar o restauro em milissegundos. Um exemplo disso é um disco rígido interno ou um conjunto de discos rígidos (ligados a uma SAN). Este tipo de armazenamento é muito conveniente e rápido, mas relativamente caro. Por outro lado, sofre de alguma vulnerabilidade, pois os dados podem ser apagados ou alterados, seja por acidente ou por vírus que elimine dados.

Embebido – Característica de algo com identidade própria (que pode existir autonomamente), que faz parte ou está incluído dentro de um conjunto, que também pode ser uma entidade. Ex: um OD que está embebido noutro OD (imagem embebida num documento de texto)

Emulação – Reprodução do comportamento e resultados de um material digital obsoleto através do desenvolvimento de novo hardware e/ou software para permitir a execução da antiga aplicação de software em computadores no futuro

(fonte Interpares 2, glossary)

Emulador – Software capaz de reproduzir o comportamento de uma plataforma de hardware e/ou software numa outra que à partida seria incompatível.

(fonte: Manual de Preservação)

Encapsulamento – Preservar, juntamente com o objecto digital, a informação necessária e suficiente para permitir o futuro desenvolvimento de conversores, visualizadores ou emuladores. Esta informação poderá consistir, por exemplo, numa descrição formal e detalhada do formato do objecto preservado.

(fonte: Introdução à preservação digital)

Encriptação – conversão de dados para um código secreto (ou de texto linear em texto cifrado) para transmissão numa rede pública

(fonte Interpares 2, glossary; adapt.)

Processo de execução de uma transformação complexa de um objecto digital de modo a que este não possa ser apresentado por uma aplicação de uma forma legível ou inteligível, a não ser que a transformação correspondente de descryptografia seja aplicada.

(fonte: *Moreq*)

Escalabilidade – característica do sistema ou equipamento que pode crescer em escala, isto é, que possibilita incrementos de capacidade ou funcionalidades acompanhando as necessidades dos usuários (aumento da capacidade de armazenamento do sistema através da substituição ou instalação de novos discos rígidos; versões mais actuais do formato PDF que permitem a inclusão de informação de dados CAD).

(fonte: *Wikipédia Consult. Dez. 2007*)

Ethernet – Arquitectura de redes locais que suporta um tráfego de alto débito. Nota: A Ethernet é um padrão usual de rede local.

(fonte: *Wikipédia Consult. Dez. 2007*)

Ficheiro – Colecção completa de informações com uma designação, como seja um programa, um conjunto de dados utilizados por um programa ou um documento criado pelo utilizador. O ficheiro constitui a unidade de armazenamento básica que permite ao computador distinguir um conjunto de informações de outro. É uma unidade coerente que o utilizador pode obter, alterar, eliminar, guardar ou enviar para um dispositivo de saída

(fonte *Dicionário prático de informática - McGraw Hill*)

Ficheiro linear – ficheiros cuja informação é passível de ser transferida integralmente para formato analógico.

Fora de Linha (*Off-line*) - Similar ao perto de linha, embora necessite de intervenção humana para disponibilizar as unidades de suporte. Isto pode incluir simplesmente o armazenamento das tapes de cópias de segurança nas estantes. O tempo de acesso deste suporte é superior a uma hora.

Formato – Em geral, designa a estrutura ou aspecto de uma unidade de dados. 2) Disposição de dados num ficheiro que normalmente permite que o documento seja lido ou alterado por uma determinada aplicação.

(fonte: *Dicionário prático de informática - McGraw Hill*)

Organização sequencial de dados em termos dos seus componentes.

(fonte: *OAIS*)

Formato proprietário - formatos de ficheiros com patente ou *copyright*

(fonte: *Wikipédia Consult. Dez. 2007*)

Formato normalizado – Formatos cujas definições e características estão especificadas numa norma, ou é consensualmente utilizado pela comunidade cibernética,

(fonte: *Wikipédia Consult. Dez. 2007*)

Formato aberto - Formato com especificações públicas para armazenamento, geralmente mantido por organizações de normas não proprietárias e de uso livre de restrições legais.

(fonte: Wikipédia Consult. Dez. 2007)

Gestor – ver Administrador

Horizonte de obsolescência – Prazo contado a partir da data de produção de um objecto digital a partir do qual, não se verificando acções de preservação, a possibilidade deste se tornar obsoleto incrementa numa escala geométrica.

(fonte: Wikipédia Consult. Dez. 2007)

Integridade – A integridade de um documento de arquivo refere-se a este permanecer completo e inalterado. É necessário que os documentos sejam protegidos contra alterações não autorizadas. As políticas de gestão de documentos de arquivo devem especificar que tipo de adições ou anotações podem ser feitas a um documento depois da sua produção, em que circunstâncias essas alterações podem ser autorizadas, e quem está autorizado a fazê-las. Qualquer alteração autorizada a um documento de arquivo deve ser explicitamente indicada e reconhecível enquanto tal.

(fonte: NP 4438-1, p. 15)

Latência (em inglês *lag*) refere-se ao atrasos que se podem verificar na comunicação entre computadores (internet, por exemplo) e comunicações via satélite. Em computação, *lag* refere-se ao tempo que um pacote de dados leva a completar um percurso de ida e volta entre um computador local e o seu destino.

(fonte: Wikipédia Consult. Dez. 2007)

Metadados – Ver metainformação

Metainformação – (no contexto da gestão de documentos de arquivo) Informação estruturada ou semiestruturada que permite a produção, gestão e utilização de documentos de arquivo ao longo do tempo, assim como nos e através dos domínios em que são produzidos.

(fonte: Wikipédia Consult. Dez. 2007)

Migração - Processo de transferência de um documento de arquivo de um sistema ou suporte de armazenamento para outro, de forma a assegurar a continua acessibilidade à medida que o sistema ou suporte onde se encontrava de degrada ou se torna obsoleto.

Mirror – Procedimento de protecção de dados e de acesso aos mesmos nos equipamentos informáticos. Este sistema implementa-se na tecnologia de RAID 1. Consiste basicamente em ter 2 discos rígidos ligados. Um é o principal e o outro guarda uma cópia exacta do principal armazenando qualquer alteração que se faça em tempo real. Desta forma consegue ter-se 2 discos rígidos idênticos e que permite, se tudo estiver bem configurado, que perante a falha do disco principal, o segundo assuma a sua função impedindo a quebra do sistema e a perda de dados armazenados

(fonte: Wikipédia Consult. Dez. 2007)

NFS – (Network File System), é um modelo de sistema de ficheiros, que tem como função centralizar ficheiros num servidor, formando assim uma pasta virtual. É utilizado para sistemas de ficheiros distribuídos numa rede de computadores de área local. Possibilita que diferentes sistemas ligados a uma mesma rede acedam a ficheiros remotos como se fossem ficheiros locais.

(fonte: Wikipédia Consult. Dez. 2007)

Objecto Digital (OD) – Unidade de informação digital que inclui propriedades do objecto e poderá incluir também métodos de execução de operações no objecto.

(fonte Interpares 2, glossary)

Unidade discreta de informação em formato digital.

(fonte RODA Relatório Final)

Obsolescência – a perda de valor que um bem sofre em resultado do progresso técnico ou da evolução.

(fonte: Priberam)

Pacote de aplicações – conjunto de programas de computador que tem por objectivo o desempenho de tarefas de índole prática, em geral ligadas ao processamento de dados, como o trabalho em escritório ou empresarial. A sua natureza é diferente da de outros tipos de software, como sistemas operativos e ferramentas a eles ligadas, jogos e outros softwares lúdicos, entre outros. (Microsoft Office, Open Office)

(fonte: Wikipédia Consult. Dez. 2007)

Perto de Linha (Near-line) - Normalmente trata-se de um tipo de armazenamento de dados menos acessível, mas mais barato, do que o armazenamento em linha, sendo útil para o armazenamento de dados de réplicas (cópias de segurança). Um exemplo disso é um conjunto de tapes com tempos de restauro que vão desde segundos até alguns minutos. Normalmente está associado a um aparelho mecânico que move as unidades de suporte do local de armazenamento para o dispositivo de leitura e/ou gravação de dados.

(fonte: Wikipédia Consult. Dez. 2007)

Prazo de retenção – Período, previamente estabelecido para a permanência dos documentos junto da actividade produtora, antes do seu destino final.

(Dicionário de Terminologia Arquivística, adapt.)

Plano de classificação - Documento que regista o sistema de classes concebido para a organização de um arquivo corrente e as respectivas notas de aplicação, nomeadamente quanto ao âmbito das classes e à ordenação das unidades arquivísticas abrangidas por cada uma, de modo a determinar toda a estrutura de um arquivo.

(fonte: Manual para a Gestão de Documentos)

Plano de Preservação Digital - documento estratégico contendo políticas, procedimentos e práticas/actividades para a constituição de uma estrutura técnica e organizacional que permita preservar de forma continuada Objectos Digitais.

Plataforma – 1) Tecnologia de base de um sistema de computador. Dado que os computadores são dispositivos em camadas, compostos por uma camada de hardware a nível dos chips, por uma camada de controladores de hardware e de sistema operativo e por uma camada de aplicação, a camada mais baixa de um computador é normalmente designada por plataforma
2) Na utilização quotidiana, é o tipo de computador ou sistema operativo a ser utilizado

(Dicionário prático de informática - McGraw Hill)

Preservação digital – Conjunto de actividades ou processos responsáveis por garantir o acesso continuado e a longo-prazo à informação e restante património cultural existente em formatos digitais.

(fonte: Introdução à preservação digital)

Conjunto de actividades desenvolvidas com o fim de aumentar a vida útil dos OD, salvaguardando a utilização operacional, e protegê-los das falhas de suportes, perda física e obsolescência tecnológica;

Conjunto de actividades que promovem a acessibilidade aos conteúdos;

Conjunto de actividades que assistem na preservação do conteúdo intelectual, forma, estilo, aparência e funcionalidade.

Produtor – Entidade que cria os registos, produz e manipula os conteúdos.

Proprietário – Unidade Orgânica responsável pela informação/documento.

RAID – (**Redundant Array of Independent Drives** ou **Disks**) – Conjunto Redundante de Discos Independentes, é um meio de se criar uma unidade virtual de armazenamento composta por vários discos individuais, com a finalidade de aumentar a segurança e desempenho através de redundância.

(fonte: Wikipédia Consult. Dez. 2007)

Réplica – ver backup.

Repositório – local centralizado onde se armazena e mantém informação digital, habitualmente bases de dados ou ficheiros informáticos.

(fonte: Wikipédia Consult. Dez. 2007)

Restauro do sistema – Recuperação dos dados do sistema, através das cópias de segurança.

Série documental – Unidade arquivística constituída por um conjunto de documentos simples ou compostos a que, originariamente, foi dada uma ordenação sequencial, de acordo com um sistema de recuperação da informação. Em princípio, os documentos de cada série deverão

corresponder ao exercício de uma mesma função ou actividade, dentro de uma mesma área de actuação.

(fonte: *Manual para a Gestão de Documentos*)

Sistema aberto - Sistema informático concebido com uma arquitectura tal que permita a junção de componentes de *hardware* e/ou de *software* provenientes de diversos fornecedores.

(Fonte Glossário da sociedade da informação – versão 7)

Sistema de armazenamento – Dispositivo ou conjunto de dispositivos utilizados para armazenar primariamente os dados e a informação produzida.

Sistema operativo – programa ou um conjunto de programas cuja função é servir de interface entre um computador e o utilizador. (Windows, Linux; MacOS)

Software de base de um computador destinado a controlar a execução de programas, a comunicação entre dispositivos e programas, assegurando as operações de entrada-saída, a atribuição de recursos aos diferentes processos, o acesso às bibliotecas de programas e aos ficheiros, assim como a compatibilidade dos trabalhos.

Nota: O sistema operativo é o *software* mais importante a correr num computador.

(fonte: *Glossário da sociedade da informação - Versão 2007*)

SMB – (**Server Message Block**) opera como um protocolo de comunicação de rede utilizado.

(fonte: *Wikipédia Consult. Dez. 2007*)

Software livre - segundo a definição criada pela Free Software Foundation, é qualquer programa de computador que pode ser usado, copiado, estudado, modificado e redistribuído sem nenhuma restrição. A liberdade de tais directrizes é central ao conceito, o qual se opõe ao conceito de software proprietário, mas não ao software que é vendido almejando lucro (software comercial). A maneira usual de distribuição de software livre é anexar a este uma licença de software livre, e tornar o código fonte do programa disponível. O software livre também é conhecido pelo acrónimo FLOSS (do inglês *Free/Libre Open Source Software*).

(fonte: *Wikipédia Consult. Dez. 2007*)

Suporte – Material adequado ao registo de informação. Exemplos de suportes electrónicos: CD, DVD, tape, Discos rígidos).

(fonte: *Manual para a Gestão de Documentos*)

Tabela de selecção – Instrumento que regista o resultado da avaliação, apresentando-se como uma relação dos documentos de arquivo de um organismo ou administração que fixa os respectivos prazos de conservação em fase activa e semi-activa e destino final.

(fonte: *Manual para a Gestão de Documentos*)

Unidade binária: - (em inglês Binary unit, ou bit), unidade de medida de transmissão de dados usada na Computação e na Teoria da Informação.

(fonte: Wikipédia Consult. Dez. 2007)

Usabilidade – significa que um documento é utilizável ou seja, que pode ser localizado, recuperado, apresentado e interpretado.

(fonte: NP 4438-1 p.15)

USB – Sigla de universal serial bus (bus série universal). Bus série com uma largura de banda de 1.5 Mbps, utilizado para ligar periféricos (modems, impressoras, ratos, teclados, ...) a um microcomputador.

(Dicionário prático de informática - McGraw Hill)

Utilizador – ver consumidor

(fonte OAIS)

Virtualização de armazenamento (Storage Virtualization) – processo de abstracção de armazenamento lógico a partir do armazenamento físico. Este conceito é utilizado actualmente para descrever esta abstracção em qualquer camada no conjunto de software e hardware de armazenamento.

(fonte: Wikipédia Consult. Dez. 2007)