

Preservação de publicações electrónicas na Internet: os arquivos imperfeitos

PEDRO FARIA LOPES
GUSTAVO CARDOSO
MARIA VASCONCELOS MOREIRA

PALAVRAS-CHAVE

PRESERVAÇÃO

NADO-DIGITAL

DOCUMENTOS

PUBLICAÇÕES

WEB

DOCUMENTOS DINÂMICOS

R E S U M O

Cada vez mais informação é gerada num contexto digital, introduzindo uma revolução em conceitos tão familiares como *documentos* e *publicações*. Esta informação é um recurso valioso e deve ser preservada para as gerações presentes e futuras.

Para a preservação nacional de informação nado-digital deve estudar-se se a aproximação deve ser selectiva, escolhendo o conteúdo a preservar de acordo com critérios previamente definidos, ou exaustiva, recolhendo o universo nacional de *sites* na Web e guardando todos os documentos encontrados. Vários exemplos são apresentados, representativos das duas aproximações, e as suas implicações são discutidas. A preservação de documentos dinâmicos é também discutida pela sua particular relevância relativamente a jornais e publicações *online* e pela especificidade dos desafios tecnológicos envolvidos.

São apresentadas algumas recomendações para que a Biblioteca Nacional concretize a preservação de documentos e publicações nado-digital, nomeadamente: considerar a aproximação de uma recolha exaustiva de documentos estáticos, criando uma série de *fotografias* do universo nacional (domínio “.pt”) de documentos publicados na Web; considerar a viabilidade do depósito voluntário, selectivo, baseado em acordos directos a ser desenvolvido/ensaiado com os editores dos principais jornais *online*.

>

A B S T R A C T

Information is increasingly being born, or originally created, in a digital context and can only be managed effectively in a digital world, thus bringing about a revolution in such familiar concepts as *documents* and *publications*. This information is an undoubtedly valuable resource and should be preserved for the current and future generations. For preserving national digitally born information it should be considered whether the approach is to be selective, choosing the contents to be preserved according to previously defined criteria, or exhaustive, harvesting the entire universe of national sites on the Web and recording all found documents. Several examples of the above-referred approaches are presented and their implications discussed.

The preservation of dynamic documents is also discussed for their particular relevance concerning online journals and newspapers, and for the specificity of the challenge their preservation demands in technological terms. Some recommendations follow on steps forward for the Portuguese National Library to implement the preservation of digitally born publications, namely: considering the feasibility of an exhaustive approach for static documents, thus creating a series of snapshots for the national universe (".pt" domain) of Web published documents; considering the feasibility of a selective voluntary and agreement-based approach for dynamic documents to be tested/developed with the publishers of main online newspapers.

INTRODUÇÃO

O presente artigo resulta de um estudo e respectivo relatório¹, levado a cabo por solicitação da Biblioteca Nacional relativamente às condicionantes e recomendações a ter em conta para a preservação do património nado-digital português presente na World Wide Web. Com base nos dados apresentados por LAWRENCE e GILES², e no estudo da ISC³ sobre a evolução de *hosts* Internet, entre 1999 e 2001, é possível estimar um valor para a actual dimensão da Internet que rondará os 2 biliões de páginas Web⁴. Esta riqueza enquanto espaço produtor/distribuidor de informação é também a primeira dificuldade quando se pretenda definir o campo de análise da Internet enquanto arquivo⁵. Neste artigo enunciam-se as características da comunicação existentes na Internet e como deverão ser consideradas num processo de criação de um sistema arquivístico que se pretende tão completo quanto o possível.

A primeira questão é saber sobre que dimensão deverá incidir a nossa atenção. As opções podem ser semelhantes ao sistema analógico, ou seja, o produzido em Portugal, e estaríamos condicionados à informação disponibilizada no domínio “.pt” – o nosso universo poderia ser o registo de domínios existente na FCCN⁶ – ou, em alternativa, englobar também todos os registos realizados a partir de qualquer fornecedor de acesso à Internet registado na entidade reguladora para as telecomunicações, o ICP⁷. Poderá ainda optar-se por uma abordagem mais vasta, a constituição do universo autoral em função da nacionalidade dos autores, o que implicaria um registo da informação independentemente do domínio ou dos fornecedores de acesso à Internet, baseado na produção escrita ou não em português desde que com origem em sujeitos nacionais ou entidades baseadas no território nacional.

Falar de arquivos é, como refere Fausto COLOMBO⁸, afirmar a existência de uma dimensão de imperfeição associada aos arquivos, daí que esta análise retenha como orientação esse princípio. A Internet, dadas as suas características enquanto sistema aberto, não se compatibiliza, na sua actual fase de evolução, com um registo total da produção autoral. É importante realizar escolhas que sejam realizáveis, não apenas declarações de intenção que se aproximem do ideal mas sem possibilidade de implementação prática, quer por razões tecnológicas quer pela incapacidade de influenciar os sujeitos envolvidos no processo de criação e distribuição.

Analisadas algumas das condicionantes *espaciais* importa agora questionar os condicionantes temporais. A Internet é um espaço de mutabilidade constante, o que agora está acessível pode deixar de o estar. Daí que o arquivo e posterior registo de informação tenha de ter presente essas condicionantes, qualquer lógica de indexação deve procurar reflectir um ponto de comparabilidade temporal. O registo de informação deve ser capaz de produzir um historial da evolução dessa mesma informação,

A proliferação de documentos digitais, nomeadamente os que apenas se encontram disponíveis na Internet, coloca maiores desafios à preservação da memória dos tempos actuais em prol das gerações futuras. É preciso ter em conta que certas previsões apontam para que, em certos sectores, dentro de aproximadamente 10 anos, as publicações electrónicas poderão constituir o grupo predominante de publicações⁹. Falhar na sua estratégia de preservação implica que se perca no tempo a história e a identidade de cada nação. Torna-se assim necessário prever a protecção e a preservação de documentos digitais, acompanhando as tendências verificadas na legislação internacional e garantindo a correcta preservação do património cultural no nosso país.

Numa primeira fase abordamos o ponto de situação quanto à implementação da recolha sistematizada de informação na Internet, e do seu arquivo e indexação, a nível internacional. Segue-se uma análise da realidade portuguesa, concluindo com a apresentação de um conjunto de propostas para a criação de uma memória para os conteúdos nado-digitais portugueses.

O novo património nado-digital

Desde a década de 70 que o desenvolvimento tecnológico informático tem tido um impacto significativo na indústria da publicação e no funcionamento das bibliotecas. A utilização cada vez mais frequente do computador para a criação de documentos, bem como de novos suportes ou veículos de publicação, são disso exemplos elucidativos. Para este fenómeno contribui o custo relativamente baixo da sua disponibilização num mercado *online* cuja dimensão favorece e amplifica o que Kate SHARP¹⁰ designa por democratização da informação, ou seja, o acesso massificado a um conjunto imenso de recursos, livres da filtragem levada a cabo pelos editores.

Esta necessidade de uma maior liberdade de publicação e acesso – na origem, por exemplo, da criação, em 1995, da IuK Kommission¹¹ – é também considerada como estando na base do crescimento exponencial na colocação de material na World Wide Web. Em Fevereiro de 1999, a revista *Nature* estimava em cerca de oitocentos milhões o número de páginas existentes, das quais se estima que apenas cerca de dezasseis por cento estejam registadas no conjunto dos motores de busca existentes¹². É a dimensão destes dados que leva Friedrich GEISSELMAN a considerar o mundo das publicações na Internet como sendo mais variado que o das publicações impressas¹³.

Cada vez mais é criada informação nado-digital, ou seja, criada originalmente no meio digital¹⁴ e cujo manuseio efectivo só pode, em alguns casos, ser feito no mesmo meio¹⁵. Na opinião de Jeff ROTHENBERG, as tecnologias de informação estão, portanto, a revolucionar as concepções de *documento* e *registo* a um nível tão grande quanto o da própria introdução da imprensa, transformando hábitos de literacia – nomeadamente pela introdução de práticas de metaleitura¹⁶. Por esta razão, considera o mesmo autor que a geração actual de registos digitais possui um significado histórico único¹⁷. Johan MANNERHEIM, por seu turno, considera a informação disponibilizada *online* como uma parte da herança cultural dos povos¹⁸, enquanto que dois outros autores, MIKSA e DOTY, atribuem a todas as fontes de informação o carácter de *obra*, representando como tal manifestações únicas de expressão intelectual¹⁹ – algo igualmente demonstrado pelo facto de cada vez menos registos da laboração de organizações serem vistos como dispensáveis, uma vez que constituem não só a sua memória enquanto instituição mas também um património e um bem em si mesmos²⁰.

No entanto, dada a sua natureza, esta informação é particularmente frágil²¹. Uma das características das publicações *online* é mesmo, para MANNERHEIM, a sua curta duração, estimada em cerca de alguns meses²². Por outro lado, o próprio ciclo de vida do equipamento (*hardware*) e programas (*software*) actualmente em utilização é, segundo Sara GOULD e Marie-Therèse VARLAMOFF, de apenas dezoito meses²³. Daí que, e paradoxalmente, apesar de a sua reprodutibilidade tornar a informação digital teoricamente invulnerável ao tempo, os suportes dos quais se serve estão longé de ser eternos²⁴. A menos que se proceda a uma recolha dessa informação no mais curto espaço de tempo²⁵, se não forem tomadas medidas concretas que permitam a preservação

de conteúdos representativos de todo um período da história humana – sejam eles de natureza governamental, organizacional, legal, financeira, tecnológica, ou pessoal²⁶ – muita dessa memória pode perder-se.

As utilizações potenciais desta informação são, na opinião de Sara GOULD, variadas, incontáveis e imprevisíveis, quer no presente quer no futuro.

Do número de páginas estimadas, muitas não terão interesse actualmente para a maior parte dos indivíduos; no entanto, não é garantido que tais materiais não venham a ser necessários para as próximas gerações. Como nos diz Jeff ROTHENBERG, o significado histórico de muita da nossa informação digital – que podemos por ora não considerar relevante – pode apenas vir a revelar-se muito depois da informação ter sido perdida²⁷.

A questão da preservação de colecções e registos foi desde sempre uma das principais preocupações das bibliotecas²⁸. Uma vez que as bibliotecas nacionais têm a seu cargo a responsabilidade de proceder à recolha e disponibilização de quase tudo o que é publicado em cada país – ou seja, textos impressos distribuídos ao público²⁹ – faz sentido considerar então que nenhuma outra instituição possui, como elas, condições para levar a cabo esse trabalho sistemático de recolha de longo alcance, razão pela qual se pode considerar que desempenham ainda hoje um papel fundamental na preservação da memória digital dos povos³⁰. Deixar esta tarefa a cargo das bibliotecas nacionais, alargando o raio da sua acção às publicações digitais, parece ser a melhor solução na maior parte dos casos³¹. Por último, o carácter transnacional de certas páginas parece favorecer a abordagem internacional, ainda que exista a esse respeito um longo caminho a percorrer até que surja uma instituição com existência e financiamento estáveis. Parece mais sensato apostar por agora na cooperação entre instituições e iniciativas nacionais, capitalizando as sinergias daí resultantes em termos de experiência e orientações futuras como, aliás, é o caso das iniciativas promovidas pelo grupo CoBRA+, em parte financiadas pela Comissão Europeia e em sintonia com a Conferência Europeia de Bibliotecas Nacionais (*Conference of European National Libraries – CENL*). É, pois, responsabilidade da comunidade bibliotecária recolher e preservar pelo menos parte deste património para permitir a sua consulta no futuro³². De acordo com KUNY e CLEVELAND, importa enfatizar a natureza de *biblioteca* nas bibliotecas digitais, de modo a constituir e preservar colecções de materiais nado-digitais para gerações futuras, sob pena de o nosso tempo poder ser considerado

como uma *idade das trevas digital* durante a qual o registo da actividade humana se perdeu³³.

Enquadramento terminológico

Importa precisar com maior exactidão alguns dos termos utilizados ao longo deste documento.

Por objecto **digital** podem entender-se essencialmente duas coisas: por um lado, aquele objecto originalmente criado no meio informático, que se serve de um suporte digital e cuja riqueza só se concretiza efectivamente quando manuseado nesse mesmo meio; por outro lado, o objecto enquanto representação digitalizada de um outro, qualquer que seja a sua natureza.^{34 35} Destes dois apenas o primeiro será alvo de abordagem neste documento.

Por **informação** podem igualmente ser entendidos dois significados: por um lado, a informação enquanto valor-conteúdo de objectos reunidos e disponibilizados *online*; por outro lado, informação enquanto processo de comunicação concretizado na criação e manuseio desses objectos³⁶. Também neste caso será o primeiro entendimento o abordado.

No que diz respeito ao conceito de **documento** vimos já que ao mesmo pode ser dado o significado de **obra** no sentido de expressão intelectual de um agente social. No entanto, algo mais pode ser adiantado. Documentos, segundo LEVY e MARSHALL, são artefactos através dos quais é fixada ou estabelecida uma intenção de partilha de um significado, mediante a utilização do que poderíamos chamar de tecnologias de fixação de significado. Contudo, considerar que é fixado a cada documento um significado não implica que o mesmo se mantenha imutável. Como nos dizem os mesmos autores, todos os documentos podem ter um significado fixo durante um período de tempo e fluido durante outro. Por outro lado, a cada documento podem ser atribuídas diferentes utilidades por diferentes períodos de tempo, podendo ser considerados como permanentes se excedem a duração da vida humana ou transitórios nos casos em que isso não acontece³⁷.

Finalmente, por **comunidade** deve entender-se no âmbito deste trabalho o conjunto interdependente de agentes sociais cujas características podem revestir-se de suficiente durabilidade temporal, de forma a serem observados³⁸.

MODELOS, BENEFÍCIOS E CUSTOS DE BIBLIOTECAS DIGITAIS DE DEPÓSITO

Para Stephen HARTER, há pouco consenso sobre a natureza de uma biblioteca digital, particularmente no que diz respeito ao seu conteúdo, o qual depende da adopção de uma visão mais abrangente ou mais redutora. Factores como 1) a política de colecção da biblioteca, 2) os conteúdos que os seus utilizadores pretendem que ela disponibilize, 3) o que é técnica e economicamente viável, determinam em grande medida o figurino de cada uma. Destas, a questão que mais circunscreve o âmbito de uma biblioteca digital não pode deixar de ser a relação estabelecida entre os conteúdos que podem ser efectivamente suportados em termos financeiros e o que é de facto necessário disponibilizar.

Em termos concretos, configuram-se actualmente dois eixos orientadores de práticas de preservação de conteúdos nado-digitais: por um lado, o eixo da natureza nacional ou internacional de determinada informação; por outro lado, o eixo da relevância ou qualidade atribuídas no presente ao seu conteúdo.

No que diz respeito ao primeiro eixo, a divisão observável nas práticas de preservação de informação nado-digital dos nossos dias é clara: na esmagadora maioria dos casos é atribuída às bibliotecas nacionais de cada país a recolha dos materiais de autores nacionais – alojados em servidores e domínios nacionais ou estrangeiros – e de materiais que retratem ou digam respeito ao país em causa. Apenas num caso, o do Internet Archive (iniciado em 1996), essa recolha é feita sem atender à nacionalidade dos conteúdos³⁹.

Em relação à questão da relevância e qualidade dos mesmos, existem hoje em dia essencialmente duas abordagens. A primeira destas abordagens prende-se com a **recolha selectiva** de conteúdos, ilustrada por um lado pelo projecto Pandora da Biblioteca Nacional Australiana (iniciado em 1996) e pelo Electronic Publications Pilot Project da Biblioteca Nacional Canadiana (iniciado em 1994) e, por outro lado, pela prática da Biblioteca Nacional Dinamarquesa (DanBib, iniciado em 1998), a qual define um conjunto de critérios mais restritivos do que os projectos australiano e canadiano.

A segunda abordagem preconiza a **recolha exhaustiva** de conteúdos, ilustrada pelo projecto sueco Kulturarw³ (iniciado em 1996), pelo projecto finlandês EVA (iniciado em 1997) e pelo Internet Archive: nestes casos, a prática é a de recolher tudo o que é publicado na Internet, o que compreende já milhões de documentos⁴⁰.

Esquemáticamente temos então a seguinte configuração, atendendo a ambos os eixos:

RELEVÂNCIA /NATUREZA	CONTEÚDO NACIONAL	CONTEÚDO TRANSNACIONAL
Recolha selectiva	Pandora Electronic Publications Pilot Project	
Depósito	DanBib	
Recolha exhaustiva	Kulturarw ³ EVA	Internet Archive

A recolha selectiva e o depósito

Partindo do pressuposto já referido de que qualquer informação publicada na Internet é uma *publicação*, LAW defende que as bibliotecas que tenham a seu cargo responsabilidades no capítulo do depósito nacional devem aperceber-se de que a World Wide Web é um espaço em si mesmo, no qual um legado cultural e documental importante de objectos digitais está a ser criado e disponibilizado, pelo que deve ser considerado como uma parte da herança bibliográfica nacional⁴¹. No entanto, e como afirma John GARRETT, «*data is drowning us*»⁴²; no novo domínio da informação digital a dificuldade reside, não na existência de informação disponível, mas sim no conseguir descobrir, por entre as inúmeras fontes e referências obtidas, o que se procura e com a relevância e qualidade pretendidas.

Uma das principais questões que se levanta quanto à falta destes mecanismos de publicação aquando da colocação de informação na Internet é mesmo a da sua autenticidade. Livres do funil editorial, os documentos digitais podem agora com maior facilidade e rapidez ser acedidos⁴³; mas valerá a pena conservar informação cuja validade se desconhece⁴⁴? A resposta para esta questão é, para MICHELSON e ROTHENBERG, o cruzamento sistemático de registos no seio de uma biblioteca digital ou entre várias, caso essa rede exista⁴⁵.

De acordo com a equipa responsável pelo projecto australiano de preservação de informação nada-digital Pandora, o volume potencial de material digital publicado na Internet a ser tratado pelas bibliotecas é considerável, a que acresce o facto de as publicações disponibilizadas na Internet carecerem muitas vezes de qualquer mecanismo de filtragem qualitativa; assim sendo, um grau elevado de selectividade é necessário⁴⁶. Esta selectividade é, aliás, reapropriada para o domínio das bibliotecas digitais, uma vez que políticas de gestão de uma colecção de registos baseadas em critérios de catalogação⁴⁷ e selecção por relevância e qualidade são métodos já utilizados e considerados essenciais para a organização da informação impressa, que podem, na opinião de Kate SHARP⁴⁸ e do PADI – Preserving Access to Digital Information – ser aplicadas a objectos digitais⁴⁹. Opinião semelhante possui HANSEN, para quem existe o risco real de muitos utilizadores de informação conhecerem sérias dificuldades em encontrar a informação de que necessitam caso os documentos nado-digitais não sejam sujeitos a um controlo bibliográfico⁵⁰. Um dos argumentos mais utilizados na defesa da prática da selecção é o de que não devem ser gastos os poucos e limitados recursos disponíveis para preservar documentos cujo interesse futuro não é seguro⁵¹.

Este controlo coloca, por seu turno, outra questão, a da sua concretização, a qual só é possível, segundo vários autores, pelo reconhecimento e incorporação do contributo dos bibliotecários, uma vez que, dado o amplo capital profissional que a sua prática lhes confere, a importância se mantém ou mesmo se vê reforçada como pedra-de-toque na era da informação, desde que readaptada às novas tecnologias⁵².

Para ALFORD, o papel do selector permanece fundamental no que toca a assegurar que o conteúdo de uma colecção de registos é apropriado às necessidades dos utilizadores, justificando os custos envolvidos na sua manutenção⁵³. No entanto, uma vez que essa colecção se reveste de um carácter cada vez mais pluridisciplinar, e porque a própria natureza dos objectos digitais a isso obriga, o processo de selecção não pode estar a cargo de um único agente social, antes necessitando do contributo conjunto de vários especialistas em diversas áreas.

Em suma, e de acordo com HEDSTROM, as bibliotecas e os arquivos apenas cumprirão o papel que historicamente lhes é atribuído se disponibilizarem os conteúdos requeridos pelos seus utilizadores⁵⁴, razão pela qual deve adaptar

esses conteúdos bem como o serviço que prestam, ao contexto em que se encontram inseridos⁵⁵.

Remetemos o leitor interessado na descrição detalhada dos projectos Pandora⁵⁶, DanBib⁵⁷ e Publications Pilot Project⁵⁸ para o relatório completo⁵⁹.

A recolha exaustiva

A abordagem selectiva levanta, como acabamos de ver, algumas questões, nomeadamente a da qualidade e relevância dos materiais nada-digitais disponibilizados na Internet, bem como a das limitações orçamentais impostas a este tipo de prática; questões para as quais alguns autores e os responsáveis por projectos de recolha exaustiva têm respostas diferentes.

Abordemos em primeiro lugar a questão das limitações orçamentais.

De acordo com o relatório final de um *workshop* realizado no âmbito da IFLA⁶⁰ intitulado “Social Aspects of Digital Libraries” o mundo real das bibliotecas digitais enfrenta, e é em grande medida condicionado por, constrangimentos de ordem orçamental⁶¹. Também HEDSTROM defende este diagnóstico, afirmando que, se é certo que as bibliotecas e arquivos não cumprem o seu papel se não forem ao encontro das necessidades dos seus utilizadores, não é menos improvável que efectivamente o consigam, dados os constrangimentos que enfrentam a nível de recursos económicos⁶². Quais são então os custos mais elevados a suportar?

Paradoxalmente, enquanto o espaço para armazenamento em computador é cada vez mais barato, os custos com pessoal qualificado para proceder a essa selecção são cada vez maiores⁶³. Apesar de, e segundo um estudo realizado pela British Library, preservar uma publicação digital custar vinte e cinco vezes mais que preservar a mesma publicação em papel⁶⁴ – o que, de acordo com KUNY e CLEVELAND, deita por terra o mito de a biblioteca digital ser menos onerosa que a tradicional⁶⁵ – são os custos infraestruturais e com o pessoal os que se revelam mais proibitivos e constrangedores da prática de recolha bibliográfica nada-digital. O custo de armazenamento tem vindo a decrescer; contudo, a preservação de objectos digitais, a ser realizada por processos de selecção que impliquem o recurso a pessoal especializado, tem tendência a onerar toda a prática de recolha⁶⁶. São, portanto, os projectos selectivos os que incorrem em mais encargos⁶⁷.

Abordemos agora a questão da qualidade e relevância futuras dos objectos digitais como fundamento para a sua recolha. Um dos argumentos utilizados para defender a tese da selecção é o de que não devem ser gastos os parcos e limitados recursos disponíveis para preservar documentos sem interesse. No entanto, os critérios que assistem a essa selecção são questionáveis tanto agora como, e mais decisivamente, no futuro. De acordo com MANNERHEIM, mesmo a mais louvável das selecções deita a perder informação importante⁶⁸. Também RUSSELL manifesta algumas reservas quanto à validade dos critérios de selecção baseados nos interesses e valências que assistem à investigação e valorização actuais de recursos digitais, afirmando que uma vez que muitos dos objectos existentes são recentes, pode incorrer-se no erro de ajuizar incorrectamente a sua real importância e interesse para as gerações futuras⁶⁹.

Remetemos o leitor interessado na descrição detalhada dos projectos Kurturaw⁷⁰, EVA⁷¹ e Internet Archive⁷² para o relatório completo⁷³.

Portais temáticos e documentos dinâmicos

Abordam-se agora, pela sua especificidade, os portais temáticos e os documentos dinâmicos.

Os portais temáticos

Os portais temáticos⁷⁴ são instrumentos de consulta individual de registos existentes em países como o Reino Unido, a Holanda, a Finlândia, a Suécia ou a Dinamarca. Estes portais possibilitam a consulta de registos cuja qualidade é controlada. Pelo garante de qualidade da informação disponibilizada, são naturais contribuintes para integrar a informação nado-digital a preservar em políticas de recolhas selectivas.

Documentos dinâmicos

A preservação dos documentos dinâmicos coloca problemas específicos. Por documentos dinâmicos entendem-se os documentos que apresentam a informação segundo um conjunto de opções, do leitor ou dos responsáveis pela publicação. São, regra geral, gerados em tempo real a partir de sistemas activos.

O exemplo paradigmático de documentos dinâmicos é o jornal *online*.

a utilização de sistemas informáticos na produção de jornais permitiu o aparecimento de jornais *online*, já que os ficheiros gerados podiam ser utilizados para outro tipo de disseminação⁷⁵.

Com o advento de sistemas de bases de dados interactivas e capazes de armazenar e disponibilizar grandes quantidades de informação *online*, assistiu-se a um significativo crescimento de publicações periódicas de todos os tipos, mas principalmente de jornais, através da Internet. Nem sempre estas publicações se limitam a reproduzir em formato digital as versões em papel, apresentando grafismos e conteúdos próprios inexistentes nas versões em papel. Tendo em conta a importância que este tipo de publicações tem, nomeadamente para investigadores e historiadores, para a preservação do conhecimento é necessário estabelecer critérios e procedimentos próprios para a sua recolha e preservação.

A questão é que os conteúdos deste tipo de publicações são alimentados por bases de dados dinâmicas e extremamente interactivas que podem ser actualizadas em questão de minutos e personalizadas à medida de cada utilizador que acede a um *site* deste tipo. Os sistemas utilizados para disponibilizar os jornais *online* permitem a actualização sempre que necessário ao longo do dia e dão origem a versões permanentemente actualizadas e facilmente acessíveis mas também mais efémeras do que as versões originais do mesmo produto em papel. Este tipo de publicação, que apenas existe no momento em que é visualizado por cada utilizador, é na realidade constituído por uma estrutura de página, na qual está definido o grafismo e a paginação das notícias, e pelos respectivos conteúdos, conjuntos de notícias inseridas em bases de dados que vão sendo carregadas na estrutura de página pré-definida à medida que o utilizador as solicita. Em certo sentido, as versões electrónicas dos jornais não existem como um produto completo em si mesmo a não ser no momento em que são visionadas por cada utilizador.

Neste âmbito levantam-se questões difíceis de contornar e que reforçam a necessidade de uma estratégia própria para este tipo de publicações:

Como se preservam conteúdos que são actualizados em minutos?

O formato digital tem o mesmo impacto do que o formato em papel?

Que informação passa para o leitor em ambos os casos? O conteúdo efémero de um jornal digital tem o mesmo valor do que o mesmo jornal em papel,

nomeadamente no que se refere à responsabilidade legal? Como fazer prova da existência desses conteúdos? Qual a responsabilidade dos editores relativamente aos conteúdos específicos das versões electrónicas?

Na realidade, as estratégias apontadas para a captura e preservação das outras publicações existentes na Internet raramente se adequarão à captura de jornais digitais já que apenas permitem capturar *fotografias (snapshots)* do que se está a visualizar num momento mas não capturam o conjunto total de todos os conteúdos das publicações em questão.

É ainda necessário ter em conta que, muitas vezes, o acesso a estas bases de dados depende de aplicações informáticas que poderão ser proprietárias e que não pertencem aos editores dos jornais, não podendo estes ser responsabilizados pelo depósito dessas mesmas aplicações. Esta questão deverá ser activamente debatida entre todas as partes interessadas, dado que também não é claro que as bases de dados sejam obrigatoriamente consideradas publicações abrangidas pelo depósito legal.

Não existem ainda normas de preservação internacionalmente aceites para a preservação de objectos dinâmicos de sistemas complexos. Estão ainda a ser trabalhados modelos conceptuais e processos técnicos para a preservação de obras multimédia, hipermédia interactiva, diálogos *online* e para muitas das novas formas electrónicas de comunicação que estão actualmente em fase de criação⁷⁶. Não se torna assim possível apontar inequivocamente uma direcção a seguir relativamente à preservação das versões electrónicas dos jornais. As iniciativas neste âmbito têm-se dirigido mais para projectos de microfilmagem das versões originais em papel do que para projectos de preservação das versões electrónicas. No entanto têm sido consideradas questões orientadas para várias abordagens. ROTHENBERG⁷⁷ refere a possibilidade do encapsulamento dos conteúdos num envelope virtual com indicações precisas para a recuperação e visualização dos conteúdos de forma a mantê-los independentes do *software* que pode mudar em menos de um geração. Mas uma solução deste tipo coloca problemas de escalabilidade dado o enorme universo de diferentes tipos de suportes para as inúmeras publicações deste tipo.

No entanto, a solução mais interessante parece ser a utilização de sistemas electrónicos de arquivo que são capazes de armazenar e utilizar toda a informação

utilizada na produção electrónica dos jornais e que em vez de armazenarem apenas o texto, armazenam também as fotografias, gráficos, informação gráfica, esquemas de cores e toda a informação que permite criar a estrutura das páginas onde os conteúdos são posteriormente dispostos. Esta solução permite armazenar e disponibilizar totalmente a informação, com a qualidade visual original, partindo do mesmo pressuposto que a página apenas é criada para ser visualizada e ocupando assim muito menos espaço do que a digitalização de cada página para além de manter intacta a interactividade inerente ao produto original.⁷⁸

Uma possibilidade para as estratégias de recolha selectiva do conteúdo, nestes casos, passa pela configuração de um *robot harvester* para recolha da obra/site, segundo determinados parâmetros que permitirão reproduzir no depósito uma ou várias facetas da mesma, ou por um modelo de depósito de conteúdo tal como existente na origem (base de dados, XML,...).

Para o caso nacional a solução deverá ser baseada num acordo entre a Biblioteca Nacional e os editores dos principais jornais nacionais negociando, por exemplo, o fornecimento à BN da informação que permita gerar a estrutura das páginas e uma actualização periódica das bases de dados. Para evitar no futuro a necessidade de migrações muitas vezes complexas e dispendiosas, deverá ser salvaguardado o armazenamento das bases de dados em formatos *standard* e compatíveis com SQL (*Structured Query Language*) já que a indústria de *software*, apesar da volatilidade que lhe é característica, tenderá a manter a compatibilidade dos seus produtos com os formatos normalizados de hoje em dia.

A REALIDADE PORTUGUESA

Em Portugal, tanto quanto sabemos, não foi até agora desenvolvido um sistema de arquivo sistemático, exaustivo ou selectivo, de páginas Web ou de qualquer outra dimensão de publicação na Internet, excepto o desenvolvimento de tecnologias de busca e indexação para pesquisa na Internet, i.e. os motores de busca.

Deveremos constituir como objecto toda a Internet, incluindo as páginas não acessíveis ao público em geral (páginas protegidas por códigos ou autorizações prévias) ou apenas a parte pública da Internet? A par das decisões tomadas pela maioria das entidades que trabalha na preservação de documentos nados-digitais, e segundo os objectivos dos modelos de depósito legal em geral (que incidem sobre aquilo que é posto à disposição do público)

também nós entendemos que se deverá considerar apenas a parte directamente acessível da Internet e, como tal, passível de ser recolhida e indexada.

Não existindo informação de origem nacional, ou de entidades internacionais, quanto à possível dimensão da produção e tipo de conteúdos Internet com origem em Portugal, procedemos a uma estimativa para a situação portuguesa tomando como ponto de partida o estudo desenvolvido por Steve LAWRENCE e Lee GILES⁷⁹, em 1999.

Segundo a informação disponibilizada pelo Instituto de Comunicações de Portugal (ICP), sobre o registo de domínios efectuado através dos *Internet Service Providers* (ISP) nacionais, podemos concluir que mais de 75% dos conteúdos nacionais se encontram albergados no domínio “.pt”, tendência que se tem mantido quase constante ao longo dos últimos 4 anos.

Pela análise que realizámos,⁸⁰ constata-se que o intervalo credível para a dimensão da Web no domínio “.pt” se situará entre as 929 000 e as 3 426 153 páginas.

Quanto ao tipo de informação disponibilizada, com base numa classificação dos registos temáticos acessíveis em Abril de 2001 em dois dos maiores portais existentes no domínio “.pt”, Sapo e Clix, constata-se que as páginas pessoais representam cerca de 18% da totalidade de páginas disponíveis cabendo os restantes 82% a entidades públicas, organizações não lucrativas ou comerciais, sendo as áreas de maior volume a comercial (15%) e a de lazer (15%), seguindo-se a de “sociedade, política e cultura” (14%) e a dedicada à Internet e computadores (11%).

CONSIDERAÇÕES FINAIS

Após a caracterização das soluções para a conservação da informação existente na Internet, e a análise do caso português, propomos a abordagem que pensamos constituir a melhor resposta para a preservação do património nado-digital em Portugal.

Os projectos aqui analisados centram a sua atenção apenas na World Wide Web, não abordando outras dimensões como a *usenet* e as *mailing-lists*, já que a distribuição de mensagens deverá obedecer a outros critérios: por exemplo, a opção por parte do emissor em submeter a mensagem à entidade de arquivo

responsável pela preservação digital, por considerar que a mensagem constitui uma mais-valia para a compreensão de uma dada página existente no acervo recolhido automaticamente.

No âmbito das publicações em suportes tradicionais, o sistema de depósito legal é habitualmente considerado como o garante mais sólido para a preservação completa do arquivo nacional publicado. As iniciativas voluntárias de depósito de material electrónico, apesar de muito importantes, não asseguram a representatividade do material nado-digital. Assim, pesados os custos/benefícios, a solução de recolha automática e exaustiva parece ser a mais sólida e permanente para as publicações nado-digitais estáticas. Dada a facilidade de publicação *online* torna-se necessário apostar na recolha automática em intervalos regulares. No entanto, dada a profusão de documentos existentes, para tornar a sua catalogação exequível há que apostar igualmente na utilização de metadata por parte dos próprios autores/editores (ex: Dublin Core)⁸¹ o que constitui um desafio ainda não resolvido, pois implicará a partilha, a nível global, de regras de catalogação para a Internet, uniformes ou pelo menos compatíveis, que estará sempre dependente de critérios subjectivos enquanto a sua criação couber aos próprios criadores de páginas.

Resta decidir onde recolher os conteúdos, o que na Internet corresponde a determinar em que domínios e *hosts* tal se deverá efectuar. O ponto de partida deverá ser o domínio “.pt” e todos os domínios onde o registo tiver ocorrido com base num endereço ou número de telefone nacional.⁸² Assim se obterá um acervo que tenderá a englobar grande parte da produção realizada por nacionais ou por entidades sediadas em Portugal. No entanto, será conveniente que, antecedendo qualquer indexação dos conteúdos obtidos, se proceda a uma pesquisa complementar não automática que permita regularmente adicionar endereços que não se situem nos endereços pesquisados mas que contenham conteúdos nacionais.

No nosso entender, e com base na implementação da experiência dinamarquesa, a legislação sobre o depósito legal não deverá ser estendida aos documentos nado-digitais. Reserva-se uma excepção para os documentos publicados em papel que possuam igualmente idêntica edição electrónica. Estes deveriam ser submetidos através da Internet para o acervo electrónico da Biblioteca Nacional, permitindo aos utilizadores estabelecer uma relação directa entre os dois tipos de suportes, papel e digital.

No que diz respeito às publicações nado-digitais dinâmicas, dada a complexidade tecnológica e heterogeneidade dos sistemas de publicação utilizados hoje em dia, torna-se mais relevante a opção por estratégias de depósito voluntário em detrimento de leis de depósito legal, mas, a este nível, através de acordo entre a Biblioteca Nacional e os proprietários da informação (por exemplo os editores dos principais jornais ou revistas nacionais) que permita a implementação de um sistema de recolha selectiva, negociando, por exemplo, o fornecimento à Biblioteca Nacional da informação que permita gerar a estrutura das páginas e a sua actualização periódica.

NOTAS

- ¹ LOPES, Pedro Faria, CARDOSO, Gustavo, MOREIRA, Maria Vasconcelos – *Preservação de publicações electrónicas na Internet – os arquivos imperfeitos*, Relatório, Lisboa: ADETTI/ISCTE, Biblioteca Nacional, 2001.
- ² LAWRENCE, GILES, *Accessibility of information on the Web*, Nature, vol. 40, 8 Julho 1999.
- ³ <http://www.isc.org/ds/WWW-200101/index.html>
- ⁴ LOPES, Pedro Faria, CARDOSO, Gustavo, MOREIRA, Maria Vasconcelos; op. cit.
- ⁵ O termo *arquivo* é utilizado aqui numa perspectiva lata, relacionada com as temáticas da recolha e preservação digital de documentos/publicações digitais na Web, e não necessariamente no contexto concreto, e por isso mais restrito, da arquivística.
- ⁶ Fundação para o Cálculo Científico Nacional, <http://www.fccn.pt>
- ⁷ Instituto de Comunicações de Portugal, <http://www.icp.pt>
- ⁸ COLOMBO, Fausto – “Uma Memória para a Tecnologia”. *Sociologia Problemas e Práticas*, 2000, n.º 32. Lisboa, CIES/ISCTE.
- ⁹ UNESCO, op. cit.
- ¹⁰ SHARP, Kate. “Internet Librarianship: Traditional roles in a New Environment”. 66.ª CONFERÊNCIA IFLA (International Federation of Library Associations and Institutions) - Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/005-120e.htm>
- ¹¹ A qual integrava diversos cientistas alemães precisamente com o propósito de partilhar mais rápida e facilmente conhecimentos científicos sem se sujeitarem às demoras e recusas processuais da indústria livreira. GEISSELMANN, Friedrich. “The indexing of electronic publications – Ways out of heterogeneity”; 66.ª CONFERÊNCIA IFLA (International Federation of Library Associations and Institutions) – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/173-181e.htm>
- ¹² SHARON, Taly, FRANK, Ariel; *Digital libraries on the Internet*; 66.ª Conferência IFLA (International Federation of Library Associations and Institutions) – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/029-142e.htm>
- ¹³ GEISSELMAN, Friedrich; op. cit.
- ¹⁴ Também considerado como *digital*, toda a digitalização ou reprodução digital de materiais não criados originalmente no meio digital não será considerada neste documento, como se explica no ponto relativo ao enquadramento terminológico.
- ¹⁵ ERSHOVA, Tatiana, HOHLOV, Yuri; op. cit.
- ¹⁶ Leitura e incorporação de conhecimentos não concentrados num mesmo documento. ERSHOVA, Tatiana, HOHLOV, Yuri; op. cit.
- ¹⁷ ROTHENBERG, Jeff; *Ensuring the Longevity of Digital Information*; RAND, 1999; <http://www.clir.org/pubs/archives/ensuring.pdf>
- ¹⁸ MANNERHEIM, Johan; op. cit.
- ¹⁹ MIKSA, Francis, e DOTY, Philip; *Intellectual Realities and the Digital Library*; <http://www.csdl.tamu.edu/DL94/paper/miksa.html>
- ²⁰ *Social Aspects of Digital Libraries – Final Report; Workshop* realizado pelo Department of Information Studies, Graduate Scholl of Education & Information Studies, University of California, Los Angeles; 1996; <http://is.geis.ucla.edu/research/dl/index.html>
- ²¹ KRANCH, Douglas; *Preserving Electronic Documents*, Proceedings of the third ACM conference on Digital Libraries, ACM Press, 1998

- 22 MANNERHEIM, Johan; op. cit.
- 23 GOULD, Sara, VARLAMOFF, Marie-Thérèse; *Digital Disappearances*; The UNESCO Courier; October 2000
- 24 ROTHENBERG, Jeff; op. cit.
- 25 MANNERHEIM, Johan; op. cit.
- 26 ROTHENBERG, Jeff; op. cit.
- 27 Idem
- 28 MANNERHEIM, Johan; op. cit.
- 29 Idem
- 30 ERSHOVA, Tatiana, HOHLOV, Yuri; op. cit.
- 31 MANNERHEIM, Johan; op. cit.
- 32 Idem
- 33 KUNY, Terry, CLEVELAND, Gary; op. cit.
- 34 *Social Aspects of Digital Libraries – Final Report, Workshop* realizado pelo Department of Information Studies, Graduate Scholl of Education & Information Studies, University of California, Los Angeles; 1996; <http://is.gseis.ucla.edu/research/dl/index.html>
- 35 HEDSTROM, Margaret; *Digital Preservation: a time bomb for Digital Libraries*; <http://www.uky.edu/~kiernan/DL/hedstrom.html>
- 36 *Social Aspects of Digital Libraries – Final Report, Workshop* realizado pelo Department of Information Studies, Graduate Scholl of Education & Information Studies, University of California, Los Angeles; 1996; <http://is.gseis.ucla.edu/research/dl/index.html>
- 37 LEVY, David, MARSHALL, Catherine; *Washington's White Horse? A Look at Assumptions Underlying Digital Libraries*; <http://www.csdl.tamu.edu/DL94/paper/levy.html>
- 38 *Social Aspects of Digital Libraries – Final Report, Workshop* realizado pelo Department of Information Studies, Graduate Scholl of Education & Information Studies, University of California, Los Angeles; 1996; <http://is.gseis.ucla.edu/research/dl/index.html>
- 39 <http://www.archive.org/index.php>
- 40 MANNERHEIM, Johan, op. cit.
- 41 LAW, Cliff; *PANDORA – Towards a National Collection of Selected Australian Online Publications*; 66.^a Conferência IFLA (International Federation of Library Associations and Institutions – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/174-157e.htm>
- 42 GARRETT, John; *Digital Libraries: The Grand Challenges*, EDUCOM Review, Julho/Agosto 1993, Volume 28, n.º 4; <http://www.ifla.org/documents/libraries/net/garrett.txt>
- 43 GEISSELMANN, Friedrich; op. cit.
- 44 HARTER, Stephen; op. Cit
- 45 HEDSTROM, Margaret; op. cit.
- 46 PANDORA Archive; *Guidelines for the Seleccion of Online Australian Publications Intended for Preservation by the National Library of Australia*; 2001; <http://pandora.nla.gov.au/selectionguidelines.html>
- 47 Com recurso a metainformação descritiva incorporada no próprio documento – a chamada *metadata*, de que a o protocolo Dublin Core é o exemplo mais disseminado – KUNY, Terry, e CLEVELAND, Gary; op. cit.
- 48 SHARP, Kate; op. cit.
- 49 PADI – Preserving Access to Digital Information; 2001; <http://www.nla.gov.au/padi/topics/9.html>
- 50 HANSEN, Randi; *Net Publications and Bibliographic Control – Seen from Denmark with a view to Sweden*; 65.^a Conferência IFLA (International Federation of Library Associations and Institutions – Bangucoque, Tailândia, 1999; <http://www.ifla.org/IV/ifla65/papers/120-153e.htm>
- 51 MANNERHEIM, Johan; op. cit.
- 52 Processo ao qual Philip Agre dispensa particular atenção.
- 53 ALFORD, Larry; *The impact of digital resources on organization and management of collection development and acquisitions*; 66.^a Conferência IFLA (International Federation of Library Associations and Institutions – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/168-180e.htm>
- 54 HEDSTROM, Margaret; op. cit.
- 55 *Social Aspects of Digital Libraries – Final Report, Workshop* realizado pelo Department of Information Studies, Graduate Scholl of Education & Information Studies, University of California, Los Angeles; 1996; <http://is.gseis.ucla.edu/research/dl/index.html>
- 56 PADI – Preserving Access to Digital Information; 2001; <http://www.nla.gov.au/padi/topics/9.html>
- 57 HANSEN, Randi; op. cit.
- 58 Electronic Publications Pilot Project (EPPP) – Summary of the Final Report; 1996; <http://www.nlc-bnc.ca/e-coll-e/ereport.htm>
- 59 LOPES, Pedro Faria, CARDOSO, Gustavo, MOREIRA, Maria Vasconcelos; op. cit.
- 60 International Federation of Library Associations and Institutions
- 61 *Social Aspects of Digital Libraries – Final Report, Workshop* realizado pelo Department of Information Studies, Graduate Scholl of Education & Information Studies, University of California, Los Angeles; 1996; <http://is.gseis.ucla.edu/research/dl/index.html>

- 62 HEDSTROM, Margaret; op. cit.
- 63 MANNERHEIM, Johan; op. cit.
- 64 GOULD, Sara e VARLAMOFF, Marie-Thérèse; op. cit.
- 65 KUNY, Terry e CLEVELAND, Gary; op. cit.
- 66 RUSSELL, Kelly; *Why Can't We Preserve Everything? Report for The Cedars Project*, 1999; <http://www.leeds.ac.uk/cedars/documents/ABS01.htm>
- 67 MANNERHEIM, Johan; op. cit.
- 68 Idem
- 69 RUSSELL, Kelly; op. cit.
- 70 ARVIDSON, Allan, PERSSON, Krister, MANNERHEIM, Johan; *The Kulturarw³ Project – The Royal Swedish Web Archiw³e: An example of "complete" collection of Web pages*, 66.ª Conferência IFLA (International Federation of Library Associations and Institutions – Jerusalém, Israel, 2000
- 71 EVA; *The Acquisition and archiving of electronic network publications*, 1997; <http://www.lib.helsinki.fi/eva/english.html>
- 72 The Internet Archive; *Why the Archive is Building an "Internet Library"*; 2001; <http://www.archive.org/about/index.html>
- 73 LOPES, Pedro Faria, CARDOSO, Gustavo, MOREIRA, Maria Vasconcelos; op. cit.
- 74 Também conhecidos por "Internet Subject Gateways" ou "Points d'Accès Par Sujet".
- 75 SMITH, G *Access to newspaper collections and content in a time of change*, 61st IFLA General Conference – Conference Proceedings – August 20-25, 1995; <http://www.ifla.org/IV/ifla61/61-smig.htm>
- 76 HEDSTROM, M.; op. cit.
- 77 ROTHENBERG, Jeff; op. cit.
- 78 SMITH, G; op.cit.
- 79 LAWRENCE e GILES, *Acessibility of information on the Web*, Nature, Vol 40, 8 Julho 1999.
- 80 LOPES, Pedro Faria, CARDOSO, Gustavo, MOREIRA, Maria Vasconcelos; op. cit.
- 81 MANNERHEIM, Johan; *Le World Wide Web et la préservation de notre patrimoine numérique – De nouvelles missions pour les bibliothèques*, 66.ª Conferência IFLA (International Federation of Library Associations and Institutions – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/158-157f.htm>

82 Por referência ao número de telefone não se entende aqui a recuperação ou manutenção de práticas antigas de acesso a BBS (Bulletin Board Systems), os sistemas pré-era Internet e *World Wide Web*. Neste contexto o número de telefone tem a ver com a possibilidade de identificação nacional na medida em que existam

registos em que o campo do número de telefone é preenchido, em bases de dados ou páginas Web: este elemento, o número de telefone nacional, pode ser um identificador da origem portuguesa da informação.

BIBLIOGRAFIA

- ALFORD, Larry; *The impact of digital resources on organization and management of collection development and acquisitions*, 66.ª Conferência IFLA (International Federation of Library Associations and Institutions – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/168-180e.htm>
- ARVIDSON, Allan, PERSSON, Krister, MANNERHEIM, Johan; *The Kulturarw³ Project – The Royal Swedish Web Archiw³e: An example of "complete" collection of Web pages*, 66.ª Conferência IFLA (International Federation of Library Associations and Institutions – Jerusalém, Israel, 2000
- CAMERON, Robert, e FRASER, Simon; *To Link or To Copy? – Four Principles for Materials Acquisition in Internet Electronic Libraries*, 1994; <http://elib.cs.sfu.ca/project/papers/e-lib-links.html>
- COLOMBO, Fausto, (2000), "Uma Memória para a Tecnologia", *Sociologia Problemas e Práticas*, n.º 32, Lisboa, CIES/ISCTE.
- COLOMBO, Fausto, (2000), "Uma Memória para a Tecnologia", *Sociologia Problemas e Práticas*, n.º 32, Lisboa, CIES/ISCTE.
- Electronic Publications Pilot Project (EPPP) – Summary of the Final Report; 1996; <http://www.nlc-bnc.ca/e-coll-e/ereport.htm>
- ERSHOVA, Tatiana, e HOHLOV, Yuri; *Migrating from the library of today to the library of tomorrow: re- or e-volution?*, 66.ª Conferência IFLA (International Federation of Library Associations and Institutions – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/063-111e.htm>
- EVA; *The Acquisition and archiving of electronic network publications*, 1997; <http://www.lib.helsinki.fi/eva/english.html>
- FOX, Edward; *Digital Libraries Initiative (DLI) Projects 1994-1999*, Bulletin of the American Society for Information Science; Outubro/Novembro, 1999.
- GARRETT, John; *Digital Libraries: The Grand Challenges*, EDUCOM Review, Julho/Agosto 1993, vol. 28, n.º 4; <http://www.ifla.org/documents/libraries/net/garrett.txt>
- GEISSELMANN, Friedrich; *The indexing of electronic publications – Ways out of heterogeneity*, 66.ª Conferência IFLA (International Federation of Library Associations and Institutions – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/173-181e.htm>
- GOULD, Sara, VARLAMOFF, Marie-Thérèse; *Digital Disappearances*, The UNESCO Courier; October 2000.

- HANSEN, Randi; *Net Publications and Bibliographic Control – Seen from Denmark with a view to Sweden*; 65.^a Conferência IFLA (International Federation of Library Associations and Institutions – Bangucoque, Tailândia, 1999; <http://www.ifla.org/IV/ifla65/papers/120-153e.htm>)
- HARTER, Stephen; *What is a Digital Library? Definitions, Content, and Issues*; Conferência Internacional KOLISS DL 1996, Seul, Coreia; <http://php.indiana.edu/%7Eharter/korea-paper.htm>
- HEDSTROM, Margaret; *Digital Preservation: a time bomb for Digital Libraries*; <http://www.uky.edu/~kiernan/DL/hedstrom.html>
- HODGE, G. M., “Best practices for digital archiving”, *D-Lib Magazine*, vol. 6, n.º 1, January 2000.
- KENNY, Anthony, chairman, *Report of the working party on legal deposit*, British Library, 1998; <http://www.bl.uk/>
- KRANCH, Douglas; *Preserving Electronic Documents*, Proceedings of the third ACM conference on Digital Libraries, ACM Press, 1998
- KULTURARW³; *Kulturaw³ Heritage Project*; <http://kulturarw3.kb.se/html/projectdescription.html>
- KUNY, Terry, CLEVELAND, Gary; *The Digital Library: Myths and Challenges*, IFLA Journal, vol. 24, N.º 2, 1998; <http://www.ifla.org/IV/ifla62/62-kuny.pdf>
- LAW, Cliff; *PANDORA – Towards a National Collection of Selected Australian Online Publications*; 66.^a Conferência IFLA (International Federation of Library Associations and Institutions) – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/174-157e.htm>
- LAWRENCE ET GILES, *Accessibility of information on the Web*, Nature, vol 40, 8 Julho 1999.
- LEVY, David, MARSHALL, Catherine; *Washington's White Horse? A Look at Assumptions Underlying Digital Libraries*; <http://www.csd.tamu.edu/DL94/paper/levy.html>
- LOPES, Pedro Faria, CARDOSO, Gustavo, MOREIRA, Maria Vasconcelos – *Preservação de publicações electrónicas na Internet – os arquivos imperfeitos*, Relatório, Lisboa, ADETTI/ISCTE, Biblioteca Nacional, 2001 (apresentado no Encontro sobre “Preservação Digital – Experiências e Estratégias”, 25 de Novembro de 2002, organização conjunta da Biblioteca Nacional e ECPA – European Commission on Preservation and Access, com participação de DELOS – Network of Excellence on Digital Libraries, e CLIR – Council on Library and Information Resources).
- MANNERHEIM, Johan; *Le World Wide Web et la préservation de notre patrimoine numérique – De nouvelles missions pour les bibliothèques*; 66.^a Conferência IFLA (International Federation of Library Associations and Institutions) – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/158-157f.htm>
- MIKSA, Francis, e DOTY, Philip; *Intellectual Realities and the Digital Library*; <http://www.csd.tamu.edu/DL94/paper/miksa.html>
- PADI – Preserving Access to Digital Information; 2001; <http://www.nla.gov.au/padi/topics/9.html>
- PANDORA Archive; *Guidelines for the Selection of Online Australian Publications Intended for Preservation by the National Library of Australia*; 2001; <http://pandora.nla.gov.au/selectionguidelines.html>
- PLACE, Emma; *Internet et la collaboration internationale autour des points d'accès par sujet*; 65.^a Conferência IFLA (International Federation of Library Associations and Institutions) – Bangucoque, Tailândia, 1999; <http://www.ifla.org/IV/ifla65/papers/009-143f.htm>
- ROTHENBERG, Jeff; *Ensuring the Longevity of Digital Information*; RAND, 1999; <http://www.clir.org/pubs/archives/ensuring.pdf>
- RUSSELL, Kelly; *Why Can't We Preserve Everything?*; Report for The Cedars Project, 1999; <http://www.leeds.ac.uk/cedars/documents/ABS01.htm>
- SHARON, Taly, FRANK, Ariel; *Digital libraries on the Internet*; 66.^a Conferência IFLA (International Federation of Library Associations and Institutions) – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/029-142e.htm>
- SHARP, Kate; *Internet Librarianship: Traditional roles in a New Environment*; 66.^a Conferência IFLA (International Federation of Library Associations and Institutions) – Jerusalém, Israel, 2000; <http://www.ifla.org/IV/ifla66/papers/005-120e.htm>
- SMITH, G.; *Access to newspaper collections and content in a time of change*, 61st IFLA General Conference – Conference Proceedings – August 20-25, 1995; <http://www.ifla.org/IV/ifla61/61-smig.htm>
- Social Aspects of Digital Libraries – Final Report*; Workshop realizado pelo Department of Information Studies, Graduate Scholl of Education & Information Studies, University of California, Los Angeles; 1996; <http://is.geis.ucla.edu/research/dl/index.html>
- TFADI; *Preserving Digital Information, report of the Task Force on Archiving of Digital Information, commissioned by The Commission on Preservation and Access and The Research Libraries Group, Inc.*, May 1996; <http://www.rlg.org/ArchTF/tfadi.index.htm>
- The Center for Research Libraries; *Preservation: microfilm and digital options*, March 1999; <http://www.crl.uchicago.edu/info/icon/proicon.htm#Preservation Microfilm and Digital Options>
- The Internet Archive; *Why the Archive is Building an "Internet Library"*; 2001; <http://www.archive.org/about/index.html>
- UNESCO; *The Legal Deposit of Electronic Publications*, 1998; <http://www.unesco.org/Webworld/memory/legaldep.htm>