



Crítérios de recolha da Web Portuguesa pelo Arquivo.pt

Ricardo Basílio^a

^aFCCN serviços digitais da FCT, Portugal, ricardo.basilio@fccn.pt

Resumo

Nesta apresentação descreve-se os critérios de recolha de conteúdos Web pelo Arquivo.pt, relacionando-os com a visão de respeito pela pluralidade, inclusão e cobertura de todas as manifestações da vida social em democracia.

Consideramos três aspetos: i) a Web como espaço de publicação sem limites; ii) critérios de recolha utilizados pelo Arquivo.pt; iii) três casos de recolhas especiais focadas na inclusão, atenção a uma comunidade de artistas e cobertura de conteúdos regionais.

Conclui-se, sugerindo a utilização do Arquivo.pt, o qual é uma estrutura pública para a preservação da Web nacional. É um serviço que pode ser incluído na estratégia de preservação de websites pelas organizações. Assim, espera-se colocar a audiência a par dos critérios de recolha e também das limitações. Não é possível arquivar a totalidade da Web.

Espera-se, no entanto, que todas as comunidades, minoritárias, periféricas, assim como nichos profissionais, em Portugal, possam ter as suas páginas web preservadas. pois assim podem contribuir ativamente para melhorar um histórico do Arquivo.pt.

Palavras-chave: Websites, Preservação digital, Memória institucional, Inclusão, Democracia.

Nesta apresentação descreve-se os critérios de recolha de conteúdos Web pelo Arquivo.pt¹, relacionando-os com a visão de respeito pela pluralidade, inclusão, abrangência de todas as manifestações da vida social em democracia.

Consideramos três aspetos: i) a Web como espaço de publicação sem limites; ii) critérios de recolha utilizados pelo Arquivo.pt; iii) referência a três casos de recolhas especiais focadas na inclusão, atenção a uma comunidade e cobertura de conteúdos regionais.

i) A Web como espaço de publicação sem limites. Diariamente, são publicados todo o tipo de conteúdos, tais como, informação institucional, serviços das empresas e da indústria, conteúdos a título individual, notícias dos media certificados, conteúdos de veracidade duvidosa, falsas notícias, conteúdos de entretenimento e também conteúdos impróprios e fraudulentos, alguns deles criminosos. Todos os conteúdos são resultado da vida social

¹ O Arquivo.pt é um serviço digital da FCT, através da sua unidade FCCN (outubro de 2025). Tem por missão “promover a preservação de conteúdos disponíveis na Internet nacional, garantindo a disponibilização deste à comunidade científica e ao público em geral” (Decreto-Lei nº 55/2013).

atual. É possível recolhê-los no arquivo da Web? Com que limites? Em outubro de 2025, o DNS.pt tinha ultrapassado os 700 mil nomes de domínio registados sob o .pt. Além desses, há muitos sites portugueses registados fora do domínio .pt (em domínios .com, .net, etc.).

ii) Critérios de recolha utilizados pelo Arquivo.pt. O primeiro critério é prioridade para os websites do domínio .PT e seus respetivos subsites ou subdomínios. O segundo, é a recolha dos websites portugueses em outros domínios, identificados e sugeridos pela comunidade. O terceiro, é a recolha de websites de interesse nacional, tais como websites de instituições e projetos da União Europeia, seleção de websites de Países de Língua Oficial Portuguesa. O quarto critério, são coleções temáticas baseadas numa seleção de sites no âmbito de projetos de colaboração com a comunidade. É neste âmbito que se privilegia a inclusão, a atenção a minorias ou cobertura a conteúdos regionais, por exemplo. No Arquivo.pt não se censura a recolha dos conteúdos, com exceção daqueles proibidos por lei (Decreto-Lei n.º 48/95) os quais são removidos. O quinto e último critério compreende as limitações das tecnologias de gravação da web. Nem tudo se consegue gravar, capturar no formato arquivo da Web, apesar de o conteúdo ser do maior interesse. Assim, há limites no tempo de gravação de uma página (timeout), no número de links a guardar em cada website (10.000) e

no tamanho do ficheiro (10 Mb). Outros limites são os formatos não suportados, a periodicidade com que se grava e a limitação aos conteúdos públicos, sem paywall ou login. Como resultado óbvio destes critérios, as páginas web arquivadas por vezes são representações fragmentadas do que foi publicado em determinada data na Web.

iii) Três casos de recolhas especiais focadas na inclusão e na cobertura de conteúdos regionais.

O primeiro caso é a coleção especial LGBT, a pedido de investigadores de uma universidade portuguesa. Os investigadores encontraram conteúdos úteis para o seu estudo no Arquivo.pt e forneceram informação sobre outros sites relevantes, que o Arquivo.pt gravou e acrescentou ao acervo.

O segundo caso é a lista de recolhas diárias, focado sobretudo conteúdos noticiosos. A lista foi incrementada em 2019 com websites de jornais regionais e de rádios locais que fornecem notícias sobre o dia-a-dia das localidades. Dessa forma, o Arquivo.pt procurou melhorar a cobertura territorial de conteúdos no seu acervo.

O terceiro caso foi o dos sites dos artistas e galeristas portugueses. Numa colaboração com a Biblioteca de Arte da Fundação Calouste Gulbenkian, uma investigadora identificou os artistas e galeristas portugueses e os respectivos websites. A partir de uma lista curada de websites foi possível melhorar a gravação e a representação dessa comunidade no acervo do Arquivo.pt.

De referir que as coleções especiais têm em vista a colaboração, um movimento nos dois sentidos. O Arquivo.pt dedica mais recursos para recolher uma coleção especial de websites e a comunidade interessada, assim se espera, passa a ser mais ativa em cuidar da preservação dos seus sites, alertando o Arquivo.pt, sugerindo novos websites, gravando por sua iniciativa diretamente no Arquivo.pt com o serviço ArchivePageNow (<https://arquivo.pt/archivepagenow>).

Para concluir, é fundamental para todas as pessoas com responsabilidade na gestão e edição de websites conhecerem os critérios e limites com que são preservados os conteúdos da Web portuguesa, pois assim podem contribuir ativamente para criar um histórico dos seus websites. Não é possível arquivar a totalidade da Web. Espera-se, no entanto, que todas as comunidades, minoritárias, periféricas e outros nichos profissionais, em

Portugal, possam ter as suas páginas web preservadas com qualidade no Arquivo.pt. Finalmente, espera-se que o Arquivo.pt enquanto estrutura nacional de preservação seja peça útil na estratégia de preservação da Web.