





# Introdução à Preservação Digital

---

Conceitos, estratégias e actuais consensos

MIGUEL FERREIRA  
2006

<b>Título</b>	Introdução à preservação digital – Conceitos, estratégias e actuais consensos
<b>Autor</b>	Miguel Ferreira
<b>Ano</b>	2006
<b>ISBN</b>	972-8692-30-7 978-972-8692-30-8
<b>Editora</b>	Escola de Engenharia da Universidade do Minho
<b>Local</b>	Guimarães, Portugal
<b>Capa</b>	Miguel Ferreira
<b>Suporte</b>	Edição electrónica
<b>Dimensões de impressão</b>	A: 297mm, L: 210mm (A4)
<b>Revisão científica</b>	Ana Alice Baptista (Universidade do Minho)
	Cristina Ribeiro (Faculdade de Engenharia da Universidade do Porto)
	Francisco Barbedo (Instituto dos Arquivos Nacionais/Torre do Tombo)
	José Carlos Ramalho (Universidade do Minho)
	Paulo Cortez (Universidade do Minho)
<b>Idioma</b>	Português

**Citação.** M. Ferreira, Introdução à preservação digital – Conceitos, estratégias e actuais consensos. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006.

Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Uso Não-Comercial-Vedada a Criação de Obras Derivadas 2.5 Portugal. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-nc-nd/2.5/pt/> ou envie uma carta para Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

Dedicado à Marta...  
...a minha mais que tudo.



## AGRADECIMENTOS

Foram várias as pessoas que contribuíram para a edificação deste livro. A todos os que directa ou indirectamente participaram ou influenciaram a realização deste trabalho, o meu mais sincero obrigado.

Há, contudo, um conjunto de pessoas que pela forma incisiva como colaboraram na concepção deste projecto merece a minha particular atenção. Gostaria, assim, de agradecer de forma especial às seguintes individualidades:

À Ana Alice Baptista e ao José Carlos Ramalho, meus mentores e orientadores de doutoramento que sempre me indicaram o caminho mais acertado, tanto no trabalho como na vida.

À equipa de revisores, Cristina Ribeiro e Paulo Cortez, pelos comentários, correcções e sugestões fornecidas durante a apresentação pública do meu trabalho de doutoramento.

Ao sempre eloquente Francisco Barbedo, pelas discussões filosóficas em torno de semiótica, comunicação e arquivística.

Ao Rui Castro e Duarte Duque, companheiros de sempre, pelas inúmeras correcções de linguagem e conselhos técnicos que tão acentuadamente elevaram a qualidade final desta obra.

Ao Dr. Eloy Rodrigues, por ter aceite o meu humilde convite para prefaciар este livro.

A todas estas pessoas, o meu mais profundo obrigado.





## SOBRE O AUTOR

Licenciado em Engenharia de Sistemas e Informática pela Universidade do Minho, iniciou a sua carreira profissional na Philips Research em Eindhoven, Holanda, onde trabalhou como investigador no domínio da domótica e *Ambient Intelligence*. Aí, fez parte da equipa de desenvolvimento de um dispositivo capaz de armazenar e navegar nas memórias colectivas de uma comunidade, e.g. fotografias, vídeos, sons, aromas.

Regressado a Portugal, mudou radicalmente o rumo da sua vida profissional ao ingressar a equipa do projecto DigitArq. Um projecto multifacetado realizado no Arquivo Distrital do Porto, sob coordenação científica da Universidade do Minho, do qual resultou a atribuição do prémio Fernandes Costa – Agência para a Sociedade do Conhecimento, por ter sido considerado o trabalho que melhor respondeu à “inovação e contributo para o desenvolvimento da Sociedade da Informação” em Portugal no ano de 2004.

Regressou à Universidade do Minho onde iniciou a sua carreira como investigador na área dos Arquivos e Bibliotecas Digitais. Durante um ano explorou novas formas de navegar sobre repositórios digitais. Parte desses desenvolvimentos vieram a ser integrados na distribuição oficial da plataforma DSpace e são agora mantidos pelo Massachusetts Institute of Technology (MIT) e pela Hewlett-Packard (HP).

Em 2005 iniciou o seu programa de doutoramento na área da Preservação Digital, do qual resultou esta publicação.



# Prefácio

---

É para mim um grande prazer, e também uma honra, prefaciá-la esta obra. Em primeiro lugar, pela sua relevância, oportunidade e utilidade.

No mundo de hoje é diariamente produzido um volume gigantesco de informação, registada e/ou transmitida sobre diversos suportes e formatos. Em resultado da evolução social à escala planetária no último século (crescimento demográfico, evolução do acesso à educação e às tecnologias, desenvolvimento e “enriquecimento” de múltiplas regiões e grupos sociais) e da revolução tecnológica dos últimos 20 anos (em particular da informática e das comunicações), é provável que as gerações actuais produzam e registem mais informação, do que toda a informação registada pelas milhares de gerações que nos precederam.

Uma parte crescente desta informação é produzida, registada e transmitida em suportes e formatos digitais. Claro que, tal como no passado, não é possível, nem é útil ou relevante, guardar e preservar toda a informação hoje criada. Mas, do mesmo modo que hoje podemos aceder e consultar um registo significativo da informação produzida pelas gerações que nos

antecederam (em particular as dos últimos cinco séculos), é necessário garantir que as gerações futuras irão ter acesso a um registo igualmente significativo e relevante da produção informativa contemporânea.

Por isso, a preservação digital, que de forma feliz é definida neste livro como a actividade que garante que a “comunicação entre um emissor e um receptor é possível, não só através do espaço, como também através do tempo”, assume uma importância fundamental no actual contexto social e tecnológico. E daí, a extrema oportunidade e relevância deste livro.

O tema da preservação digital é, ao mesmo tempo, um tema novo, vasto e complexo. É um tema novo porque ele apenas se autonomizou e desenvolveu de forma visível há pouco mais de dez anos. É um tema vasto e complexo porque o conjunto de questões e problemas, quer de natureza conceptual e teórica, quer de natureza prática e tecnológica, é imenso. O volume crescente, a heterogeneidade e as características da informação digital (que, do ponto de vista dos seus utilizadores, não apenas pode ser independente dos suportes, como dos formatos), tais como a facilidade de manipulação, interligação e reutilização vão traduzir-se certamente no alargamento do campo da preservação digital.

A grande utilidade deste livro resulta precisamente de apresentar e mapear, de forma simples mas nem por isso menos rigorosa, o território novo, vasto e complexo da preservação digital.

O livro começa, de forma muito adequada, por apresentar e discutir a anatomia dos objectos digitais, cuja preservação se pretende garantir, chamando a atenção para a existência de vários níveis a que podem ser observados, como que diferentes “camadas” de que se compõem.

Uma das características essenciais dos objectos digitais, que os distingue de objectos informativos anteriores, como os livros e outros documentos registados em papel em que a informação pode ser acedida directamente do

objecto, é que eles exigem “camadas” de intermediação tecnológica (hardware/equipamento, e software/formato), sem as quais a informação que contém não pode ser acedida e usada. E apesar de os leitores/utilizadores só interagirem com a última “camada” (no livro designada como objecto conceptual), esta depende das “camadas” anteriores, muito vulneráveis à obsolescência tecnológica.

Partindo deste útil enquadramento conceptual, o livro aborda nos capítulos seguintes a problemática específica da preservação digital, apresentando de um modo sintético, e ao mesmo tempo claro e de fácil compreensão, o modelo de referência OAIS, as diferentes estratégias de preservação digital, os directórios de formatos, a autenticidade, a metainformação de preservação, concluindo com uma muito útil avaliação de estratégias de preservação.

Este livro fornece portanto uma panorâmica abrangente e introdutória às questões da preservação digital que será certamente preciosa para todos quantos se queiram iniciar, ou actualizar conhecimentos, neste domínio.

Em segundo lugar, é para mim também um enorme prazer prefaciar um livro que o seu autor disponibiliza em Acesso Livre (Open Access).

Como se sabe, o Acesso Livre aplica-se primariamente à versão final (após peer-review) de artigos de revistas (postprints), mas também inclui versões não revistas (preprints) que os investigadores queiram divulgar para alertar sobre novos resultados ou para estabelecer a primazia. Mas o Acesso Livre pode aplicar-se naturalmente a todos os trabalhos, como livros e monografias especializadas, working papers, e outros, dos quais os autores não esperem obter rendimento (apesar de serem cada vez mais frequentes exemplos de livros com versões disponíveis em Acesso Livre e, simultaneamente, com versões vendidas no circuito comercial).

Este livro, que o autor agora generosamente disponibiliza, constitui um dos primeiros exemplos nacionais de livros acessíveis em Acesso Livre. Espero que, também sob esse ponto de vista, este seja um livro pioneiro abrindo caminho a muitos outros que se editarão em Acesso Livre nos próximos anos em Portugal.

Finalmente, é também um prazer prefaciá-lo pelo conhecimento e estima pessoal que tenho pelo Miguel Ferreira, com quem tenho tido o privilégio de interagir e colaborar, ainda que de forma esporádica e fugaz, nos últimos dois anos. Ambos somos membros de uma “comunidade” que se tem vindo a constituir, de forma mais ou menos informal, na Universidade do Minho, mas com fortes ligações nacionais e internacionais, em torno da informação digital, das bibliotecas e arquivos digitais, dos repositórios institucionais, dos “arquivos” abertos (Open Archives Initiative) e do Acesso Livre à literatura científica.

Para finalizar este prefácio, deixo um convite e um desafio. O convite é a todos os que se deram ao trabalho de ler estas linhas, para que leiam, usem e aproveitem intensamente este livro, não se esquecendo do dever básico do reconhecimento e atribuição da autoria.

O desafio é para o autor: para que esta seja apenas a primeira versão de um livro, que através de múltiplas revisões, acrescentos e actualizações, se transforme numa obra de referência sobre a preservação digital em língua portuguesa.

Guimarães, Novembro de 2006

Eloy Rodrigues

## CONTEÚDO

INTRODUÇÃO.....	17
A ANATOMIA DE UM OBJECTO DIGITAL.....	21
O MODELO DE REFERÊNCIA OAIS .....	27
ESTRATÉGIAS DE PRESERVAÇÃO DIGITAL.....	31
Preservação de tecnologia.....	32
Refrescamento .....	33
Emulação.....	33
Migração/conversão.....	36
Migração para suportes analógicos.....	37
Actualização de versões .....	37
Conversão para formatos concorrentes.....	38
Normalização .....	38
Migração a-pedido .....	40
Migração distribuída .....	41
Encapsulamento.....	43
A Pedra de Rosetta digital.....	44
DIRECTÓRIOS DE FORMATOS .....	46
AUTENTICIDADE.....	49
METAINFORMAÇÃO DE PRESERVAÇÃO.....	54
AVALIAÇÃO DE ESTRATÉGIAS DE PRESERVAÇÃO.....	58
SÍNTESE E CONCLUSÃO .....	62
GLOSSÁRIO.....	68
REFERÊNCIAS.....	74
ÍNDICE .....	84





# Introdução

---

Desde a invenção da escrita que existe uma manifesta preocupação pela preservação dos artefactos que resultam de processos intelectuais e criativos do ser humano [1]. A preservação desses artefactos permite às gerações futuras compreender e contextualizar a história e a cultura dos seus povos [2]. Os museus, as bibliotecas e os arquivos assumem neste contexto um papel determinante, responsabilizando-se pela preservação e longevidade desses artefactos.

Nos dias de hoje, uma parte significativa da produção intelectual é realizada com o auxílio de ferramentas digitais. A simplicidade com que o material digital pode ser criado e disseminado através das modernas redes de comunicação e a qualidade dos resultados obtidos são factores determinantes na adopção deste tipo de ferramentas.

No entanto, o material digital carrega consigo um problema estrutural que coloca em risco a sua longevidade. Embora um documento digital possa ser copiado infinitas vezes sem qualquer perda de qualidade, este exige a presença de um contexto tecnológico para que possa ser consumido de

forma inteligível por um ser humano. Esta dependência tecnológica torna-o vulnerável à rápida obsolescência a que geralmente a tecnologia está sujeita [3].

O curso da história tem revelado inúmeros exemplos fatídicos de obsolescência tecnológica. Na década de 1970 a multinacional japonesa Sony introduziu um formato de vídeo designado Betamax (Fig. 1 – a). Comparativamente ao comum VHS (Video Home System) (Fig. 1 – b), a cassette Betamax era de menores dimensões e oferecia uma qualidade de imagem superior. O pico da sua popularidade foi atingido em 1983 quando cerca de um terço do mercado de vídeo doméstico era dominado por este formato [4].

Apesar do seu sucesso comercial, o facto de a Sony não facilitar o licenciamento de produção a terceiros foi decisivo para que uma viragem radical ocorresse no mercado e os consumidores passassem a utilizar massivamente o formato VHS. Em escassos anos, o formato Betamax desapareceu do mercado europeu e norte-americano, sendo hoje em dia praticamente impossível encontrar um dispositivo capaz de ler a informação armazenada numa dessas cassetes [4].



Fig. 1 – a) Cassete Betamax; b) Cassete VHS.

Um exemplo mais recente de obsolescência tecnológica, desta vez no domínio digital, reporta-se ao uso das populares disquetes de 3.5 polegadas ( Fig. 2 - a). Em Março de 2003, o fabricante Dell Computer Corporation anunciou que os seus computadores deixariam de integrar dispositivos capazes de ler este tipo de suportes (Fig. 2 - b). Vários fabricantes seguiram de imediato o seu exemplo [5].

Actualmente, é ainda possível adquirir dispositivos capazes de ler disquetes de 3.5 polegadas. No entanto, o mercado inclina-se rapidamente para o uso de DVD e flash-drives.



Fig. 2 – a) Disquete de 3.5 polegadas; b) Leitor de disquetes de 3.5 polegadas.

A obsolescência tecnológica não se manifesta somente ao nível dos suportes físicos. No domínio digital, todo o tipo de material tem obrigatoriamente de respeitar as regras de um determinado formato. Isto permite que as aplicações de software sejam capazes de abrir e interpretar adequadamente a informação armazenada. À medida que o software vai evoluindo, também os formatos por ele produzidos vão sofrendo alterações.

É bastante comum encontrar aplicações de software capazes de carregar os ficheiros produzidos por versões anteriores dessa mesma aplicação. No entanto, essa capacidade raramente vai além das duas versões precedentes [5].

No mundo actual, onde cada vez mais organizações dependem da informação digital que produzem, torna-se premente a implementação de técnicas e de políticas concertadas que vão no sentido de garantir a perenidade e a acessibilidade a este tipo de informação.

Designa-se, assim, por preservação digital o conjunto de actividades ou processos responsáveis por garantir o acesso continuado a longo-prazo à informação e restante património cultural existente em formatos digitais [6]. A preservação digital consiste na capacidade de garantir que a informação digital permanece acessível e com qualidades de autenticidade suficientes para que possa ser interpretada no futuro recorrendo a uma plataforma tecnológica diferente da utilizada no momento da sua criação.

Ao longo dos últimos 10 anos, foram muitos os projectos e iniciativas que contribuíram para a edificação da base de conhecimento que actualmente suporta o domínio científico da preservação digital. Desses projectos resultaram ideias, conceitos e estratégias que conduziram ao reconhecimento universal do problema e à elaboração de possíveis soluções.

Nos próximos capítulos serão apresentados os conceitos e as iniciativas de maior relevo no domínio da preservação digital.

# A anatomia de um objecto digital

---

Um objecto digital pode ser definido como todo e qualquer objecto de informação que possa ser representado através de uma sequência de dígitos binários<sup>1</sup> [7]. Esta definição é suficientemente lata para acomodar tanto, informação nascida num contexto tecnológico digital (objectos nado-digitais), como informação digital obtida a partir de suportes analógicos (objectos digitalizados).

Documentos de texto, fotografias digitais, diagramas vectoriais, bases de dados, sequências de vídeo e áudio, modelos de realidade virtual, páginas *Web* e aplicações de software são apenas alguns exemplos do que podemos considerar um objecto digital.

---

<sup>1</sup> Do inglês *bit stream*.

De modo a promover a compreensão e o enquadramento dos diferentes processos envolvidos na preservação de objectos digitais, torna-se fundamental analisar as diferentes formas como os podemos observar.

Para que um ser humano seja capaz de decifrar um objecto digital, há um conjunto de transformações que têm necessariamente de ocorrer. Um objecto digital começa por ser um objecto físico, i.e. um conjunto de símbolos ou sinais inscritos num suporte físico (e.g. disco rígido, CD, DVD, disquete).

O suporte físico define o domínio dos símbolos a utilizar. Vejamos um exemplo. Uma fotografia digital pode encontrar-se inscrita numa vasta gama de suportes físicos. Os símbolos, ou sinais físicos, utilizados para representar essa fotografia num CD-ROM diferem substancialmente dos símbolos utilizados para a representar num disco rígido [7]. No primeiro, os símbolos utilizados são essencialmente pequenos orifícios reflectores dispostos em espiral sobre uma base de policarbonato. No segundo, são utilizados padrões magnéticos sobre um prato metálico.

O objecto físico constitui aquilo que, geralmente, o hardware é capaz de interpretar (Fig. 3). O hardware assume aqui a responsabilidade de transformar os símbolos inscritos no suporte físico num conjunto de dados que o software será capaz de manipular. Esse conjunto de dados encontra-se geralmente organizado segundo as regras decretadas pelo software que foi utilizado para produzir o objecto digital. Essas regras ou estruturas de dados constituem aquilo que vulgarmente se designa por *f o r m a t o* de objecto digital [7]. Essas estruturas constituem o nível de abstracção lógico, ou *s i n t á c t i c o*, do objecto digital.

O software assume então a responsabilidade de preparar o objecto lógico para que este seja devidamente apresentado a um receptor humano. Nesta fase, os sinais digitais manipulados no interior do computador são

transformados em sinais analógicos que serão veiculados até ao receptor humano através de um periférico de saída (Fig. 4).

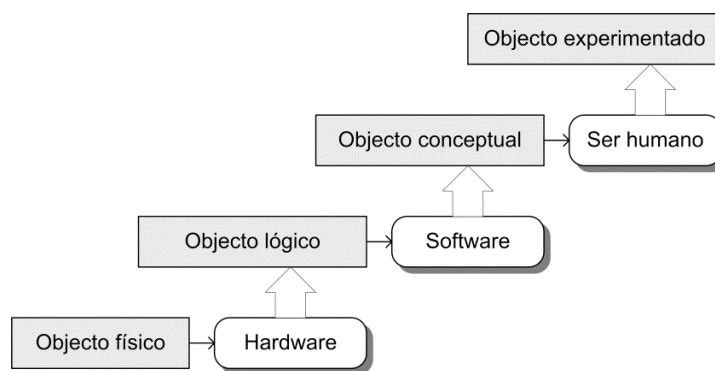


Fig. 3 - Diferentes níveis de abstracção de um objecto digital.

A imagem que posteriormente se forma na mente do receptor constitui o que vulgarmente se designa por um objecto conceptual ou objecto semântico (Fig. 3).

Os objectos conceptuais assumem formas ou concepções familiares aos seres humanos, i.e. formas que existem no mundo real e que lhes são conhecidas, como livros, filmes ou fotografias. Do ponto de vista do ser humano, o objecto conceptual constitui aquilo que deve ser preservado.

Não obstante, cada ser humano acaba por fazer uma interpretação individual do objecto recebido. Essa interpretação será aqui designada por objecto experimentado (Fig. 3). Apesar de teoricamente ser possível captar e preservar o objecto experimentado, nenhuma das estratégias de preservação apresentadas ao longo deste livro abordam seriamente esta questão.

De modo análogo, quando um ser humano assume o papel de emissor (ou produtor de informação), este mesmo conjunto de transformações é

realizado, mas em sentido inverso. Nesta situação, o objecto conceptual que ganhou forma no cérebro do emissor é codificado numa qualquer linguagem passível de ser comunicada (e.g. língua portuguesa, linguagem gráfica, etc.). Essa linguagem poderá então ser transmitida ou armazenada num suporte físico adequado à sua retenção, passando inevitavelmente por um processo intermédio de codificação que transforma a linguagem “humana” em códigos binários capazes de ser processados pelo computador.

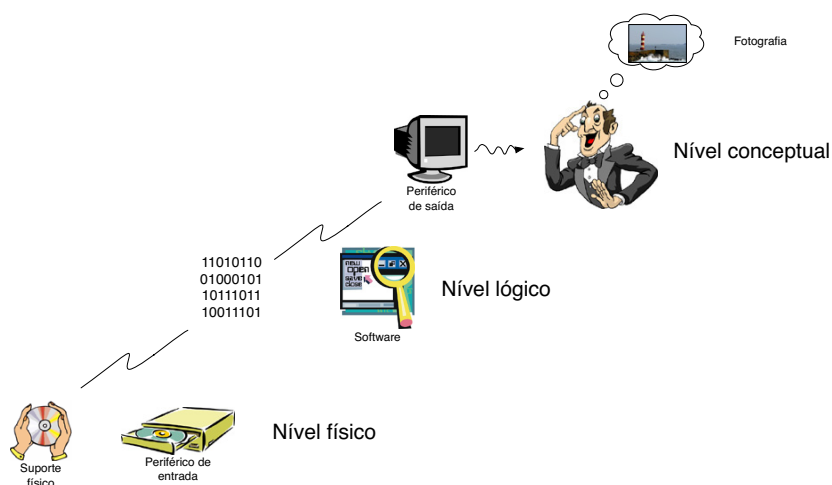


Fig. 4 - Cadeia de interpretação do nível físico ao conceptual.

Numa situação ideal, o objecto conceptual formado na mente do emissor será em tudo semelhante ao objecto conceptual concebido pelo receptor. Somente nessa situação a comunicação poderá ser considerada perfeita.

A preservação digital é a actividade responsável por garantir que a comunicação entre um emissor e um receptor é possível, não só através do espaço mas também através do tempo.

Para que a preservação de um objecto digital seja possível, é necessário assegurar que todos os níveis de abstracção anteriormente descritos se encontram acessíveis e interpretáveis. Se a cadeia de interpretação que



permite elevar um objecto digital desde o seu nível físico até ao nível conceptual for rompida, a comunicação deixa de ser possível e o objecto perder-se-á para sempre [8, 9].

Visto numa outra perspectiva, um mesmo objecto conceptual pode ser representado em diversos formatos lógicos, podendo cada um destes ser suportado por um sem-número de representações físicas [10]. Voltando ao exemplo da fotografia digital, facilmente podemos constatar que poderá ser codificada em diversos formatos, e.g. TIFF, JPEG, PNG. Não obstante, cada um destes formatos pode, por sua vez, ser inscrito numa multitude de suportes físicos, e.g. DVD, disco rígido, flash-drive. (Fig. 5).

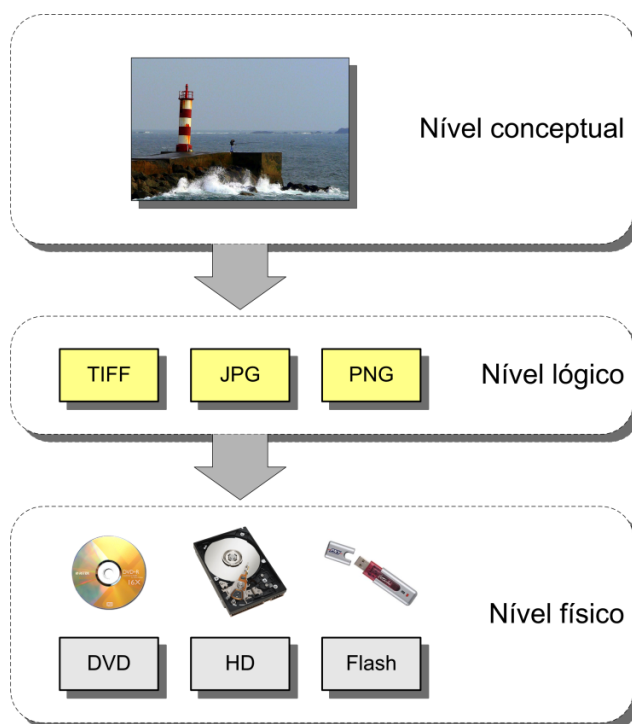


Fig. 5 - Objecto digital observado a diferentes níveis de abstracção.

Esta dissecação do conceito objecto digital à luz da semiótica, i.e. recorrendo a diferentes níveis de abstracção, permite um melhor enquadramento das

diversas estratégias de preservação que serão apresentadas ao longo deste livro.

# O modelo de referência OAIS

---

Em 1990, o Consultative Committee for Space Data Systems (CCSDS) iniciou um esforço conjunto com a International Organization for Standardization (ISO) a fim de desenvolver um conjunto de normas capazes de regular o armazenamento a longo-prazo de informação digital produzida no âmbito de missões espaciais.

Deste esforço nasceu o modelo de referência OAIS (Open Archival Information System), um modelo conceptual que visa identificar os componentes funcionais que deverão fazer parte de um sistema de informação dedicado à preservação digital [11, 12]. O modelo descreve ainda as interfaces internas e externas do sistema e os objectos de informação que são manipulados no seu interior [11]. O modelo foi aprovado como uma norma internacional em 2003 – ISO Standard 14721:2003 [12].

Um dos contributos mais notáveis desta iniciativa foi a definição de uma terminologia própria que viria a facilitar a comunicação entre os diversos intervenientes envolvidos na preservação de objectos digitais [13].

A Fig. 6 ilustra os diferentes componentes funcionais, assim como os pacotes de informação trocados no interior de um repositório digital compatível com o modelo de referência OAIS.

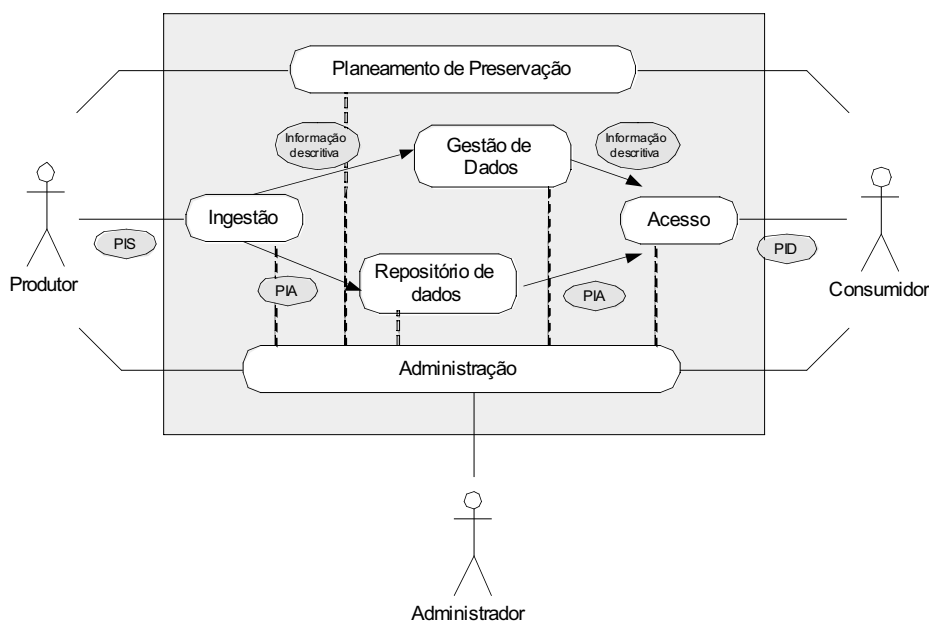


Fig. 6 - Modelo de referência Open Archival Information System (OAIS).

O *Produtor* deverá ser entendido como a entidade externa ao repositório que se responsabiliza pela submissão de material. O material submetido a arquivo é aqui representado pelo *Pacote de Informação de Submissão*<sup>2</sup> (PIS).

Durante o processo de submissão ou incorporação, designado neste contexto por *Ingestão*, o repositório é responsável por garantir a integridade da

<sup>2</sup> Do inglês *Submission Information Package* (SIP).

informação recebida. Ainda nesta fase, é produzida toda a Informação Descritiva que irá suportar a descoberta e localização do material depositado. Essa informação descritiva (ou metainformação) é armazenada e gerida pelo componente Gestão de Dados<sup>3</sup>. O material a preservar (i.e. Pacote de Informação de Arquivo<sup>4</sup> ou PIA) será conservado no Repositório de Dados<sup>5</sup>.

O componente de ingestão constitui, assim, a interface entre o arquivo OAIS e os respectivos produtores de informação [11].

O componente Planeamento de Preservação encarrega-se da definição de políticas de preservação. Este serviço é responsável pela monitorização do ambiente externo ao repositório e por desencadear eventos de preservação sempre que necessário. É, por exemplo, da responsabilidade deste componente definir as estratégias de preservação a utilizar no interior do repositório, monitorizar as tendências comportamentais da sua comunidade de interesse ou identificar formatos que se encontram na iminência de se tornar obsoletos [11].

O componente Acesso estabelece a ponte entre o repositório e a sua comunidade de interesse<sup>6</sup>, i.e. o conjunto de potenciais Consumidores do material custodiado. Este componente é responsável por facilitar a

---

<sup>3</sup> Do inglês *Data Management*.

<sup>4</sup> Do inglês *Archival Information Package* (AIP).

<sup>5</sup> Do inglês *Archival Storage*.

<sup>6</sup> Também conhecido por população potencialmente utilizadora. De notar que o conceito de comunidade de interesse deverá ser entendido no seu sentido mais lato. Trata-se de um conceito por vezes associado a centros de documentação e bibliotecas especializadas, como é o caso de certas bibliotecas universitárias (e.g. Biblioteca de Física da Universidade do Minho). Em bibliotecas de carácter geral, como bibliotecas públicas ou nacionais, e na generalidade dos arquivos este conceito não é aplicável ou apenas o será se considerarmos que a comunidade de interesse coincide com a totalidade da população.

descoberta e localização dos objectos digitais, bem como preparar os mesmos para entrega ao consumidor.

Os pacotes que são entregues ao consumidor assumem a forma de Pacotes de Informação de Disseminação<sup>7</sup> – PID [11]. É de realçar o facto de os Pacotes de Informação de Disseminação poderem ser diferentes dos Pacotes de Informação de Arquivo. A informação que é entregue ao consumidor poderá ser apenas um subconjunto da informação arquivada ou até uma versão transformada da mesma (ver Migração/conversão na página 36).

Por último, o componente Administração é responsável pelas operações diárias de manutenção e sobretudo pela parametrização e monitorização dos processos desencadeados no interior do repositório. Este componente interage com todos os restantes de modo a assegurar o correcto funcionamento do mesmo [11].

---

<sup>7</sup> Do inglês *Dissemination Information Package* (DIP).

# Estratégias de preservação digital

---

Ao longo dos últimos anos têm vindo a ser propostas inúmeras estratégias no sentido de solucionar o problema da preservação digital. Segundo Lee e seus colaboradores, estas estratégias podem ser agrupadas em três classes fundamentais: emulação, migração e encapsulamento [2].

Thibodeau, por sua vez, organiza as diferentes estratégias propostas num mapa bidimensional, posicionando no seu extremo esquerdo as estratégias centradas na preservação do objecto físico/lógico<sup>8</sup> e no extremo oposto as estratégias centradas na preservação do objecto conceptual (Fig. 7). No eixo vertical as diversas estratégias são dispostas mediante o seu grau de especificidade, isto é, se são estratégias apenas aplicáveis a uma dada classe de objectos digitais ou se se tratam de estratégias genéricas, passíveis de ser administradas a qualquer classe de objectos digitais [7].

---

<sup>8</sup> Também designada na literatura por preservação de tecnologia.

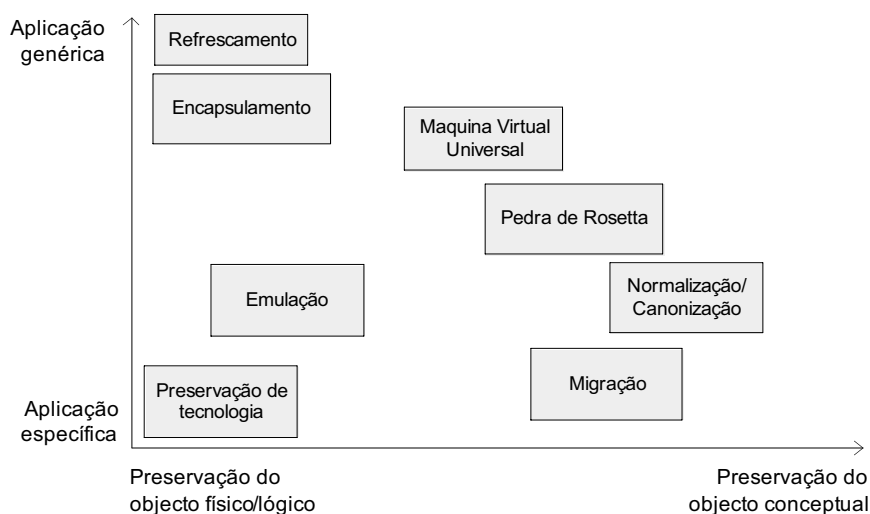


Fig. 7 - Classificação das diferentes estratégias de preservação digital.

## Preservação de tecnologia

Uma das primeiras estratégias de preservação a ser proposta consiste na conservação do contexto tecnológico utilizado originalmente na concepção dos objectos digitais que se procuram preservar. Esta estratégia consiste, essencialmente, na conservação e manutenção de todo o hardware e software necessários à correcta apresentação dos objectos digitais [14-17].

Trata-se sobretudo da criação de museus de tecnologia. Aqui, o foco da preservação não se concentra no objecto conceptual, mas sim na preservação do objecto digital na sua forma original. Os impulsionadores desta estratégia consideram-na a única forma suficientemente eficaz para assegurar que os objectos digitais são experimentados de forma fidedigna [2].

Contudo, a história da computação tem vindo a demonstrar que qualquer plataforma tecnológica, mesmo a mais popular, acaba inevitavelmente por se tornar obsoleta, acabando frequentemente por desaparecer sem deixar qualquer rasto [17]. Este tipo de estratégias introduz dificuldades ao nível da gestão do espaço físico, manutenção e custo de operação, tornando-as



inadequadas para aplicação a longo-prazo [2]. Outras desvantagens assinaláveis deste tipo de estratégias têm que ver com o facto de o acesso à informação ficar confinado a apenas alguns locais físicos do globo e com condicionalismos acrescidos ao nível da reutilização de informação [18].

## Refrescamento

Um objecto digital torna-se persistente no momento em que é inscrito num suporte físico de armazenamento (e.g. disquete, disco rígido, CD-ROM). Garantir a integridade do suporte é fundamental para que a informação nele armazenada possa ser correctamente interpretada. Se o suporte físico se deteriorar ou se se tornar obsoleto a ponto de deixarem de existir periféricos capazes de extrair a informação nele armazenada, incorremos no sério risco dessa informação se perder para sempre [17].

O *refrescamento* de suporte consiste na transferência de informação de um suporte físico de armazenamento para outro mais actual antes que o primeiro se deteriore ou se torne irremediavelmente obsoleto [15, 17, 19, 20].

O refrescamento atempado de suporte não constitui uma estratégia de preservação por si só. Deverá, em vez disso, ser entendido como um pré-requisito para o sucesso de qualquer estratégia de preservação [21]. A frequente verificação da integridade dos suportes físicos, assim como o seu refrescamento periódico, são consideradas actividades vitais num contexto de preservação digital.

## Emulação

As estratégias de emulação baseiam-se essencialmente na utilização de um software, designado *emulador*, capaz de reproduzir o comportamento de uma plataforma de hardware e/ou software, numa outra que à partida seria incompatível [18]. A grande vantagem desta abordagem está na capacidade

de preservar, com um elevado grau de fidelidade, as características e as funcionalidades do objecto digital original [2].

Tal como acontece em estratégias baseadas na preservação de tecnologia, as técnicas de emulação centram-se na preservação do objecto lógico no seu formato original. No entanto, este tipo de estratégias não sofre de alguns dos problemas anteriormente enunciados, como por exemplo, o envelhecimento do hardware.

Existem essencialmente dois tipos de emuladores: emuladores de sistemas operativos e de hardware. Os primeiros focam-se na reprodução de um sistema operativo por completo permitindo, deste modo, a execução de diversas aplicações no contexto de um único emulador (e.g. Wine [22]). Os segundos visam mimar o comportamento de uma plataforma de hardware, possibilitando que vários sistemas operativos e correspondentes aplicações possam ser executados no contexto de um único emulador (e.g. VMware Workstation [23], Parallels Desktop [24]) [7, 25]. Existem actualmente vários exemplos de emuladores de plataformas consideradas obsoletas, e.g. ZX Spectrum [26], Nintendo NES [27], entre outras.

Rothenberg, um dos principais promotores deste tipo de abordagens, defende um modelo teórico capaz de emular plataformas actuais em computadores futuros. O modelo consiste na conservação do objecto digital juntamente com todo o software necessário à sua execução/apresentação (incluindo o sistema operativo) e na criação de uma especificação abstracta da plataforma de hardware que suporta a execução desse software. Essa especificação deverá ser escrita numa linguagem independente da plataforma e ser suficientemente rica para que um emulador possa ser construído automaticamente num qualquer computador do futuro [18].

Hendley considera que a emulação apenas deveria ser utilizada em contextos em que a comunidade de interesse valoriza a preservação do ambiente tecnológico original ou ainda em situações em que os objectos digitais não

são passíveis de ser convertidos para formatos contemporâneos [17]. Outros autores consideram potencialmente arriscado confiar no software original como forma de preservar objectos digitais, uma vez que este pode ser portador de vírus ou *bugs* que poderão, no futuro, resultar em perdas substanciais de informação [7, 28].

É importante reconhecer, também, que a criação de especificações capazes de descrever transversalmente plataformas de hardware não é uma tarefa simples de concretizar. Geralmente, implica recorrer a mão-de-obra altamente especializada, o que por si só poderá constituir um obstáculo considerável para a maioria das organizações [7, 25, 29]. Para além disso, a criação de especificações imprecisas ou incompletas poderá impossibilitar a construção futura dos respectivos emuladores [30]. É também importante salientar que, com o tempo, o próprio emulador irá sofrer de obsolescência, havendo então necessidade de o converter para uma nova plataforma ou desenvolver um novo emulador capaz de emular o primeiro [7].

O uso de emuladores parte também do pressuposto que os utilizadores do futuro serão capazes de operar adequadamente aplicações e sistemas operativos há muito desaparecidos. Por exemplo, num futuro próximo será difícil conceber que os utilizadores estejam aptos a enfrentar as particularidades do sistema operativo MS-DOS [31].

Apesar dos problemas apresentados, as estratégias de emulação continuam a assumir um papel importante na preservação de objectos digitais. Determinados tipos de objectos, especialmente aqueles dotados de características dinâmicas e/ou interactivas, poderão exigir o recurso a emuladores como única forma de assegurar uma experimentação fidedigna [32]. As estratégias de emulação são particularmente relevantes em contextos em que o objecto que se pretende preservar é uma aplicação de software, tal como acontece actualmente com um número crescente de jogos de computador considerados de valor histórico assinalável.

## Migração/conversão

A migração consiste na “(...) transferência periódica de material digital de uma dada configuração de hardware/software para uma outra, ou de uma geração de tecnologia para outra subsequente” [15].

Como referido anteriormente, os objectos digitais são constituídos por elementos estruturais e elementos de informação. O formato de um objecto digital constitui a estrutura pela qual estes elementos de informação se encontram organizados. Neste contexto, a migração pode ser vista como o processo responsável pela reorganização dos elementos de informação que constituem um objecto digital [33].

Ao contrário das estratégias de preservação já apresentadas que procuram cristalizar o objecto digital no seu formato original, as estratégias baseadas em migração centram-se sobretudo na preservação do seu conteúdo intelectual, ou seja, na preservação do objecto conceptual [34].

A migração tem como objectivo manter os objectos digitais compatíveis com tecnologias actuais de modo a que um utilizador comum seja capaz de os interpretar sem necessidade de recorrer a artefactos menos convencionais, como por exemplo, emuladores.

No entanto, os processos de migração introduzem algumas desvantagens que devemos considerar. Neste tipo de processos existe uma grande probabilidade de algumas das propriedades que constituem os objectos digitais não serem correctamente transferidas para o formato de destino adoptado [29, 35]. Isto deve-se, sobretudo, a incompatibilidades existentes entre os formatos de origem e destino ou à utilização de conversores incapazes de realizar as suas tarefas adequadamente [33, 36].

Adicionalmente, não é espectável que uma estratégia de migração possa resolver permanentemente os problemas de preservação. O formato de

destino encontra-se, também este, sob constante ameaça de se tornar obsoleto o que significa que será apenas uma questão de tempo até que uma nova migração tenha de ser administrada. Não obstante, a migração é de longe a estratégia de preservação mais aplicada até à data e a única que tem vindo a dar provas da sua eficácia [2].

Existem diversas variantes de migração que poderão ser consideradas: migração para suportes analógicos, actualização de versões, conversão para formatos concorrentes, normalização, migração a-pedido e migração distribuída.

## **Migração para suportes analógicos**

A migração para suportes analógicos consiste na conversão de objectos para suportes não digitais com o intuito de aumentar a sua longevidade [15]. Esta estratégia consiste, essencialmente, na reprodução de um objecto digital em papel, microfilme ou qualquer outro suporte analógico de longa duração e concentrar os esforços de preservação em torno do novo suporte.

Esta estratégia, no entanto, apenas pode ser aplicada a objectos digitais que possuam uma representação aproximada em suportes analógicos, como por exemplo, documentos de texto ou imagens. Objectos interactivos e/ou dinâmicos ficam assim automaticamente excluídos deste tipo de estratégias.

## **Actualização de versões**

É bastante comum encontrar aplicações de software capazes de abrir ou importar objectos digitais produzidos por versões anteriores dessa mesma aplicação. Essas aplicações permitem geralmente gravar os objectos importados no formato mais actual produzido pela mesma. Esta operação designa-se por *actualização da versão do formato*.

A actualização de versões é, possivelmente, a estratégia de preservação mais vulgarmente utilizada pelos generalidade dos utilizadores. Essencialmente, consiste em actualizar os materiais digitais produzidos por um determinado software recorrendo a uma versão mais actual do mesmo [7].

## **Conversão para formatos concorrentes**

O processo e actualização de versões é geralmente controlado pela organização que desenvolveu a respectiva aplicação de software. A qualidade da migração depende, assim, da capacidade dos importadores fornecidos pelo fabricante e do grau de retrocompatibilidade oferecido pelo novo formato.

Idealmente, o fabricante deveria assegurar que todos os atributos presentes numa dada versão de um formato se encontram disponíveis na nova versão que o vem substituir. No entanto, independentemente do sucesso económico de um fabricante ou produto de software, os formatos estão constantemente sujeitos a descontinuidade [7]. Uma forma de garantir que os objectos digitais sobrevivem a este tipo de rupturas tecnológicas consiste em convertê-los para formatos de uma linha de produtos concorrente.

Existem, no entanto, formatos que não são dependentes de qualquer aplicação de software. Tal, acontece com grande parte dos formatos de imagem (e.g. JPEG, TIFF, PNG). Isto possibilita que os objectos sejam convertidos entre formatos análogos, independentemente da aplicação utilizada na sua criação.

## **Normalização**

A normalização tem como objectivo simplificar o processo de preservação através da redução do número de formatos distintos que se encontram no repositório de objectos digitais [2, 7]. Havendo um número controlado de formatos, uma mesma estratégia de preservação poderá ser

aplicada transversalmente a um maior número de objectos digitais, o que poderá conduzir a uma redução generalizada dos custos de preservação [37].

Vejam os um exemplo. Existe um leque variado de opções no que diz respeito a formatos para representação de imagens bidimensionais (e.g. BMP, GIF, JPEG, PNG, TARGA). Se durante o processo de ingestão, todas as imagens digitais forem convertidas para um único formato, futuras intervenções ao nível da sua preservação poderão ser realizadas de forma mais simples e, conseqüentemente, mais económica.

A escolha do formato de normalização é um factor determinante no sucesso desta estratégia. Sempre que possível, deverão ser escolhidos formatos conhecidos pela comunidade de interesse e baseados em normas internacionais abertas [29]. Isto poderá evitar futuras complicações a nível de direitos de autor ou pagamento de royalties [38]. Paralelamente, o formato de normalização deverá ser suficientemente rico para que as características fundamentais dos vários formatos possam ser devidamente incorporadas.

A normalização promove, também, a interoperabilidade entre sistemas distintos. Ao serem utilizados formatos abertos e independentes da plataforma, diferentes configurações de hardware e software serão capazes de os interpretar [7, 39].

A normalização de formatos pode ser implementada de diversas formas. Determinados repositórios procedem à conversão automática dos objectos recebidos para um formato único de preservação. Outros, definem políticas de arquivo que limitam os formatos em que aceitam informação, significando isto, que cabe aos produtores da informação converter os seus objectos digitais para os formatos estipulados [40, 41]. O argumento que suporta esta abordagem assenta no pressuposto de que os produtores de informação serão as entidades mais indicadas para avaliar a qualidade da conversão efectuada.

## Migração a-pedido

O sucesso de uma migração depende, fundamentalmente, da qualidade dos conversores utilizados e da capacidade que o formato de destino possui para acomodar o conjunto de propriedades do formato de partida. Poder-se-á assumir, no entanto, que sempre que é efectuada uma migração, os objectos digitais resultantes serão de alguma forma diferentes dos objectos de partida. Ao fim de algumas iterações, os objectos preservados poderão ser substancialmente diferentes dos objectos originais (Fig. 8). Para combater este fenómeno de degradação surgiu uma estratégia designada por migração a-pedido [42].

Neste tipo de migração, ao invés de as conversões serem aplicadas ao objecto mais actual, estas são sempre aplicadas ao objecto original (Fig. 9). Deste modo, se de uma dada conversão resultar um objecto substancialmente diferente do original, numa futura conversão, o problema poderá ser resolvido recorrendo a um conversor de melhor qualidade ou a um formato de destino mais adequado.

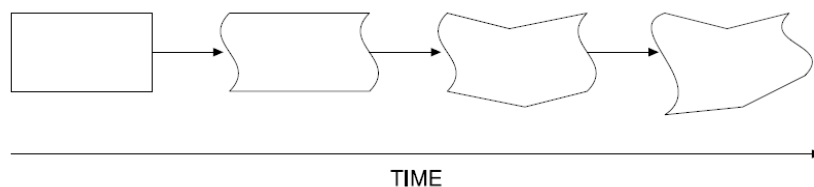


Fig. 8 - Degradação do objecto digital ao longo de sucessivas migrações.

Esta abordagem possui como principal vantagem o facto de, uma vez construído o módulo de decodificação do conversor (i.e. o módulo capaz de ler as propriedades do formato de origem), apenas ser necessário desenvolver os codificadores específicos para cada formato de saída. Não obstante, será necessário suportar ao longo do tempo um conjunto alargado de conversores de modo a garantir a capacidade de transformar os objectos armazenados



nos seus formatos originais para formatos que sirvam adequadamente as necessidades dos seus consumidores.

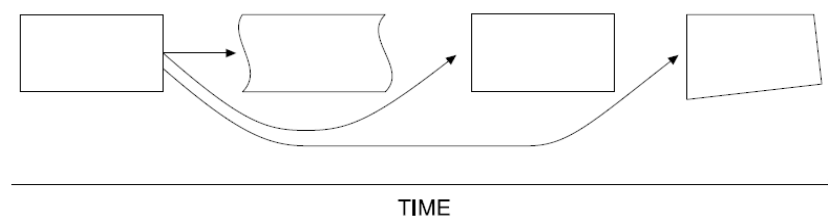


Fig. 9 - Migração a-pedido.

## Migração distribuída

Os mais recentes desenvolvimentos no contexto da migração introduzem arquitecturas distribuídas de conversores (Fig. 10). Neste tipo de migração, existe um conjunto de serviços de conversão que se encontram acessíveis através da Internet e que poderão ser invocados remotamente recorrendo a uma pequena aplicação-cliente.

Existem actualmente várias iniciativas que visam o desenvolvimento deste tipo de conversores. O Typed Objects Model (TOM) sintetiza um sistema distribuído de conversores, suportado por uma taxionomia de tipos e formatos de objectos, que recorre a agentes mediadores para descobrir e executar conversões entre formatos [43].

No Lister Hill National Center for Biomedical Communications foi desenvolvido um Serviço *Web*<sup>9</sup> que converte objectos digitais de cinquenta formatos distintos para PDF [44]. Hunter e Choudhury dão um passo em frente propondo uma rede de serviços de conversão suportada por uma descrição semântica que possibilita a sua descoberta e invocação automática por agentes de software [45].

---

<sup>9</sup> Do inglês *Web Service*.

Na Universidade do Minho está actualmente a ser desenvolvida uma Arquitectura Orientada ao Serviço (SOA) que disponibiliza várias centenas de serviços de conversão, avaliação e recomendação [3, 46-50].

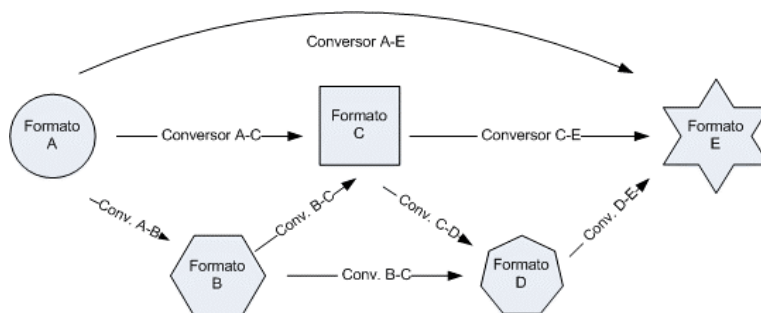


Fig. 10 - Migração distribuída baseada em Serviços *Web*.

Este tipo de migração apresenta algumas vantagens face às estratégias de migração mais convencionais, nomeadamente:

- A utilização de serviços de conversão permite esconder as especificidades de cada conversor e da plataforma que o suporta;
- A criação de serviços redundantes assegura a fiabilidade do sistema perante situações de ruptura parcial;
- A existência de múltiplos caminhos de migração permite à solução resistir ao desaparecimento gradual de parte dos conversores;
- Este tipo de abordagem é compatível com uma série de variantes de migração, como por exemplo, normalização e migração a-pedido;
- A criação de uma rede global de conversores poderá conduzir a uma redução generalizada dos custos de preservação. Qualquer organização poderá rentabilizar os seus investimentos no

desenvolvimento de conversores, publicando-os na rede de serviços e cobrando uma pequena taxa pela sua utilização.

Apesar das vantagens apresentadas, a migração distribuída poderá não ser adequada a todos os contextos. Um repositório de informação digital pode facilmente conter milhares de itens, atingindo níveis de armazenamento na ordem dos múltiplos Terabytes. Transferir através da Internet um volume de informação desta natureza acarreta custos que poderão ser impeditivos para muitas organizações. Para além disso, requisitos em termos de largura de banda, segurança dos dados e tempo de transferência poderão ser factores determinantes no sucesso deste tipo de estratégias.

## Encapsulamento

Por vezes, não é fácil determinar o valor intrínseco de determinados objectos digitais. Poderão passar-se vários anos até que a comunidade revele um interesse particular por uma determinada colecção de objectos [51]. Durante esse tempo, o material custodiado poderá nunca ser consultado. Neste tipo de cenários, as estratégias de preservação que carecem de uma diligência contínua (e.g. migração) poderão revelar-se demasiado onerosas. As soluções baseadas em encapsulamento procuram resolver esse problema mantendo os objectos digitais inalterados até ao momento em que se tornam efectivamente necessários.

A estratégia de 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. Esta informação poderá consistir, por exemplo, numa descrição formal e detalhada do formato do objecto preservado [52].

O Formato Universal de Preservação<sup>10</sup> (UPF) é uma iniciativa que visa criar um formato normalizado para agregar metainformação de preservação junto

---

<sup>10</sup> Do inglês *Universal Preservation Format*.

do objecto digital. Este formato será independente da aplicação, do sistema operativo e do suporte físico utilizados para criar e armazenar o objecto digital [53, 54].

Raymond Lorie propõe uma alternativa a esta estratégia substituindo a especificação formal por uma aplicação de software compilada para uma máquina virtual universal, por exemplo, para a Java Virtual Machine [55, 56]. Esta aplicação é na realidade um descodificador e tem como finalidade apresentar uma visão lógica do objecto digital, permitindo deste modo uma navegação simples através das suas propriedades. Lorie argumenta que a máquina virtual universal é suficientemente simples para que possa ser implementada em qualquer arquitectura de hardware futura.

## **A Pedra de Rosetta digital**

O povo egípcio deixou uma infindável quantidade de vestígios da sua presença na Terra. No entanto, só a partir do século XIX foi possível decifrar os seus escritos hieroglíficos. Tudo aconteceu em 1799 quando um grupo de soldados franceses descobriu no delta do Nilo um bloco de granito que ficou conhecido como a Pedra de Rosetta. Nele encontrava-se escrito em três línguas distintas (egípcio hieroglífico, egípcio cursivo e grego clássico) um decreto emitido em 196 a.C. por Ptolomeu V Epifânio. Em 1822 o paleógrafo francês Jean-François Champollion descodificou a versão egípcia do texto recorrendo aos seus conhecimentos de grego clássico, um idioma bem conhecido dos historiadores da época [57, 58]. Esta descoberta conduziu à descodificação de inúmeros outros textos egípcios encontrados nos mais variados locais e suportes (e.g. monumentos, rochas, papiros).

Heminger e Robertson propõem a utilização de uma estratégia semelhante para recuperar objectos digitais para os quais não existe informação suficiente sobre o seu formato [51]. Nesta estratégia, em vez de se preservar as regras que permitem descodificar o objecto digital, são reunidas amostras de objectos que sejam representativas do formato que se pretende recuperar.

Estas amostras deverão existir num formato que possa ser directamente interpretado pelo ser humano. Trata-se do conjunto de referência, i.e., a versão grega do decreto inscrito na Pedra de Rosetta. Com esta informação seria possível inferir as regras necessárias para traduzir/converter o objecto original para um qualquer formato contemporâneo [7, 59].

Um exemplo de aplicação desta estratégia consiste em imprimir em papel um conjunto representativo de documentos de texto juntamente com a sua representação binária. No futuro, as regras necessárias interpretar e migrar os objectos para um novo formato poderiam ser inferidas, comparando os documentos impressos com a sua representação binária [7].

Esta estratégia deverá ser considerada apenas em situações em que todos os esforços de preservação falharam. Trata-se sobretudo de uma ferramenta de arqueologia digital e não propriamente de uma estratégia de base para preservação de objectos digitais [51].

# Directórios de formatos

---

Uma das formas de minimizar a ansiedade de todos os agentes envolvidos em processos de preservação digital consiste na criação de directórios centralizados de informação técnica sobre formatos digitais. Esta informação inclui, por exemplo, a identificação dos produtores de um dado formato, a sua data de criação, informação sobre as aplicações que o suportam, especificações técnicas, grau de obsolescência, entre outros.

Para além de disponibilizar este tipo informação, os directórios de formatos poderão prestar serviços avançados de apoio à preservação digital. Por exemplo, um directório de formatos poderá disponibilizar ferramentas capazes de identificar o formato de um dado objecto digital e promover esquemas normalizados para a sua descrição. Poderá ainda fornecer especificações técnicas sobre formatos, permitindo a qualquer instituição desenvolver descodificadores para os mesmos, bem como disponibilizar um conjunto de informações relevantes para apoio a actividades de preservação digital, como por exemplo, informação sobre a cota de mercado de um dado

formato ou produzir recomendações quanto aos formatos mais apropriados para preservação a longo-prazo.

Existem actualmente diversas iniciativas que visam a construção de directórios deste tipo, como por exemplo, o PRONOM [60], Global Digital Format Registry [61] e o Typed Object Model [43].

O PRONOM<sup>11</sup> é uma iniciativa dos UK National Archives<sup>12</sup> que visa a concentração de uma grande quantidade de informação sobre produtos de software e seus formatos associados [60, 62]. O PRONOM disponibiliza ainda uma ferramenta que facilita a identificação de formatos (Droid [63]) e promove a utilização de identificadores únicos para os mesmos.

O Global Digital Format Registry<sup>13</sup> (GDFR) apresenta-se como uma alternativa aos actuais MIME Media Types<sup>14</sup> introduzindo um mecanismo de identificação de formatos mais preciso e rigoroso. O GDFR possui ainda como objectivo a reunião de informação sobre a sintaxe e semântica dos diversos formatos digitais existentes. A sua criação está a cargo de um grupo de trabalho internacional constituído por membros de diversas bibliotecas e arquivos nacionais, assim como bibliotecas académicas, num total de 18 instituições [61].

---

<sup>11</sup> <http://www.nationalarchives.gov.uk/PRONOM/>

<sup>12</sup> Arquivos Nacionais do Reino Unido.

<sup>13</sup> <http://hul.harvard.edu/gdfr/>

<sup>14</sup> Actualmente, o sistema de identificação de formatos mais utilizado é o MIME Media Types. Este sistema é amplamente utilizado na Internet, nomeadamente, para especificar as regras de codificação/descodificação de documentos anexados a mensagens de correio electrónico e para identificar os formatos de dados trocados durante as transferências entre browsers e servidores Web. Este sistema, no entanto, não possui a granularidade necessária para identificar de forma unívoca todos os formatos existentes. Por exemplo, todas as versões da família PDF – da versão 1.0 à 1.4, PDF/X da versão 1 à 3, e PDF/A – são identificadas pelo mesmo descritor: application/pdf.

O projecto Typed Object Model<sup>15</sup> (TOM) assenta no pressuposto de que todos os formatos digitais podem ser vistos como objectos (i.e. possuem propriedades e métodos) e, como tal, será possível construir uma arquitectura baseada em herança, capaz de descrever a estrutura de cada formato de dados, as suas instâncias e as relações existentes entre os mesmos [43]. Este projecto introduz uma taxionomia classificativa de formatos e um sistema distribuído de conversores baseado em agentes mediadores. No entanto, não se antevê que se possa vir a tornar numa norma *de facto*, pois a sua implementação é bastante residual.

---

<sup>15</sup> <http://tom.library.upenn.edu/>



# Autenticidade

---

O conceito de autenticidade está longe de ser consensual entre os profissionais da preservação. Este poderá assumir significados consideravelmente diferentes consoante a comunidade que o manipula. Para um historiador um objecto é autêntico se a sua identidade e integridade não foram comprometidas, i.e., se for possível aferir que um objecto é realmente aquilo que se propõe ser [64]. Esta definição pressupõe que o seu conteúdo é verdadeiro e que o seu contexto histórico se encontra devidamente identificado. Em suma, um objecto é autêntico se estiver conforme o original e se a sua história custodial tiver sido devidamente documentada ao longo do tempo.

Do ponto de vista de um arquivista, a autenticidade de um documento não pressupõe uma legitimação da sua veracidade ou até mesmo utilidade. Um arquivista preocupa-se, sobretudo, com a prova que um documento poderá constituir. Este poderá conter incorrecções, erros ou até falsidades, mas isso não invalida a sua importância como testemunho de que algo aconteceu [65]. Um documento falsificado, por exemplo, pode ser considerado autêntico uma vez que constitui prova de que alguém falsificou um documento [66].

Definições mais abrangentes de autenticidade giram em torno dos conceitos de autenticação, integridade, completude, veracidade, validade, conformidade com o original, significância e adequabilidade ao fim a que se destina [67].

Em termos genéricos, o conceito de autenticidade traduz a capacidade de identificar os elementos diplomáticos que permitem aferir se um dado objecto é autêntico. Trata-se da identificação do “porquê”, do “quando”, do “onde” e do “por quem” de um objecto digital [10]. A autenticidade tem que ver com a capacidade de se conseguir demonstrar que um objecto digital é aquilo que se propõe ser [37, 68-72]. Para atingir esse objectivo é fundamental documentar devidamente a proveniência do objecto, contextualizar a sua existência, descrever a sua história custodial e atestar que a sua integridade não foi comprometida, i.e. provar que existe um conjunto de propriedades, consideradas significativas, que foram correctamente preservadas ao longo do tempo [68, 73].

Os problemas associados à determinação da autenticidade de um recurso não estão limitados à documentação digital. Na idade média, por exemplo, a reprodução de livros era realizada manualmente. Cada cópia de um livro apresentava, geralmente, um conjunto de diferenças face ao original. A maior parte dessas diferenças resultavam de infelizes erros de transcrição. No entanto, não seriam raras as vezes em que escrivães mais perspicazes introduziam deliberadamente “melhorias” durante o processo de transcrição do documento [74].

No contexto digital, os problemas relacionados com a autenticidade são em tudo semelhantes aos do mundo analógico. Contudo, a simplicidade com que alterações podem ser introduzidas, a rapidez com que estas podem ser disseminadas e a dificuldade inerente à sua detecção tornam o problema sensivelmente mais complexo.

No contexto analógico, o conteúdo e o suporte são geralmente duas entidades inseparáveis. As propriedades físicas que caracterizam o suporte

forneem, geralmente, pistas suficientes para que a autenticidade do seu conteúdo possa ser aferida [66]. No mundo digital este tipo de pistas não existe. O ambiente tecnológico é propício à introdução de modificações provocando um clima generalizado de desconfiança em relação à autenticidade deste tipo de material [69, 71, 74, 75].

A preservação de informação no domínio analógico manifesta-se, sobretudo, pela tentativa de conservar o suporte inalterado durante o máximo de tempo possível. No contexto digital, a preservação do suporte ou da sequência de *bits* que constitui o objecto, não é condição suficiente para garantir que a informação permanece acessível, reutilizável e autêntica ao longo do tempo [76]. Preservar informação digital consiste, por vezes, em modificar ou transformar deliberadamente o objecto físico ou lógico que transporta a mensagem (ver Migração/conversão). Para que essa transformação não produza uma mensagem exageradamente degradada, é fundamental definir quais as propriedades da mensagem que deverão ser asseguradas durante o processo de transformação.

Paralelamente, a informação não é um conceito ou substância concreta. Esta materializa-se através de um processo de transformação que eleva um conjunto de símbolos a algo com significado [77]. Para além disso, a interpretação desse significante está sujeita a influências adicionais. O hardware e o software que servem de mediadores nesse processo podem diferir substancialmente de consumidor para consumidor [73]. Neste contexto, a definição de *essência*<sup>16</sup> de um objecto digital é de extrema importância, pois caracteriza o conjunto de propriedades que deverão ser mantidas e preservadas de forma intacta para que o objecto possa ser considerado autêntico [66].

---

<sup>16</sup> Propriedades significativas e *essência* de um objecto digital são duas expressões vulgarmente utilizadas para representar o mesmo conceito.

O conjunto das propriedades significativas não é absoluto nem estático. A sua definição deverá ter em conta a natureza da organização responsável pela preservação, as características da colecção e, acima de tudo, os requisitos e exigências da comunidade de interesse [66, 78]. É a definição da essência de um objecto digital que deverá determinar a forma como o material deve ser preservado [76].

A definição das propriedades significativas de um objecto digital influencia directamente a forma como este deverá ser preservado. Quanto maior for o número de propriedades significativas, maiores serão os requisitos relativamente à infra-estrutura tecnológica necessária para suportar a preservação [79].

Embora desejável, a definição de um conjunto de propriedades significativas para cada objecto digital existente num arquivo não é economicamente viável. Torna-se, assim, necessária a criação de políticas de preservação que exprimam, para cada classe de objectos digitais, o conjunto das propriedades significativas que serão asseguradas pelo repositório [79].

A título de exemplo, considere-se uma biblioteca responsável por preservar artigos científicos (o repositório institucional de uma universidade, por exemplo). Se a sua política de preservação apenas especificar a propriedade significativa: preservação do conteúdo textual dos artigos científicos; então, estes estarão a ser adequadamente preservados se apenas os caracteres ASCII<sup>17</sup> que os constituem forem conservados. Se por outro lado a política de preservação especificar propriedades significativas adicionais como a disposição do texto na página ou a sua formatação, então a preservação dos caracteres ASCII deixa de ser

---

<sup>17</sup> American Standard Code for Information Interchange. Trata-se de um conjunto de códigos capaz de representar letras, dígitos e outros símbolos, amplamente utilizado por computadores para troca de informação textual.

suficiente, passando a ser necessário recorrer a formatos mais complexos, como por exemplo o PDF.

Ainda neste contexto surge o conceito de *canonização*. Lynch apresenta-o como uma forma de avaliar o sucesso de uma migração [80]. O formato canónico tem como objectivo representar de forma unívoca as características essenciais de uma classe de objectos digitais. O formato canónico especifica a ordem e a estrutura das propriedades que constituem os objectos digitais e assume valores por omissão para as propriedades sem valor. O método funciona comparando os objectos canónicos obtidos a partir dos objectos original e convertido (Fig. 11).

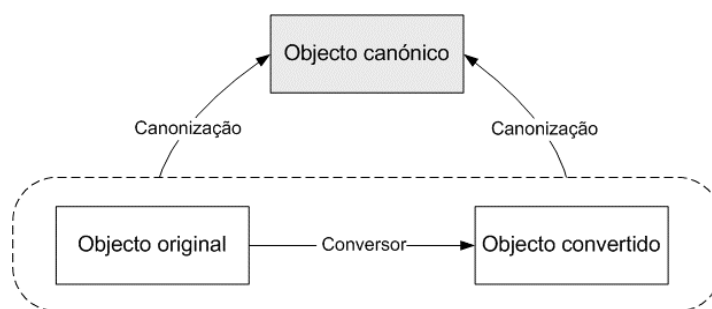


Fig. 11 - Verificação da qualidade de uma migração através de canonização.

Considere-se dois objectos digitais: o original e o convertido. Poder-se-á afirmar que o objecto convertido preserva as características essenciais do objecto original, se os objectos canónicos obtidos a partir de ambos forem em tudo semelhantes [80].

# Metainformação de preservação

---

A metainformação de preservação tem como objectivo descrever e documentar os processos e actividades relacionadas com a preservação de materiais digitais. Ou seja, a metainformação de preservação é responsável por reunir, junto do material custodiado, informação detalhada sobre a sua proveniência, autenticidade, actividades de preservação, ambiente tecnológico e condicionantes legais [68].

No que diz respeito à proveniência, a metainformação de preservação procura descrever a história custodial dos materiais, i.e. o caminho percorrido por estes materiais desde a sua criação até à sua incorporação no repositório [68]. Esta assume também a responsabilidade de garantir a autenticidade dos materiais. Para tal, agrega um conjunto de metainformação que descreve detalhadamente todo o tipo de actividades desenvolvidas no interior do repositório [68].

A metainformação de preservação tem também como objectivo descrever o ambiente tecnológico (i.e. hardware, sistemas operativos e software que sustenta a correcta apresentação dos objectos), bem como incluir informação sobre todas as contingências legais que afectam o material custodiado [68].

O modelo de referência OAIS constituiu um ponto de partida para a discussão em torno da necessidade de criar um conjunto de elementos de metainformação capazes de dar suporte às actividades relacionadas com a preservação digital [12, 68]. Desde então, diversas instituições têm vindo a propor dicionários de metainformação que reflectem as necessidades individuais dos projectos em que estão ou estiveram envolvidas [81-83].

Em 2002, o consórcio Online Computer Library Center e Research Libraries Group (OCLC/RLG) consolidou este conhecimento acumulado num único documento onde se enumeram as diversas classes de informação que deverão estar presentes num esquema de metainformação de preservação [84].

A OCLC/RLG constituiu um segundo grupo de trabalho designado PREMIS (PREservation Metadata: Implementation Strategies) com o objectivo de continuar o desenvolvimento deste esquema de metainformação. O grupo de trabalho foi constituído por um comité internacional com mais de trinta especialistas em preservação digital. Deste trabalho resultou o Dicionário de Dados PREMIS<sup>18</sup>, um documento que identifica e descreve um conjunto básico de elementos de metainformação de suporte à preservação digital, bem como um conjunto de recomendações quanto à forma como estes deverão ser utilizados no contexto de um arquivo digital [85].

---

<sup>18</sup> Do inglês *PREMIS Data Dictionary*.

O Dicionário de Dados PREMIS descreve cinco entidades fundamentais: Entidades Intelectuais, Agentes, Eventos, Objectos e Direitos<sup>19</sup> (Fig. 12).

Uma Entidade Intelectual é um conjunto coerente de informação que pode ser identificado e descrito como uma unidade. Um livro, uma fotografia ou uma base de dados são exemplos do que podemos considerar Entidades Intelectuais.

Uma Entidade Intelectual pode conter outras Entidades Intelectuais no seu interior. Um sítio *Web* pode ser constituído por várias páginas *Web* e cada uma destas pode conter diversas fotografias. É também importante realçar que uma Entidade Intelectual pode assumir mais do que uma representação. Uma fotografia, por exemplo, pode existir em formato digital, impressa em papel fotográfico ou estampada numa T-shirt. A Entidade Intelectual<sup>20</sup> que a descreve é sempre a mesma, apesar da sua manifestação física variar substancialmente.

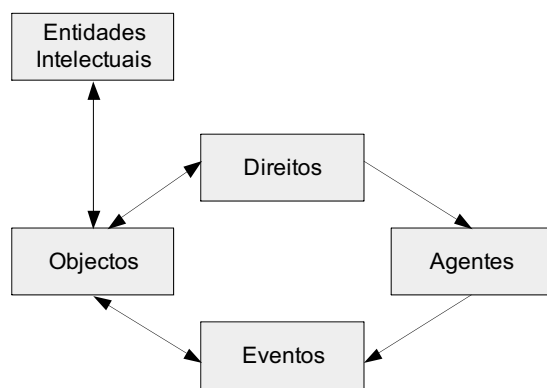


Fig. 12 - Entidades do Dicionário de Dados PREMIS.

---

<sup>19</sup> Do inglês *Intellectual Entities, Agents, Events, Objects e Rights*.

<sup>20</sup> De notar o paralelismo entre o conceito de Entidade Intelectual e Objecto Conceptual.



A entidade *Agente* descreve qualquer pessoa, organização ou aplicação de software que esteja envolvida num Evento de preservação. Por sua vez, um *Evento* agrega informação sobre as acções de preservação realizadas em torno de um *Objecto* (e.g. verificação de integridade, migração, etc.). O registo das acções de preservação, especialmente aquelas que têm como objectivo modificar o objecto digital, é considerado uma actividade fundamental para assegurar a autenticidade do mesmo.

A entidade *Objecto* é responsável por descrever o conjunto das representações ou manifestações físicas de uma Entidade Intelectual. Um livro, por exemplo, poderá ser representado de diversas formas: através de um conjunto de imagens TIFF, em PDF, etc. Um *Objecto* pode ainda ser dividido de três subtipos: *Representação*, *Ficheiro* ou *Sequência de bits*<sup>21</sup>.

Uma *Representação* é um conjunto de ficheiros com metainformação estrutural associada que permite que um dado objecto intelectual possa ser consumido recorrendo a software apropriado. Um *Ficheiro* trata-se de um conjunto ordenado de bytes reconhecido por um sistema operativo. Um ficheiro pode ter propriedades associadas tais como: tamanho, data de criação/modificação, permissões de acesso, etc. Por último, uma *Sequência de bits* é um conjunto de dados coeso e com particular interesse de preservação que pode ser identificado e extraído a partir de um ficheiro, e.g. a faixa de áudio de um ficheiro de vídeo.

A entidade *Direitos* reúne informação sobre os direitos de propriedade intelectual e permissões associadas ao objecto digital.

---

<sup>21</sup> Do inglês *Representation, File e Bitstream*.

# Avaliação de estratégias de preservação

---

Apesar do número de estratégias de preservação não parar de aumentar, nenhuma destas foi até ao momento seriamente validada ou universalmente aceite [86]. A preferência por uma alternativa de preservação exige, geralmente, a ponderação de diversos factores, como por exemplo: as características da colecção, a satisfação da comunidade de interesse ou os custos associados ao processo de preservação [86].

Rauch e Rauber desenvolveram um método capaz de comparar e seleccionar alternativas de preservação tendo em conta as necessidades individuais de cada organização ou utilizador [86, 87]. O seu trabalho é baseado em conceitos de Análise de Utilidade [88], um método originalmente desenvolvido para o auxílio à tomada de decisão em projectos complexos no domínio da construção civil e economia.

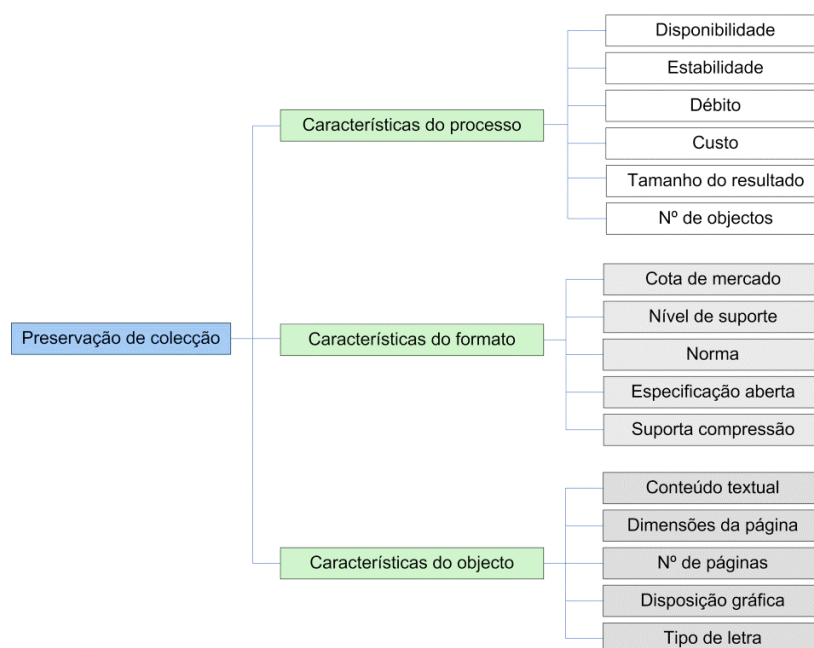


Fig. 13 - Exemplo de uma árvore-objectivo ponderada.

O método segue um processo composto pelas seguintes etapas:

1. Inicialmente é construída uma árvore-objectivo, i.e. um conjunto de critérios organizados de forma hierárquica que servirão para aferir o grau de adequabilidade de uma dada estratégia de preservação (Fig. 13);
2. Numa segunda fase são associadas unidades de medida a cada um desses critérios, e.g. milímetro, segundo, Mb/s, EURO;
3. A terceira fase consiste na selecção de um conjunto de alternativas que poderão ser utilizadas para preservar a colecção de objectos em causa. Estas alternativas serão comparadas e ordenadas de acordo com as necessidades específicas da organização;
4. No quarto passo, cada uma das alternativas é executada face a um conjunto representativo de objectos digitais. O resultado de cada

- intervenção será então avaliado à luz de cada um dos critérios que constam da árvore-objectivo (Fig. 14 – 1);
5. No quinto passo, os resultados das avaliações são normalizados, i.e. transformados em unidades numéricas comparáveis (Fig. 14 – 2);
  6. No sexto, são atribuídas ponderações percentuais a cada um dos critérios que constituem a árvore-objectivo. Os pesos atribuídos representam os requisitos ou preferências de preservação da organização e irão determinar a estratégia a adoptar (Fig. 14 – 3);
  7. O passo sete consiste na agregação de valores parciais e totais obtidos a partir das experiências (Fig. 14 – 4);
  8. Finalmente, todas as alternativas são ordenadas mediante o seu grau de adequação face aos requisitos manifestados pela organização.

	Critérios de avaliação	E1	E2		E1	E2		Pesos		E1	E2
Características do Processo	Disponibilidade	0.95	0.9		1.0	0.0		0.2		0.2	0.0
	Estabilidade	0.8	0.9		0.0	1.0		0.2		0.0	0.2
	Débito	1 Kb/s	2 Kb/s		0.0	1.0		0.2		0.0	0.2
	Custo	0.01 €	0.05 €		1.0	0.0		0.3		0.3	0.0
	Tamanho do resultado	1.2	0.9		0.0	1.0		0.07		0.0	0.07
	Nº de objectos	1.0	1.0		1.0	1.0		0.03		0.03	0.03
Características do Formato	Cota de mercado	0.5	0.9		0.0	1.0		0.2		0.0	0.2
	Nível de suporte	1.0	0.2		1.0	0.0		0.2		0.2	0.0
	Norma	1.0	0.0		1.0	0.0		0.2		0.2	0.0
	Especificação aberta	1.0	1.0		1.0	1.0		0.3		0.3	0.3
	Suporta compressão	0.0	1.0		0.0	1.0		0.1		0.0	0.1
...	...	...		...	...		...		...	...	
Características do Objecto	Conteúdo textual	0.9	0.95		0.0	1.0		0.4		0.0	0.4
	Dimensões da página	0.6	0.9		0.0	1.0		0.2		0.0	0.2
	Nº páginas	1.0	1.0		1.0	1.0		0.2		0.2	0.2
	Disposição gráfica	0.8	0.6		1.0	0.0		0.1		0.1	0.0
	Tipo de letra	0.5	0.8		0.0	1.0		0.1		0.0	0.1
...	...	...		...	...		...		...	...	
										1,53	2.0

(1)
(2)
(3)
(4)  
 Critérios avaliados por cada estratégia      Valores normalizados [0-1]      Atribuição de pesos a cada critério      Resultado final

Fig. 14 - Processo de selecção de estratégias de preservação. E1 e E2 representam duas estratégias de preservação distintas.

É importante realçar que a construção da árvore-objectivo é por si só uma tarefa complexa e morosa, geralmente envolvendo profissionais da área tecnológica, arquivística, produtores de informação e respectivos consumidores.

Rauch e Rauber têm promovido a construção de árvores-objectivo para diversas classes de objectos digitais através da realização de *workshops* no seio de organizações detentoras de informação digital. Durante esses *workshops*, um conjunto de pessoas é convidado a sugerir critérios de avaliação que consideram importantes no sentido de garantir a preservação de um conjunto concreto de objectos digitais. Estes critérios são então organizados em classes e subclasses de forma constituir uma árvore-objectivo semelhante à apresentada na Fig. 13.

A árvore-objectivo da Fig. 13 reúne um conjunto de critérios para a avaliação de estratégias de preservação aplicadas a documentos de texto. Nela podemos encontrar critérios relativos ao processo de preservação (e.g. disponibilidade, estabilidade, débito, custo, etc.), aos formatos envolvidos na preservação (e.g. cota ou taxa de penetração no mercado, nível de suporte, se se trata de um formato normalizado, etc.) e critérios relativos aos objectos propriamente ditos (e.g. conteúdo textual, dimensões da página, nº de páginas, etc). O último conjunto de critérios pode ser entendido como o conjunto das propriedades significativas relevantes para esta classe de objectos digitais, i.e. documentos de texto.

Este método de avaliação pode vir a ser utilizado para construir sistemas automáticos de recomendação que possibilitem a pequenas organizações identificar e implementar estratégias de preservação adequadas à conservação dos objectos digitais [47, 48, 50].

# Síntese e conclusão

---

Este livro tem como objectivo descrever e contextualizar as principais actividades que têm vindo a ser realizadas internacionalmente no domínio da preservação digital. O livro começa por definir preservação digital e introduzir o conceito de objecto digital. Este é apresentado numa perspectiva semiótica, sendo feita uma análise dos diferentes níveis de abstracção a que pode ser observado: físico, lógico e conceptual. Esta visão multidimensional do objecto digital promove uma melhor compreensão e enquadramento das diferentes estratégias de preservação apresentadas ao longo desta publicação.

De seguida foi apresentado o modelo de referência OAIS (Open Archival Information System), uma norma internacional que visa a identificação dos principais componentes funcionais e objectos de informação presentes num sistema de arquivo com pretensões de preservação a longo-prazo. O modelo de referência serviu fundamentalmente para introduzir alguma da notação utilizada ao longo do livro.

Foram também descritas e contextualizadas as principais estratégias de preservação que têm vindo a ser propostas pela comunidade científica, bem como as mais relevantes iniciativas no que toca a directórios de formatos.

O livro engloba ainda uma breve discussão sobre autenticidade, salientando-se a importância do conceito de propriedade significativa na elaboração de políticas de preservação. Paralelamente, procurar-se-á realçar a necessidade da utilização de metainformação como meio para assegurar a autenticidade dos materiais custodiados, dando especial ênfase ao Dicionário de Dados PREMIS.

Finalmente, foi apresentado um método que permite avaliar e comparar diferentes estratégias de preservação, ordenando-as de acordo com o seu grau de adequação aos requisitos individuais da instituição que pretende realizar a preservação digital.

É fundamental realçar que existem muitos outros aspectos da preservação que não foram descritos nem tão pouco mencionados ao longo desta publicação. Questões como a definição de políticas de preservação, modelos de financiamento, normas de metainformação (descritiva, técnica, estrutural, administrativa, etc.), requisitos para a certificação de repositórios digitais, ferramentas de preservação, disposições legais e projectos relevantes a nível nacional e internacional ficaram de fora desta edição. Não obstante, a informação presente neste livro é suficiente para que o leitor seja capaz de identificar e compreender os principais conceitos associados a esta temática, bem como divisar estratégias adequadas à preservação de classes distintas de objectos digitais.

Para além do disposto, o leitor deverá compreender que, apesar do número elevado de alternativas de preservação digital apresentadas, continuam a não existir provas conclusivas quanto à eficácia de cada uma delas. Para muitas, só o tempo o dirá. Mesmo assim, tem-se assistido a uma batalha ideológica entre aqueles que defendem estratégias baseadas em migração e os que

defendem a utilização de estratégias de emulação. Esta discussão tem na sua origem, questões relacionadas com a autenticidade e integridade dos materiais.

Nas estratégias derivadas da migração, assume-se que os objectos digitais irão sofrer modificações ao longo do tempo. Determinadas migrações poderão mesmo originar perdas substanciais de informação. Para os defensores da emulação, assumir de antemão que a informação que se procura preservar será sistematicamente adulterada ao longo do tempo viola os pressupostos mais elementares da preservação [18].

Esta questão, no entanto, não está confinada ao domínio digital. Na arquivística tradicional, há quem defenda que a preservação do material no seu estado original deverá ser considerada como a única medida de sucesso. Há, no entanto, quem opte por transferir os seus materiais para suportes menos volumosos, como por exemplo o microfilme, tomando uma decisão explícita pela poupança de espaço em detrimento da originalidade [89]. O material digital, no entanto, possui características que fazem com que estas questões acabem por ser de certa forma amplificadas. O material digital é estruturalmente mais complexo que o seu equivalente analógico. Diferentes tipos de informação podem ser combinados num único objecto (e.g. texto, vídeo, som) e este, pode ainda, exibir capacidades dinâmicas e/ou interactivas. Para além disso, este pode facilmente ser modificado, desconstruído e re combinado de formas inovadoras usando software adequado [89].

Apesar de tudo, a preocupação obstinada pela originalidade tem vindo a diminuir à medida que aumenta a compreensão generalizada acerca dos processos de preservação. Começa-se a difundir a ideia de que o foco da preservação não deverá estar na retenção do objecto físico original, mas na conservação da experiência sensorial que é produzida por esse objecto [29].



Neste contexto, Burkel questiona-se sobre o papel da tecnologia no processo de interpretação de informação digital – “(...) as entradas e saídas de qualquer sistema digital são na forma de linguagens humanas. A tecnologia e as suas linguagens próprias apenas asseguram um processamento mais eficiente dessa informação no interior do computador” [90].

Reforçando esta ideia, Thibodeau argumenta que no futuro, tal como hoje, os consumidores desejam servir-se das tecnologias mais modernas ou daquelas que melhor conhecem para manipular mais eficientemente a informação com que trabalham. A opção por uma estratégia de emulação poderá conduzir ao incumprimento desta necessidade básica [7, 89].

A batalha ideológica – migração versus emulação – tem, assim, tendência a esgotar-se. Instala-se o reconhecimento generalizado de que diferentes estratégias de preservação deverão ser implementadas mediante o contexto específico da organização preservadora e o tipo de objectos que se pretendem preservar [91]. A selecção de estratégias de preservação deve ter em conta diversos factores como as características intrínsecas dos objectos, o custo de implementação e manutenção, os interesses do arquivo e da sua comunidade de interesse, entre outros. Para diversos autores este último ponto é de extrema importância. A informação terá pouca utilidade se não for preservada e disseminada de acordo com as necessidades da sua comunidade de interesse [40, 89, 92].

A tendência actual vai no sentido de combinar um conjunto de técnicas como o refrescamento automático do suporte, a normalização para formatos de preservação durante o processo de ingestão, a conservação do objecto original (como salvaguarda e para fins arqueológicos) e migração a-pedido para adaptar os formatos de preservação (i.e. os Pacotes de Informação de Arquivo) a formatos mais adequados à sua disseminação (i.e. Pacotes de Informação de Disseminação) [7]. Os objectos digitais predominantemente dinâmicos ou interactivos são geralmente preservados nos seus formatos originais e são utilizadas estratégias de emulação para a sua interpretação [41].

Apesar do recente aparecimento de ferramentas de software que auxiliam o processo de arquivo e preservação (e.g. OCLC Digital Archive<sup>22</sup>, DSpace<sup>23</sup>, LOCKSS<sup>24</sup>, Fedora<sup>25</sup>, Eprints<sup>26</sup>, PANDAS<sup>27</sup>, Digital Information Archiving System<sup>28</sup> da IBM, JHove<sup>29</sup>, Droid<sup>30</sup>, Xena<sup>31</sup>) ainda se verifica uma escassez assinalável de produtos comerciais com capacidades de preservação [41]. Isto faz com que cada organização se sinta de certa forma responsável pelo desenvolvimento e implementação do seu próprio sistema de preservação, bem como pela definição de políticas de preservação adequadas.

A definição de uma política de preservação envolve, geralmente, todas as facetas de um arquivo. Implica a criação de políticas de avaliação e selecção de materiais, a identificação de esquemas de metainformação apropriados (e.g. metainformação descritiva, técnica, de disseminação, estrutural e de preservação), a definição de estratégias de preservação adequadas a cada classe de objectos digitais, a criação de planos de sucessão para a eventualidade da organização detentora da informação interromper a sua actividade, a utilização de modelos sustentáveis de financiamento, entre outros.

Uma política de preservação deverá descrever claramente as estratégias adoptadas para assegurar a preservação dos materiais a cada um dos níveis de abstracção a que estes podem ser considerados, i.e. físico, lógico e conceptual, mas também a níveis superiores, como o social, o económico e o organizacional. A tabela que se segue enumera possíveis estratégias que

---

<sup>22</sup> <http://www.oclc.org/digitalarchive/>

<sup>23</sup> <http://www.dspace.org>

<sup>24</sup> <http://www.lockss.org>

<sup>25</sup> <http://www.fedora.info>

<sup>26</sup> <http://www.eprints.org/software/>

<sup>27</sup> <http://pandora.nla.gov.au/pandas.html>

<sup>28</sup> <http://www-5.ibm.com/nl/dias/>

<sup>29</sup> <http://hul.harvard.edu/jhove/>

<sup>30</sup> <http://www.nationalarchives.gov.uk/aboutapps/pronom/droid.htm>

<sup>31</sup> <http://sourceforge.net/projects/xena>

poderão ser utilizadas para preservar cada um destes níveis de abstracção (Tabela 1).

Nível de abstracção	Estratégias a aplicar
Físico	Acondicionamento adequado dos suportes físicos, utilização de suportes de longa duração, salas de prevenção contra desastres naturais, etc.
Lógico	Refrescamento, <i>backup</i> , replicação local e/ou remota, etc.
Conceptual	Migração, Emulação, Encapsulamento, etc.
Social	O sistema de preservação deverá ser capaz de impedir ou de corrigir a ocorrência de erros provocados por operadores ou atacantes externos, e.g. implementação de mecanismos de <i>undo</i> , registo de actividades, autenticação e gestão de permissões, etc.
Económico	Definição de modelos de financiamento sustentáveis. As despesas com a preservação deverão fazer parte dos orçamentos de base das organizações.
Organizacional	Definição de planos de sucessão que garantam a sobrevivência dos materiais face à eventual de cessação de actividade por parte da organização detentora.

Tabela 1 – Possíveis estratégias de preservação para cada nível de abstracção.

Não obstante, para além da definição de uma política de preservação e do estabelecimento de estratégias de preservação adequadas, é fundamental adoptar um sistema de arquivo digital, i.e. um repositório capaz de albergar os objectos, bem como facilitar a implementação dessas políticas e respectivas estratégias de preservação.

Até à data, nenhum dos principais repositórios digitais (e.g. DSpace, Fedora, Eprints) facilitam a implementação de políticas de preservação de forma transversal, nem tão pouco suportam esquemas de metainformação de preservação, tão essenciais para garantir a autenticidade dos materiais. No entanto, oferecem já a capacidade de armazenar, organizar, descrever e disseminar esses materiais. Será portanto espectável que a curto prazo estas plataformas passem a incorporar funcionalidades de preservação digital que permitam preservar plenamente e a longo-prazo os materiais digitais armazenados.

# Glossário

---

**Arquitectura Orientada ao Serviço.** Arquitectura de software onde vários agentes disponibilizam recursos aos restantes participantes da rede sob a forma de serviços independentes, invocáveis de forma normalizada (ver Serviço *Web*).

**Arquivo.** Organização responsável por garantir o acesso continuado à informação custodiada.

**ASCII.** American Standard Code for Information Interchange. Conjunto de códigos capaz de representar letras, dígitos e outros símbolos, amplamente utilizado por computadores para troca de informação textual.

**Autenticação.** Processo responsável por assegurar que um utilizador, serviço ou recurso é exactamente aquilo que se propõe ser.

**CD-ROM.** Compact Disc Read-Only Memory. Suporte físico de armazenamento baseado em tecnologia óptica.

**Comunidade de interesse.** Conjunto identificável de consumidores de informação de um repositório.

**Conversão.** Ver Migração.

**Digitalização.** Processo responsável pela transformação de informação analógica para formato digital.

**Disco rígido.** Suporte de armazenamento de informação digital baseado em tecnologia magnética.

**DVD.** Digital Versatile Disk. Suporte físico de armazenamento baseado em tecnologia óptica. Fisicamente semelhante a um CD-ROM mas com uma capacidade de armazenamento várias vezes superior.

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

**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.

**Estratégia de preservação digital.** Abordagem técnica que garante o acesso continuado à informação existente em formatos digitais (ver Migração, Emulador ou Encapsulamento).

**Flash-drive.** Dispositivo que combina uma memória flash com uma interface USB, vulgarmente utilizado para armazenar informação digital.

**GIF.** Graphics Interchange Format. Formato matricial para representação de imagens digitais.

**Ingestão.** Processo ou componente responsável pela recepção de material de arquivo.

**Internet.** Rede global de comunicação baseada em protocolos TCP/IP.

**Java.** Linguagem de programação orientada a objectos desenvolvida na década de 90. Contrariamente às linguagens de programação convencionais, que são compiladas para código nativo, a linguagem Java é compilada para *bytecode* que é executado por uma máquina virtual.

**JPEG.** Joint Photographic Experts Group. Formato matricial para representação de imagens digitais.

**Material digital.** Conjunto de informação ou objectos digitais.

**Metadados.** Ver Metainformação.

**Metainformação.** Informação utilizada para descrever um determinado objecto ou recurso.

**Migração.** Transferência periódica de material digital de uma configuração de hardware/software para outra, ou de uma geração de tecnologia para outra subsequente.

**Objecto digital.** Todo e qualquer objecto de informação que possa ser representado através de uma sequência de dígitos binários (*bitstream*). Documentos de texto, fotografias digitais, diagramas vectoriais, bases de dados, sequências de vídeo e áudio, modelos de realidade virtual, páginas *Web* e jogos ou aplicações de software são apenas alguns exemplos do que pode ser considerado um objecto digital.

**Objecto nado-digital.** Objecto criado recorrendo apenas a ferramentas ou processos digitais.

**PDF.** Portable Document Format. Formato digital vulgarmente utilizado para representar documentos de texto estruturado, i.e., documentos com tabelas, formatação, imagens, etc.

**PNG.** Portable Network Graphics. Formato matricial para representar imagens digitais.

**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.

**Propriedade significativa.** Característica técnica ou atributo que caracteriza um objecto digital, considerada relevante para efeitos de preservação.

**Refreshamento.** Processo que consiste na cópia de informação de um suporte físico de armazenamento para outro do mesmo tipo.

**Repositório digital.** Sistema de informação responsável por gerir e armazenar material digital.

**Service Oriented Architecture (SOA).** Ver Arquitectura Orientada ao Serviço.

**Serviço *Web*.** Forma de partilhar informação onde são utilizados protocolos de ligação e formatos de mensagens normalizados baseados em XML (SOAP). De modo a facilitar a descoberta de serviços, estes são geralmente publicados em directórios, vulgarmente designados por UDDI (Universal Description, Discovery and Integration).

**TARGA.** Truevision TGA. Formato matricial utilizado para representar imagens digitais.

**TIFF.** Tagged Image File Format. Formato matricial vulgarmente utilizado para representar imagens digitais.

**Web service.** Ver Serviço *Web*.





# Referências

---

- [1] A. Proença and S. Lopes, "Digital Preservation," Departamento de Informática da Universidade da Beira Interior, Covilhã, Monography, 2004.
- [2] K.-H. Lee, O. Slattery, R. Lu, X. Tang and V. McCrary, "The State of the Art and Practice in Digital Preservation," *Journal of Research of the National Institute of Standards and Technology*, vol. 107, no. 1, pp. 93-106, 2002.
- [3] M. Ferreira, A. A. Baptista and J. C. Ramalho, "Avaliação Automática de Migração em Redes Distribuídas de Conversores," presented at Conferência da Associação Portuguesa de Sistemas de Informação (CAPSI), Bragança, Portugal, 2005.
- [4] Wikipedia contributors, "Betamax," in Wikipedia, The Free Encyclopedia, 2006. [Online]. Available: <http://en.wikipedia.org/w/index.php?title=Betamax&oldid=60719745>.
- [5] A. R. Kenney, N. Y. McGovern, R. Entlich, W. R. Kehoe and E. Olsen, "Digital Preservation Management," 2003. [Online]. Available: <http://www.library.cornell.edu/iris/tutorial/dpm/>.

- [6] C. Webb, "Guidelines for the Preservation of Digital Heritage," United Nations Educational Scientific and Cultural Organization - Information Society Division, 2003.
- [7] K. Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," presented at The State of Digital Preservation: An International Perspective, Washington D.C., 2002.
- [8] E. Oltmans, R. J. v. Diessen and H. v. Wijngaarden, "Preservation Functionality in a Digital Archive," presented at Joint ACM/IEEE Conference on Digital Libraries (JCDL'04), 2004.
- [9] T. v. d. Werf, "Our digital heritage: how authentic should it be?," presented at Victorian Association for Library Automation Inc., Melbourne, 2002.
- [10] H. Hofman, "A global issue: preservation of digital objects," presented at Korean Association of Archives Management, Seoul, Korea, 2002.
- [11] B. F. Lavoie, "The Open Archival Information System Reference Model: Introductory Guide," Digital Preservation Coalition, Dublin, USA, Technology Watch Report Watch Series Report 04-01, 2004.
- [12] Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS) - Blue Book*. Washington: National Aeronautics and Space Administration, 2002.
- [13] M. d. L. Saramago, "Metadados para preservação digital e aplicação do modelo OAIS," presented at VIII Congresso da BAD, Estoril, Portugal, 2004.
- [14] D. Bearman, "Collecting Software: A New challenge for Archives & Museums," *Archival Informatics* 1, 1987.
- [15] Task Force on Archiving of Digital Information, Commission on Preservation and Access and Research Libraries Group, *Preserving digital information: report of the Task Force on Archiving of Digital*

- Information*. Washington, D.C.: Commission on Preservation and Access, 1996.
- [16] D. Swade, "Preserving Software in an Object-Centred Culture," in *In History and Electronic Artefacts*, E. Higgs, Ed. Oxford: Clarendon Press, 1998, pp. 195-206.
- [17] T. Hendley, "Comparison of Methods & Costs of Digital Preservation," British Library Research and Innovation Center, West Yorkshire 106, 1998.
- [18] J. Rothenberg, Commission on Preservation and Access and Council on Library and Information Resources, *Avoiding technological quicksand: finding a viable technical foundation for digital preservation: a report to the Council on Library and Information Resources*. Washington, DC: Council on Library and Information Resources, 1999.
- [19] D. Bearman, "Archival Methods," Archives and Museum Informatics, Pittsburgh, Technical Report 1, 1989.
- [20] D. Woodyard, "Farewell my Floppy: a strategy for migration of digital information," 1998. [Online]. Available: <http://www.nla.gov.au/nla/staffpaper/valadw.html>.
- [21] H. Besser, "Digital Preservation of Moving Image Material?," *The Journal of the Association of Moving Image Archivists*, vol. 1, no. 2, pp. 39-55, 2001.
- [22] Wine development team, "Wine HQ Web site," 1997. [Online]. Available: <http://www.winehq.com>. [Accessed 2006-10-11]
- [23] VMWare, "VMWare Workstation Web site," 1998. [Online]. Available: <http://www.vmware.com/>.
- [24] Parallels, "Parallels Desktop Web site," 1995. [Online]. Available: <http://www.parallels.com>. [Accessed 2006-10-12]
- [25] S. Granger, "Emulation as a Digital Preservation Strategy," *D-Lib Magazine*, vol. 6, no. 10, 2000.

- [26] A. Davidson and A. Pollard, "Jasper - ZX Spectrum Emulator," 2005. [Online]. Available: <http://www.spectrum.lovely.net/>.
- [27] G. Krijgsman, "Emulator Zone," 2005. [Online]. Available: <http://www.emulator-zone.com>.
- [28] A. Waugh, R. Wilkinson, B. Hills and J. Dell'oro, "Preserving Digital Information Forever," presented at Fifth ACM Conference on Digital Libraries, San Antonio, Texas, 2000.
- [29] H. Heslop, S. Davis and A. Wilson, "An Approach to the Preservation of Digital Records." Canberra, Australia: National Archives of Australia, 2002.
- [30] D. Holdsworth and P. Wheatley, "Emulation, Preservation and Abstraction," *DigiNews, Research Library Group*, vol. 5, no. 4, 2001.
- [31] Microsoft Corporation, "MS-DOS," 1.0 ed, 1981.
- [32] D. Woodyard, "Digital Preservation: The Australian Experience," presented at Third Conference Digital Library: Positioning the Fountain of Knowledge, Malaysia, 2000.
- [33] G. W. Lawrence, W. R. Kehoe, O. Y. Rieger, W. H. Walters and A. R. Kenney, "Risk Management of Digital Information: A file format investigation," Council on Library and Information Resources, Washington, DC, 2000.
- [34] K. Russell, "Digital Preservation and the CEDARS Project Experience," presented at International Conference on Preservation and Long Term Accessibility of Digital Materials, York, England, 2000.
- [35] M. Hedstrom, "Digital Preservation: Problems and Prospects," *Digital Library Network (DLnet)*, no. 20, 2001.
- [36] A. Rauber and A. Aschenbrenner, "Part of Our Culture is Born Digital - On Efforts to Preserve it for Future Generations," *TRANS - On-line Journal for Cultural Studies.*, vol. 10, 2001.

- [37] H. Hofman, "How to keep digital records understandable and usable through time?," presented at Long-Term Preservation of Electronic Records, Paris, France, 2001.
- [38] C. Ayre and A. Muir, "The Right to Preserve," *D-Lib Magazine*, vol. 10, no. 3, 2004.
- [39] A. G. Howel, "Preserving Digital Information: Challenges and Solutions," Victorian Academic Libraries, Victorian university libraries and State Library of Victoria, 2004.
- [40] M. Hedstrom, "Digital Preservation: A time bomb for digital libraries," *Computers and the Humanities*, vol. 31, pp. 189-202, 1998.
- [41] G. Hodge and E. Frangakis, "Digital Preservation and Permanent Access to Scientific Information: The State of the Practice," International Council for Scientific and Technical Information & CENDI, Report 2004-3: Rev. 05/04, 2004.
- [42] P. Mellor, P. Wheatley and D. M. Sergeant, "Migration on Request, a Practical Technique for Preservation," presented at ECDL '02: 6th European Conference on Research and Advanced Technology for Digital Libraries, London, UK, 2002.
- [43] J. M. Ockerbloom, "Mediating Among Diverse Data Formats," in *School of Computer Science*. Pittsburg: Carnegie Mellon University, 1998, pp. 164.
- [44] F. L. Walker and G. R. Thoma, "A Web-Based Paradigm for File Migration," presented at IS&T's 2004 Archiving Conference, San Antonio, Texas, USA, 2004.
- [45] J. Hunter and S. Choudhury, "A Semi-Automated Digital Preservation System based on Semantic Web Services," presented at Joint ACM/IEEE Conference on Digital Libraries (JCDL'04), 2004.
- [46] M. Ferreira, "Automatic Evaluation of Migration Quality in Distributed Networks of Converters," presented at Doctoral Consortium of the 9th European Conference on Research and

Advanced Technology for Digital Libraries (ECDL), Vienna, Austria, 2005.

- [47] M. Ferreira, A. A. Baptista and J. C. Ramalho, "CRiB: A service oriented architecture for digital preservation outsourcing," presented at XATA - XML: Aplicações e Tecnologias Associadas, Portalegre, Portugal, 2006.
- [48] M. Ferreira, "Automatic Evaluation of Migration Quality in Distributed Networks of Converters," *Bulletin of the IEEE Technical Committee on Digital Libraries (TCDL)*, vol. 2, no. 1, 2006.
- [49] M. Ferreira, "CRiB - Conversion and Recommendation of Digital Object Formats Web site," 2006. [Online]. Available: <http://crib.dsi.uminho.pt>.
- [50] M. Ferreira, A. A. Baptista and J. C. Ramalho, "A Foundation for Automatic Digital Preservation," *Ariadne*, no. 48, 2006.
- [51] A. R. Heminger and S. B. Robertson, "A Delphi Assessment of the Digital Rosetta Stone Model," presented at 37th Annual Hawaii International Conference on System Sciences (HICSS'04), Big Island, Hawaii, 2004.
- [52] Digital Preservation Testbed, "Migration: Context and Current Status," The Hague, White Paper, 2001.
- [53] T. Shepard and D. MacCarn, "The Universal Preservation Format: Background and Fundamentals," presented at Sixth DELOS Workshop, Tomar, Portugal, 1998.
- [54] T. Shepard and D. MacCarn, "The Universal Preservation Format: A Recommended Practice for Archiving Media and Electronic Records," Boston, 1999.
- [55] R. A. Lorie, "A Methodology and System for Preserving Digital Data," presented at Second ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'02), Portland, Oregon, 2002.

- [56] R. A. Lorie, "Long Term Preservation of Digital Information," presented at First ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'01), Roanoke, Virginia, USA, 2001.
- [57] Wikipedia, "Rosetta Stone," 2005. [Online]. Available: [http://en.wikipedia.org/wiki/Rosetta\\_stone](http://en.wikipedia.org/wiki/Rosetta_stone).
- [58] Wikipedia, "Jean-François Champollion," 2005. [Online]. Available: [http://en.wikipedia.org/wiki/Jean-Fran%EAois\\_Champollion](http://en.wikipedia.org/wiki/Jean-Fran%EAois_Champollion).
- [59] A. R. Heminger and S. B. Robertson, "Digital Rosetta Stone: A Conceptual Model for Maintaining Long-Term Access to Digital Documents," presented at 6th DELOS Workshop, Tomar, Portugal, 1998.
- [60] UK National Archives, "PRONOM - The file format registry," 2002. [Online]. Available: <http://www.nationalarchives.gov.uk/pronom/>.
- [61] S. L. Abrams and D. Seaman, "Towards a global digital format registry," presented at World Library and Information Congress: 69th IFLA General Conference and Council, 2003.
- [62] J. Darlington, "PRONOM - A Practical Online Compendium of File Formats," *RLG DigiNews*, vol. 7, no. 5, 2003.
- [63] UK National Archives, "Droid: Digital Record Object Identification," 1.0 ed. Surrey: UK National Archives, 2005.
- [64] C. T. Cullen, "Authentication of Digital Objects: Lessons from a Historian's Research," in *Authenticity in a Digital Environment*. Washington, DC: Council on Library and Information Resources, 2000.
- [65] P. B. Hirtle, "Archival Authenticity in a Digital Age," in *Authenticity in a Digital Environment*. Washington, DC: Council on Library and Information Resources, 2000.



- [66] H. Hofman, "Can Bits and Bytes be Authentic? Preserving the Authenticity of Digital Objects," presented at International Federation of Library Associations Conference, Glasgow, 2002.
- [67] J. Rothenberg, "Preserving Authentic Digital Information," in *Authenticity in a Digital Environment*. Washington, DC: Council on Library and Information Resources, 2000.
- [68] B. Lavoie and R. Gartner, "Technology Watch Report - Preservation Metadata," Online Computer Library Center Inc., Oxford University Library Services and Digital Preservation Coalition 05-01, 2005.
- [69] H. MacNeil, C. Wei, L. Duranti, A. Gilliland-Swetland, M. Guercio, Y. Hackett, B. Hamidzadeh, L. Iacovino, B. Lee, S. McKemmish, J. Roeder, S. Ross, W.-k. Wan and Z. Z. Xiu, "Authenticity Task Force Report," InterPARES Project, Vancouver, Canada, 2001.
- [70] L. Millar, "Authenticity of electronic records: a report prepared for UNESCO and the International Council on Archives," International Council on Archives, London, UK, 2004.
- [71] C. Lynch, "Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust," in *Authenticity in a Digital Environment*. Washington, DC: Council on Library and Information Resources, 2000.
- [72] Authenticity Task Force, "Requirements for Assessing and Maintaining the Authenticity of Electronic Records," InterPARES Project, Vancouver, Canada, 2002.
- [73] R. J. v. Diessen and T. v. d. Werf-Davelaar, "Authenticity in a digital environment," Koninklijke Bibliotheek & IBM, Amsterdam, The Netherlands, Report 2, 2002.
- [74] P. Akester, "Internet law - authenticity of works. Authorship and authenticity in cyberspace," *Computer Law & Security Report*, vol. 20, no. 6, 2004.

- [75] P. Graham, "Issues in Digital Archiving," in *Preservation: Issues and Planning*, R. Pilette and P. Banks, Eds. Chicago: IL: American Library Association, 2000.
- [76] The Cedars Project Team, "The Cedars Project Report," Consortium of University Research Libraries, UK, 2001.
- [77] R. J. v. Diessen, "Model Driven Object-Oriented Development of Systems: A Behavioural-Oriented Approach." Hilversum, The Netherlands, 1997.
- [78] N. Beagrie, M. Bellinger, R. Dale, M. Doerr, M. Hedstrom, M. Jones, A. Kenney, C. Lupovici, K. Russell, C. Webb and D. Woodyard, "Trusted Digital Repositories: Attributes and Responsibilities," Research Libraries Group & Online Computer Library Center, Report, 2002.
- [79] A. Rusbridge, "Migration on Request," University of Edinburgh - Division of Informatics, 4th Year Project Report, 2003.
- [80] C. Lynch, "Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information," *D-Lib Magazine*, vol. 5, no. 9, 1999.
- [81] C. Lupovici and J. Masanès, "Metadata for the Long Term Preservation of Electronic Publications," NEDLIB Consortium, The Hague, The Netherlands 2, 2000.
- [82] The Cedars Project Team, "Cedars Guide to Preservation Metadata," The Cedars Project, 2002.
- [83] National Library of Australia, "Preservation Metadata for Digital Collections," 1999. [Online]. Available: <http://www.nla.gov.au/preserve/pmeta.html>.
- [84] OCLC/RLG Preservation Metadata Working Group, "A Metadata Framework to Support the Preservation of Digital Objects," OCLC Online Computer Library Center, Inc., Dublin, USA, 2002.

- [85] PREMIS Working Group, "Data dictionary for preservation metadata: final report of the PREMIS Working Group," OCLC Online Computer Library Center & Research Libraries Group, Dublin, Ohio, USA, Final report, 2005.
- [86] C. Rauch and A. Rauber, "Preserving Digital Media: Towards a Preservation Solution Evaluation Metric," presented at International Conference on Asian Digital Libraries, Shanghai, China, 2004.
- [87] C. Rauch, F. Pavuza, S. Strodl and A. Rauber, "Evaluating preservation strategies for audio and video files," presented at DELOS Digital Repositories Workshop, Heraklion, Crete, 2005.
- [88] P. Weirich, B. Skyrms, E. W. Adams, K. Binmore, J. Butterfield, P. Diaconis and W. L. Harper, *Decision Space: Multidimensional Utility Analysis*. Cambridge, 2001.
- [89] B. F. Lavoie and L. Dempsey, "Thirteen Ways of Looking at... Digital Preservation," *D-Lib Magazine*, vol. 10, no. 7/8, 2004.
- [90] R. Burkel, "The Role of Microfilm in Information Management," *Information Management Journal*, vol. 37, no. 1, pp. 58-65, 2003.
- [91] D. Waters, "Good Archives Make Good Scholars: Reflections on Recent Steps Toward the Archiving of Digital Information," presented at The State of Digital Preservation: An International Perspective, Washington D.C., 2002.
- [92] J. C. Bennett, "A Framework of Data Types and Formats, And Issues Affecting the Long Term Preservation of Digital Material," British Library Research and Innovation Centre, West Yorkshire, UK, Report 50, 1997.

# Índice

---

- Acesso, 26
- actualização de versões, 35
- Agente, 54
- Análise de Utilidade, 55
- aplicações, 17, 19, 31, 32, 34, 43, 67
- arqueologia digital, 42
- árvore-objectivo, 56, 57
- ASCII, 49
- áudio, 19, 67
- autenticidade, xii, 46, 47, 51, 54, 60, 61
- bases de dados, 19, 67
- canonização, 50
- características essenciais, 50
- CCSDS, 24
- CD, 20, 30, 65, 66
- comunidade de interesse, 26
- diagramas vectoriais, 19, 67
- Direitos, 54
- disco rígido, 20, 23, 30
- disquete, 20, 30
- Documentos de texto, 19, 67
- DVD, 17, 20, 23, 66
- emulação, 31
- emulador, 30, 31, 32
- encapsulamento, 28, 40
- Entidade Intelectual, 53
- essência de um objecto digital, 49
- Evento, 54
- Ficheiro, 54
- formato, 20
- formato canónico, 50
- formato de preservação, 36
- Formato Universal de Preservação, 40
- fotografias digitais, 19, 67
- Global Digital Format Registry, 44
- hardware, 20, 29, 30, 31, 32, 33, 36, 41, 48, 52, 66, 67
- incorporação, 25, 51

informação, xi, xii, 16, 17, 18, 19,  
     24, 25, 26, 27, 30, 32,  
     33, 36, 39, 40, 41, 43,  
     44, 48, 51, 52, 54, 59,  
     61, 62, 65, 66, 67, 68  
 Informação Descritiva, 26  
 Ingestão, 25, 67  
 ISO, 24  
 Java Virtual Machine, 41  
 JPEG, 23, 35, 36, 67  
 máquina virtual universal, 41  
 metainformação, 40  
 metainformação de preservação,  
     51  
 migração, 28, 33, 34, 35, 37, 38,  
     39, 40, 50, 60, 61, 62  
 migração a-pedido, 34, 37, 39, 62  
 migração para suportes  
     analógicos, 34  
 MIME, 44  
 MS-DOS, 32  
 normalização, 34, 35, 36, 39, 62  
 OAIS, xi, 24, 25, 26, 52, 59  
 Objecto, 54  
 objecto conceptual, 21, 22, 23,  
     28, 29, 33  
 objecto digital, xi, 19, 20, 21, 22,  
     23, 29, 30, 31, 33, 34,  
     37, 40, 41, 47, 48, 49,  
     54, 59, 66, 67, 68  
 objecto experimentado, 21  
 objecto físico, 20, 28, 48, 61  
 objecto lógico, 20  
 objecto semântico, 21  
 objectos conceptuais, 21  
 objectos digitais, 20, 27, 28, 29,  
     32, 33, 34, 35, 36, 37,  
     38, 40, 41, 42, 49, 50,  
     56, 61, 62, 63, 67  
 OCLC/RLG, 52  
 Pacote de Informação de  
     Arquivo, 26  
 Pacotes de Informação de  
     Disseminação, 27  
 PDF, 38, 44, 49, 68  
 Pedra de Rosetta, 41  
 Planeamento de Preservação, 26  
 PNG, 23, 35, 36, 68  
 políticas, xii, 26, 36, 49, 60, 63  
 população potencialmente  
     utilizadora, 26  
 PREMIS, xii, 52, 53, 60  
 preservação digital, iv, xi, 18, 22,  
     24, 28, 29, 30, 32, 43,  
     52, 66  
 propriedades significativas, 48, 49  
 proveniência, 51  
 realidade virtual, 19, 67  
 refrescamento, 30  
 repositório, 25, 26, 27, 35, 39, 49,  
     51  
 Repositório de Dados, 26  
 Representação, 54  
 Sequência de bits, 54  
 Serviços, 39  
 software, 17, 19, 20, 29, 30, 31,  
     32, 33, 34, 35, 36, 38,  
     41, 44, 48, 52, 54, 63,  
     65, 66, 67  
 submissão, 25  
 suporte físico, 20, 22, 30, 40, 68  
 Thibodeau, 28, 62  
 TIFF, 23, 35, 69  
 TOM, 38, 45  
 Typed Objects Model, 38  
 Universidade do Minho, iv, ix, 38  
 vídeo, 16, 19, 61, 67  
*Web*, 19, 38, 39, 44, 65, 67, 68, 69





