

Mensagens de correio electrónico: conteúdos digitais em rede?

Almerinda Graça

Biblioteca Nacional de Portugal
Arquivo de Cultura Portuguesa Contemporânea
1749-081 Lisboa
Tel: 217982042
E-mail: agraca@bn.pt

RESUMO

Num universo documental em que muita da produção escrita nasce digital, é urgente e necessário que a comunidade profissional BAD aprofunde conhecimentos nesta área de modo a responder às necessidades no tratamento documental da informação em suporte electrónico. No contexto do Arquivo de Cultura Portuguesa Contemporânea, em cujos espólios as cartas desempenham um papel primordial, é natural que as preocupações nesta área surjam. As mensagens de correio electrónico, podendo evidenciar o valor de uma carta, podem também consubstanciar um universo diferente de prosa. Do ponto de vista da investigação, esta realidade despoleta novas necessidades relativamente ao tratamento da informação, aqui gerada em rede, e resultando em novos produtos documentais. Esta comunicação apresenta algum trabalho de investigação já que a autora entrevistou quatro pessoas simultaneamente utentes reais ou potenciais do ACPC, e produtores de conteúdos eventualmente significativos no quadro da actual produção cultural portuguesa: um investigador em filosofia, uma escritora de literatura infantil, um biólogo e uma investigadora em literatura portuguesa. A procura de soluções para o tratamento arquivístico de correspondência por correio electrónico tem sido amplamente estudada em alguns países. Do ponto de vista especialmente tecnológico, a problemática subjacente às mensagens por correio electrónico desdobra-se em múltiplas frentes: é o serviço Internet (*groupware*) com maior utilização mundial; é utilizado não só por indivíduos enquanto cidadãos mas também como trabalhadores em empresas ou no Estado. Há já soluções testadas em países como a Holanda, Canadá, Austrália, Nova Zelândia, entre outros, vocacionadas para o arquivo de mensagens de correio electrónico respondendo tecnologicamente a preocupações próprias da arquivística, isto é a questões de autenticidade, inviolabilidade e acesso continuado. Estas soluções são geralmente desenvolvidas por equipas de investigação directa ou indirectamente ligadas a instituições governamentais, sobretudo académicas, se em países como os EUA e o Reino Unido.

PALAVRAS-CHAVE: Espólios, mensagens de correio electrónico, arquivo definitivo, tecnologias, normas, investigação, património cultural

INTRODUÇÃO

É muito provável que em breve aporem ao Arquivo de Cultura Portuguesa Contemporânea (ACPC) espólios com correspondência por correio electrónico. Se não houver a resposta técnica e normativa adequada, o tratamento arquivístico desta correspondência estará em causa, em contradição com o que constitui uma prática já em curso em alguns países.

O ACPC reúne até hoje cento e cinquenta e cinco espólios, colecções e depósitos incluídos, manuscritos avulsos à parte. Neste universo documental, em suporte papel, cartas expedidas e recebidas por personalidades da cultura portuguesa contemporânea constituem parte significativa. Do ponto de vista da investigação, a realidade das mensagens de correio electrónico despoleta novas necessidades relativamente ao tratamento desta informação gerada em rede, e resultando em novos produtos documentais. Neste contexto, as mensagens por correio electrónico não podem ser vistas como simples sucedâneos das cartas.

Esta comunicação apresenta algum trabalho de investigação. A autora entrevistou quatro pessoas, simultaneamente utentes reais ou potenciais do ACPC, e produtores de conteúdos eventualmente significativos no quadro da actual produção cultural portuguesa: um investigador em filosofia, uma escritora de literatura infantil, um biólogo e uma investigadora em literatura portuguesa.

À pergunta «Confia no ACPC para levar a cabo essa missão [cuidar da correspondência em suporte electrónico]?», a resposta foi invariavelmente hesitante. Porquê? Pela novidade? Pela inexperiência ainda patenteada?

PROBLEMÁTICA

Para lá do valor intrínseco da correspondência por carta, quando num universo documental como é o que constitui o ACPC, uma mensagem por correio electrónico, podendo evidenciar esse mesmo valor, pode também consubstanciar um universo diferente de prosa, apresentando-se como que a ponta de um icebergue que lhe está subjacente. Assim, uma mensagem de correio electrónico pode ser o ponto de chegada de um percurso de raciocínio interactivo, ou à volta de um tema, ou o ponto de partida para outro. As

mensagens de correio electrónico podem deter não só a formalidade das cartas tradicionais, como deixar adivinhar o rasto de outros passos, alguns implícitos, outros subtis, devido ao seu tom coloquial por um lado, mas também à profundidade da mensagem. Esta é possibilitada não só pela “proximidade” entre os intervenientes – emissor e receptor –, como pela instantaneidade da escrita, coloquial, num face a face virtual: ocorrem mensagens nos dois sentidos e em sequência como os elos de uma cadeia.

Do ponto de vista do uso das mensagens por correio electrónico pelos investigadores, elas não podem ser vistas como simples sucedâneos das cartas. Por exemplo, e a título exemplificativo, pesquisas no *Google* podem estar subjacentes à escrita; hiper ligações implícitas e/ou explícitas podem gerar leituras multidireccionais. Nestes casos, uma sucessão de mensagens de correio electrónico pode conduzir à elaboração de um manuscrito em suporte electrónico, susceptível de acabar por ser publicado pela via tradicional, i.e. em papel, caso o investigador veja nisso outro proveito.

Para tomar o pulso à realidade, por amostragem, foram feitas quatro entrevistas, a prováveis utentes do ACPC. A grelha de perguntas foi:

1. Na sua troca de correspondência, costuma recorrer ao correio electrónico?
2. Costuma receber cartas pela via tradicional?
3. Qual imagina ser a relação de importância de uma via – a analógica – relativamente à outra – a digital –, na sua própria correspondência, nos anos de 2009 a 2006?
4. Como vê a possibilidade de ser garantido o acesso continuado e inviolável à sua correspondência por correio electrónico?
5. Confia no ACPC para levar a cabo essa missão?

Tendo como ponto de partida o correio electrónico, questões e perplexidades foram sendo analisadas, como que abrindo a porta para um mundo documental novo a exigir respostas adequadas no terreno arquivístico e tecnológico, relativamente a estes documentos, nascidos digitais. Os investigadores nas diferentes áreas do conhecimento – filosofia, literatura, biologia, história, etc. – cujos objectos de investigação podem constituir por sua vez arquivos, literários ou não, descobrem novas potencialidades de investigação a partir dos documentos digitais, aqui incluídos os que são produzidos por eles próprios: em que ponto, no seu ciclo de vida, “congelar” um documento digital tendo em vista o seu tratamento arquivístico? E como manter ou devolver-lhe contexto?

A procura de soluções para o tratamento arquivístico de correspondência por correio electrónico tem sido amplamente estudada em alguns países. A «macroavaliação» não é abordada neste trabalho pela simples razão de que estamos preocupados, não com a memória pública (produção de mensagens no contexto da administração pública dos Estados), mas com a memória colectiva, i.e. com o património cultural.

MÉTODO DE TRABALHO

Uma vez que este tema vem sendo recorrente nas nossas preocupações, e em vários cenários de acção, e também pelas suas características, socorremo-nos principalmente de pesquisas na ‘net’. As referências bibliográficas no final do trabalho têm explicitamente o objectivo de patentear o percurso das pesquisas realizadas como ponto de partida para desenvolvimentos futuros.

ABORDAGEM TECNOLÓGICA

Do ponto de vista estritamente tecnológico, o nosso posicionamento é o seguinte: questões como esta só se resolvem, no terreno, com equipas multi, inter e transdisciplinares, assentes em estruturas institucionais que devem estar alicerçadas em estratégias e políticas de informação, neste caso políticas de âmbito arquivístico, se a nível nacional, ou específico, para o caso pouco importa. O que aqui consideramos essencial é que, em qualquer momento, reunidas as circunstâncias, sobretudo as estruturais, o *know how* exista podendo resultar rapidamente em produção visível. Quanto mais sensibilidade houver também para questões tecnológicas, por parte de arquivistas, e também para questões normativas, por parte de informáticos, mais rapidamente se avançará. Tirar partido dos que já estão noutra patamar tecnológico nesta área pode constituir uma postura consequente. É neste contexto que a autora, pretendendo ser uma permanente aprendiz de arquivista, se aventura à abordagem de questões tecnológicas.

Formatos de ficheiros de correio electrónico

Na produção de documentos textuais, há três hipóteses:

1. Texto sem formato ou *plaintext*, com a extensão .txt.;
2. Texto com formato, ou *richtext* format, com a extensão .rtf;
3. Hiper texto, com a extensão .html (*Hiper Text Markup Language*), que permite a criação de páginas Web.

Segundo a norma RFC (Request For Comments) 2822, uma mensagem de correio electrónico deve conter o cabeçalho (*head*), e o corpo (*body*) da mensagem. O cabeçalho é utilizado para a sua etiquetagem, se se lhe juntar informação útil para o seu posterior tratamento: origem, destinatário(s), data de envio e assunto. Opcionalmente pode surgir outra informação, como seja controlo de SPAM.

Os ficheiros de correio electrónico, são trocados entre MTA (*Message Transfer Agent*) e UA (*User Agent*) e podem ser analisados em quatro cenários:

1. À entrada (do PC cliente)
Pode apresentar um de três formatos possíveis: 1. texto simples, ou *plain text*; 2. HTML; 3. RTF. De notar que nem sempre a opção pela formatação HTML pode ser a adequada, quando o cliente não dispõe de uma aplicação capaz de a visualizar correctamente.
2. No ambiente do PC
Com a aplicação de correio electrónico de que o computador pessoal dispuser, incluindo o respectivo editor de texto, as mensagens de correio electrónico

podem ter as extensões .txt, .htm, ou .rtf. No entanto, nos casos em que a aplicação de correio electrónico é proprietária (ex. *Microsoft Office Outlook*), esta grava as mensagens de acordo com as suas especificações próprias.

A título de exemplo, o *Outlook* arquiva as mensagens com as extensões .eml., colocando-as todas dentro de um único ficheiro perfeitamente identificado; com a extensão .msg salvaguarda as mensagens no ficheiro .pst, onde também são gravados outros dados de “agenda” (*Personal Information Manager*). Para que os arquivos anexos a uma mensagem sejam correctamente copiados, é preferível guardá-los fora da mensagem que lhes serviu de invólucro para assegurar a sua correcta leitura, independentemente da aplicação que lhe serviu de hospedeiro.

3. Transporte

Durante o transporte da mensagem, ela pode necessitar de percorrer vários MTAs até alcançar o UA de destino. Neste processo é necessário salvaguardar a sua integridade, independentemente das configurações específicas que cada MTA possa ter. Na maior parte das vezes cada MTA vai adicionando informação ao cabeçalho da mensagem tornando-a maior e, em muitos casos, é avaliado o seu conteúdo no que diz respeito à existência de vírus, formatos não permitidos (em função da política de segurança em uso na organização) ou mesmo detecção e controlo de SPAM. Para assegurar este requisito, à saída, os ficheiros de mensagens de correio electrónico podem obedecer às normas:

MIME – *Multipurpose Internet Mail Extensions*. Formato garante da integridade da mensagem durante o transporte pela rede. O envio de mensagens de correio em *plaintext* ou html, recorre a este formato. Este formato é hoje um sistema de reconhecimento de virtualmente todos os ficheiros;

S/MIME – *Secure / Multipurpose Internet Mail Extensions*. Define um formato para permitir a assinatura de mensagens de correio encapsuladas com o formato MIME. De entre as funcionalidades asseguradas, destacam-se a capacidade de autenticação, garantia de integridade dos dados e não-repudição;

TNEF – *Transport Neutral Encapsulation Format*. Formato que está na base do método especial de compactação de informações a incorporar em mensagens através da Internet. Utilizado no *Microsoft Outlook 2003, 2007* e no posto cliente do *Microsoft Exchange*. Este método é o veículo de transporte utilizado no envio de mensagens com o formato RTF;

IPv6 – Muitas vezes, as mensagens de correio necessitam de transpor a rede de endereçamento IPv4 (versão 4 da formatação de endereços utilizada na internet), para alcançarem MTAs que já estão em redes de endereçamento IPv6 (versão 6), maioritariamente redes académicas. Com o previsível esgotamento do espaço IPv4, rapidamente os MTAs terão de ser capazes de suportar o formato IP versão 4 e 6, em simultâneo.

Tecnologias de arquivo de correio electrónico

O correio electrónico, um dos serviços Internet de comunicação grupal (*groupware*), implica questões de confidencialidade e privacidade, aqui não afluadas. Há uma grande diferença entre aplicações “comerciais” disponíveis no mercado para arquivo de correio electrónico, e soluções informáticas resultantes da investigação por parte de Instituições, geralmente estatais, para a garantia do acesso continuado a mensagens de correio electrónico. É a segunda abordagem – a da preservação digital – a que preocupa os arquivistas. No entanto, a título exemplificativo, afluam-se pela rama algumas soluções tecnológicas comerciais.

Arquivo de correio electrónico pessoal

«Correio pessoal» é o de um utilizador que dispõe de uma caixa de correio para seu uso, privado, ou no âmbito das suas funções enquanto trabalhador de uma Organização. Neste segundo caso, e uma vez que só ele tem acesso às mensagens da sua caixa de correio, as tecnologias de arquivo de correio electrónico são as que estão implicadas no acesso, implicando os protocolos POP e IMAP, e a tecnologia Webmail, que tem por base o protocolo HTTP/HTTPS.

Em que consiste cada uma destas tecnologias?

POP – *Post Office Protocol*. Possibilita que o utilizador “puxe” o correio do servidor para o seu posto de trabalho. Este processo tem a vantagem de permitir a leitura *offline* para os utilizadores que não podem estar sempre *online*. Tem as desvantagens de ocupar espaço no posto cliente e de o acesso *offline* depender deste;

IMAP – *Internet Message Access Protocol* – usado nos casos em que o correio de um utilizador individual está num servidor. Esta solução é adequada para os utilizadores que precisam deslocar-se geograficamente, variando o ponto a partir do qual acedem ao seu correio: o posto de trabalho na Empresa, o seu *laptop* em casa. Além disto, o IMAP está por trás de tudo o que é webmail: podemos aceder-lhe através de um Web browser e ler as mensagens sem as transferir para o posto de trabalho. A grande vantagem numa leitura webmail/IMAP no entanto consiste em limpar no servidor o que não interessa e só chamar para o arquivo respectivo o que realmente é importante guardar;

Webmail – a gestão da caixa de correio pode ser feita através da Web, sendo naturalmente o acesso feito também pela Web. Disponibiliza ao utilizador menos funcionalidades, mas possibilita a leitura, resposta, reencaminhamento e eliminação de mensagens. Este tipo de acesso é disponibilizado pelos grandes motores de busca como o *Google* (plataforma *GMail*), *Yahoo*, *Hotmail*, entre outros.

O interesse destes Fornecedores de Serviços Internet consiste na publicidade, sempre presente quando se lhes acede. Quase todos, senão todos estes serviços, possibilitam um acesso POP. Na Web portuguesa existem serviços de correio electrónico gratuitos, como o *clix*, *iol*, *hotmail*, *portugalmail*. A questão que se coloca, para efeitos patrimoniais, que é o móbil

principal desta comunicação, é se os utilizadores destes serviços cuidam do arquivamento das suas mensagens nos seus próprios recursos tecnológicos, garantindo que ficam na sua posse mesmo se o Fornecedor de Serviços Internet desaparece no éter.

Arquivo de correio electrónico institucional, empresarial ou subordinado a um tema

Nos países tecnologicamente avançados, pelo menos 70% das Empresas armazenam a sua informação electronicamente. Para o negócio é vital: rápida localização de mensagens electrónicas; esquema de segurança avançado; armazenamento à prova de delapidação; disponibilização adequada das mensagens com vista a questões de auditoria e legais. Neste cenário, as Empresas são confrontadas com a necessidade fundamental de definir os métodos de arquivo relativamente a tamanho, tipo, classe e grau de retenção das mensagens, períodos de arquivamento, limite do tamanho da caixa de correio, e gestão da quota de cada trabalhador no servidor de arquivo (*backup*).

As soluções já existentes no mercado preocupam-se em princípio com o arquivo “primário” e “secundário”, de acordo com o “tempo de retenção” definido pelo *Chief Information Officer* (CIO). Baseiam-se na gestão do ciclo de vida da informação num contexto de *backoffice*.

Por outro lado, e uma vez que as empresas hoje se inserem num mundo global, é fundamental que, estando distribuídas geograficamente, tenham um sistema de gestão de correio electrónico distribuído, capaz de, recorrendo a tecnologia tipo *Post Process Management*, garantir a inviolabilidade do sistema de arquivo empresarial. Cumulativamente deve ter características de gestão *workflow*, e de interacção com outros sistemas clientes, ao mesmo tempo que deve ser capaz de indexar metadados, sobretudo com vista ao arquivo secundário.

Em resumo: um bom sistema de arquivo de correio electrónico empresarial deve assumir os seguintes pontos críticos:

1. métodos de arquivo de correio electrónico;
2. sistema de gestão dos dados em caso de desastre;
3. importador .pst ou outros (ex. *Linux*);
4. gestão global de correio electrónico;
5. fácil localização da informação relevante;
6. sistema de gestão de acordo com imposições legislativas sobre a Empresa e/ou o negócio;
7. processo de endereçamento do correio electrónico (*EMail Post Process*);
8. facilidade e transparência para o utilizador.

Existem já disponíveis no mercado soluções «chave-na-mão» que aqui não desenvolvemos.

Outro cenário é o que consiste nas situações em que os grupos são definidos à volta de um tema ou de um contexto comum, rotulando-se estas situações de “grupais”: Empresas, alunos de determinado grau de ensino, membros de partidos políticos ou de organizações à volta de um tema de interesse comum,

etc.

Nestes casos as mensagens não se destinam a um utilizador, mas a vários, variando a forma como as mensagens são armazenadas, sendo possível quase sempre aceder-lhes via *Web*, tendo ao dispor ferramentas de pesquisa algo sofisticadas. Há dois grandes grupos de estruturas neste contexto, que tecnologicamente acabam por ter uma estrutura muito semelhante: as distribuição de mensagens mais ou menos longas (*listserv*) e a distribuição de mensagens curtas, normalmente apenas com um *link* de algo que seja de interesse (por exemplo, mediante subscrição de um RSS. Nestes casos poderemos, ou não, ver as mensagens mais tarde.

A ARQUIVÍSTICA E AS TECNOLOGIAS DE ARQUIVO DE CORREIO ELECTRÓNICO

Há já tantas frentes de desenvolvimento tecnológico no campo do acesso continuado a documentos electrónicos, quantos os objectivos a que urge dar resposta. Um desses objectivos prende-se com as actividades a nível dos Governos, em cujo seio a informação produzida já é, numa enorme percentagem, exclusivamente electrónica.

Questões de macroavaliação emergem e, no contexto desta, questões tecnológicas de preservação digital. Aqui, no entanto, escolhemos como preocupação, não o acesso continuado à produção documental no seio da Administração Pública, mas unicamente o acesso continuado a mensagens de correio electrónico. Este constitui um ponto de partida tecnicamente adequado, por um lado, e, por outro, vai ao encontro de necessidades prementes, a nosso ver, do *Arquivo de Cultura Portuguesa Contemporânea*, bem como de instituições que tenham por missão o tratamento de arquivos electrónicos.

Por força do seu estatuto legal o ACPC tem a missão de arquivar, tratar documentalmente e garantir o acesso aos espólios de personalidades da cultura portuguesa. Ora a bem curto prazo, a correspondência pode já não ser em papel, mas exclusivamente em suporte electrónico, isto é cingir-se a mensagens de correio electrónico. O mesmo acontecerá com outro tipo de documentos, mas para se resolver um problema amplo é melhor política começar por tentar analisar um dos ângulos.

A era *Google*, o risco de iliteracia informacional sentam-nos ao computador a *googlar* procurando informação relevante, independentemente do tempo e do lugar, rapidamente tornada obsoleta na maior parte dos casos. Assim, o risco de obsolescência e a necessidade de refrescamento são constantes, ciclo ininterrupto que exige a aprendizagem ao longo da vida como atitude incontornável de cada cidadão. As pesquisas na ‘net’ à procura de soluções tecnológicas já testadas em alguns dos países com tradição de tecnologia de ponta no campo da arquivística digital podem, quando o 10.º Congresso se concretizar, já estar ultrapassadas, pelo menos em parte. Mas é sempre possível visitar, na *Web* “profunda” quais as novidades.

Estados Unidos da América. North Carolina State Archives – Preservation of Electronic Mail Collaboration Initiative

Em parceria, os Arquivos da Carolina do Norte, Kentucky e Pennsylvania, têm trabalhado intensivamente na frente tecnológica relativa à preservação digital de mensagens de correio electrónico, a fim de estabelecer as melhores práticas e desenvolver orientações técnicas, em termos arquivísticos, para arquivo continuado de mensagens de correio electrónico. Preocupando-se também com as aplicações informáticas, ao nível dos postos de clientes de correio electrónico, procuram implementar uma aplicação única a nível estadual, o sistema DoD/NARA 5015.2.

Desenvolveram também o Projecto [EMCAP](#) cujo sistema tem as seguintes características gerais:

1. Faz uso de servidores *eMail open source*, com base no protocolo IMAP;
2. Dispõe de um suporte lógico parametrizado para a fase de colecta e “ingestão” de correio electrónico;
3. Tem a possibilidade de criar pastas no servidor de arquivo (SA);
4. As mensagens que forem assinaladas como tendo valor arquivístico, podem facilmente ser copiadas e arquivadas no SA, usando um posto cliente;
5. É capaz de copiar a estrutura completa de uma pasta para uma pasta do SA;
6. Pode importar directamente um ficheiro .pst, uma *mbox*, ou um directório de mensagens.

Mensagens: a cópia (intocada) de cada mensagem original é arquivada em ficheiros *mbox*, uma cópia para cada pasta de caixa de correio; as mensagens são então convertidas para XML, havendo um ficheiro XML para cada conta (*account*); o ficheiro XML condensado (*digest*) é guardado em local seguro. É assegurado um armazenamento de *backup*, o que implica a conservação de conjuntos de pastas em duplicado.

Ingestão de mensagens: as mensagens podem ser deslocadas no SA; não podem ser eliminadas ou duplicadas se já foram colectadas; a definição da gestão da colecção de ficheiros pode ser determinada por *batch*, para o período nocturno ou semanalmente.

Ficheiros anexos às mensagens: os ficheiros anexos são extraídos automaticamente e salvaguardados no sistema de ficheiros sendo, sempre que possível posterior e automaticamente convertidos para o formato .pdf. O ficheiro XML (com o corpo da mensagem) inclui um modo de estabelecer relações dos ficheiros de mensagens *mbox* para cada pasta e dos anexos para cada mensagem: a) utilizando *links* de relação entre a mensagem e o anexo; b) armazenando um ficheiro *digest* separadamente.

Técnica: cada conta é arquivada separadamente, recebendo um identificador único global (IUG); a cada mensagem numa conta (*account*) é atribuído um IUG; as cópias não tocadas, originais, de cada mensagem, são salvaguardadas em ficheiros *mbox*, um para cada pasta

de caixa de correio; as mensagens são convertidas para XML, havendo um ficheiro XML para cada conta; a solução EMCAP está concebida para trabalhar com um repositório digital, funcionando de servidor colectador.

Metadados: é feita a extracção de metadados e pesquisa; são armazenados os seguintes componentes das mensagens: a) Para; b) De; c) Data; d) endereço de correio electrónico; e) informação sobre o modo de transmissão da mensagem.

Arquivo Nacional da Austrália

Sendo um dos países pioneiros em matéria de arquivo, a Austrália, publicou já, através dos seus Arquivos Nacionais, o que entende por preservação digital de mensagens de correio electrónico com vista à sua preservação, num contexto de desenvolvimento tecnológico. Inclui um excelente arquivo de textos sobre o tema.

Os «National Archives of Australia» têm estado na frente de desenvolvimento tecnológico relacionado com a preservação digital no contexto da Administração Pública, abordagem neste trabalho não contemplada. No entanto, não podemos deixar de registar o que os ANA entendem por preservação de registos electrónicos: «... as mensagens electrónicas armazenadas resultantes das transacções governamentais da *Commonwealth* são consideradas registos ... As mensagens electrónicas tomadas como evidências da actividade governamental devem ser capturadas num sistema de preservação digital de registos electrónicos. Se este processo não puder ser executado, uma funcionalidade de preservação digital (*record keeping*) deve ser construída nos sistemas electrónicos de mensagens de correio, ou estas, com anexação do contexto respectivo e apropriado, devem ser impressas e arquivadas em sistemas de conservação, tendo como suporte o papel».

Por outro lado o projecto *PADI – Preserving Access to Digital Information* inclui o correio electrónico como linha de acção. Para os Arquivos Nacionais da Austrália a publicação [Managing Electronic Messages as Records](#) (1997 May), relativamente às mensagens de correio electrónico diz: «To maintain their value as evidence, electronic messages as records must be inviolate, that is, not able to be altered or manipulated. Measures must be taken to ensure their authenticity».

Sobre o correio electrónico doméstico, na mesma obra: «Much *eMail* is generated from personal computers at home. Such *eMail* is potentially part of the personal records of an individual or family, comparable to the letters and notes that form a considerable component of any manuscript collection. Whether or not long-term access to these records is preserved depends on the individual creator, or their family if the creator has died. Access to these records is more likely to be preserved if professionals (such as records managers, archivists and manuscript librarians) provide advice to the record creators about preserving access to their *eMail*».

Holanda. Arquivos Nacionais – Dutch Digital preservation test-bed Project: practical experiences

Os Arquivos Nacionais da Holanda incluem muita informação sobre a experiência holandesa. O sítio Web respectivo reporta um encontro que juntou especialistas internacionais em I&D sobre preservação digital de correios electrónicos. O projecto governamental em título (2000-2003), teve como objectivo responder às seguintes questões de investigação:

1. Quais as vantagens e desvantagens de implementar abordagens diferentes sobre medidas de preservação?
2. Como pode cada acção ser medida e demonstrada?
3. Quais são os factores que afectam a eficácia e a pertinência de cada abordagem técnica? Por exemplo, custo? Tipo de registo? Especificações de autenticidade e períodos de retenção?
4. Quais são as especificações mínimas para funções de preservação? Por exemplo, quais são as especificações para aceder e recuperar registos a partir da função de preservação?

As questões sobre metadados passam por:

1. Que factores afectam os metadados necessários para preservação? E como? Por exemplo, o tipo de registo e a abordagem de preservação.
2. Quais as opções para associar metadados com registos?

A «test-bed» classifica os registos electrónicos de acordo com os 5 atributos definidos por Rothenberg Bikson (1999): conteúdo, contexto, estrutura, aspecto e comportamento. Tais aspectos espelham-se nas questões:

- Quais são as opções para preservar os atributos dos registos?
- Qual é a relação entre a preservação de atributos específicos e o custo da preservação?
- Perspectivas de preservação a longo prazo: 1. migração; 2. XML; 3. emulação.

Para a preservação de correio electrónico, a Holanda optou por fazer experiências com a linguagem XML. Com base em investigação de secretária (*desktop*), os registos *correio electrónico* deram provas de ser o registo típico para tratamento com XML. Assim, há muitas semelhanças entre os formatos *correio electrónico* e esta linguagem, sendo a conversão entre ambos bastante exequível já que ambos estão fortemente estruturados.

Por outro lado, as mensagens de correio electrónico devem seguir o formato das mensagens Internet, de modo a poderem ser compatíveis em plataformas diferentes. Este formato está muito bem estabelecido definindo as partes componentes de um ficheiro de transmissão de *correio electrónico*. A norma geralmente usada nos *correio electrónicos* é a RFC 2822, com as extensões MIME especificadas na norma RFC 2045-2049 (Ver também RFC 5321, 5322). Estas normas são controladas por uma organização não lucrativa, a *Internet Engineering Task Force* (IETF) – <http://www.ietf.org/>).

XML, por sua vez, é um formato normalizado, como uma linguagem *mark-up*. Está muito bem definida e também é controlada por uma organização não lucrativa – o *WWW Consortium* (<http://www.w3.org/>). As similitudes entre ambas significam que a conversão é um procedimento relativamente fácil. Todas as secções individuais estão completamente identificadas (marcadas) na transmissão do ficheiro de correio electrónico e podem facilmente ser convertidas num ficheiro XML, semelhante. Não podemos no entanto deixar de registar que o XML tem problemas com os diacríticos o que talvez implique ainda necessidade de ir observando desenvolvimentos futuros.

Há dois cenários possíveis na conversão para XML:

Post-use (conversão posterior para XML);
Pre-use (geração directa em XML).

O 1.º cenário é o que deve ser usado com as mensagens existentes (já enviadas ou já recebidas) que devem ser conservadas por tempo indeterminado. No 2.º cenário, na fonte, as mensagens são geradas directamente em XML.

Reino Unido – The Paradigm Project

Divulgado por uma arquivista, este projecto contextualiza estratégica e tecnicamente o problema. Pretendia desenvolver políticas e procedimentos, testar ferramentas e suportes lógicos e publicar um livro de trabalho.

Este constitui um excelente instrumento para quem quer concretizar um projecto nesta área, pois inclui, além de uma excelente bibliografia, normas e ferramentas implicadas nestes processos de desenvolvimento tecnológico que os arquivistas, não sendo informáticos, devem dominar por se tratar de uma frente de desenvolvimento em que há que intersectar saberes.

É necessário ter, entre outras, noções de metadados, de normas - METS (*Metadata Encoding and Transmission Standard*), OAIS (*Open Archival Information Systems*), ISAD, EAD, MODS, normas de indexação LCSH, NCA, ISAAR (CPF) -, de estratégias sobre preservação e conhecimentos de gestão (intervir no problema a montante, i.e junto dos criadores de conteúdos), noções de direitos de propriedade intelectual, de legislação sobre protecção de dados, de suportes lógicos *open source*, conhecimentos sobre repositórios digitais - por exemplo, suportes lógicos Dspace e Fedora -, de processos na ingestão de material digital vs. custos de arquivamento, de protocolos de cooperação interinstitucional, de ferramentas de extracção de metadados (e.g. JHOVE), formatos de ficheiros, de acesso à informação vs. descrição por níveis e de problemas técnicos como migração, emulação e arqueologia digital.

Maureen Pennock, na sua obra *Instalment on curating eMails: a life-cycle approach to management and preservation of eMail messages*, dá informações

estratégicas sobre aspectos práticos na preservação a longo prazo de mensagens de correio electrónico. Explica o porquê da necessidade da preservação de mensagens de correio electrónico e contextualiza os casos em que se devem estabelecer políticas de preservação neste tipo de documentos digitais.

Bibliothèque et Archives Canada – Stratégie canadienne sur l’information numérique

Sendo o Canadá um país pioneiro, no sítio Web da Biblioteca Nacional e dos Arquivos canadianos estão definidas as linhas estratégicas que darão continuidade à frente tecnológica que desenvolve, considerando que é a definição atempada de estratégias consequentes e pragmáticas o caminho indispensável para garantir a preservação da memória colectiva.

Logo na introdução do seu plano estratégico, o governo canadiano defende: «It is clear that the nations that nurture their digital information assets and infrastructure will prosper; those that do not will fall behind. Canada must act quickly and decisively. We must ensure that the needs of all Canadians – private citizens, scientists, creators, industry, students, and workers – are met. We must also make certain that the fundamental values of our nation, such as bilingualism, multiculturalism, inclusiveness, and equity, are reflected in the digital realm. This can only be accomplished with a strategic approach; one that is highly coordinated and involves all of those engaged in the creation, preservation and dissemination of digital information (Date modified: 2008-02-12)».

Arquivos da Nova Zelândia – Programa Continuum

Programa *Continuum: create and maintain*

Programa governamental integrado, que tem em vista a preservação digital de longo prazo.

Foi desenhado para facultar as ferramentas e serviços governamentais de modo a viabilizar boas práticas e normas na preservação digital. Pretende colaborar na construção de uma estrutura forte que inclua parceiros diversos de modo a mútua e cooperativamente poderem partilhar benefícios. É o programa que dá suporte à Lei de Preservação de Documentos Digitais Neozelandeses (2005).

CONCLUSÃO

O manuscrito autógrafo é o objecto por excelência da *Arquivística Literária*. A crítica textual, a génese da escrita, o imaginário literário, entre outras áreas do conhecimento, são as que ocupam o investigador tipo do ACPC. Os investigadores estão a mudar os seus paradigmas sendo utilizadores e consumidores das Tecnologias da Informação e também eles próprios geradores de arquivos electrónicos, isto é de conteúdos digitais.

Partindo tecnologicamente do tratamento de mensagens de correio electrónico pode ser uma das vias para aplanar o caminho do tratamento arquivístico de outro tipo de documentos em suporte electrónico. Não porque a ideia seja responder às necessidades arquetípicas da investigação literária (especialmente as dos

investigadores que tomam como objecto de estudo os manuscritos), mas porque compreendemos que eles próprios vêm levantar-se ante os seus olhos outras realidades, novas leituras, outro potencial, novas direcções por onde conduzir investigação que conta com respostas documentais a documentos de uma outra natureza.

No âmbito do trabalho citado na introdução – *O correio electrónico como possível veículo de mudança de olhar - do suporte papel ao suporte electrónico - no Arquivo de Cultura Portuguesa Contemporânea (BNP)* –, os quatro interlocutores foram unânimes em concordar que em qualquer ramo da investigação – Literatura, Biologia, Filosofia – o acesso continuado a mensagens de correio electrónico não só é necessário como incontornável. Há que garanti-lo também para efeitos de estudo e de investigação e, por consequência, com preocupações de memória colectiva e de património cultural e científico.

Circunscrevendo com rigor a questão, as mensagens de correio electrónico, nascidas em ambiente digital e em rede, são conteúdos digitais que urge preservar em muitos casos. Como profissional BAD e como utilizadora de recursos informáticos, incluindo de serviços de correio electrónico (com um endereço assente na infra-estrutura tecnológica da BNP, e outro nos serviços Google (Gmail)), interrogo-me seriamente sobre por onde anda o nosso património: quem, de entre os utentes portugueses dos Fornecedores de Serviços Internet se preocupa com recolher as suas próprias mensagens, a sua “correspondência”, e dela fazer uma cópia de salvaguarda própria e doméstica?

Na página principal do sítio Web da «Agência para a Sociedade do Conhecimento», integrada no «Ministério da Ciência, Tecnologia e Ensino Superior» pode ler-se: «Visão»: «Mobilizar a Sociedade da Informação e do conhecimento - o desenvolvimento tecnológico ao serviço de uma cidadania moderna e de progresso exige uma sociedade da informação inclusiva onde o conhecimento é um valor ético, social, cultural e económico fundamental que promove a criação de riqueza e emprego, a qualidade de vida e o desenvolvimento social».

Revendo o que existe em Portugal (até onde conseguimos chegar), RODA é o nome de baptismo do *Repositório de Objectos Digitais Autênticos*. No correspondente sítio Web é feito o ponto da situação a propósito do acesso continuado (preservação digital) a objectos digitais da Administração pública. Aí se encontram normas, instrumentos e livros de referência utilizados ao longo do projecto.

Estando conscientes da realidade portuguesa, tão estimulante, subscrevemos o que defende Jeanette Martin (EUA), que se identifica como «arquivista digital». A propósito da análise à salvaguarda de mensagens de correio electrónico do ponto de vista arquivístico, ela diz, entre outras afirmações: «Os arquivistas têm agora a oportunidade de intervir no próprio momento da criação dos registos de arquivo». É

essa a intenção desta nossa comunicação ao 10.º Congresso BAD.

REFERÊNCIAS

BIBLIOTHÈQUE ET ARCHIVES CANADA – *Stratégie canadienne sur l'information numérique*. [Consult. 15 Fev. 2010] Disponível na Internet: <http://collectionscanada.ca/scin/index-f.html>

DUTCH NATIONAL ARCHIVE [Consult. 11 Fev. 2010] Disponível na Internet: www.digitaleduurzaamheid.nl/

GRAÇA, Almerinda – *O correio electrónico como possível veículo de mudança de olhar - do suporte papel ao suporte electrónico - no Arquivo de Cultura Portuguesa Contemporânea (BNP)* (Trabalho apresentado na disciplina de «Teoria e Metodologia da Investigação», UNL, 2009 Fev.)

GRAÇA, Almerinda; MACHADO, Aurora; PINTO, Laurindo – *Formatos de ficheiros de correio electrónico* (Trabalho apresentado na disciplina de «Tecnologias da Informação e da Comunicação», UNL, 2008 Dez.)

GRAÇA, Almerinda; MACHADO, Aurora; PINTO, Laurindo – *Tecnologias de arquivo de correio electrónico* (Trabalho apresentado na disciplina de «Tecnologias da Informação e da Comunicação», UNL, 2008 Dez.)

INTERNATIONAL COUNCIL ON ARCHIVES. (ICA) *Australasian Digital Record keeping Initiative – Principles and functional requirements for records in electronic office environments* [Consult. 11 Fev. 2010] Disponível na Internet: <http://www.naa.gov.au/records-management/publications/Digital-recordkeeping-guidelines.aspx>

MARTIN, Jeanette – *The Paradigm Project: the University of Oxford & the University of Manchester* [Consult. 11 Fev. 2010] Disponível na Internet: http://www.paradigm.ac.uk/projectdocs/papers/200507a_rchiveshub.pdf

THE NATIONAL ARCHIVES OF AUSTRALIA – *Managing electronic messages as records* [Consult. 11 Fev. 2010] Disponível na Internet: http://pandora.nla.gov.au/pan/22371/20011105-0000/www.naa.gov.au/recordkeeping/er/elec_messages/contents.html

NATIONAL LIBRARY OF AUSTRALIA – *PADI: Preserving Access to digital information* [Consult. 15 Fev. 2010] Disponível na Internet: <http://www.nla.gov.au/padi/>

NORTH CAROLINA STATE ARCHIVES – *Email account XML schema document* http://www.records.ncdcr.gov/emailpreservation/mail-account/mail-account_docs.html

[NORTH CAROLINA STATE ARCHIVES – Preservation of electronic mail collaboration initiative \[EMCAP\] \[Consult. 11 Fev. 2010\] Disponível na Internet: http://www.records.ncdcr.gov/EmailPreservation/](http://www.records.ncdcr.gov/EmailPreservation/)

NOVA ZELÂNDIA. Arquivos Nacionais *Continuum: create and maintain [program]: eMail archiving solutions* [Consult. 11 Fev. 2010] Disponível na Internet: <http://continuum.archives.govt.nz/files/file/factsheets/fl4.html>

THE PARADIGM PROJECT – *Workbook on digital private papers* [Consult. 11 Fev. 2010] Disponível na Internet: <http://www.paradigm.ac.uk/workbook/index.html>

PENNOCK, Maureen - *Instalment on curating eMails: a life-cycle approach to management and preservation of eMail messages* [Consult. 11 Fev. 2010] Disponível na Internet: <http://www.dcc.ac.uk/resource/curation-manual/chapters/curating-e-mails/>

PORTUGAL. Direcção-Geral de Arquivos; UNIVERSIDADE DO MINHO – *RODA: Repositório de Objectos Digitais Autênticos: relatório final*. Projecto 613/2006 POAP – [Lisboa, Braga] : DGARQ e Universidade do Minho Consult. 11 Fev. 2010] Disponível na Internet: <http://roda.iannt.pt/pt/system/files/roda-relatorioFinal.pdf>

UMIC – AGÊNCIA PARA A SOCIEDADE DO CONHECIMENTO Consult. 11 Fev. 2010] Disponível na Internet: <http://www.unic.pt/>

VERDEGEM, Remco; SLATS, Jacqueline - *Practical experiences of the Dutch digital preservation test-bed* [Consult. 11 Fev. 2010] Disponível na Internet: http://www.digitaleduurzaamheid.nl/bibliotheek/docs/Article_in_VINE_2004.pdf