

Refrescando a memória – arquivo e gestão da informação

Alberto Sá¹

Introdução

Nos últimos anos, a comunidade científica tem-se debruçado sobre a problemática do arquivo e gestão da informação, consciente da fragilidade dos meios de suporte informativo tradicionais e da necessidade de adoptar estratégias no presente para salvaguarda da informação no futuro. O cepticismo instalou-se. Questionaram-se a vulnerabilidade e a longevidade dos suportes do registo informativo, diagnosticou-se a dependência dos processos digitais de arquivo de dados face aos programas (*software*) geradores, percepcionou-se a obsolescência do equipamento tecnológico (*hardware*) responsável pela leitura da cadeia de bits e pela sobrevivência do código de linguagem inerente (Hedstrom, s.d.; Rothenberg, 1999b).

Desde sempre foi preocupação do Homem a ideia de arquivar, guardar e preservar, mantendo, ainda que inconscientemente nos primórdios da Humanidade, uma perspectiva de transmissão enquanto herança. Guardar agora para mostrar amanhã, perpetuando significados, vivendo-o em função de um futuro incerto. Foi desde que o Homem tomou consciência de si próprio e do que o rodeava que procurou registar na pedra as formas do seu imaginário por meio das pinturas ou gravuras. Registava informações, construindo memórias que ainda hoje servem de narrativas de um espaço e de um tempo concretos, como é o caso da arte paleolítica. Significava eternizar algo que era importante, construindo mitos acerca da natureza que o envolvia.

Uma dificuldade sempre sentida foi a da natureza da informação a transmitir. A de carácter físico, aquela que existe por si, facilita o acto de preservar e arquivar, dado o seu perfil material - um edifício, uma estátua, uma construção, os documentos escritos, os mapas, as pinturas. Neste sentido, surgiram os arquivos como guardiões

do passado e as bibliotecas como depositários de um saber acumulado que o Homem foi produzindo na forma escrita. Contudo, o desenvolvimento tecnológico ocorrido sobretudo em meados do século passado trouxe uma nova concepção de informação e de conhecimento.

Os processos de transformação da informação analógica em formato digital reduziram a panóplia de dados sensoriais à expressão encadeada de “zeros” e de “uns”. Deste modo se constitui um outro tipo de informação, de carácter não-físico, virtual, entendível pela máquina.

Se, no passado, a consulta da informação pressupunha um suporte físico em papel - e a própria informação era constantemente “reiventada” por novas e sucessivas interpretações de forma a produzir conhecimento, no presente, novas questões se levantam: como registar a informação expressa por código binário; como guardar e preservar documentos do tipo e-mail, páginas *web*, bases de dados, ou mesmo telefonemas e vídeos em formato digital. Ou seja, a questão de fundo é a de se saber como guardar os *bits* que constituem a informação.

Os tradicionais depositários do saber da Humanidade, as bibliotecas e os arquivos, ganharam a companhia de outra entidade armazenadora de informação: o disco-duro, sem dúvida o suporte físico (*hardware*) que viabiliza o arquivo, gestão, organização e posterior processamento de todos os dados armazenados. Esta tendência foi reveladora de uma nova atitude, a do recurso às novas tecnologias da informação e da comunicação, aproximando-se da noção da *Biblioteca de Babel*, de Jorge Luis Borges, ilimitada no seu acervo, contendo todos os livros possíveis.

Sem qualquer paralelo em alguma época histórica, a sociedade actual lida diariamente com uma constante produção de informação. Num passado não muito remoto, o suporte

tradicional de informação – o livro - trazia inerente a sua conservação. O acto de publicação a isso “obrigava”. Este procedimento consistia na realização de um conjunto de etapas de validação da informação: os momentos do revisor, do corrector e do editor. Os custos deste percurso exigiam o seu posterior arquivo e conservação, expressos na existência de um Depósito Legal que garantisse a sua sobrevivência.

Actualmente, os processos tecnológicos permitem saltar directamente da etapa da redacção para a publicação, sem passar pelos processos intermediários de validação da informação. Não existe um arquivo central oficial, um organismo que funcione como *Depósito Legal*, mas antes vários e distintos processos de arquivo, sem, contudo, haver a orientar qualquer actividade reguladora – precisamente, um dos motivos para o sucesso da *web*. Porque o suporte de registo é volátil, as formas de preservação da informação passam pelo armazenamento em vários espaços. Um à escala global: a *Internet*, exposta aos “radares” dos motores de pesquisa. Outro à escala local: o disco-duro do sistema, operando por processos de disseminação realizados em suporte magnético ou na forma óptica de cd-rom e dvd.

Na verdade, a *Internet* permite ao cidadão anónimo tornar-se facilmente um editor, que produz os seus próprios escritos e os publica num qualquer *site*, gratuitamente disponível mas residente em parte geograficamente incerta e irrelevante. Os baixos custos associados à publicação *online* permitiram a denominada “democratização da informação”.

«Existo, logo, armazeno»

Se o “presente” já se vai registando na forma de código binário, a tendência para a digitalização de tudo aquilo que constitui o passado vem aumentando, fazendo passar pelo crivo óptico todo o tipo de documentos existentes em formato não-digital: livros e obras de referência, periódicos, jornais, manuscritos, cartografia, entre outros, formando autênticas bibliotecas virtuais.

Algumas iniciativas têm correspondido a grandes projectos nacionais², fazendo apelo à necessidade de proteger o património,

enquanto herança cultural, para as gerações do presente e do futuro.

Outras propostas têm em vista fins de natureza militar, “fazendo dos computadores uma ferramenta muito mais eficaz em situações de guerra”³, pela alimentação de gigantescas bases de dados com informações individuais, esperando-se, com isso, elaborar padrões de comportamento e percursos de vida dos criminosos.

Também o meio académico e universitário têm apresentado propostas de colecções digitais para armazenamento, preservação e divulgação da propriedade intelectual da comunidade científica, com a tónica assente num modelo menos centralizado e mais distribuído da comunicação: os repositórios institucionais⁴. O conjunto das ferramentas informáticas utilizadas permite ultrapassar o problema complexo da integração dos diferentes processos de depósito, tão necessários à eficiência de um sistema multidisciplinar, recorrendo ao uso de tecnologias *open source*, que facilitam a exposição da informação na *Internet*, conjuntamente com os seus metadados⁵. Os repositórios estão definidos por políticas de acesso restritas de modo a controlar cada acção do processo, prevendo a existência dos *submitters*, dos *reviewers*, dos *metadata editors*, entre outros. De modo a ser possível encontrar um recurso depositado, é crucial que as citações permaneçam válidas por longos períodos de tempo, recorrendo-se à criação de identificadores persistentes para cada item – o *Uniform Resource Identifier* (URI)⁶.

A ideia de tudo arquivar pode também pertencer ao domínio do individual⁷, alcançando-se verdadeiros projectos de vida digital, perpetuando a memória de toda uma existência. A crescente digitalização do quotidiano moderno tem a vantagem de permitir guardar a informação sobre o nosso dia-a-dia de uma maneira fácil e barata: mensagens de correio electrónico, fotos e vídeos pessoais, documentos textuais, cálculo da economia doméstica, mas também todo o registo das nossas acções efectuadas no contexto económico-social. Por exemplo, os movimentos bancários, os registos de comunicação móvel, o registo do dentista, as escolhas no clube de vídeo, ente tantos outros exemplos.

Em busca do paradigma linguístico

A cultura do computador criou novos sistemas cognitivos diferentes daqueles a que estávamos habituados. Numa sociedade em constante mutação, onde todo o conhecimento é fluido e dinâmico, a informação é assimilada instantaneamente e o pensamento que a acompanha espartilhado e fragmentado: “a aceleração tecnológica e social súbita sem preparação pode na verdade levar à desintegração (...). Os nossos computadores estão a acelerar as nossas respostas psicológicas e os nossos tempos de reacção muito mais do que fizeram os aviões, os comboios e os automóveis” (Kerckhove, 1997b, págs. 118-119).

Um ponto sensível é o da questão da compatibilidade das linguagens. Tal motiva o estado de dúvida acerca da capacidade de interpretação das gerações futuras sobre os processos de informação desenvolvidos na actualidade.

A velocidade frenética das sociedades actuais levou à renovação das interfaces e das linguagens da programação, o que implica criar processos não só de descodificação, mas também da manutenção dos códigos linguísticos.

A compatibilidade da informação exige a preservação do código que a formula. A comunidade científica deu conta de que as vantagens que promoveram a tendência generalizada para a digitalização camuflavam problemas às gerações vindouras (Rothenberg, 1999b). Ninguém duvida dos benefícios da digitalização documental por razões de preservação, facilidade de armazenamento, processo de cópia, redução de custos e reutilização em novas e mais avançadas ferramentas informáticas: “o novo torna possível uma utilização mais específica daquilo que já existe” (Luhmann, 1992, p.153).

No entanto, já menos perceptíveis - e aqui reside o problema -, são as relações de dependência estabelecidas durante o processo de digitalização: a codificação só é interpretável pelo programa gerador, que, por sua vez, depende do sistema operativo que o acolheu e que, em conjunto, todos dependem do equipamento que os executou (Rothenberg, 2001).

É possível agrupar os problemas em dois conjuntos: um primeiro diz respeito à preservação e revitalização dos antigos documentos (*Arqueologia Digital*); um outro, de teor profilático, diz respeito à adopção, na actualidade, de um conjunto de procedimentos para que, no futuro, não se tenha ainda que resolver aquilo que hoje se tenta solucionar (McCray, 2001).

Da panóplia de soluções encontradas, nenhuma parece satisfazer plenamente todas as condicionantes em jogo, visto a comunidade científica ter desenvolvido várias estratégias de preservação digital, com resultados satisfatórios em situações concretas, mas sem atingir um nível de eficácia tal que seja capaz de responder em absoluto ao problema (Hedstrom, s.d.).

Estes procedimentos podem passar pela adopção de estratégias singulares ou, quando muito, mistas, que incluem a impressão dos documentos digitais em suporte físico, a uniformização dos procedimentos de codificação, a criação de um espaço museológico de computadores com vista à recriação de ambientes informáticos extintos, a conversão/migração dos formatos digitais como forma de sustentabilidade e, finalmente, a emulação dos dados conjuntamente com o programa que originalmente os criou (Rothenberg, 1999a; Lorie, 2001).

Tarefa nada fácil por depender de várias variáveis, desde aquelas de carácter técnico até às organizacionais, passando pelos comportamentos sociais. Em conjunto, pretendem evitar o designado pela “improbabilidade da comunicação” (Luhmann, 1992, p. 42), assegurando processos de compreensão em função de um contexto próprio, de forma a aceder aos receptores, obtendo resultados comunicativos desejados.

A questão da compatibilidade, da migração e da emulação tem, por isso, enorme relevância no contexto da preservação da memória digital, assumindo particular destaque a sobrevivência do código linguístico enquanto factor de evolução.

Um exemplo recorrentemente citado é o da *Pedra da Roseta* e o seu contributo para a interpretação da escrita ideográfica egípcia. Neste caso, a preservação do conhecimento foi possível pela recolha e recuperação do código linguístico, através da compa-

ração entre a escrita hieroglífica antiga, o grego e a escrita egípcia coeva.

Outro exemplo, que nos é mais próximo, aponta precisamente para as dificuldades que as gerações sucessivas encontram na transmissão dos códigos de linguagem, e tal diz respeito à reforma arquivística que D. Manuel empreendeu (e que foi continuado depois com D. João III), a qual ficou conhecida pelo nome de *Leitura Nova*⁸: “sabendo o passado hordenariam melhor o presente”.

Tratou-se de uma medida reformadora e organizacional da Torre do Tombo (à altura, Real Arquivo, no castelo de S. Jorge, em Lisboa). Na prática, resultou na cópia, no século XVI, de documentos dispersos, quer de chancelarias régias, quer das “Gavetas”, de milhares de cartas que, à altura, pareceram de interesse histórico perpetuar com o intuito de facilitar a sua leitura. Muitos desses códices encontravam-se já quase ilegíveis e incapazes de decifrar, para uma média de funcionários. Em suma, eram “escritas” com séculos de distância. A intenção era a de agilizar a “burocracia” já que, quando alguém pedia uma cópia de um documento (“certidão”), tal revelava-se extremamente moroso, pois poucos sabiam já ler escritos a séculos de distância. Assim, procedeu-se à transcrição de muitos documentos, agora classificados e arrumados com novos critérios.

O grande problema colocou-se na deficiência de muitas transcrições, cujos erros foram causados pela ignorância de alguns copistas. Em princípio, tratava-se apenas de “copiar”, pelo que a estrutura da língua deveria manter-se medievá. No entanto, o que se verificou, em grande parte, foi o acrescentar de fenómenos linguísticos introduzidos pelos diversos copistas, com níveis de cultura e erudição muito díspares⁹.

Tal como o referenciado hoje em dia, ainda que com menores implicações, também há quinhentos anos atrás se colocou a problemática da migração dos paradigmas linguísticos.

Escapar ao naufrágio

A tão propalada “avalanche informativa” proporcionada pela *Internet* forçou à consciencialização de que era preciso fazer algo para evitar o soterramento no turbilhão informativo.

À primeira vista, o progresso tecnológico verificado nas unidades de armazenamento não parece aliviar esse afogamento. A indústria do sector, recorrendo à nanotecnologia, tem procurado formas capazes de condensar maior informação em espaços mais compactos com vista ao incremento da capacidade das unidades. No entanto, estudos laboratoriais apontam para uma nova concepção de armazenamento mediante a utilização do átomo como elemento de representação, atingindo-se o estádio *quantum*¹⁰, capaz de guardar o bit um milhão de vezes mais densamente do que aquele actualmente presente num vulgar cd-rom (Bennewitz *et alii*, 2002).

Todo este progresso convida a um certo laxismo no utilizador: a perspectiva infindável de espaço em disco parece tornar irresistível a tendência para tudo armazenar, sem olhar a meios, numa terrível “obsessão pelo disco-duro”. Armazenamos, porque podemos. E se cada vez mais pudermos, mais queremos armazenar. Já não são apenas documentos em texto, mas toda uma variedade produtiva, de características e pesos (em bits) diferentes: mensagens de correio electrónico, fotos digitais, registos de vídeo, comentários em *weblogs* e/ou fóruns, entre outros. A limpeza do disco torna-se uma tarefa cada vez menos recorrente, porque o limiar do espaço insuficiente se encontra distante.

Nesta perspectiva, há o perigo de muitos dos registos caírem no esquecimento, camuflados por entre desmultiplicações de pastas de arquivo. A já referida proliferação informativa faz com que o tempo da assimilação do saber se esgote rapidamente. Dado existir um hiato temporal entre a capacidade humana de assimilação informativa e o caudal de informações que desagua por vários meios no quotidiano social, verifica-se a tendência para arquivar tudo o que se encontra e o que se pensa ser, posteriormente, objecto de interesse: na dúvida, armazenamos!, remetendo a validação crítica do conhecimento para um *a posteriori*.

Estamos, pois, em presença de dois universos muito amplos, o do armazenamento e o da pesquisa. Nesse sentido, a preservação digital não se resume à tarefa de armazenar, a uma musealização passiva e estática. Ao facilitismo proporcionado pelo desenvolvi-

mento tecnológico referido (maior capacidade a menor custo), deve contrapor-se uma lógica inteligente de organização e de gestão dinâmica da informação.

Tal mostra-se sobretudo premente quando fazemos da *Internet* o campo de recolha de informação. Quem, por certo, já experimentou arquivar informação com base em pesquisas na *Internet* não se terá deparado tanto com a exiguidade do espaço em disco (cada vez se compra mais por menor preço¹¹), mas antes com um problema de orientação entre as centenas de ficheiros arquivados. Quem já não terá experimentado uma sensação mista de surpresa e de apreensão, ao encontrar algum ficheiro importante escondido numa pasta recôndita?

A solução imediata é a da elaboração de bases de dados sobre a informação arquivada. Mas, se optarmos por uma abordagem mais profissional - logo, mais intensa mas também mais eficaz e exigente -, esta tarefa parece insolúvel.

Para a descrição dos recursos arquivados tem-se procurado recorrer ao uso de metadados, que consistem, basicamente, na formulação de dados sobre os dados. A ideia é a de desenvolver uma forma eficaz de descrever os recursos electrónicos, algo já incontornável no ambiente caótico da *Internet*, ao qual os sistemas de indexação e de recuperação da informação tradicionais não permitem alcançar níveis satisfatórios (Baptista, e Machado, 2001).

A aplicação de metadados tem sido experimentada no campo dos media, concretamente no domínio da informação noticiosa digital¹², onde se tem procurado estabelecer um conjunto padronizado de metadados de modo a fornecer uma plataforma comum para a análise dos artigos noticiosos em formato digital, produzidos por agentes noticiosos *online* (Yaginuma *et alii*, 2003). Este sistema parte de duas premissas fundamentais em torno do ficheiro-base: por um lado, que ele contenha o texto propriamente dito (a notícia), e, por outro, que ele inclua na sua código-estrutura os metadados - estes, para além de descreverem a notícia, permitirão que sobre eles se aplique tratamento informático adequado através da tecnologia de descrição dos metadados extraídos. Tanto num como em outro caso, são várias as abordagens ex-

perimentais possíveis, pelo que daqui resulta que as próprias soluções tentadas carecem do antídoto que pretendem anular: a padronização, tão essencial para a utilização eficaz dos metadados¹³.

Ora, apesar da progressiva conversão das agências noticiosas *online* a esta tecnologia, o processo ainda se encontra em fase experimental, o que significa que apenas uma pequena parte das notícias estarão preparadas para suportar estes procedimentos. Daí que os projectos de investigação que incidam sobre recolha da informação pela *Internet* enfrentem algumas dificuldades no manuseamento dos dados que dificilmente são ultrapassáveis - o processo de classificação da notícia tem que ser feito no momento, sem excepção.

Partindo deste campo de recolha, e partilhando de muitas das dificuldades descritas, o Projecto *Mediascópio*¹⁴ pretende estudar a comunicação e os media, designadamente aqueles publicados em agências e jornais impressos e electrónicos, nacionais e estrangeiros. Esses textos podem estender-se a géneros diversos: notícias breves e desenvolvidas, entrevistas, reportagens, dossiers, textos opinativos (editoriais, colunas, análises, opiniões, cartas).

A ideia da base de dados, neste caso, surgiu de uma constatação e de uma necessidade: constatação, porque, ao reunir os materiais informativos para elaborar o registo dos eventos do campo da comunicação e dos media, verificava-se a existência de documentos relevantes para a memória sobre o campo; necessidade, porque, para promover as leituras sectoriais e globais, essa documentação mostrava-se essencial.

Ao todo, estão compulsados cerca de 6600 registos, compreendidos entre inícios de 2000 e finais de 2003. A proveniência das fontes advém, em cerca de 85%, de jornais impressos, sendo o restante dividido entre revistas e panfletos, em formato *online*, quando possível, ou através da versão papel, recorrendo-se à digitalização por *scanner*, em complemento.

As actividades no âmbito do projecto são, assim, escalonadas em duas fases: uma primeira, a da produção digital da informação recolhida, e uma segunda, a da catalogação desses registos.

Até ao momento, o projecto tem sido, igualmente, um laboratório de análise e testagem ao desafio da criação de instrumentos eficientes de gestão e pesquisa da informação. A solução a adoptar ainda aguarda pela estabilização das ferramentas de extracção e descrição dos metadados.

Conclusão

O desígnio tecnológico, em muito explicado pela célebre *Lei de Moore*, teve o dom de aproximar povos e culturas, impulsionando a comunicação e a partilha de informação multimédia. A maior capacidade informática daí decorrente potencializou a rapidez e a largura de banda da transmissão de dados. Ao mesmo tempo, vai aumentando, de forma espectacular, a capacidade de armazenamento digital e descobrindo formas mais eficazes na compressão dos dados que contribui para a redução do tamanho ocupado, em bits. Como consequência, a produção e a transmissão da informação ficou facilitada: produção de *websites* institucionais e privados, mensagens de e-mail, *weblogs*, conteúdos digitais multimédia diversos (fotos, vídeo, animações).

Parece, ainda que algo paradoxalmente, que a sociedade tecnologicamente desenvolvida criou um «monstro»: como gerir a avalanche informativa crescente e evitar o soterramento? Simultaneamente, como conferir utilidade ao oceano de dados?

A aproximação que se dá entre o sujeito e o conhecimento informático processa-se por moldes diferentes dos tradicionais, que remetiam para a relação com os livros: “este novo salto na forma de adquirir e transmitir informações (...) certamente trará modificações às demais formas tradicionais – orais e escritas – de se lidar com o saber” (Kenski, 1999, p. 173).

A ideia da *World Wide Web*, de Tim Berners-Lee e dos seus colegas do CERN, era a de integrar todos os conteúdos de qualquer servidor em qualquer parte do mundo com outro computador *online*. O caminho para alcançar esta convergência é o da digitalização de todos os conteúdos, mas cuidando na promoção da interconectividade entre todas as redes e a humanização do *software* e do *hardware*, atendendo aos efeitos à escala globalizante dos satélites. Na opi-

nião de Kerckhove, tal pressupunha uma “nova ecologia das redes”, baseada na interactividade, na hipertextualidade e na conectividade (Kerckhove, 1997a). Mas, a estas, há que assegurar o objectivo da interoperabilidade entre os sistemas, melhorando pela adopção da indexação e da interconexão, condição imprescindível para o carácter funcional dos arquivos: “uma memória está morta se não for catalogada, disponível, transmissível, criticada e eventualmente reinterpretada” (Hoog, 2003, p. 173).

Antes mesmo da questão do que conservar e do que transmitir, é preciso encontrar formas comuns de preservação da informação já existente: “a história da memória deve ser também a história dos seus suportes” (Hoog, 2003, p. 170).

Será, de facto, possível, construir-se uma memória a partir ciberespaço? Alguns entraves estão diagnosticados: a obsolescência dos suportes de registo informativo, do código linguístico e do seu respectivo equipamento; o carácter volátil e imaterial dos conteúdos; a perenidade dos *links* que inter-relacionam a informação na rede. Em conjunto, concorrem para a urgência de, no presente, promover processos de produzir informação digital que já incluam na estrutura-código a forma de interpretação – um *dna* digital. A vantagem é a da capacidade de proporcionar imediato armazenamento, indexação e catalogação, através de programas estandardizados apropriados. As ferramentas informáticas devem corresponder às características da *open-source*, de modo a ser concretizável o objectivo da interoperabilidade com outros sistemas, permitindo, assim, a ampla disponibilização dos ficheiros através da *web*, por meio de um sistema de procura e recuperação da informação.

De um ponto de vista do interesse individual ou colectivo, todo o arquivo é património. A tecnologia assim o vai permitindo, e permitirá cada vez melhor. Não nos cabe, agentes do presente, decidir o que sobre nós deverão saber, no futuro. Da mesma maneira que um arqueólogo exulta quando vê num artefacto um sobrevivente da amnésia do tempo.

Refrescando a memória, porque a era tecnológica se esforça por permitir aos suportes do conhecimento a durabilidade e transmissão infinitas, sem degradação nem perdas.

Bibliografia

Baptista, Ana Alice e **Machado**, Altamiro Barbosa, “Um gato preto num quarto escuro: falando sobre metadados”, *Revista de Biblioteconomia de Brasília*, vol. 25, nº 1, 2001, URI: <http://hdl.handle.net/1822/380>.

Bennewitz, R *et alii*, “Atomic scale memory at a silicon surface”, *Nanotechnology* vol. 13, nº 4, 2002, págs. 499-502.

Hedstrom, Margaret. *Mass storage and long-term preservation*, disponível em <http://www.uky.edu/~kiernan/DL/hedstrom.html> (20.Abril.2004).

Hoog, Emmanuel, “Tout garder? Les dilemmes de la mémoire à l’âge médiatique”, *Le débat*, nº 125 (maio-agosto 2003), Paris, Ed. Gallimard, 2003, págs. 168-189.

Kenski, Vani Moreira, “Memória e conhecimento na era tecnológica”, *Revista de Comunicação e Linguagens*, 25-26, Lisboa, Ed. Cosmos, 1999, págs. 165-175.

Kerckhove, Derrick de, *Connected intelligence. The arrival of the web society*, Toronto, Somerville House Books Limited, 1997a.

Kerckhove, Derrick de, *A pele da cultura (uma investigação sobre a nova realidade electrónica)*, Lisboa, Relógio d’Água Editores, 1997b.

Lorie, Raymond A., “Long Term Preservation of Digital Information”, *Proceedings of the first ACM/IEEE-CS Joint Conference on Digital Libraries, Roanoke, Virginia*, New York, 2001, ACM-Press, p. 346-352, disponível em: <http://doi.acm.org/10.1145/379437.379726>.

Luhmann, Niklas, *A improbabilidade da comunicação*, Lisboa, Vega, 1992.

McCray, Alexa T.; **Gallagher**, Marie E., “Principles for digital library development”, *Communications of the ACM*, Vol. 44, No. 5, May, 2001.

Nunes, Eduardo Borges, *Abreviaturas paleográficas portuguesas*, Lisboa, Faculdade de Letras, 1981.

Oliveira Marques, A. H., “Leitura nova”, Joel Serrão (dir.), *Dicionário de História de Portugal*, vol. III, Porto, Livraria Figueirinhas, 1971, págs. 475-476.

Rothenberg, Jeff, “Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital

Preservation”, 1999a, pdf, disponível em: <http://www.clir.org/pubs/reports/rothenberg/pub77.pdf>.

Rothenberg, Jeff, “Ensuring the Longevity of Digital Information”, *Scientific American*, 42-47, 1999b, disponível em: <http://www.clir.org/pubs/archives/ensuring.pdf>.

Rothenberg, Jeff. *Digital Information Lasts Forever - Or Five Years, Whichever Comes First*, disponível em <http://www.amibusiness.com/dps/rothenberg-arma.pdf> (20.Abril.2004).

Santos, Maria José Azevedo, *Da visigótica à carolina. A escrita em Portugal de 882 a 1172 (aspectos técnicos e culturais)*, Dissertação de doutoramento em História, Universidade de Coimbra, 1988.

Yaginuma, Tomoko *et alii*, “Implementation of Metadata for OmniPaper RDF Prototype”, *International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society 2004 (DLKC'04)*, University of Tsukuba - Japan, 2004, disponível em: http://canada.esat.kuleuven.ac.be/omnipaper/downloads/WP7_DLKC04Paper_1.0.pdf.

Yaginuma, Tomoko *et alii*, “Metadata elements for digital news resource description”, *Congresso Luso-Moçambicano de Engenharia*, 3 — “*Informática e tecnologias da informação*”, Maputo, 2003, p. 1317-1326; URI: <http://hdl.handle.net/1822/279>.

¹ Departamento de Ciências da Comunicação, Universidade do Minho.

² Servem apenas de referência, para além de tantos outros exemplos, os projectos pioneiros *Electronic Publication Pilot Project* da Biblioteca Nacional do Canadá, iniciado em 1994, e o equivalente australiano *Pandora*, a partir de Junho de 1994, e o congénere sueco *Kulturarw*, a partir de 1996. De igual menção, o projecto *Gallica* (Biblioteca digital da *Bibliothèque Nationale de France*), a *Library of Congress* (parte integrante do projecto *American Memory*, dirigido pelo Congresso Americano), e, entre nós, a *BNDigital* (projecto lançado pela *Biblioteca Nacional (Lisboa)*).

³ Tal é o caso do entretanto suspenso Projecto *Lifelog*, da responsabilidade da *Agência Nacional de Defesa Americana (DARPA)* [<http://www.darpa.mil/ipto/Programs/lifelog/index.htm>].

⁴ Entre outros exemplos, o *RepositoriUM*, repositório institucional da Universidade do Minho, “organizado por comunidades científicas, armazena, preserva, divulga e dá acesso à produção intelectual desta universidade em formato digital” [<http://repositorium.sdum.uminho.pt>].

⁵ Cf. o protocolo Open Archives Initiative Protocol for Metadata Harvesting [<http://www.openarchives.org>].

⁶ Como o sistema desenvolvido pela *Corporation for National Research Initiatives* [<http://www.handle.net>]

⁷ Cite-se, da Microsoft, o *MyLifeBits* [<http://research.microsoft.com/barc/MediaPresence/MyLifeBits.aspx>] baseada na visão pioneira de Vannevar Bush, que, em 1945, antecipava a possibilidade de se criar um dispositivo capaz de tudo «ciberizar», isto é, registar todos os elementos da vida de uma pessoa. Eram ideias visionárias, integradas num projecto pessoal denominado *Memex*, que consistia num “aparelho com o qual um indivíduo guardaria todos os seus livros, registos, comunicações, numa forma mecânica, pelo que tudo poderia ser consultado com extrema rapidez e flexibilidade”. [<http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>].

⁸ A *Leitura Nova* consiste na reescrita de um conjunto dos documentos legais e administrativos portugueses, copiados por ordem do rei D. Manuel I entre 1504 e 1552 em letra vigente na época de sua transcrição, com o intuito de facilitar sua leitura e evitar a sua perda (Oliveira Marques, 1971).

⁹ Ainda no mesmo âmbito, e em jeito de complemento, refira-se que muitas e graves imprecisões foram tomadas na tradução do X’ (“xis” aspadado). Este numeral, que deriva do XL, com valor igual a 40, como sabido, foi muito utilizado na Península Ibérica até ao séc. XIV, mas nos séculos XV e XVI foi-se tornando raro ao ponto de muitos escribas o ignorarem e, por

isso, o copiarem mal, com o valor 10 (X romano), levando a erros de transcrição (Santos, 1988). Em contrapartida, nos séculos XV e mesmo XVI era muito vulgar o R com valor de 40, que era uma deturpação do x aspadado (Nunes, 1981).

¹⁰ Cf. “Quantum Information”, *IBM Almaden Research Center* [http://www.almaden.ibm.com/st/quantum_information/qio/index.shtml].

¹¹ À data, os 300 Gb. do novo “Ultrastar 10K300” da *Hitachi* converteu-se no disco-duro com maior capacidade no mercado comercial [<http://www.hitachi.com/New/cnews/E/2003/0106e/>].

¹² Uma iniciativa da *Information Society Technologies* – o projecto *Omnipaper* (*Smart Access to European Newspapers*) [<http://canada.esat.kuleuven.ac.be/omnipaper/>] — “pretende investigar formas de promover o acesso a diferentes tipos de fontes de informação distribuída, permitindo aos utilizadores um acesso estruturado, personalizado e multilingue a todo o conjunto de artigos de notícias” (Yaginuma *et alii*, 2004).

¹³ Por exemplo, o *Dublin Core Metadata Elements Set* e o *Resource Description Framework* (RDF), ambas recomendadas por organismos amplamente reconhecidos a nível mundial, tanto pela comunidade científica, como pela comunidade empresarial (a *DCMI – Dublin Core Metadata Initiative*, no primeiro caso, e a *World Wide Web Consortium*, no outro). Igualmente se referenciam outros formatos estandarizados de notícias, o NITF (*News Industry Text Format*) e o NewsML, implementados pela *International Press Telecommunications Council* (IPTC), e o XMLNews, desenvolvido pelo XMLNews.org (Yaginuma *et alii*, 2004).

¹⁴ *Estudo da Reconfiguração do Campo da Comunicação e dos Media em Portugal* (Núcleo de Estudos de Comunicação e Sociedade da Universidade do Minho), projecto apoiado pela FCT (POCTI/COM/41888/2001) – <http://www.necs.ics.uminho.pt>.