

Preservação Digital

Francisco Barbedo

Licenciado em História (arte e arqueologia), pós graduado em ciências documentais (opção arquivo), e mestre em gestão de informação. Foi subdiretor da Direção Geral de Arquivos entre 2007 e 2012. É presentemente diretor de serviços na área de inovação e administração eletrónica, na Direção Geral do Livro, dos Arquivos e das Bibliotecas.

Sílvia Saraiva Carvalho Martins

Mestre em Engenharia Eletrónica e Telecomunicações pela Universidade de Aveiro, Especialista de Informática no Instituto de Segurança Social desde 1999, tendo desempenhado cargos de chefia e direção desde 2005. Auditora do Curso de Defesa Nacional 2009.

Resumo

As sociedades atuais estão completamente suportadas e dependentes das Tecnologias de Informação, documentos e conteúdos digitais. Tudo aparentemente à distância de um simples clique, porém, para que se tenha acesso à informação digital é sempre necessário utilizar um interlocutor (hardware e software), que fruto da sua evolução exponencial rapidamente se torna obsoleto. Acontece, que quando se fala em objetos e documentos digitais só se preservarão os que se conservarem, tudo o resto se perderá. São aliás conhecidos casos de formatos, suporte e software que foram descontinuados, sem que hoje seja possível recuperar essa informação.

As instituições sejam do setor público ou privado, parecem ainda não ter acordado para esta realidade que pode por em causa o seu normal e bom funcionamento. É pois pertinente que se salvasse informação crítica para as organizações, pessoas e sociedade em geral e, esteja disponível, autêntica, fidedigna, utilizável e com caráter evidencial durante o período de tempo que dela se necessite. É importante que a nível nacional se tome consciência desta problemática sob pena de se perder irrecuperavelmente informação, criando uma política de preservação digital.

Abstract

Digital Preservation

Currently our society fully relies on ITC and digital content. However digital information is also dependent of an intermediary system, which is basically the software and hardware used to produce the information in the first place. The dynamism of the information technology industry quickly turns this intermediary system obsolete, creating dramatic problems to the retrieval and preservation of information on the medium and long terms. If no preservation actions are considered, digital information will become unreadable and therefore will be lost.

Institutions, both public and private, seem to have problems grasping this reality that can indeed jeopardize their regular activities. In order to achieve it several measures are necessary: one of them consist on the development of digital preservation plans that may help organizations to identify their digital information assets and prescribe the best strategies to preserve them. Is of utmost importance to acquire conscience of this problem and what exactly is at stake in order to develop a national level policy for information that includes its preservation

A crescente dependência das tecnologias de informação, atualmente o principal suporte para a produção e armazenamento de informação, a par da rápida taxa de obsolescência tecnológica característica da indústria informática, levanta às organizações problemas críticos de preservação de informação digital. Esta tem de ser mantida durante o tempo que for operacionalmente necessária, ou seja, enquanto a organização necessitar de dispor dessa informação por razões de execução da sua missão e para o exercício das suas atividades.

O tempo durante o qual é necessário guardar a informação é variável pois depende, ou deveria depender, de critérios estabelecidos com o máximo possível de objetividade, relacionados com necessidades de caráter administrativo e operacional e também com razões de caráter histórico e social.

Nem toda a informação é idêntica, ou tem propósitos funcionais similares. Assim o período de conservação dessa informação irá mudar em função de variáveis que têm de ser analisadas, tendo como objetivos:

- Guardar de forma autêntica e acessível a informação durante os períodos de tempo que a organização estime como necessários;
- Eliminar a informação assim que ela deixa de ser necessária à organização (pois a sua preservação implica elevados custos);
- Assegurar que a informação com interesse histórico e social é identificada e preservada;
- Otimizar os custos necessários para manter uma infraestrutura capaz de assegurar a preservação e utilização da informação digital.

De entre as várias razões acima indicadas para a realização desta análise discriminativa, chama-se a atenção para o elevado e permanente custo exigido para preservar informação digital de forma autêntica, e acessível ao longo do tempo.

Informação Digital

A era digital trouxe consigo a produção exponencial de informação. Há quem afirme que se produziu mais informação nos últimos anos do que em toda a história conhecida da humanidade. Saliente-se que a informação digital tem como principais vantagens a sua fácil criação, manipulação, acesso, partilha, disseminação, duplicação e transmissão e como principal desvantagem a sua volatilidade inerente à sua dependência da tecnologia, que rapidamente se torna obsoleta.

Importa também referir neste contexto, o conceito de objeto digital que se define como “todo e qualquer objeto de informação que possa ser representado através de uma sequência de dígitos binários” (Ferreira, 2006). São exemplos de objetos digitais: texto, som, vídeo, radiografias, fotografias, bases de dados, aplicações de *software*, páginas *Web*, modelos de realidade virtual, etc.. Um único objeto digital pode ser constituído por vários tipos de informação, contendo capacidades dinâmicas e interativas.

A distinção entre formato e suporte torna-se evidente, o mesmo documento pode ser representado ou reproduzido em suportes diferentes. Assim como, o mesmo objeto digital pode ser representado em formatos diferentes.

Preservação Digital

Atualmente, a informação toma um valor singular, aparentemente sempre disponível e fácil de aceder, à distância de um clique, sem que o seu utilizador final tenha consciência da sua localização ou da necessidade da sua disponibilidade ao longo do tempo. A forma fácil e confortável como toda a informação de que se depende no quotidiano tem sido rapidamente transferida para *bits* e *bytes* é espantosa e assustadora, assim como é espantosa a forma como as infraestruturas críticas estão assentes e dependentes desta tecnologia.

Acontece que este mundo espantoso tem a particularidade de a informação só estar acessível e poder ser lida através de um interlocutor tecnológico com características específicas, leia-se *software* e *hardware*. As Tecnologias da Informação e Comunicação (TIC), como são denominadas, evoluem exponencialmente ficando rapidamente obsoletas, não sendo invulgar alguma informação crítica ficar refém de formatos e suportes que entretanto foram descontinuados.

A obsolescência é portanto o fator crítico e determinante na preservação, autenticação e arquivo digital, tanto mais que o ciclo de vida de um formato/suporte digital está estimado de 5 a 7 anos, e a informação a preservar será a de conservação de médio prazo (10 anos) e longo prazo (mais de 10 anos), classificada como de preservação permanente.

É importante que se salvguarde informação crítica para as organizações, pessoas e sociedade em geral e, esteja disponível, autêntica, fidedigna, utilizável e com caráter evidencial durante o período de tempo que dela se necessite.

O objetivo da preservação digital é garantir a acessibilidade sustentada e fiável à informação. Como definição, “a preservação digital consiste na capacidade de garantir que a informação digital permanece acessível e com qualidades de autenticidade suficientes para que possa ser interpretada no futuro recorrendo a uma plataforma tecnológica diferente da utilizada no momento da sua criação.” (Ferreira, 2006).

Estratégias de Preservação Digital

Acompanhando a evolução tecnológica foram desenvolvidos dois tipos de estratégias de preservação digital: as de preservação tecnológica, centradas no objeto físico/lógico e as de preservação de informação centrada no objeto conceptual. O primeiro preserva a informação nos seus formatos físicos e/ou lógicos originais, utilizando a tecnologia inicialmente associada aos objetos digitais, garantindo-lhes assim o acesso. O segundo tipo está focado na preservação das características essenciais dos objetos digitais, sendo independente do *software* ou *hardware* utilizado (Ferreira, 2008).

As estratégias de preservação digital são a emulação, a migração e o encapsulamento. A emulação consiste na utilização de um *software* capaz de reproduzir o comportamento de uma determinada plataforma de *hardware* e/ou *software* numa outra que, à partida, seria incompatível. Este *software* tem o nome de emulador (Ferreira, 2006). É uma estratégia utilizada em objetos digitais com características dinâmicas e/ou interativas como aplicações de *software*, animações *flash*, por exemplo o antigo Windows, ou jogos do antigo Spectrum. A sua vantagem é a capacidade de preservar com grande fidelidade as características e funcionalidades originais do objeto. A grande desvantagem é o facto dos próprios emuladores se tornarem obsoletos.

A migração converte objetos digitais de determinado formato *hardware* e/ou *software* para outro mais recente. É uma estratégia utilizada em objetos digitais com características não dinâmicas como documentos de texto, imagens e base de dados que se divide em vários tipos: retro-compatibilidade, interoperabilidade, conversão para *standard* e a migração para suportes analógicos.

A retro-compatibilidade (ou atualização da versão do formato) é definida como a possibilidade de interpretar e reproduzir um ficheiro que foi criado numa versão antiga de uma determinada aplicação informática, numa nova versão da aplicação. Por exemplo, é possível criar, ler e gravar ficheiros Excel 97 na versão Excel 2010. Esta possibilidade tem como desvantagens permitir um número limitado de compatibilidade de formatos, normalmente dois ou três; o fato dos registos permanecerem dependentes do fornecedor que é proprietário do *software* e muitas vezes o objeto conceptual ser alterado ligeiramente, o que pode afetar a autenticidade e a integridade de registos digitais (Verdegem e Slats, 2004). Este tipo de migração é mais indicado para o curto prazo, pois com as sucessivas migrações corre o risco de acumular pequenos erros ou mutações.

A interoperabilidade é entendida, neste contexto, como a possibilidade de um ficheiro poder ser transferido de uma plataforma ou aplicação para outra diferente mantendo a possibilidade de reprodução na mesma aplicação ou outra similar. Por exemplo, um ficheiro Excel pode ser lido em Lotus 1-2-3 e pode ser gravado/

transformado num ficheiro Lotus. Neste tipo de migração corre-se o risco de se perderem características essenciais dos registos digitais, em particular se se tratar de um *layout* complexo ou conteúdo multimédia (Verdegem e Slats, 2004).

A conversão para *standards* é a migração de um formato proprietário (fechado) para um formato *standard* publicado (público ou proprietário). A vantagem desta migração é o facto dos registos digitais deixarem de depender do *hardware* e *software* original, que os poderiam levar à obsolescência (Verdegem e Slats, 2004).

A migração para suportes analógicos consiste na conversão de objetos digitais para suportes analógicos (não digitais) de longa duração, como por exemplo, a reprodução de um objeto em papel ou microfilme, cuja finalidade é aumentar a longevidade do objeto digital. Tem como principal desvantagem, o facto de só poder ser adotada para objetos digitais que tenham características passíveis de serem representadas em suportes analógicos, como texto ou imagem, não se aplicando a objetos digitais interativos e/ou dinâmicos (Ferreira, 2006).

O encapsulamento preserva juntamente o objeto digital e toda a informação necessária e suficiente para no futuro permitir o desenvolvimento de emuladores, conversores e visualizadores e tem como objetivo a conservação do formato original. Foram desenvolvidos alguns formatos que são independentes da aplicação informática a utilizar, ou dos suportes físicos onde os objetos digitais são criados ou armazenados.

Quando se fala em preservação digital é ainda importante realçar os conceitos de refrescamento e meta-dados. O primeiro consiste na transferência periódica de suporte físico de armazenamento para um mais recente antes que o anterior se torne obsoleto ou se deteriore. O segundo é definido, de forma simplista, como a informação sobre a informação, ou, informação estruturada para descrever dados digitais utilizando *standards* específicos, ou de outra forma, é a descrição exaustiva e estruturada da informação digital. Embora não constituam estratégias de preservação em si mesmas contribuem para o seu sucesso.

Afinal qual a melhor estratégia de preservação a seguir? A resposta é simples. É a que melhor se adequa ao contexto específico da instituição preservadora e às características do objeto digital a preservar. Ao escolher a estratégia de preservação deve ter-se em consideração além do já referido, o interesse em fazer o arquivo e sua comunidade de interesse, o tipo de material, o custo da manutenção e implementação.

O mercado oferece várias soluções informáticas, vocacionadas essencialmente para a gestão de documentos eletrónicos, que permitem gerir de forma eficaz a produção, circulação e armazenamento de documentos, mas não a conservação digital, não comportando funcionalidades de preservação dos documentos por períodos de tempo prolongados. Esta lacuna implica que as organizações tenham muitas vezes que desenvolver o seu próprio *software*.

A preservação digital deve ser planeada desde o momento da criação da informação digital, pois dada a sua complexidade e pertinência não se pode correr o risco de se perder informação crítica para as instituições ou para os cidadãos. É determinante que seja equacionada logo na fase de requisitos e implementação das aplicações informáticas, o que raramente acontece.

Plano de Preservação Digital

Objetivos Genéricos

O projeto de elaboração de um Plano de Preservação Digital (PPD) pretende produzir um documento estratégico que:

- Identifique e caracterize a informação digital produzida pelo organismo, categorizando-a de acordo com necessidades operacionais e conseqüentemente, atribuindo prazos de conservação das classes de informação identificadas;
- Determine quais os procedimentos a realizar para evitar obsolescência tecnológica e a conseqüente probabilidade de perder informação, indicando os procedimentos que devem ser empreendidos para cada classe de informação identificada;
- Indique e clarifique responsabilidades dentro da organização relativamente à execução desses mesmos procedimentos.

Complementarmente, a elaboração de um plano de preservação digital pretende dotar o organismo de ferramentas conceptuais e operacionais que permitam a gestão da sua informação digital de forma a mantê-la utilizável, garantindo os seus propósitos administrativos e eventualmente patrimoniais pelo tempo considerado necessário.

Objetivos Específicos

Para conseguir os propósitos enunciados, a elaboração de um PPD segue uma metodologia que pretende essencialmente recolher de forma pertinente, sistemática e direcionada, a informação necessária para a formulação de procedimentos práticos.

Esta informação incide sobre diversas entidades e objetos que podem ser assim categorizados:

- Sistemas. São recolhidos dados identificativos sobre os sistemas onde a informação digital é produzida e/ou guardada. Com este primeiro nível de informação ficar-se-á a saber que sistemas existem, onde existem (física e

logicamente), quais os atores que agem sobre esses sistemas e em que papel o fazem. Cada sistema será identificado de forma unívoca dentro da organização através da atribuição de um identificador de projeto.

- **Avaliação.** É recolhida informação específica sobre o conteúdo funcional de cada sistema identificado (e da informação nele contida), a relação informacional e funcional que cada sistema tem em relação aos outros; será feita uma estimativa de prazo de conservação dessa informação (sistema) e do seu destino final. Sempre que possível procurar-se-á sustentar esta análise com diplomas jurídicos ou normativos.
- **Plataforma tecnológica.** Neste nível é recolhida informação técnica sobre a plataforma tecnológica subjacente a cada um dos sistemas identificados. Esta plataforma integra o sistema intermediário (*software*) do *hardware*, as estruturas de armazenamento, de *backup*. Pretende-se ainda localizar procedimentos de *software* conducentes a aumentar a segurança do sistema (existência de criptografia, rotinas de auditoria e respetivos registos, etc.). Ainda nesta etapa serão identificados os formatos em que se encontra registada a informação.

Todos estes dados serão necessários para propor procedimentos e plataformas alternativas para assegurar a preservação da informação digital.

Após a análise desta informação, a qual deverá ser recolhida da forma mais completa, precisa e sistemática possível, é realizado um quadro síntese que inclui as estratégias de preservação bem como os procedimentos específicos para a informação contida em cada sistema identificado.

A concretização do Plano de Preservação Digital implica uma atitude organizacional compaginável com a gestão da mudança. A atribuição de novas responsabilidades para a concretização do plano bem como as responsabilidades imputadas aos serviços e pessoas implica a assimilação dessas novas tarefas e esquemas interpretativos associados, numa lógica de atividade corrente e continuada da organização.

Benefícios

Evitar a obsolescência tecnológica terá como consequência a diminuição da probabilidade de perder informação. Atuando antes da informação digital se tornar obsoleta e consequentemente inutilizável estar-se-á em melhor situação para gastar significativamente menos recursos e conseguir melhores resultados na gestão de informação.

Pretende-se que os organismos sejam capazes de gerir a sua informação digital de forma a mantê-la utilizável garantindo os seus propósitos operacionais.

Se o Plano for corretamente elaborado, executado e monitorizado as vantagens serão:

- A utilização racional e continuada da informação digital mantendo a operacionalidade da organização;
- O conhecimento exato dos sistemas existentes, a sua relação com o contexto funcional e organizacional, bem como a clarificação dos agentes que sobre eles atuam;
- A criação de conhecimento e infraestruturas apropriadas para a preservação digital;
- A redução de custos de armazenamento e gestão provocadas pela eliminação de volumes de informação desnecessária;
- A aquisição do conhecimento e das informações necessárias para a implementação de outros projetos como a segurança informática (e.g. COBIT).

A Preservação Digital em Portugal

Em Portugal temos vindo a acompanhar a tendência de desmaterialização de processos e crescente intermediação tecnológica. O *Simplex* e a modernização administrativa são exemplo paradigmático dessa tendência. No entanto, nenhum dos projetos desenvolvidos neste âmbito contempla qualquer ação de planeamento ou implementação relativa a preservação digital, ou seja, são orientados para o resultado imediato sem levar em linha de conta as necessidades decorrentes da manutenção dos produtos obtidos a médio e longo prazo. Os próprios sistemas desenvolvidos subjacentes a interfaces virados para o utilizador cliente, não consideram fatores destinados a acautelar a usabilidade da informação produzida e armazenada a médio ou longo prazo. Esta situação tem implicações graves que nos parecem óbvias.

Sem o desenvolvimento de medidas de proteção e salvaguarda da informação que lhe permita resistir mantendo a sua autenticidade e usabilidade ao longo do tempo, o resultado provável é essa informação tornar-se obsoleta acompanhando assim a tendência natural de desatualização dos sistemas de informação e das tecnologias informáticas. Não se questiona a existência de ações de atualização de *software* e que estas tenham sido planeadas. No entanto, estas medidas não incidem ou podem não incidir sobre informação proveniente de transações passadas e que não têm utilidade administrativa operacional mas mantêm ainda valor informativo e probatório.

Nota-se uma atitude de desvalorização desta componente em favor de outros fatores considerados prioritários, como a rapidez de desenvolvimento e disponibilização ao cliente de um determinado serviço. A nosso ver esta atitude compor-

tamental e metodológica apresenta graves riscos no que respeita à preservação do património arquivístico digital.

Constata-se a existência de falta de consciência relativamente ao risco da memória social digital se perder. De uma forma geral as pessoas encontram-se ainda mentalmente enquadradas, no que respeita a preservação e memória, ao ambiente tradicional de suporte papel. A perenidade e estabilidade que são qualidades intrínsecas deste suporte levam a pensar que o mesmo se passa no que respeita a informação digital. Considerações sobre quem preserva, como se preserva não são óbvias nas ações e planeamento dos agentes organizacionais e sociais. E se em papel esta atitude era similar, as suas consequências transportadas para o contexto volátil das TIC, resulta provavelmente no desaparecimento da memória.

Existem casos constatados de perda de registos em determinados sistemas de informação desenvolvidos no contexto do plano tecnológico, sistemas esses que são críticos do ponto de vista social e probatório, justificativos de conservação permanente. É, por exemplo, o caso do sistema SIRIC (Sistema de Informação do Registo Civil) que guarda os registos vitais dos portugueses – nascimentos, casamentos, óbitos, divórcios –, ou o CITIUS o atual sistema de suporte à atividade judicial. Estas perdas podem ainda não ser consideradas críticas, na medida em que, sendo o sistema recente, as pessoas sobre as quais os registos dizem respeito ainda são presumivelmente vivas pelo que é possível refazer esses registos, mas não há garantia que no futuro o mesmo não possa voltar a ocorrer. Não existe inclusivamente qualquer estudo publicamente divulgado que nos permita saber a dimensão da perda nem tão pouco o impacto que essa perda pode vir a ter tanto sob o ponto de vista de prova como de memória nacional.

A preservação digital não se inicia, tal como acontece com documentos em papel, na sua fase terminal de vida operacional. Um documento eletrónico torna-se potencialmente ilegível e portanto inutilizável ao fim de um período de tempo variável mas que pode ser tão curto como apenas sete anos. Nestas circunstâncias, a preservação digital é necessária mesmo dentro da organização produtora para que a informação produzida possa manter a sua usabilidade e utilidade probatória e operacional.

Considerações Finais

O facto de o papel ter uma durabilidade de mais de cem anos, leva o ser humano a não ter urgência na sua preservação. Aliás, a preocupação é o expurgo por questões de segurança e espaço. Mas, o paradigma mudou, e agora à medida que as aplicações, formatos e tecnologia vão sendo descontinuados, se tornam obsoletos ou irrecuperáveis, tudo se perderá à exceção do que se preservar.

As instituições não estão sensibilizadas para a pertinência dos planos de preservação digital. Embora as instituições possam ter em alguns casos um plano de preservação documental para suportes analógicos, não sentem necessidade de desenvolver o mesmo tipo de estratégia para a informação do universo digital. Muitas instituições já só dispõem de determinados documentos em formato digital sem que tenham a mínima sensibilidade para a sua preservação de longo prazo. Este facto pode implicar que esta informação se perca com as óbvias consequências ao nível de desempenho operacional traduzido em todas as vertentes da atividade organizacional (económica, salvaguarda de direitos, bem social). Muitos dos sistemas implementados estão já em fim de vida sem que as organizações e os seus gestores estejam sensibilizados para este tema, correndo sérios riscos de se perder informação crítica que será irrecuperável.

As infraestruturas de informação críticas para a segurança nacional deveriam obrigatoriamente ter planos para garantir a continuidade operacional da informação. E não nos referimos apenas à informação residente em sistemas ativos mas aquela que tendo ultrapassado um determinado prazo de vigência operacional ainda possui importância estratégica sob o ponto de vista de conhecimento, ou seja, a informação residente em *backups* não monitorizados e sistemas em vias de isolamento tecnológico.

A identificação da informação digital categorizada como crítica, independentemente da sua área funcional – saúde, obras públicas, etc. –, incluindo a sua evolução temporal associada ao respetivo grau de criticidade bem como a definição de medidas que assegurem continuamente o acesso e exploração dessa informação, deveria ser uma prioridade na definição de uma política de informação nacional.

O desenvolvimento da sociedade está de certa forma relacionado com a memória de que essa mesma sociedade dispõe. A coesão social apenas é possível se a memória de um país e das suas estruturas individuais e coletivas for mantida viva, dinâmica e acessível atuando então como referencial identitário do país. No entanto, a memória radica na informação e se esta desaparecer, as consequências serão dificilmente mensuráveis, mas facilmente perceptíveis como graves.

Afinal de contas de que serve a segurança nacional se não houver Nação?

Referências Bibliográficas

Ferreira, M. (2006). *Introdução à Preservação Digital: Conceitos, Estratégias e Atuais Consensos*. Guimarães: Escola de Engenharia da Universidade do Minho.

Ferreira, M. (2008). *Preservação Digital: Um Problema Multidimensional*. Paper apresentado no Seminário de Preservação Digital. Disponível em <http://repositorium.sdum.uminho.pt/bitstream/1822/8431/1/SMD-0.2.pdf>

Verdegem, R. e J. Slats (2004). "Practical Experiences of the Dutch Digital Preservation Test-bed". *The Journal of Information and Knowledge Management System* n.º 34.