



Divulgação/Nasa



Braço robótico da sonda Viking 1, que pousou em Marte em maio de 1976: as imagens estão preservadas, mas os dados se perderam porque não há software capaz de ler o que foi produzido

PESQUISA

## O esforço de preservar o que nasceu efêmero

Nada é mais passageiro do que a internet. Mas bibliotecas e centros de estudos pelo mundo estão se esforçando para armazenar um pouco do conteúdo gerado dia a dia

Publicado em 03/05/2010 | [AGÊNCIA ESTADO](#)

Em 2001, o cientista Joseph Miller pediu à Nasa dados coletados pela sonda Viking em Marte, nos anos 70. A Nasa achou as fitas, mas os dados gravados não puderam ser abertos. O software que os lia não existia mais, e, como disse Miller na época à agência de notícias Reuters, os técnicos que conheciam o formato estavam todos mortos.

Essa é uma história. Há muitas outras. Parte do conhecimento produzido de maneira digital já era. De dados científicos a modinhas da internet. “Temos poucos serviços de preservação da história da cultura digital e muito conteúdo já se perdeu ao longo dos últimos anos”, diz Roberto Taddei, coordenador do Simpósio Internacional de Políticas Públicas para Acervos Digitais, que discutiu o tema em São Paulo, na semana passada.

Se um pergaminho pode ser desenrolado por qualquer pessoa, um cartão perfurado, bisavô do disquete, não se deixa abrir facilmente. Carlos Augusto Ditadi, da Câmara Técnica de Documentos Eletrônicos do Conselho Nacional de Arquivo (Conarq, que já cuida de parte do patrimônio digital do país), dá a medida da encrenca: “O disco depende do driver, que depende do computador, que depende do software, que depende do sistema operacional: isso se chama interdependência”. Some a essa equação o fato de computadores ficarem obsoletos, programas saírem de linha e linguagens caírem em desuso: o cartão perfurável fica tão indecifrável quanto hieróglifos egípcios.

A preocupação com o patrimônio digital é recente. Em 2002, foi apresentada pela Unesco a Carta pela

Preservação do Patrimônio Digital. Diz o documento: “Muitas dessas fontes têm valor e relevância duradouros e, assim, constituem um patrimônio a ser preservado”. A organização criou o órgão E-Heritage, dedicado, sobretudo, à conscientização de governos e à capacitação de arquivistas. É um bom começo, mas o patrimônio digital tem lá seus obstáculos específicos.

Eric Chan



Sala de leitura da Biblioteca do Congresso: preservação de sites selecionados

## Europa

### Portugal tenta guardar tudo

Portugal é um dos países em que toda a internet está sendo arquivada. “É tão importante guardar as publicações que estão na web quanto as impressas”, disse à reportagem Daniel Gomes, coordenador do Arquivo da Web Portuguesa. O projeto, capitaneado pela Fundação para a Computação Científica Nacional, já guardou 14 terabytes de dados em servidores próprios e em computadores de cidadãos comuns que doam pelo menos 100 MB de espaço para o projeto. Não há seleção – tudo que está sob domínio .pt é guardado. A maior dificuldade, porém, não é o espaço: é garantir uma busca eficiente. “As pessoas estão acostumadas ao Google. Temos que dar as respostas em poucos segundos, senão os utilizadores ficam frustrados”.

### Usuário, a solução

A interdependência é um deles. E, nesse caso, uma das melhores soluções veio de um jeito que é a cara da web: dos usuários. “A primeira geração de gamers percebeu, nos anos 90, que não tinha mais acesso a jogos da infância. Eles foram os primeiros a usar emuladores, que sempre existiram, como ferramentas de preservação. Graças a eles há emuladores para quase qualquer plataforma computacional”, diz Andreas Lange, diretor do Museu de Jogos de Computador, em Berlim, que tenta evitar o desaparecimento de games. O emulador é um programa que recria qualquer ambiente de computador: softwares extintos, consoles não mais fabricados etc.

Outro desafio evidente é o volume. Em 2009, de acordo com o Instituto de Pesquisas IDC, a humanidade produziu 750 bilhões de GB de informação. Como escolher o que preservar? “Não fazemos nenhuma seleção. Tentamos fazer o registro mais exaustivo. Arquivamos tudo o que encontramos sob o domínio .pt”, diz Daniel Gomes, coordenador do projeto Arquivo da Web Portuguesa. A declaração da Unesco sugere: “Os principais critérios devem ser significância e durabilidade (cultural, científica). Materiais ‘nativos digitais’ devem ter prioridade”.

Decidido o que guardar, falta definir como guardar e arrumar dinheiro para isso. Duas questões nada simples. Segundo Ditadi, o site é das coisas mais difíceis de preservar. “Ele deve permanecer navegável, mas como garantir os links? E eles levam a coisas protegidas por direitos autorais. É um registro muito dinâmico.” E o armazenamento custa caro. É preciso fazer uma cópia no formato nativo, chamada cópia de testemunho, que é a garantia de que aquele documento é real. Então, é feita a versão de preservação, em uma extensão mais duradoura – quase sempre um formato aberto, baseado em software livre. Daí, grava-se a cópia de acesso, aquela que fica disponível para consulta. Multiplique, portanto, tudo por três.

Por essas e outras, muitas vezes a memória da web é preservada justo por quem a alimenta. De novo, o usuário. Mas daí não há novidade. “Muitas bibliotecas foram montadas por usuários e depois doadas a instituições ou bibliotecas”, lembra Taddei.

### **Biblioteca já tem 147 terabytes arquivados**

A Biblioteca do Congresso dos Estados Unidos é a maior do mundo e há dez anos arquiva a internet. Já são 167 terabytes de dados (cada TB equivale a 1.000 gigabytes) – e a eles em breve se juntará todo o acervo do Twitter, adquirido no início do mês. Segundo a responsável pelo Projeto de Mídias Digitais da biblioteca, Abigail Grotke, o desafio maior é a seleção daquilo que deve ser preservado. Mas há ainda questões legais e de direito autorais. “Nos EUA, as instituições precisam dar permissão para arquivar, preservar e tornar o acervo acessível aos pesquisadores”, diz ela. Além disso, ferramentas especiais são necessárias para arquivar o conteúdo da web e esse é um campo relativamente novo.

Abigail sabe que outros países – como Portugal, Suécia, Noruega, Dinamarca e França decidiram preservar todo o conteúdo publicado em seus domínios. Nos Estados Unidos, entretanto, isso é mais complicado. “Adotamos uma abordagem seletiva: preservamos os sites selecionados pelos nossos especialistas que são relacionados a temas ou eventos, como as eleições ou o 11 de setembro. Também arquivamos sites produzidos pelo Congresso e os de interesse para nossa biblioteca legal”, explica.

E há o caso do Twitter, que ela classifica como parte de um registro histórico de comunicação, divulgação de notícias e tendências sociais. “É um arquivo direto de importantes eventos como a eleição americana de 2008 e a revolução no Irã”, analisa. “Individualmente, os tweets podem soar insignificantes. Coletivamente, são um recurso para futuras gerações entenderem a vida no século 21.”