Ciclo de Webinars

Património cultural na *Web*:

presença *online* dos museus

Rede Portuguesa de Museus

Formação 2022

O Arquivo.pt e a preservação da memória digital

Bem publicar, para bem preservar

Arquivar a Web: faça-você-mesmo!

Iniciamos às 15h30

Arquivar a Web: faça-você-mesmo!

ricardo.basilio@fccn.pt

Objetivo

- Utilizar ferramentas do Webrecorder.net para **gravar páginas** Web no **próprio computador** num formato normalizado

Agenda

Parte I - Webrecorder.net – tutorial com o ArchiveWeb.page

Parte II - Porque utilizamos o Webrecorder.net

Parte III - Sobre o formato WARC

Parte IV - SavePage*Now* para gravar no Arquivo.pt

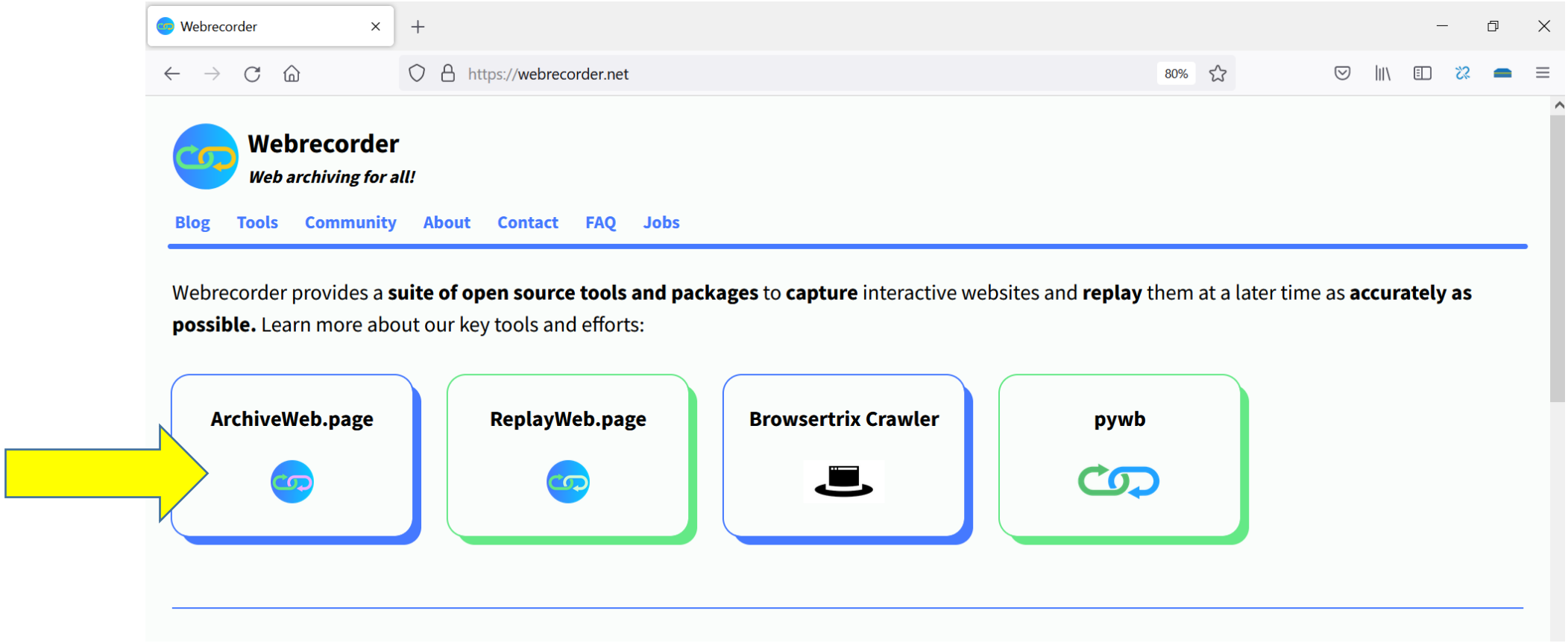
Parte V - Browsertrix para gravar site inteiro

Parte I

Webrecorder.net

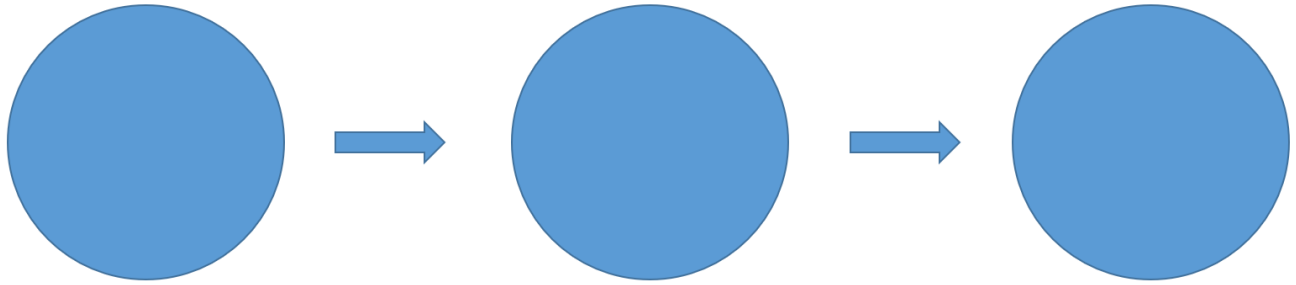
Tutorial com o ArchiveWeb.page

Ferramentas do Webrecorder.net



Webrecorder.net

3 fases da Preservação da Web



Gravar, Armazenar e Reproduzir são, de forma simplificada, as três fases da preservação (poderíamos falar de outras, tal como a seleção, controle de qualidade).
O ArchiveWeb.page cumpre as três fases: 1) grava 2) gera um WARC e/ou WACZ para armazenar em qualquer lugar 3) reproduz o conteúdo.
É isso que os arquivos da web também fazem. Por isso é muito adequado utilizar o ArchiveWeb.page

Tutorial – casos de uso do ArchiveWeb.page

Aceder aos vídeos demonstrativos

Poderá encontrar os vídeos descritos nos slides anteriores em

- https://youtu.be/mh6e5vsCivI

Instalar a extensão ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=337

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=337

Instalar a extensão ArchiveWeb.page

Demora um minuto.
Pré-requisitos: ter um destes browsers – Chrome, Edge, Opera, Brave. Não funciona no Safari e no Firefox.
Se usa sistema operativo iOS, instale primeiro o Edge e no Edge instale a extensão ArchiveWeb.page
Se usa Linux, instale o Chromium.
Há quem não goste de instalar extensões no browser. Para este tutorial vale a pena abrir uma exceção e instalar o ArchiveWeb.page. Remover é tão fácil como instalar.

Descrição do vídeo:

Neste vídeo demonstra-se como instalar o ArchiveWeb.page no browser Chrome:
1- encontrar o site https://webrecorder.net
2- destaque para a máxima “Archiving for all”; quatro ferramentas “open source”.
3- escolher o ArchiveWeb.page
4- “Install extension from Chrome Web Store” >> Adicionar
5- Verificar ícone do ArchiveWeb.page na barra do browser
6- Marcar o alfinete “pin” para manter o ícone sempre visível
7- Abrir no símbolo da “Homepage” ou “casinha” e pronto para começar

Para remover a extensão
1- ícone semelhante a peça de puzzle canto superior direito >> Clicar >> Manage extensions ou Gerir extensões
2- botão remover

Gravar uma página única

https://youtu.be/mh6e5vsCivI?t=528

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=527

Gravar uma página única.

Esta demo explica o processo de fazer um arquivo da Web local
- cria uma simples pasta no ambiente de trabalho “MeuArquivoDaWeb” para armazenar os ficheiros
- explica que o ArchiveWeb.page vai gravar páginas, vai exportar ficheiros para a pasta que criei, e que os ficheiros vão ser reproduzidos onde e quando se quiser pelo ReplayWeb.page ou pelo próprio ArchiveWeb.page

Descrição do vídeo

Esta demo explica o processo de iniciar um arquivo da Web local
1- cria uma simples pasta no ambiente de trabalho “MeuArquivoWeb” para armazenar os ficheiros
2- explica que o ArchiveWeb.page vai gravar páginas, vai exportar ficheiros para a pasta que criei, e que os ficheiros vão ser reproduzidos onde e quando se quiser pelo ReplayWeb.page ou pelo próprio ArchiveWeb.page
3- mostra o percurso que vamos adotar em todas as gravações: grava, verifica a gravação, faz download do WARC/WACZ

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=527

Gravar uma página única.

Gravar:
1- entramos no ArchiveWeb.page sempre pela homepage, ícone casinha
2- criamos uma pasta para a sessão de gravação, “demo-gravar-uma-pagina”
3- clicamos no botão azul para gravar
4- indicamos o endereço de partida a gravar (ex. https://exemplo.com) e ok
5- basta o conteúdo descarregar no browser; a página está gravada pelo ArchiveWeb.page
6- no canto superior-direito posso acompanhar a evolução da gravação e parar a gravação, stop

Verificar:
1- clico na casinha, vou à sessão de gravação
2- reproduzo a página e verifico se está bem gravada
3- experimento um clic num qualquer botão só para lembrar que só grava os links em que clicamos

Download:
1- clico na “casinha” e vou à sessão de gravação (este percurso é bom para não nos perdermos; depois cada pessoa usará o ArchiveWeb.page como entender
2- adquiro para o meu “MeuArquivoWeb” no meu ambiente de trabalho os ficheiros em WARC e WACZ

Diferença entre WARC e WACZ:

WARC é o formato standard que serve, por exemplo, para integrar em arquivos da Web como o Arquivo.pt

WACZ (Web Archive Collection Zipped) é um ficheiro comprimido ideal para usar com o ArchiveWeb.page. Dentro de um WACZ vamos encontrar um ficheiro sempre um ficheiro standard WARC e ainda material adicioanal, como por exemplo um índice dos URLs que ajuda a reprodução das páginas no ArchiveWeb.page.
É só isso.

Gravar várias páginas

https://youtu.be/mh6e5vsCivI?t=697

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=697

Gravar várias páginas

Descrição do vídeo:

Este vídeo segue os passos do anterior e reforça o procedimento para gravar páginas.

Em vez de uma pagina, vamos gravar três páginas, dando mais dois cliques

Assim, gravar:
1- abre o ArchiveWeb.page a partir da “casinha”,
2- cria uma pasta “Demo-gravar-várias-páginas”
3- inicia a gravação da página de partida http://www.museudearteantiga.pt/
4- dá mais 2 cliques. Por exemplo, em coleções e exposições
5- Controlar gravação no canto superior direito e parar.

Verificar gravação:
1- partir da “casinha” “home”
2- encontrar a sessão de gravação
3- verifica-se que foram gravadas 3 páginas: a página de partida e outras duas
4- Se eu for verificar outro link em que não cliquei, não está gravado

Download:
1- voltar à “casinha”, “homepage” (para reforçar a rotina de gravação nesta nova interface)
2- encontrar sessão de gravação
3- fazer download do WARC/WACZ – adquirir os conteúdos para o “MeuArquivoWeb”

Gravar muitas páginas

https://youtu.be/mh6e5vsCivI?t=832

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=832

Gravar muitas páginas

Descrição do vídeo:

Desta vez vamos gravar muitas páginas, não três, mas cinco e depois outras cinco, só para exemplificar.

Escolhemos como ponto de partida a página das coleções: http://www.museudearteantiga.pt/colecoes
Quantos links tem esta página? Muitos.

Como fazer para gravar muitas páginas de forma sistemática, progressiva e sem me perder na navegação

Gravação de cinco notícias:
1- Os passos são os mesmos do que nos vídeos anteriores: abre, cria sessão de gravação, grava
2- Usa esta página específica como ponto de partida: http://www.museudearteantiga.pt/colecoes
3- Em cada notícia, abrir novo separador, sucessivamente sem fechar nenhum
5- Gravar 5 notícias para este exemplo
6- O habitual… parar gravação, verificar

Continuar a gravação mais notícias, acrescentando-as a esta sessão:
Botão gravar e…
7- Novamente fornecer o endereço de partida http://www.museudearteantiga.pt/colecoes
8- Em cada notícia, abrir novo separador, sucessivamente sem fechar nenhum
9- Quando terminar, botão direito do rato num dos separadores abertos e fechar todas os separadores à direita
10- Verificar gravação e fazer download do WARC/WACZ

Este procedimento é bom para fazer gravações sistemáticas e de forma controlada em páginas com muitos links e páginas constantemente alimentadas com novos conteúdos, como por exemplo, as do arquivo de notícias num site.

Gravar redes sociais
Twitter

https://youtu.be/mh6e5vsCivI

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI

Gravar redes sociais – Twitter

Descrição do vídeo:

Processo como nos anteriores
Abre, cria sessão de gravação, gravar,

Neste caso, gravar a página do Twitter institucional, de fora para dentro, ou seja, sem estar "logado".
Endereço de partida https://twitter.com/mnaa_lisboa

A página do Twitter apresenta os conteúdos verticalmente, à medida que se faz scroll down.
O ArchiveWeb.page tem o botão "Auto pilot", piloto automático, que rola automaticamente a página e descarrega os conteúdos do post. Não segue links a apontar para fora do Twitter, a não ser que se faça manualmente (abrindo novo separador em cada link que cujo conteúdo queiramos gravar).

Sem login gravamos o conteúdo visível publicamente, evitando recolher outros dados, como por exemplo dados de perfil.
A ter em conta: ao gravar conteúdos sem fazer login podem surgir banner à frente do conteúdo, quando se grava e quando se reproduz.
A gravação sistemática seguindo este método "de fora para dentro", semanalmente, por exemplo, seria suficiente para ter uma boa representação das publicações.

O resto do processo é como nos exemplos anteriores: verificar a gravação, fazer download.

Gravar redes sociais Twitter **com login**

https://youtu.be/mh6e5vsCivI?t=1249

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=1249

Gravar redes sociais - Twitter com login

Descrição do vídeo:

Neste vídeo mostra-se o caso da gravação de uma página do Twitter com login.

Há alguns dados de perfil (foto, canais de que é seguidor, etc). Gravamos de dentro para dentro.
Não aparecem banners, como acontecia no caso anterior em que estávamos gravar “de fora para dentro”.

No caso de gravar “de dentro para dentro”, deve fazer login antes e começar a gravar depois.
Não faça login enquanto está a gravar, porque a sua interação e os seus dados de login podem ficar escritos no ficheiro WARC.

Ao fazer gravação “de dentro para dentro” deve considerar esses conteúdos não públicos. São portanto conteúdos não adequados para acesso público. Podem ser guardados à parte para consulta interna ou pessoal.

O que se refere ao Twitter com login também se aplica a outras redes sociais.

Gravar redes sociais Facebook

https://youtu.be/mh6e5vsCivI?t=1384

Tutorial - casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=1384

Gravar redes sociais – Facebook

Descrição do vídeo:

O procedimento é semelhante ao Twitter.
O endereço de partida é https://www.facebook.com/museunacionaldearteantiga/

O Facebook é gravável "de fora para dentro", seguindo o mesmo procedimento como no caso do Twitter. No entanto, a reprodução dos conteúdos gravados é muito limitada. Essa é uma dificuldade partilhada pela comunidade dos arquivos da Web. Todos apontam da dificuldade de gravar e reproduzir páginas do Facebook.

Ponderação da questão:
- o foco do Facebook não é a preservação digital
- alimenta-se do conteúdos e das interações dos utilizadores
- há alguns uns anos era possível ao dono da conta descarregar um ficheiro HTML com todos os posts na ordem em que foram publicados
- E atualmente? Como ter uma cópia da sequência de posts publicados? Como posso obter de volta uma cópia dos conteúdos no seu contexto (página do Facebook)?

Contudo, muitas instituições ou serviços no interior das instituições têm no Facebook o seu principal canal, por vezes publicando conteúdos em exclusivo no Facebook.
A prática adequada seria as instituições publicarem versões desses conteúdos em Websites ou em plataformas que se possam gravar.

Incentivamos os mais empreendedores a gravarem pequenas porções de conteúdos do Facebook com o ArchiveWeb.page, de modo a avaliarem o resultado e a decidirem o que fazer para assegurarem a sua preservação.

Gravar vídeos embebidos do Youtube

Tutorial – casos de uso do ArchiveWeb.page

Gravar vídeos embebidos do Youtube

Demo não incluída na apresentação

Os vídeos embebidos do Youtube são muito frequentes. No entanto, o Youtube é um serviço externo e tende a ser fechado. Em tempos permitiu a gravação dos vídeos através de um software específico (chamado DLL) que depois foi retirado pois estava a ser usado por outros serviços para difundir conteúdos, violando os direitos. Apesar de não visar os arquivos da Web, estes foram afetados por essa limitação.

O que acontece ao gravar o Youtube com o ArchiveWeb.page?
Recomenda-se remover da lista de extensões pré-instaladas no browser a do Youtube, se ainda lá estiver.
Verifica-se que a gravação corre bem, que o vídeo é gravado rapidamente.
No entanto, a sua reprodução posterior tende a não funcionar. Experimente fazer download e reproduzir com o ArchiveWeb.page, por exemplo.
Não funciona, “playback error”.

Vale a pena gravar vídeos do Youtube?
Na minha opinião sim, com critério, tendo em conta os recursos e o conteúdo em causa.
Se está gravado pode haver forma de reproduzir, ficando pelo menos essa possibilidade.

O contexto em que o vídeo foi embebido e publicado num site é uma informação complementar. Mesmo que o vídeo não esteja reprodutível, fica registado o seu endereço no Youtube. Assim, se ainda estiver online é possível encontrá-lo usando esse metadado.
Tal como para outras redes sociais, é recomendável que os vídeos publicados no Youtube sejam preservados em outros lugares e com outra estratégia.

Animações 3D

https://youtu.be/mh6e5vsCivI?t=1539

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=1539

Gravar animações 3D e outras formas de apresentar conteúdo. É possível?

Descrição do vídeo:

Este vídeo mostra que podemos tentar gravar outras formas de apresentar conteúdos nos site para ver se são graváveis com o ArchiveWeb.page.

Umas vezes resulta outras não.

Gravamos, por exemplo, http://www.mudasmuseuvirtual.com/ que apresenta animação 3D. Neste caso correu bem e consegue-se reproduzir, ainda que com alguns erros.

Em outros casos pode não ser possível gravar e reproduzir.

Cabe às instituições terem em consideração a questão da preservação e balancear a publicação de conteúdos em certos formatos. Há outras formas de preservação, além do formato Web que podem resolver. O que importa em último caso, é que seja feita a preservação.

Gravar conteúdos atrás de formulários

https://youtu.be/mh6e5vsCivI?t=1774

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=1774

Gravar conteúdos atrás de formulários

Descrição do vídeo:

Procedimento semelhante, abreviados alguns passos.

Os arquivos da Web como o Arquivo.pt não gravam conteúdos atrás de formulário.
É o caso de sites em que temos que introduzir uma palavra ou uma expressão de pesquisa para obter resultados.

Com o ArchiveWeb.page é possível gravar conteúdos atrás de formulário, pois o processo é manual.

Neste exemplo, utilizamos o MatrizNet http://www.matriznet.dgpc.pt/matriznet/home.aspx e pesquisamos "berlinda". Selecionámos o registo e tudo isto ficou gravado.
Este processo é útil, por exemplo, se quisermos guardar no formato arquivo da Web (WARC) um conjunto de registos tal como estavam publicados numa determinada data.

Gravar sites com dados geográficos

https://youtu.be/mh6e5vsCivI?t=1884

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=1884

Gravar sites com dados geográficos

Descrição do vídeo:

Neste vídeo mostra-se o que acontece quando tentamos gravar partes de um website que contém dados geográficos embebidos. Por exemplo, mapas de localização do Google e de outros sistemas.

Utilizámos o exemplo do portal de dados abertos da Câmara Municipal de Lisboa, numa pesquisa sobre museus.
Como resultado, verificamos que dados geográficos embebidos dificilmente são graváveis.
No entanto, neste portal há um link alternativo para os mesmos dados numa tabela que se pode gravar e reproduzir.
É boa prática fornecer alternativas aos utilizadores para acederem aos dados.

Reproduzir os conteúdos gravados

https://youtu.be/mh6e5vsCivI?t=2014

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=2014

Reproduzir os conteúdos gravados em contexto totalmente local

Descrição do vídeo:

Como resultado das gravações feitas nos casos que mostrámos neste tutorial, a pasta “MeuArquivoWeb” contém agora vários ficheiros.

Recordamos a diferença entre os ficheiros WARC e os ficheiros WACZ:
- WARC são os ficheiros standard, iguais aos utilizados pelo Arquivo.pt ou o Internet Archive
- WACZ (Web Archive Collection Zipped) são ficheiros que têm dentro o ficheiro WARC e ainda um índice com os endereços gravados, o que é muito útil para serem reproduzidos com o ReplayWeb.page ou com o ArchiveWeb.page.

Para reproduzir conteúdos podemos usar o ReplayWeb.page – https://replayweb.page
Este serviço pode ser usado por qualquer pessoa em qualquer lugar. Se eu enviar ficheiros no formato arquivo da Web WARC ou WACZ a alguém, essa pessoa pode reproduzir os ficheiros sem ter que instalar qualquer aplicação. Basta usar o – https://replayweb.page

Importamos ou carregamos para lá o ficheiro que queremos reproduzir e já está.

Vamos demonstrar como reproduzir os conteúdos gravados em contexto totalmente local:
Para isso é necessário:
1- ir a https://webrecorder.net
2- instalar a versão ArchiveWeb.page Desktop App – é um ficheiro executável, escolher a última versão
3- abrir o ArchiveWeb.page Desktop App – verifica que é semelhante à extensão de browser que usámos para gravar
4- importar ou carregar o ficheiro que quer reproduzir (de preferência o ficheiro WACZ)
5- reproduzir

Reproduzir os conteúdos gravados em contexto completamente local

https://youtu.be/mh6e5vsCivI?t=2124

Tutorial – casos de uso do ArchiveWeb.page

https://youtu.be/mh6e5vsCivI?t=2124

Reproduzir os conteúdos gravados em contexto totalmente local

Há casos em que é útil reproduzir os ficheiros de arquivo da Web WARC ou WACZ em contexto completamente local, no meu próprio computador, sem depender de serviços externos e sem haver saída de informação para fora.

Nesse caso, basta instalar a versão Desktop App do ArchiveWeb.page e usá-la para ler os ficheiros do "MeuArquivoWeb".
Para isso é necessário:
1- ir a https://webrecorder.net
2- instalar a versão ArchiveWeb.page Desktop App – é um ficheiro executável, escolher a última versão
3- abrir o ArchiveWeb.page Desktop App – verifica que é semelhante à extensão de browser que usámos para gravar
4- importar ou carregar o ficheiro que quer reproduzir (de preferência o ficheiro WACZ)
5- reproduzir

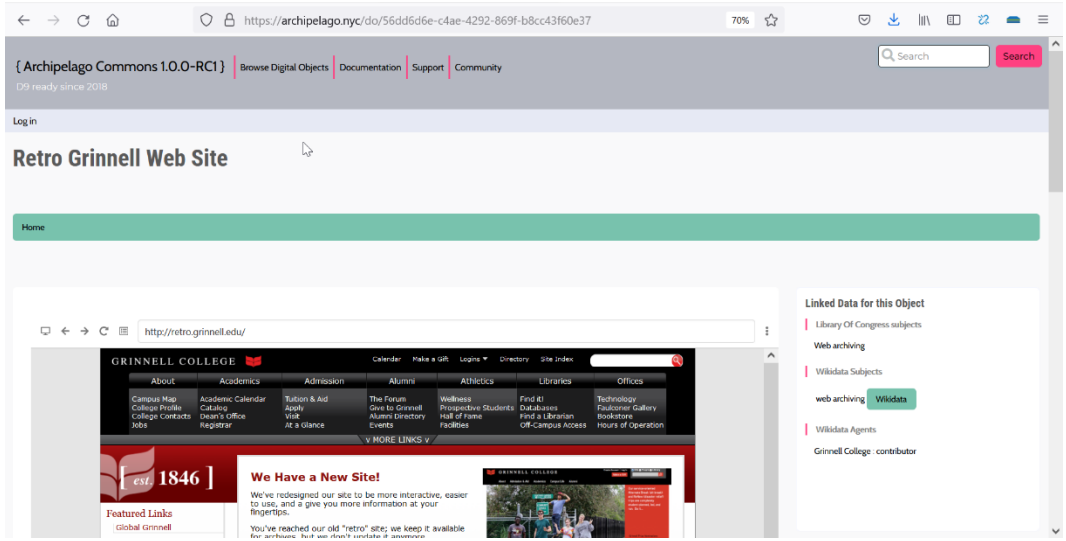
É altura ainda de ver o maior ou menor sucesso das nossas gravações. Há conteúdos cuja reprodução não funciona. Essas limitações uma vez identificadas são importantes para decidir o que fazer ou deixar de fazer.

Recorde-se que os conteúdos quando são vistos por nós num browser ou navegador estão num formato não standard (cada browser tem o seu). Quando retiramos os conteúdos gravando-os, exportando-os em formato WARC ou WACZ há conteúdos que deixam de funcionar.
Não podemos esquecer que a preservação digital é o "jogo do gato e do rato" das tecnologias. Os sites usam formas novas de apresentar conteúdos e os arquivos da Web, posteriormente, reagem aperfeiçoando as ferramentas de gravação.

A preservação deve ser feita colaborativamente, de forma complementar, entre instituições diversas, com estratégias e técnicas diversas.

Reproduzir sites preservados num repositório digital ou no meu site

Integração do ArchiveWeb.page num repositório digital



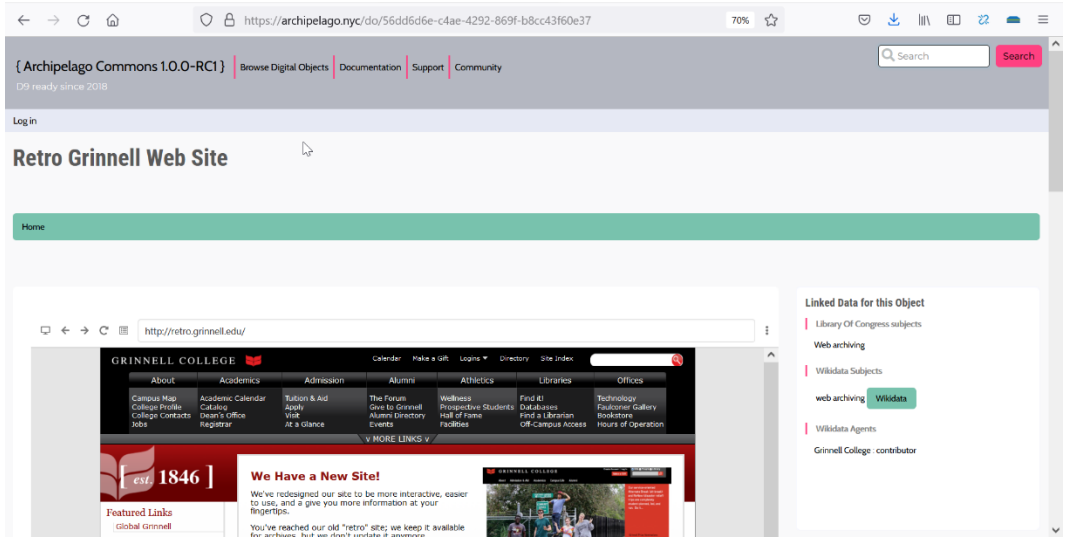
https://archipelago.nyc/

Este slide pretende mostrar que é possível incluir os novos formatos de arquivo da Web (WARC) em repositórios e bibliotecas digitais. Para isso mostramos exemplos de projetos nesse sentido.

Os websites consideram-se novos formatos, nascidos digitais. No entanto, não têm ainda lugar nos repositórios digitais, bibliotecas digitais e outros serviços semelhantes. Estes disponibilizam simultaneamente PDFs, vídeo, áudio, jogos, modelos 3D, etc, O mesmo não acontece com ficheiros WARC.

Para colmatar essa lacuna começam a surgir casos de integração do ReplayWeb.page em repositórios.
O ReplayWeb.page é o software do projeto Webrecorder.net que reproduz os ficheiros do tipo WARC e WACZ, como já referimos.

Integração do ArchiveWeb.page num repositório digital



https://archipelago.nyc/

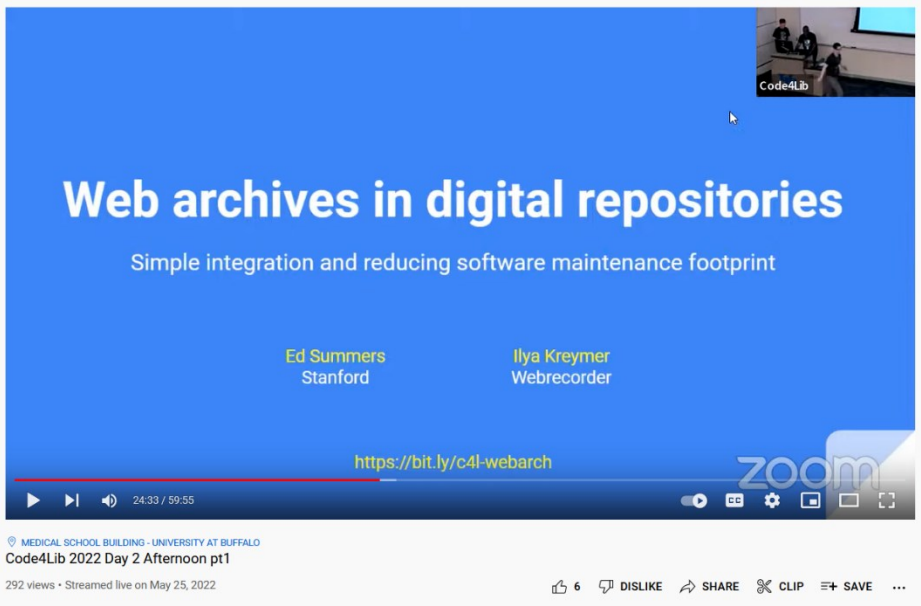
Eis um exemplo de integração: https://archipelago.nyc/ Neste repositório digital é possível aceder a sites gravados com o ArchiveWeb.page, do mesmo modo que é possível aceder a outros formatos mais habituais (PDF, JPEG, etc)

Diferentes objetos são descritos com metadados que lhes são próprios (costumizados), o mesmo acontecendo com os websites gravados. A reprodução das páginas Web gravadas é apresentada numa frame no contexto do repositório.

O grande benefício de projetos deste género é trazer para a agenda das instituições a preservação dos seus sites e de lhes reconhecer um valor cultural e patrimonial que atualmente não têm.

A integração do ArchiveWeb.page em websites está documentada no site Webrecorder.net e é relativamente simples.

Integração do ArchiveWeb.page num repositório digital



https://youtu.be/dtd5Os5t0lo

Este slide é uma referência à apresentação de Ed Summers (Univ. Stanford) e Ilya Kreymer (Webrecorder) sobre a integração do ArchiveWeb.page.

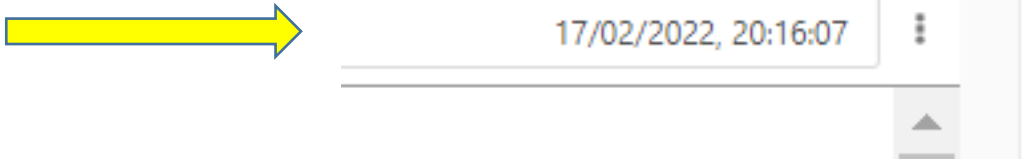
É apresentação útil para quem quiser incluir páginas preservadas num repositório institucional.

Cabe também aos desenvolvedores de produtos de repositórios e biblioteca digitais a tarefa de integrarem estas novas ferramentas.

Se for possível reproduzir um ficheiro WARC ou WACZ num repositório como se reproduz um PDF, muitas instituições passarão a a fazer arquivo dos seus websites ou de páginas desses sites.

Quando isso acontecer será um grande avanço para a preservação de conteúdos Web.

Integração do ArchiveWeb.page num repositório digital



Página gravada do site das comemorações dos 500 anos da Circum-navegação integrada num site Wordpress: https://webcurator.ddns.net/?p=291

Testámos algo mais simples num site wordpress - https://webcurator.ddns.net/?p=291

Uma página gravada pode ser disponibilizada da mesma forma que outros conteúdos embebidos em websites.

Documentação em https://replayweb.page/docs/embedding

Parte II

Porque utilizamos o Webrecorder.net

O Webrecorder

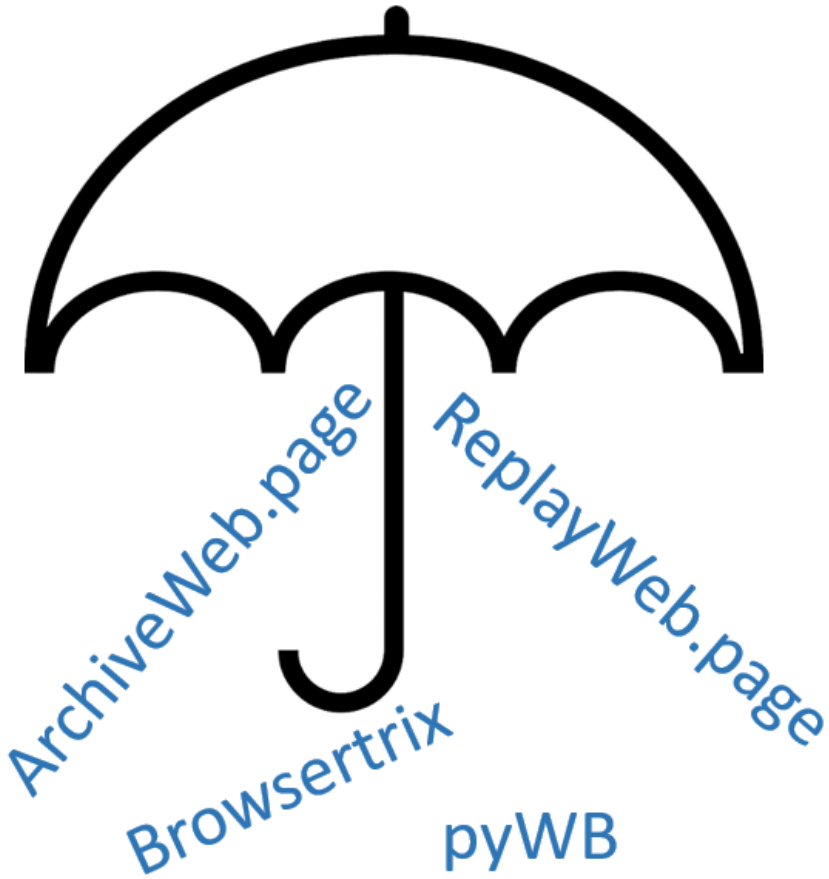
“Web archiving for all”

- 2016 - Parte do Projeto Rhizome.org (Webrecorder.io)
- 2019 - Independente do Projeto Rhizome, dá suporte ao serviço Conifer
- 2020 - Webrecorder.net

Canal Twitter: @webrecorder_io

O Webrecorder.net

- Open source
- Centrado no utilizador não especializado
- Formatos normalizados
- Compatibilidade com arquivos da Web
- Apoio do International Internet Preservation Consortium (IIPC)



O Webrecorder.net

Partners

Here are some of the partners that we work with or have collaborated with in the past:

Perma.cc

LOCKSS

RHIZOME

THE NATIONAL ARCHIVES

hypothes.is

IIPC INTERNATIONAL INTERNET PRESERVATION CONSORTIUM netpreserve.org

CLOCKSS

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

NATIONAL LIBRARY OF AUSTRALIA

KIWIX

ONF NFB

UKWA UK WEB ARCHIVE

DocNow

ARQUIVO.PT

STANFORD UNIVERSITY PRESS

If you're an institution using our tools, let us know and we can add your logo here.

O Webrecorder.net

Este slide repete o que apresentamos no início para concluir que:

- As ferramentas do Webrecorder.net que utilizámos, mais propriamente o ArchiveWeb.page dão-nos a possibilidade de cumprir as três fases da preservação: gravar, armazenar e reproduzir,
- Tudo isso formato normalizado WARC,
- Podendo ser reproduzido através de software de arquivos da Web geralmente conhecido por Wayback

Parte III

Sobre o formato WARC

Formato WARC

- ISO 28500:2017
- Independente de plataforma
- Reproduzido por software específico (geralmente designado *wayback*)

WARC e WACZ

WARC – standard para ser lido por qualquer arquivo da Web

WACZ – criado pelo Webrecorder.net para ser lido por ferramentas específicas como o ArchiveWeb.page ou o ReplayWeb.page

WARC e WACZ

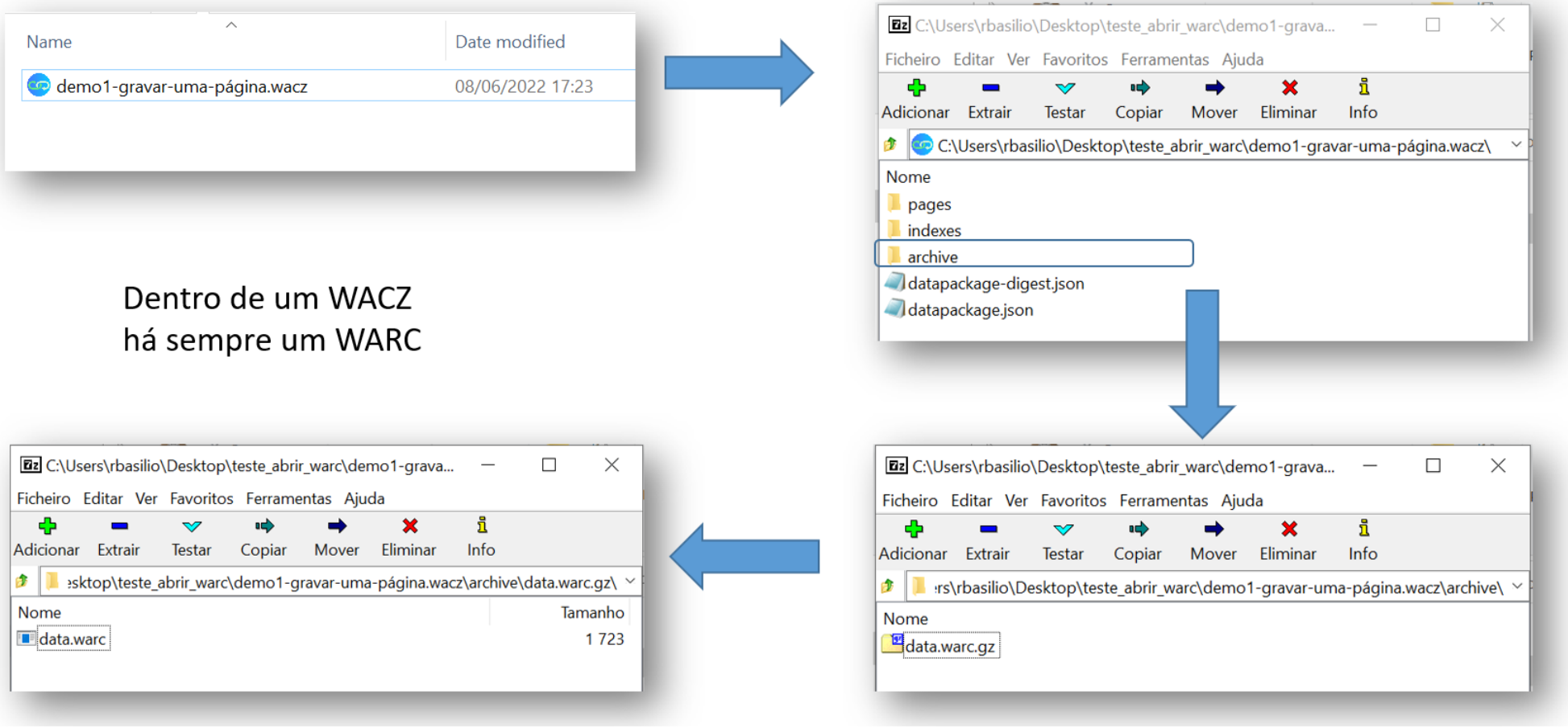
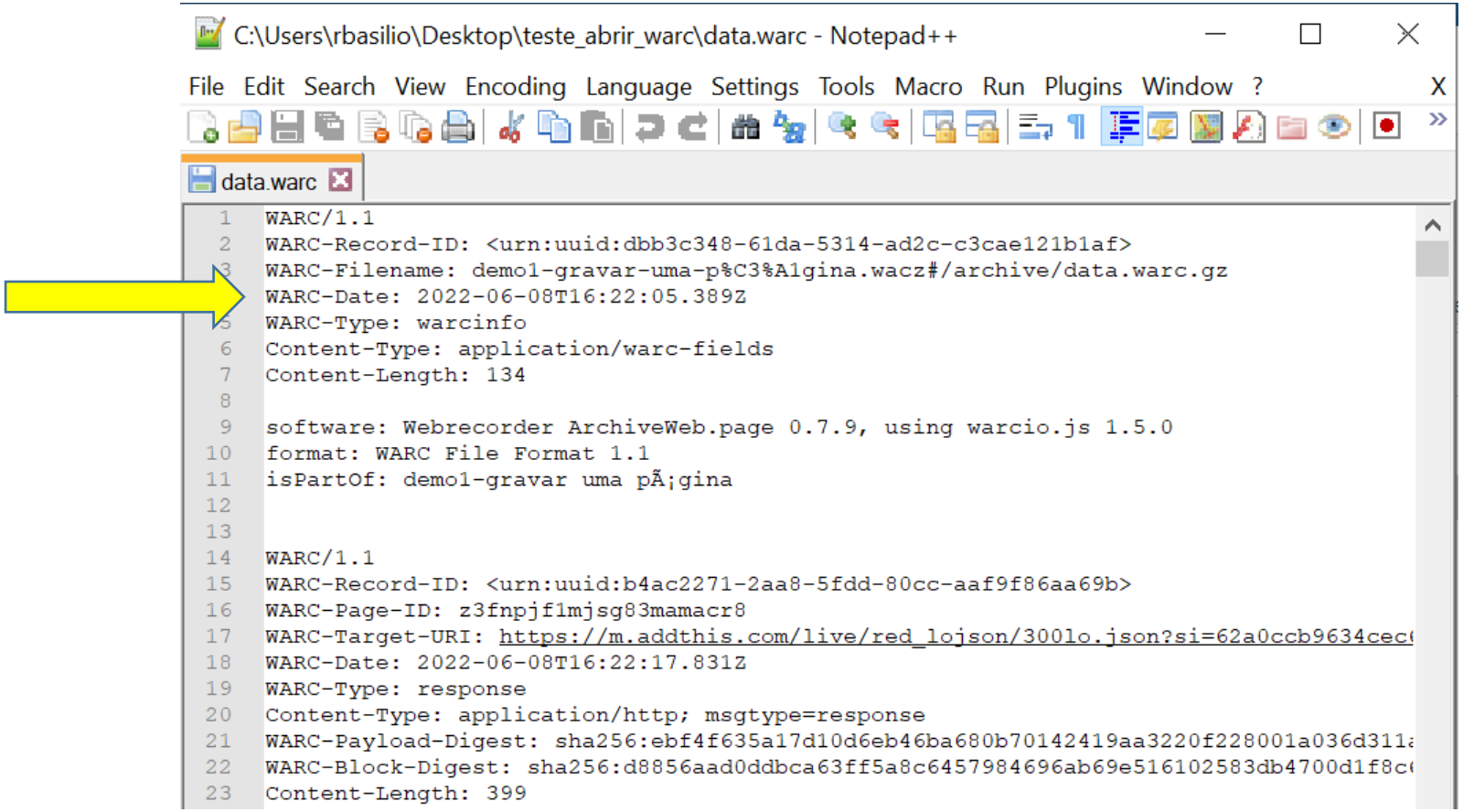


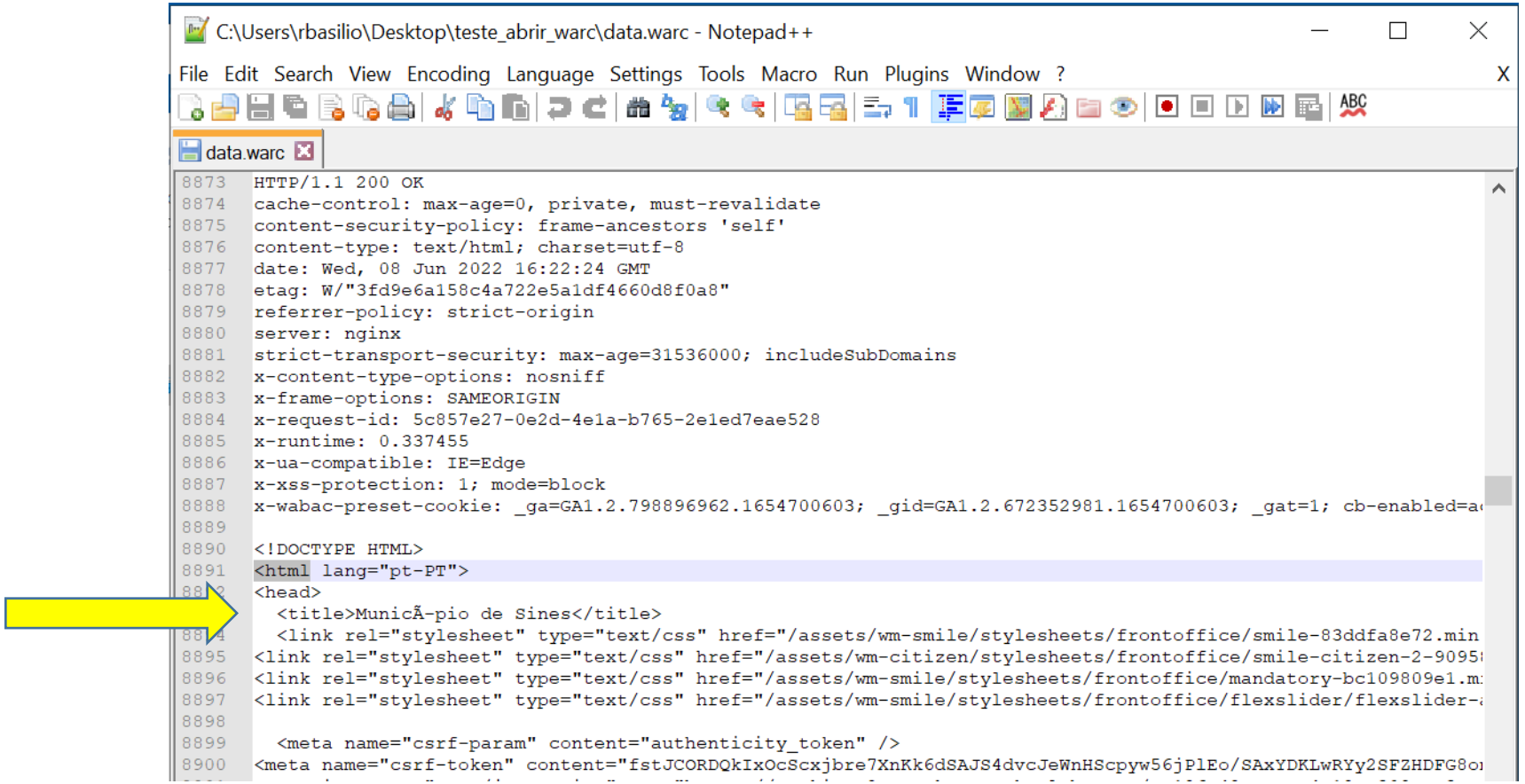
Figura que demonstra como pode encontrar o ficheiro WARC dentro de um WACZ que o ArchiveWeb.page exporta

WARC e WACZ



Ficheiro WARC aberto com NotePad++ onde se pode diversos sobre a gravação (nome do ficheiro, data, versão do software que gravou, etc)

WARC e WACZ



Ficheiro WARC aberto com NotePad++, gravação do site da Câmara Municipal de Sines com o ArchiveWeb.page, onde se pode ver por exemplo o código HTML

Parte IV

SavePage*Now* para gravar no Arquivo.pt

Parte IV – SavePage*Now* para gravar no Arquivo.pt

Depois de termos apresentado o ArchiveWeb.page, ferramenta do Webrecorder.net que permite iniciar um arquivo da Web local, vamos agora demostrar o SavePageI*Now*.

O SavePageI*Now* é um serviço de gravação de páginas Web na hora que o Arquivo.pt criou para qualquer pessoa poder usar, em qualquer local sem ter que instalar nenhuma aplicação. Qual a relação com o ArchiveWeb.page? Usa o mesmo software e grava de modo semelhante.

O SavePageI*Now*, podemos dizer desta forma, é como o ArchiveWeb.page mas do lado Arquivo.pt. O que gravarmos fica imediatamente arquivado e, em 48 horas, está disponível no Arquivo.pt.

Funciona desta forma:
- Encontrou uma página que quer gravar? Copie o link
- Abra o SavePageNow (Grave páginas) do Arquivo.pt – Encontra-o no menu lado direito ou em arquivo.pt/savepagenow
- Cole o link e comece a gravar como fez no ArchiveWeb.page
- E pronto. Os conteúdos são imediatamente gravados no Arquivo.pt e disponibilizados em 48 horas.

Atenção! Nem sempre os conteúdos de uma página Web que estamos a ver ficam gravados. O streaming de vídeos, por exemplo, foi limitado no SavePageNow para evitar carregar o sistema.

Parte IV – SavePage*Now* para gravar no Arquivo.pt

Descrição do vídeo:

Neste vídeo mostramos exemplificamos o uso SavePageI*Now*

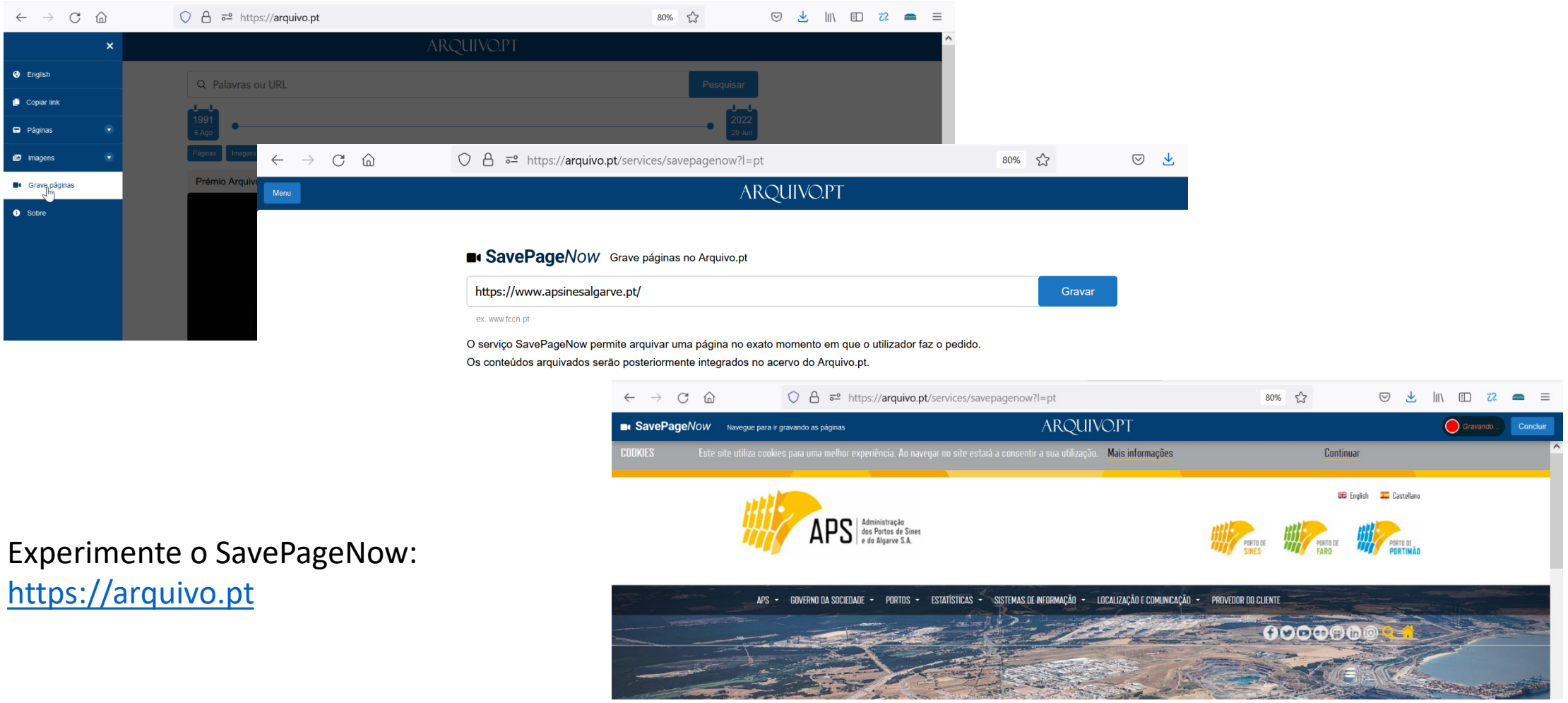
- Copiei o link http://www.museudearteantiga.pt
- Site do Arquivo.pt >> Menu lado direito SavePageNow ou Grave páginas
- Colei o link e comecei a “Gravar”
- Sigo as páginas ou os links que quero gravar, tal como no ArchiveWeb.page

- Para gravar de forma controlada, sem me perder na navegação, sigo links abrindo novo separador
- E pronto. Os conteúdos são imediatamente gravados no Arquivo.pt e disponibilizados no Arquivo.pt (pode demorar alguns dias)

Atenção! Nem sempre os conteúdos de uma página Web que estamos a ver ficam gravados. O streaming de vídeos por exemplo foi limitado no SavePageNow para evitar carregar o sistema.

Para gravar vídeos use o ArchiveWeb.page

SavePage*Now* para gravar no Arquivo.pt



Experimente o SavePageNow:
https://arquivo.pt

No seu computador ou smartphone crie um atalho para

arquivo.pt/savepagenow

Parte V

Browsertrix

para gravar site inteiro

https://youtu.be/7JNoMo302NA

Browsertrix

- Web crawler
- Baseado em Browser
- Vários perfis de recolha
  - Single page
  - Todos os links de um domínio
  - Profundidade personalizada
- Fácil de implementar para recolhas em pequena escala

Browsertrix

Parte V - Browsertrix para gravar site inteiro - introdução

O Browsertrix serve para gravar um site web inteiro de forma automática. É uma ferramenta do Webrecorder.net.

Para gravar sites com milhares de páginas o ArchiveWeb.page é pouco prático. Teríamos de clicar em todos os links e esperar que todos os conteúdos fossem descarregados. Além de demorado, certamente ficariam muitos conteúdos esquecidos.

O Browsertrix é um crawler e por isso uma das suas funcionalidades é descobrir todos os links existentes num site e proceder à sua gravação, se assim quisermos.

Browsertrix

Parte V - Browsertrix para gravar site inteiro – caraterísticas, comentário e descrição do vídeo

Dizemos que a forma de gravar a Web é *browser based* ou seja baseada num browser, diferentemente da forma mais antiga de gravar websites na qual apenas se descarregavam ficheiros e em que não era possível gravar certas interações e dinamismos que só acontecem quando se acede a um site num browser.

Destacamos os perfis ou modos de gravação que podemos personalizar. Ou seja,

- Posso gravar apenas a página cujo endereço eu indicar – por vezes é muito útil
- Posso gravar todos os links do mesmo domínio, que corresponde frequentemente a um website
- Posso definir a profundidade, indicando por exemplo, que quero gravar uma página e seguir os links lá existentes apenas uma vez

A implementação é fácil para treino e aprendizagem.
Para usar em produção num arquivo da Web local recomenda-se vivamente o envolvimento dos técnicos informáticos.

Browsertrix

Parte V - Browsertrix para gravar site inteiro – caraterísticas, comentário e descrição do vídeo

Descrição do vídeo "Recolha de website utilizando o Browsertrix (6 min):

Este vídeo exemplifica o uso do Browsertrix, neste caso, ainda numa versão antiga (2018)
Ver no Youtube com narração: https://youtu.be/7JNoMo302NA

O Browsertrix foi instalado num simples computador portátil para treino e aprendizagem.
O site da Câmara Municipal de Lisboa foi o exemplo utilizado na gravação:
Guia para não especialistas: https://tinyurl.com/instalar-browsertrix

Presentemente, há uma nova versão para Desktop e um serviço experimental em Cloud:
Documentação oficial: https://github.com/webrecorder/browsertrix

Conclusão

- Grave conteúdos em formato normalizado
- Reutilize e analise em contexto local
- Contribua (juntamente com os arquivos da Web) para a preservação da memória histórica

Obrigado.

contacto@arquivo.pt

arquivo.pt/inscrever

FCT Fundação para a Ciência e a Tecnologia

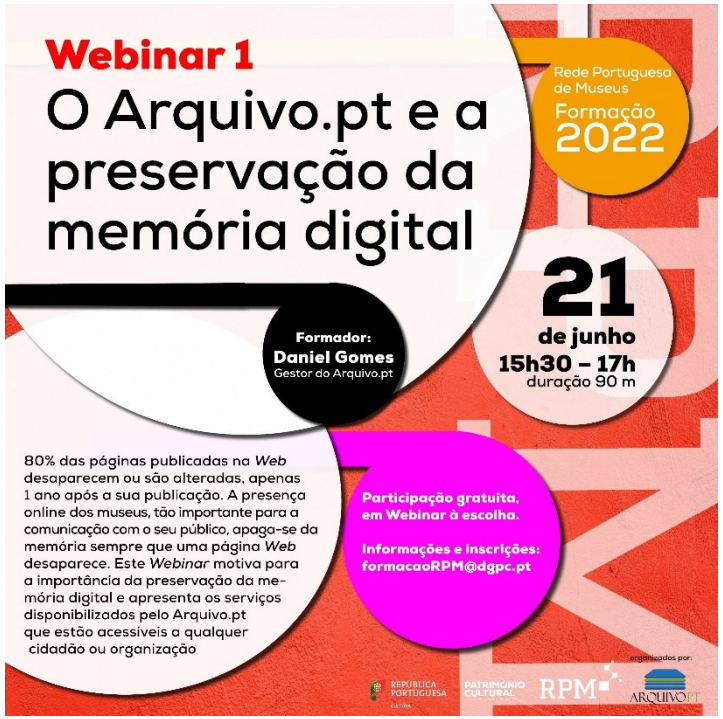
Obrigado.

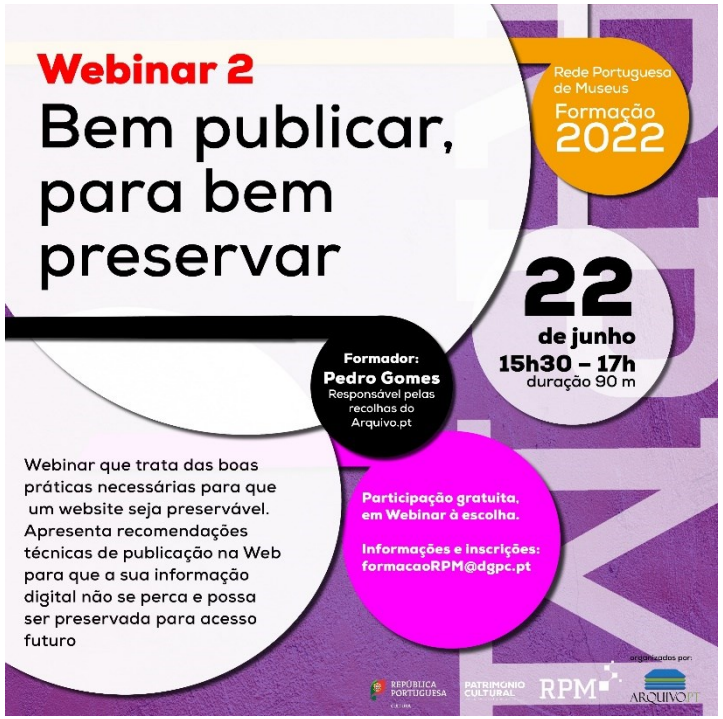
contacto@arquivo.pt

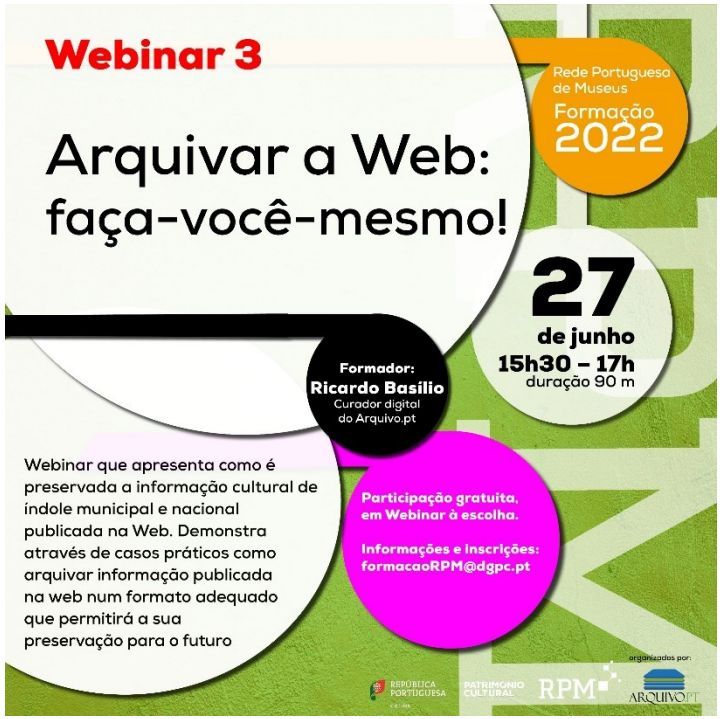
arquivo.pt/inscrever

ibict Instituto Brasileiro de Informação em Ciência e Tecnologia

Cariniana Rede Brasileira de Serviços de Preservação Digital


Webinar 1
O Arquivo.pt e a preservação da memória digital
Rede Portuguesa de Museus
Formação 2022
21 de junho
15h30 – 17h
duração 90 m
Formador:
Daniel Gomes
Gestor do Arquivo.pt
80% das páginas publicadas na Web desaparecem ou são alteradas, apenas 1 ano após a sua publicação. A presença online dos museus, tão importante para a comunicação com o seu público, apaga-se da memória sempre que uma página Web desaparece. Este Webinar motiva para a importância da preservação da memória digital e apresenta os serviços disponibilizados pelo Arquivo.pt que estão acessíveis a qualquer cidadão ou organização
Participação gratuita, em Webinar à escolha.
Informações e inscrições:
formacaoRPM@dgpc.pt
organizados por:


Webinar 2
Bem publicar, para bem preservar
Rede Portuguesa de Museus
Formação 2022
22 de junho
15h30 – 17h
duração 90 m
Formador:
Pedro Gomes
Responsável pelas recolhas do Arquivo.pt
Webinar que trata das boas práticas necessárias para que um website seja preservável. Apresenta recomendações técnicas de publicação na Web para que a sua informação digital não se perca e possa ser preservada para acesso futuro
Participação gratuita, em Webinar à escolha.
Informações e inscrições:
formacaoRPM@dgpc.pt
organizados por:


Webinar 3
Arquivar a Web: faça-você-mesmo!
Rede Portuguesa de Museus
Formação 2022
27 de junho
15h30 – 17h
duração 90 m
Formador:
Ricardo Basílio
Curador digital do Arquivo.pt
Webinar que apresenta como é preservada a informação cultural de índole municipal e nacional publicada na Web. Demonstra através de casos práticos como arquivar informação publicada na web num formato adequado que permitirá a sua preservação para o futuro
Participação gratuita, em Webinar à escolha.
Informações e inscrições:
formacaoRPM@dgpc.pt
organizados por: