

Internet, *data havens* e bases de dados virtuais: Onde está a informação?

João Paulo da Costa Amado

iDOC

Ideias em documentos
e Arquivos, Lda.

R. da Quinta da Nora, nº5, 5º Esq.
2795-558 Carnaxide

E-mail: idoc@altavista.net / jpamado@mail.com

RESUMO

A rede mundial Internet oferece hoje algumas das formas mais flexíveis e universais de difundir informação. No entanto, as suas capacidades não se esgotam aí. Têm vindo a ser desenvolvidas algumas tecnologias novas, as quais constituem enormes desafios para a actividade de profissionais da informação, como o são os arquivistas. Esta comunicação aborda algumas dessas tecnologias, as suas características principais e o seu impacto provável num futuro próximo.

PALAVRAS-CHAVE: Internet, bases de dados, *data-havens*, *peer-to-peer networking*, comunicações, segurança.

INTRODUÇÃO

Nas histórias de ficção científica, os arquivos e os arquivistas não costumam ter muito destaque. Sempre que se fala em grandes aglomerados de informação acumulados ao longo do tempo, eles costumam surgir enquadrados por cenários mais ou menos exóticos, onde o mais importante é a quantidade de braços que alguns dos protagonistas têm. Não interessa tanto a prática arquivística em si, mas apenas o conteúdo dos arquivos e a forma como ajudam a perceber o que aconteceu a eventuais extraterrestres.

No entanto, tais histórias têm uma virtude: muitas vezes servem como uma janela para especulações tecnológicas, mais ou menos realistas. Por vezes acertam.

Nesta comunicação vamos falar de tecnologias que não foram desenvolvidas a pensar em arquivos mas que, no entanto, têm tudo a ver com o armazenamento e recuperação de grandes volumes de informação. Ao mesmo tempo, a sua existência permite aproveitamentos completamente distintos daqueles para os quais elas foram inicialmente concebidas.

O QUE É UMA BASE DE DADOS?

Dependendo da forma como se entende este conceito, encontramos bases de dados em muitos sítios. Podemos entendê-lo de uma forma clássica como estando associado à recolha, armazenamento e organização de informação, utilizando na maior parte dos casos meios informáticos nestas operações. É assim que são entendidas as bases de dados que encontramos no mundo da arquivística e noutros sectores de actividade. Elas apresentam mais ou menos o mesmo tipo de componentes: tabelas para armazenar dados, formas avançadas de efectuar pesquisas, écrans de introdução e consulta de dados, relatórios para imprimir informação, possibilidades de automatização de tarefas mediante o recurso à programação.

Se não formos tão exigentes, basta olhar à nossa volta para encontrarmos conjuntos de tecnologias que também são bases de dados. Um simples PC doméstico com um dos sistemas operativos da família Windows já tem lá dentro boa parte dos componentes que compõem uma base de dados, nomeadamente:

- Espaço para armazenar informação (no disco rígido);
- Ferramentas básicas para estruturar e arrumar essa informação (o *filesystem*, os comandos e ferramentas para manipular directórios e ficheiros)
- Métodos para pesquisar e recuperar informação (outros comandos e utilitários do sistema operativo e de terceiras partes)

e ainda, finalmente, conteúdo informativo pesquisável, sob a forma de ficheiros com texto, som, imagem. A ser assim, as nossas empresas e instituições estão cheias de bases de dados, a maior parte das quais nunca chegando a ser identificadas como tal.

Habitua-mo-nos a encarar uma base de dados como uma forma de estruturar e organizar informação, com ênfase na sua normalização, manutenção de integridade, facilidade de introdução de dados, flexibilidade de recuperação - a lista de características é longa. Ao mesmo tempo, medimos factores como a quantidade de informação que suportam e a sua velocidade no processamento e resposta a pesquisas. Preocupamo-nos com corrupção de dados, com a necessidade de fazer cópias de segurança, com as escolhas a fazer sobre os dados que devem estar *on* e *offline*. As bases de dados, tal como as conhecemos hoje, foram possibilitadas pela informática. No entanto, é essa mesma informática que hoje nos fornece as mais variadas opções e tipos de bases de dados ou reuniões de informação, que pulverizam por completo as definições antigas.

A INTERNET E A WORLD WIDE WEB

Há dez anos o mundo acordou para uma realidade nova: surgiu uma maneira diferente de armazenar e recuperar informação. Entre Janeiro e Maio de 1991, foram divulgadas versões utilizáveis dos primeiros *browsers*, os programas utilizados para visualizar determinados tipos informação, armazenada em computadores ligados à rede mundial Internet, dando assim origem à *World Wide Web* (WWW). Se bem que a própria Internet tivesse já na altura cerca de 20 anos de idade, só com esta nova solução informática é que ela se transformou na verdadeira rede mundial que é actualmente.

A partir de um começo modesto, a ascensão da WWW foi fulgurante (um servidor *web* no Natal de 1990, 26 em Novembro de 1992, 200 em Novembro de 1993, quase 25 milhões no final do ano 2000). Hoje encontramos na WWW todos os tipos de informação - com e sem utilidade, de natureza legal e ilegal, polémica ou valiosa, de natureza estática ou verdadeiros festivais multimedia.

Muita coisa já foi escrita sobre a Internet e sobre o que ela significa para a raça humana. Para nós arquivistas e outros profissionais de informação, a Internet pode ser olhada como um dos maiores arquivos de informação que é possível imaginar. No entanto, esta riqueza junta-se a algumas das características estruturais da própria Internet, para criar, ao mesmo tempo, um dos maiores pesadelos que pode ser apresentados a um profissional de informação. Podem assim ser feitas ao mesmo tempo críticas e elogios. Por exemplo:

A Internet por vezes é lenta, ou de acesso difícil. Não constitui um veículo fiável para o transporte de informação.

A Internet actual é composta por uma complexa mistura de redes, computadores e *hardware* de comunicações, todos eles a correrem os mais variados programas de controlo e acesso. O simples facto de existir e de possibilitar a troca de informação entre ambientes tão díspares deve ser saudado como um dos grandes feitos do homem.

Constituindo sobretudo um canal de comunicações, a grande variedade do equipamento que lhe dá forma faz com as velocidades físicas de transporte de dados de uns lados para os outros também sejam variáveis.

Infelizmente, a alta velocidade permanente apenas está reservada para a “espinha dorsal” da Internet, a qual, como o seu nome indica, constitui um conjunto de linhas de comunicação que suportam uma elevadíssima percentagem do tráfego total:

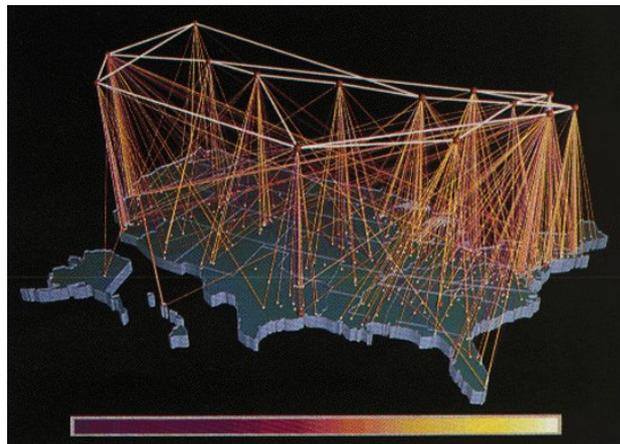


Figura 1: Acessos à Internet nos Estados Unidos. As linhas mais claras representam o backbone de alta velocidade.

A verdade é que a velocidade de transmissão é tanto mais cara quanto mais alta. Mesmo um simples utilizador doméstico tem à sua disposição uma boa variedade de linhas de comunicação de alta velocidade... pelo menos nos catálogos dos fornecedores de acesso. No entanto, o custo da maior parte destas soluções ainda assume um carácter proibitivo.

Mesmo que uma entidade tenha uma ligação dita “rápida” à Internet, os seus problemas não ficam imediatamente resolvidos. A rapidez dessa ligação só pode ser medida até à primeira peça de equipamento de comunicações que receba o seu sinal. Por exemplo:

- Um utilizador doméstico com um *modem* de 56 Kbps raramente consegue atingir essa velocidade, estando condicionado logo à partida pelas limitações de boa parte das linhas telefónicas actualmente existentes. É certo que pode estar a tentar aceder a um conjunto de páginas que estão num servidor que dispõe de linhas rápidas de comunicação. A velocidade de resposta não lhe serve de nada, pois o equipamento básico (o seu *modem* e as linhas telefónicas) limita desde logo as comunicações.
- Uma empresa pode dispor de uma linha dedicada de tecnologia HDLC com 2 Mbits de largura de banda. Esta velocidade só é garantida até ao equipamento de encaminhamento de sinais (*router*) do seu fornecedor de acesso. Pode estar a tentar aceder a um *website* que responda de forma extremamente lenta - quase toda a sua largura de banda pode estar desocupada, face à lentidão com que os dados chegam.

Não esqueçamos que as capacidades de efectuar ligações à Internet estão intimamente ligadas às capacidades da rede de comunicações existente. No caso de Portugal, existem zonas com boa acessibilidade e outras com má acessibilidade - algo que, de uma forma ou de outra, ocorre em qualquer lugar do mundo. Se uma entidade precisar de fazer o seu negócio ou actividade dependerem da Internet, convém certificar-se que a infra-estrutura de comunicações existente lhe dá resposta, ou então mudar de ramo (ou de país...)

Finalmente, a existência de fusos horários no nosso planeta também é um factor limitativo. Basta pensar que uma elevada percentagem dos conteúdos da Internet residem em servidores dos Estados Unidos. Em certos dias, a partir das 15H do nosso país, torna-se extremamente difícil aceder a *websites* americanos - lá são 9H00 e o dia de trabalho está a começar.

A Internet não tem dono, logo qualquer tentativa de estruturar e organizar o seu conteúdo está condenada desde o princípio.

De facto não tem dono, na medida em que ninguém a pode desligar completamente, ou decidir *ad-hoc* quais devem ser as suas características. No entanto, não é uma rede anárquica, e é devidamente orientada por diversos comités técnicos. Não tem um dono da totalidade da rede, mas tem muitos donos de troços específicos e de serviços centrais de encaminhamento e identificação de computadores. Estes donos parcelares estão em posição de prejudicar, voluntariamente ou não, a livre circulação dos dados (mas não de afectar a integridade dos conteúdos).

É difícil encontrar informação na Internet, porque a maior parte dela não está estruturada.

Cada entidade que disponibiliza informação na Internet fá-lo à sua maneira. As regras que todos seguem e que permitem navegar e recuperar alguma coisa, são do foro tecnológico, e não do foro arquivístico. Conclusão: os actuais motores de busca seguem abordagens de força bruta, percorrendo de facto todas as páginas que lhes surgem à frente para indexar os conteúdos.

Estas tarefas de indexação e recuperação podem ser optimizadas de duas formas:

- Havendo um grande investimento do lado humano, no momento da criação das páginas, por forma a adicionar a cada página informação descritiva e de indexação, a qual é usada pelos bons motores de busca da Internet.
- Desenvolvendo ou aproveitando o desenvolvimento das mais recentes tecnologias ao apostar em *software* que faça reconhecimentos de conteúdo e contexto e proceda à classificação automática. Esta abordagem é mais exigente que a anterior, sobretudo do ponto de vista técnico.

AO JUNTAR BASES DE DADOS E INTERNET OBTEMOS...

Um dos resultados que é possível obter constitui o sonho de qualquer profissional da informação: documentos organizados, facilmente pesquisáveis e recuperáveis, facilmente disponíveis e de acesso universal. Em qualquer

ponto do planeta, desde que haja uma ligação à Internet, torna-se possível aceder aos dados assim disponibilizados. Este resultado só é possível quando há um grande esforço inicial de preparação dos documentos a publicar.

É lícito pensar que se um arquivo disponibilizar o seu conteúdo na WWW, ficamos perante uma concretização perfeita dos princípios que enunciámos: informação disponibilizada por profissionais, de forma organizada e de uma maneira acessível por todos. No entanto, a esmagadora maioria do que existe na WWW não teve qualquer intervenção de um arquivista. Muitas vezes nem sequer a intervenção de qualquer profissional deste ramo. Tal não significa que sejam dados sem qualquer estrutura. Simplesmente, a sua organização é feita muitas vezes em função das necessidades do negócio ou actividade a que se dedica a entidade que a criou ou que a detém.

A conclusão provisória que apresentamos quanto a este tópico é que a disponibilização de uma base de dados na Internet ainda é encarada sobretudo como uma questão tecnológica, e não como uma questão do foro arquivístico.

OS ARQUIVISTAS PERANTE AS BASES DE DADOS E A INTERNET.

Até que ponto é que é desejável a intervenção dos arquivistas nestes processos de sistematização e preparação de dados, para serem disponibilizados através da Internet? Sem qualquer ironia, somos levados a fazer várias constatações.

- De um mero ponto de vista teórico, consideramos que eles são, os profissionais mais indicados para levar a cabo essas acções. Quer se trate de ajudar empresas e instituições no seu dia a dia, ou de disponibilizar para o mundo documentos acumulados ao longo dos séculos, eles têm a preparação necessária.
- Se uma base de dados estiver a cargo de arquivistas e se for necessário disponibilizá-la online via Internet, então é de todo aconselhável que eles sejam ouvidos sobre esse processo. Acontece que muitas vezes as bases de dados caem completamente fora da alçada destes profissionais, ficando nas mãos do pessoal ligado às tecnologias de informação, logo desde o momento da sua concepção.
- Quando há um esforço coerente de publicação de conteúdos a um nível alargado - por exemplo, ao nível de um país inteiro, embora isso seja raro -, costuma haver participação activa de profissionais da área da arquivística.
- Não existe qualquer intenção de impôr uma estrutura à Internet numa escala global. Os arquivistas não podem, por isso, depositar grandes esperanças em ver a sua vida facilitada no que toca à organização de conteúdos e normalização de descrições. Mesmo quando há um projecto de âmbito nacional, ele não costuma ultrapassar fronteiras.
- Infelizmente, e esta é a constatação mais triste, o grau de preparação dos arquivistas para estas questões é extremamente reduzido, sobretudo por falhas ao nível do conhecimento tecnológico.

A verdade nua e crua, pelo menos no caso do nosso país, é que não existe capacidade profissional por parte dos arquivistas para lidarem de forma profunda e sistemática com estas questões. Seja por deficiências no sector da formação, seja por limitações próprias dos meios em que os arquivistas se inserem (incompreensão das suas funções e utilidade, com a sua consequente desvalorização, insuficiência dos meios disponíveis), esta realidade está-lhes a escapar completamente da mão.

Como veremos nas alíneas seguintes, esta falta de preparação torna-se tanto mais grave quanto vão surgindo cada vez mais maneiras diferentes de partilhar e dar acesso à informação. Dizer que a Internet é uma gigantesca base de dados é uma coisa. Ver o que podem ser algumas bases de dados muito especiais na Internet é outra.

A INTERNET É ELA PRÓPRIA ARQUIVÁVEL?

Com ou sem falhas, a Internet já é o maior repositório de conhecimento da Humanidade. Neste sentido, já constitui um gigantesco arquivo de informação, práticas, artes e tecnologia. Há dados que só se encontram na Internet (importantes ou simplesmente curiosos), beneficiando das suas características de exposição universal imediata. E no entanto, até à data, têm havido muito poucas preocupações em arquivar a própria Internet, ou seja, preservar aquilo que, de alguma forma, já é um arquivo de conhecimento. Não havendo uma coordenação global, não é possível decretar a permanência dos dados. Hoje eu posso disponibilizar 40 Gb de documentos importantíssimos. Amanhã posso desligar o servidor onde eles se encontram e nunca mais o ligar. Quem me pode censurar, ou obrigar a ligá-lo novamente? Ninguém.

Associamos facilmente a palavra "arquivo" a noções como as de permanência e estabilidade. Muitas vezes, o conteúdo da Internet porta-se de maneira totalmente oposta - hoje está, amanhã já não. É extremamente difícil quantificar a Internet. No entanto, não é arriscado dizer que são criados por dia entre 4 e 8 milhões de páginas com informação (para a WWW). Ao mesmo tempo, deve desaparecer ou ser modificada uma quantidade idêntica...

Dada as dificuldades que um tal empreendimento levanta, arquivar a Internet não é um tema com grande receptividade. Tanto quanto nós saibamos, não há qualquer projecto centralizado para levar a cabo essa tarefa a um nível universal. Existem sim projectos parcelares, à escala de países, os quais servem perfeitamente como bancos de ensaio para o processo. Entre esses países contam-se a Austrália e o Reino Unido, bem como a Suécia. A França deve iniciar o seu projecto ainda este ano.

São possíveis várias abordagens metodológicas:

- Ou se arquivam todas as páginas produzidas num determinado país (abordagem sueca, visível em http://www.kb.se/ENG/Tjanst_fr.htm apenas em sueco) ou se opta logo à partida pela limitação (caso da Austrália, onde apenas é arquivado o conteúdo de *websites* seleccionados pelo seu grau de importância:

<http://pandora.nla.gov.au/pandora>)

- Quer num caso quer no outro, torna-se mais correcto falar em arquivo da informação num determinado estado e numa data específica. No caso sueco, por exemplo, o processo de recolha demora cerca de dois meses a ser concluído, sendo desencadeado apenas duas vezes por ano - não recursos ou tempo para mais. Ficam assim arquivadas "fotografias" do estado da Internet no tempo.

Para além destes esforços, de natureza institucional, podemos enquadrar nestas práticas de arquivo e preservação a consequência normal do funcionamento de determinados serviços da Internet (os servidores de *News*), ou até de motores de pesquisa e indexação como o Altavista ou o Google (pela forma como preservam e processam conteúdos textuais - basta aceder à página principal deste último serviço, em <http://www.google.com> para ver um anúncio sobre 1,326,920,000 de páginas *web* indexadas). No entanto, temos que frisar que não se tratam de serviços de arquivo, por não ser essa a sua função.

Como conclusão deste tópico, consideramos não só possível, como desejável arquivar o máximo possível do conteúdo actual da Internet. No caso português, não encontramos ainda qualquer entidade preocupada com o assunto (o qual, no entanto, chegou a ser publicamente discutido e vivamente sugerido pelo então deputado do PSD, Pacheco Pereira, numa coluna de opinião publicada na revista "Visão").

COMO COLOCAR INFORMAÇÃO ONLINE?

Falamos da Internet como um "depósito", ou um "arquivo", quando, na realidade, ela não é nem uma coisa nem outra. A Internet é a "rede das redes", uma rede informática de âmbito mundial, que interliga milhões de computadores e outras redes entre si. Sob este ponto de vista, deve ser encarada como um canal de comunicação de grande capacidade. A informação que consultamos através da Internet, encontra-se armazenada nos muitos computadores que a ela estão ligados e a ela acedem.

Quando dizemos que este ou aquele documento ou informação "está na Internet", estamos de facto a dizer que alguém se deu ao trabalho de o preparar e formatar, de arranjar espaço na rede para o disponibilizar, e de vigiar a sua integridade. Estas diferentes operações podem ser levadas a cabo de várias maneiras.

- A entidade responsável pelos dados pode ter os seus próprios meios de publicação na Internet (linha de comunicações dedicada, servidor *web*, pessoal responsável pela sua gestão, etc.)
- A mesma entidade pode optar por utilizar os recursos tecnológicos de outra empresa, que disponibilize as linhas de comunicação e os servidores. A maior parte dos fornecedores de acesso à Internet fornece este tipo de serviços. Esta constitui também a solução mais prática para particulares que pretendam publicar dados na Internet de uma forma coerente.
- Quem não quiser gastar muito dinheiro, pode optar por soluções tão baratas quanto possível,

boa parte das quais gratuitas. Sendo o espaço de armazenamento para publicação de dados na WWW um dos elementos essenciais para esta operação, existem diversas empresas que colocam à disposição de qualquer pessoa quantidade limitadas desse espaço, no qual é possível colocar páginas que depois ficam à consulta de todos.

Uma das vertentes mais recentes da disponibilização de informação através da Internet, consiste na sua utilização como depósito pessoal de dados. Neste caso, ela vai servir de facto como um depósito de informação - dizer que um documento está “guardado na Internet” passa assim a ser verdade. O ponto de partida é idêntico: uma empresa dá espaço de armazenamento a quem dele precise. Só que aqui, em vez de lá serem colocadas páginas para consulta via WWW, são lá colocados simples ficheiros, de acordo com as necessidades do utilizador. Desta forma, eles tornam-se acessíveis em qualquer lugar onde exista uma ligação à Internet.

Figura 2: Exemplos de dois websites que disponibilizam espaço de armazenamento na Internet de forma gratuita (<http://www.idrive.com> e <http://www.myspace.com/>).

No papel, esta parece ser uma solução muito aceitável. Quando foi apresentada pela primeira vez, chegou a ser apelidada de brilhante. Na prática sofre de muitas limitações. Lembramos que, na maior parte dos casos, o espaço de armazenamento é gratuito. Ora, as empresas

que fazem isto, têm que arranjar lucros desta actividade, seja com recurso a publicidade ou a outros meios. Ao mesmo tempo, procuram gastar tão pouco dinheiro quanto possível. Resultado: ninguém pode confiar a 100% na qualidade de serviço oferecida por estas empresas:

- Não há garantia de disponibilidade permanente. O objectivo de guardar dados desta maneira, é de os tornar disponíveis em qualquer local onde haja ligação à Internet. No entanto, nenhuma das empresas que dá espaço se compromete a ter os dados sempre disponíveis - no mundo real, isso implicaria gastar muito dinheiro (em pessoal, redundância de equipamento, segurança de dados), o qual dificilmente seria reembolsado.
- Não há garantia de segurança da integridade dos dados. Também aqui se aplicam as mesmas considerações que fizemos na alínea anterior.
- Não há garantia de rapidez de serviço, ou de fiabilidade nas transferências de informação. Lembramos que estas empresas trabalham de graça.
- Nesse caso, para que servem estas soluções? Quando muito, para armazenar dados de importância não vital temporariamente. Pedir mais do que isto é arriscado. Seguindo esta regra, encontramos imediatamente aproveitamentos inicialmente não previstos, sendo a pirataria informática um deles (para a distribuição de música, vídeo e *software* de forma ilegal)

COMO GUARDAR DADOS NA INTERNET?

Começámos esta comunicação a falar de ficção científica. É precisamente em várias histórias de ficção científica que encontramos uma noção radical, no que toca ao armazenamento de informação. É designada como “*data haven*”, sem tradução directa para português (um “porto seguro” para dados, em tradução livre). Consiste num depósito para dados de natureza digital, onde qualquer indivíduo ou entidade pode colocar a sua informação, mediante pagamento. Como o seu nome e tradução aproximada indicam, é data uma enorme importância à segurança e integridade desses dados. Ao mesmo tempo, é garantida sua inviolabilidade por terceiros e a resposta rápida aos verdadeiros donos dos dados. Num *data haven* não é feita qualquer distinção entre os dados - podem ser legais ou ilegais, ter qualquer dono e proveniência. Ao mesmo tempo, um *data haven* não responde a pedidos ou exigências da lei, devendo sempre manter uma rigorosa independência.

Esta é a teoria ou, pelo menos, as regras de funcionamento de um *data haven*, tal como surgem nas histórias de ficção científica. A realidade é, até ao momento, bem diferente. Analisando cada uma das características principais, constatamos o seguinte:

- Até que ponto é que um *data haven* conseguiria manter a inviolabilidade dos dados, não respondendo sequer perante a lei? Dificilmente o faria. Para funcionar bem, teria que trabalhar em zonas com recursos tecnológicos e de comunicações suficientemente avançados para o fim em vista. Essa necessidade limita grandemente os países onde poderia ser

instalado, países que dificilmente aceitariam tal comportamento.

- As necessidades de segurança e integridade dos dados podem ser abordadas e resolvidas, de um ponto de vista puramente tecnológico. Muito embora exijam investimentos avultados, não colocam desafios especiais.
- As necessidades de comunicações caem dentro da mesma situação. A alta velocidade exigida também exige investimentos avultados.
- Seria muito difícil garantir um “porto seguro” a todos os tipos de dados. A verdade é que temáticas como a pornografia infantil ou informação monetária sobre lavagens de dinheiro associadas ao tráfico de droga, não se prestam a ser impunemente escondidas e/ou armazenadas.

Com tudo isto, não duvidamos que haja um grande interesse na existência de *data havens*, com finalidade legais e ilegais. Uma empresa de alta tecnologia poderia ter vantagem em armazenar de forma segura informação sobre os seus produtos longe dos olhos da concorrência. Ao mesmo tempo, informação sobre fugas a impostos também encontraria o seu “porto seguro”.

Em meados do ano 2000, uma empresa americana chamada HavenCo (<http://www.havenco.com/>), anunciou a sua intenção de entrar neste negócio. Para tal iria constituir o seu *data haven* numa pseudo micro-nação do Mar do Norte, o autodenominado Principado de Sealand (uma antiga plataforma de defesa aérea da Segunda Guerra Mundial, ocupada por um excêntrico oficial inglês em finais da década de 60 do século passado, e que a seguir a declarou independente). Apostando numa localização “independente” mas próxima dos grandes nós de comunicações europeus, pretendiam fornecer espaço de armazenamento a todos os interessados, por bom preço. Garantiam a segurança completa dos dados (física e lógica), a sua elevada disponibilidade de um ponto de vista de comunicações e uma rigorosa independência perante investidas de governos ou forças de segurança.

Até que ponto cumpriram os seus objectivos e tiveram sucesso? Não sabemos. No *website* da empresa, praticamente a única fonte de informação sobre este assunto, a última informação disponível diz respeito ao 3º trimestre de 2000, altura em que informam os interessados que vão começar a fazer os testes iniciais de carga de servidores e de armazenamento de informação. Desde então para cá, mais nada.

Provavelmente, este é ainda um conceito demasiado radical para ter sucesso, não tanto pelas exigências tecnológicas que faz, mas sim pelas suas implicações legais e até políticas.

INTERNET E AS BASES DE DADOS VIRTUAIS

A conclusão lógica de tudo aquilo que temos vindo a dizer, surge como uma síntese perfeitas de vários conceitos. Hoje já podemos utilizar um tipo especial de bases de dados, que só podem existir porque a Internet existe.

Chamemos-lhe bases de dados “virtuais”: bases de dados que “não existem”, não podem ser apagadas e que ninguém (ou quase ninguém) sabe onde se encontram. No entanto, qualquer pessoa que tenha uma ligação à Internet e o *software* adequado (gratuito em 100% dos casos), pode utilizá-las e alimentá-las.

O processo funciona da seguinte maneira:

- É exigida uma ligação à Internet (permanente ou não) para participar e aceder à base de dados.
- Dependendo da natureza do *software* utilizado, pode existir uma coordenação central de conteúdos, com vista à sua indexação e para facilitar a pesquisa.
- Cada participante na base de dados instala o *software*, liga-se à Internet e pode de imediato aceder à base de dados.
- Em que consiste essa base de dados? Ela é formada por conteúdos que são fornecidos por cada pessoa que a ela se liga. Ou seja, quantos mais utilizadores estiverem ligados, mais conteúdo a base de dados tem.

Este processo é fantasticamente simples, mas o seu desenvolvimento e afirmação foi suficiente para lançar o pânico em muitos sectores de actividade, nomeadamente os que lidam com a publicação de conteúdos. O caso da empresa Napster é, talvez, um dos mais mediáticos e conhecidos. Essa empresa disponibiliza *software* com o mesmo nome, o qual, quando utilizado num PC ligado à Internet, permite partilhar ficheiros de música (em formato MP3, a maior parte das vezes resultantes de cópias ilegais de CDs) com quaisquer utilizadores do mesmo programa, que o tenham a correr no momento. Resultado: milhões de utilizadores deste programa, muitos milhões de faixas de música pirateadas desta maneira, grande descontentamento por parte da indústria discográfica e, finalmente, uma série de processos legais que levarão, provavelmente, ao encerramento da companhia.

Para demonstrar o grau de uso deste programa, mostramos algumas imagens resultantes de capturas de ecrã, obtidas no espaço de alguns dias em Novembro do ano 2000:

Currently 766,771 files (3,262 gigabytes)

2000/11/09 13:31

766.771 ficheiros de música partilhados
3 Tb de informação

Currently 957,358 files (4,055 gigabytes)

2000/11/09 14:52

957.358 ficheiros de música partilhados
4 Tb de informação

Currently 1,197,731 files (5,038 gigabytes)

2000/11/09 17:09

1.197.731 ficheiros de música partilhados
5 Tb de informação

Currently 1,426,962 files (6,025 gigabytes)

2000/11/13 16:08

1.426.962 ficheiros de música partilhados
6 Tb de informação

O caso Napster foi um catalisador destas tecnologias de partilha de informação, pelo menos do ponto de vista da sua exposição ao público. A empresa foi pioneira nesta área, tendo sofrido com a opção tecnológica que tomou, de ter um serviço de indexação centralizado, fixo, facilmente reconhecível e culpabilizável do ponto de vista legal.

Muito embora os usos principais desta tecnologia tenham a ver com a troca de dados obtidos de alguma forma ilegal (música e vídeos pirateados, programas copiados ilegalmente) e existam por isso várias razões para a fazer cair sob a alçada da lei, a verdade é que a sua divulgação e utilização já não podem ser paradas. E se o Napster é detectável, as principais variantes e alternativas não o são de forma alguma. A porta encontra-se, assim, completamente aberta à constituição de cada vez mais bases de dados virtuais imparáveis e indetectáveis, a fornecerem conteúdos cada vez mais diversificados e a pulverizarem, mais uma vez, noções como as de direitos de autor, protecção contra cópias, etc.

MAIS E MELHOR ESPAÇO DE ARMAZENAMENTO

Há vários anos encontrámos uma classificação de unidades de armazenamento de informação que especificava uma nova medida: o LOC, abreviatura de Library of Congress. Segundo o proponente, estimava-se em cerca de 96 Tb a quantidade de informação armazenada nesta Biblioteca. Assim, esta nova unidade, em que 1 LOC = 96 Tb, surgia como um novo paradigma de armazenamento.

Há apenas 10 anos, ter um disco de 100 Mb num PC era um luxo. Hoje já se torna difícil encontrar PCs à venda com discos com menos de 20 Gb (uma capacidade 200 vezes superior). A verdade é que hoje, repetindo o que já foi dito noutra parte desta comunicação, é possível armazenar cada vez mais informação em espaços cada vez mais reduzidos e por um preço cada vez mais baixo. E o melhor ainda está para vir.

As tecnologias mais radicais no que toca ao armazenamento de informação, as chamadas memórias moleculares, já prevêem que num espaço de tempo relativamente curto (entre 5 e 10 anos), seja possível disponibilizar centenas de Tbs de espaço de armazenamento em computadores pessoais, ocupando fisicamente um espaço irrisório. Cada utilizador poderia ter assim a totalidade da Biblioteca do Congresso e a totalidade da informação disponibilizada na Internet no seu computador.

CONCLUSÃO

Muito daquilo sobre o que falámos até agora - *data havens*, bases de dados virtuais, grandes capacidades de armazenamento - já existe, de uma forma ou de outra. Quando todas estas tecnologias se tornarem num lugar comum e olharmos para elas como hoje olhamos para um frigorífico, ou uma televisão - repare-se que não estamos a

dizer "se" -, qual é o papel que os arquivistas devem ou podem desempenhar? Até hoje, habituámo-nos a posicionar os arquivistas junto dos dados - em empresas e instituições que os criam, gerem e mantêm. Com as novas possibilidades de guardar e aceder a cada vez mais informação será que cada um tem que ser o seu próprio arquivista? Será que, finalmente, nos afogaremos em informação? Ou será que a própria profissão se torna obsoleta, por não estar a saber, já neste momento, agarrar as oportunidades que surjem e conhecer a tecnologia?

BIBLIOGRAFIA

“Archiver le Web”, *Science & Vie*, nº 1000, Janvier 2001, pp. 166-169

GARFINKEL, Simson, “The Ultimate Offshore Startup. Welcome to Sealand. Now bugger off”, *Wired*, July 2000, pp. 230-239

NAUGHTON, John, *A Brief History of the Future - The origins of the internet*, London, Weidenfeld & Nicolson, 1999

WEBSITES

<http://www.havenco.com/>

A primeira empresa que afirmou querer criar um data haven

<http://www.napster.com/>

A mais famosa das bases de dados virtuais da Internet

<http://www.idrive.com>

<http://www.myspace.com/>

Duas empresas que disponibilizam gratuitamente espaço de armazenamento da Internet

<http://www.netcraft.org>

<http://www.argreenhouse.com/netsizer>

<http://www.mit.edu/people/mkgray/net>

Três *websites* onde, de uma forma ou de outra, se encontra informação estatística sobre o estado e uso da Internet

<http://www.slashdot.org>

Um dos sites de discussão mais acedidos na Internet. Tudo o que disser respeito ao uso da Internet de formas tão radicais como as que aqui são descritas, começa por ser discutido neste site.

http://www.kb.se/ENG/Tjanst_fr.htm

Endereço da página da Biblioteca Real de Estocolmo onde se fala do projecto de arquivar *websites* da Suécia (página apenas em sueco)

<http://pandora.nla.gov.au/pandora>

Página do projecto Pandora, para arquivar *websites* da Austrália.

<http://www.google.com>

<http://www.altavista.com>

Dois dos principais motores de busca existentes na Internet.

