





"PRESERVAÇÃO DA INFORMAÇÃO DIGITAL: FERRAMENTAS E PRÁTICAS"

7 MAIO | 14H30
ANFITEATRO III
[PISO 4, FLUC]



Mediante inscrição:
<https://ls.uc.pt/index.php/324891>

Organização:
Secção de Informação

SIGA A FLUC
www.uc.pt/fluc
[@letrasfluc](https://twitter.com/letrasfluc)

Arquivar a Web: faça-você-mesmo

Ricardo Basílio

Curador digital do Arquivo.pt, FCCN-FCT

FCCN serviços
digitais
fct

fct Fundação
para a Ciência
e a Tecnologia



Objetivo

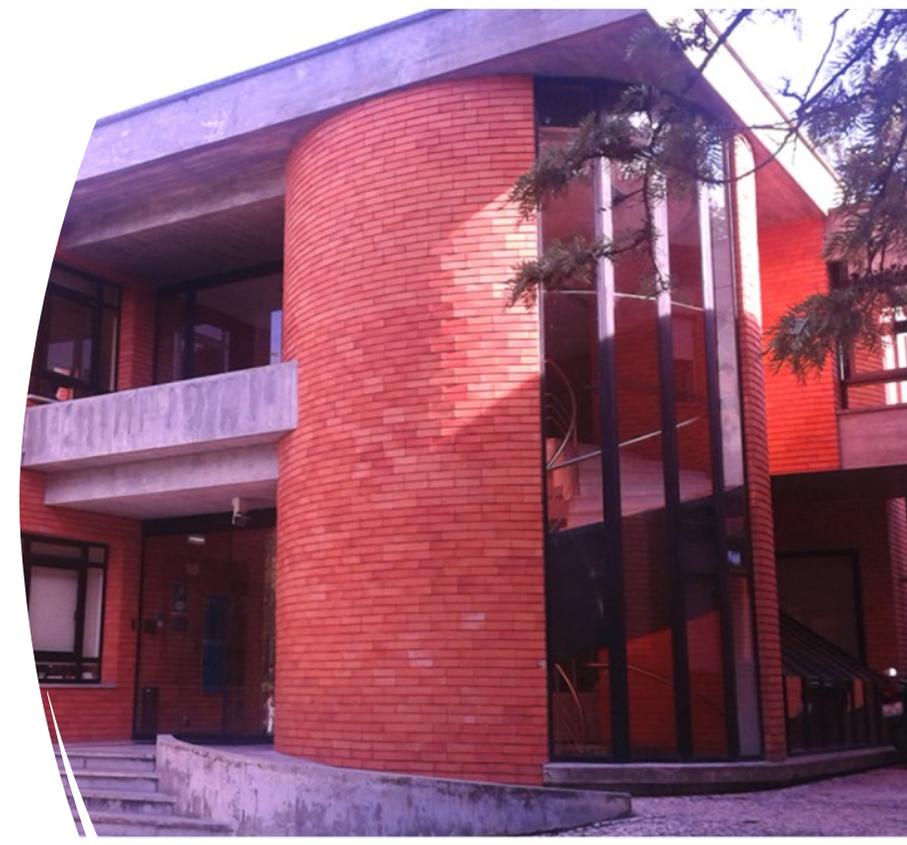
- Gravar, armazenar e reproduzir páginas Web no próprio computador num formato normalizado, utilizando ferramentas gratuitas do Webrecorder.net

Agenda

1. Introdução: O Arquivo.pt e a promoção das práticas de preservação da Web
2. Demonstração de gravação com a extensão ArchiveWeb.page
3. Extra: Gravação automática de sites inteiros com o Browsertrix-crawler

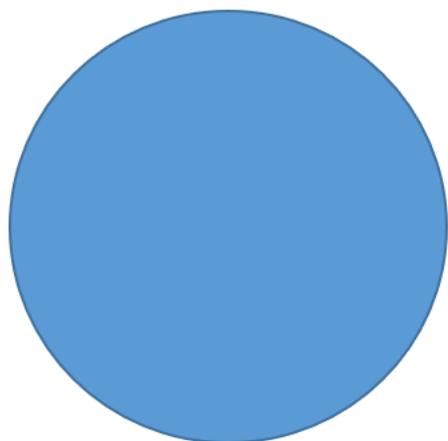
O Arquivo.pt

- Serviço público da FCCN Unidade de Computação Científica da Fundação para a Ciência e a Tecnologia, I. P.
- 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)
- Acessível na Web em <https://arquivo.pt>
- Instalações físicas – Av. do Brasil – Lisboa (Campus do LNEC)

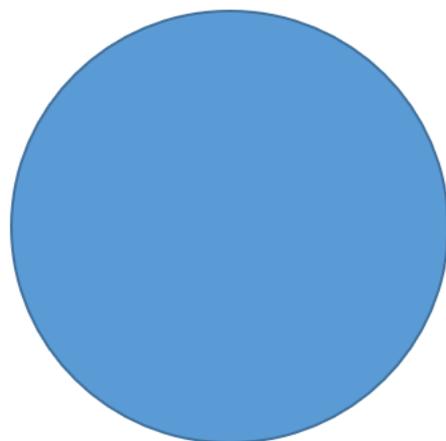


Arquivar a Web

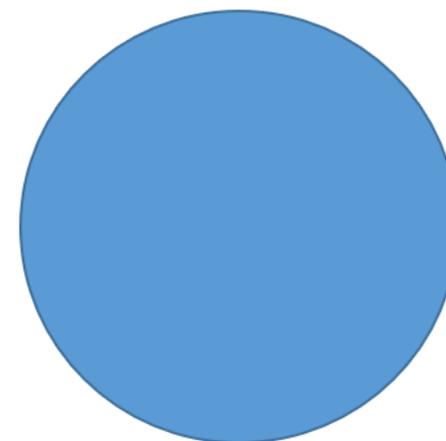
Gravar, Armazenar e Reproduzir são, de forma simplificada, as três fases da preservação (poderíamos falar de outras, tal como a seleção, controle de qualidade, etc)
O ArchiveWeb.page cumpre as três fases: 1) grava 2) gera um WARC e um WACZ para armazenar em qualquer lugar 3) reproduz o conteúdo.
É isso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



Gravar



Armazenar



Reproduzir

Arquivar a 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, etc)

O ArchiveWeb.page cumpre as três fases: 1) grava 2) gera um WARC e um 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



Webrecorder.net

“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

Webrecorder.net

- Código aberto
- Centrado no utilizador não especializado
- Formatos normalizados
- **Compatibilidade com os arquivos da Web**
- Apoio do International Internet Preservation Consortium (IIPC)

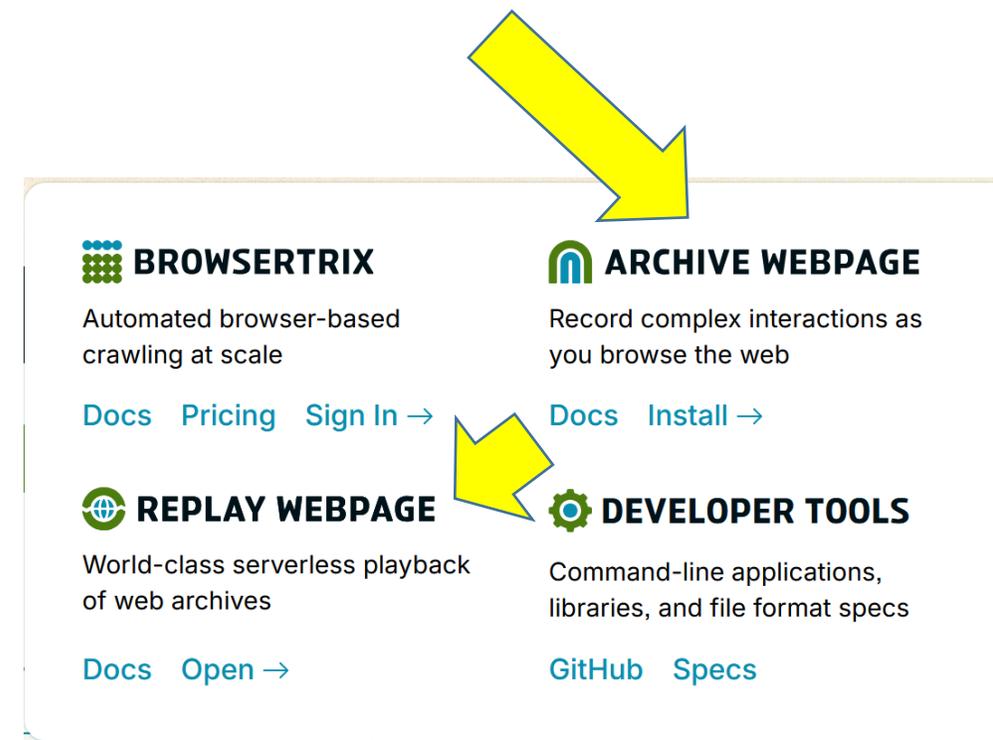
Nas ações de formação do Arquivo.pt valorizamos o que pode contribuir para a colaboração com a comunidade, independentemente das ferramentas utilizadas.

Presentemente, o Webrecorder.net tem as ferramentas que consideramos mais adequadas.

Webrecorder.net

Neste tutorial vamos usar ArchiveWeb.page na versão de extensão de browser para gravar e o ReplayWeb.page

O projeto Webrecorder.net tem um conjunto de ferramentas que todos podem usar para gravar e reproduzir páginas Web.



[Webrecorder.net](https://webrecorder.net)

Webrecorder.net

O Webrecorder.net tem o apoio de entidades internacionais importantes no domínio do Web archiving.

O Arquivo.pt, por exemplo, usa o software pywb do Webrecorder para reproduzir as páginas Web, usa-o para o novo serviço de gravação de páginas na hora SavePageNow.

O ArchiveWeb.page é utilizado em recolhas de alta qualidade de websites. O que as outras ferramentas não conseguem gravar, o ArchiveWeb.page provavelmente conseguirá.

O Webrecorder.net é utilizado para dar formação sobre preservação da Web.

Trusted by



Webrecorder software is used by libraries, universities, governments, arts organizations, and more around the world.

Página <https://webrecorder.net/community>

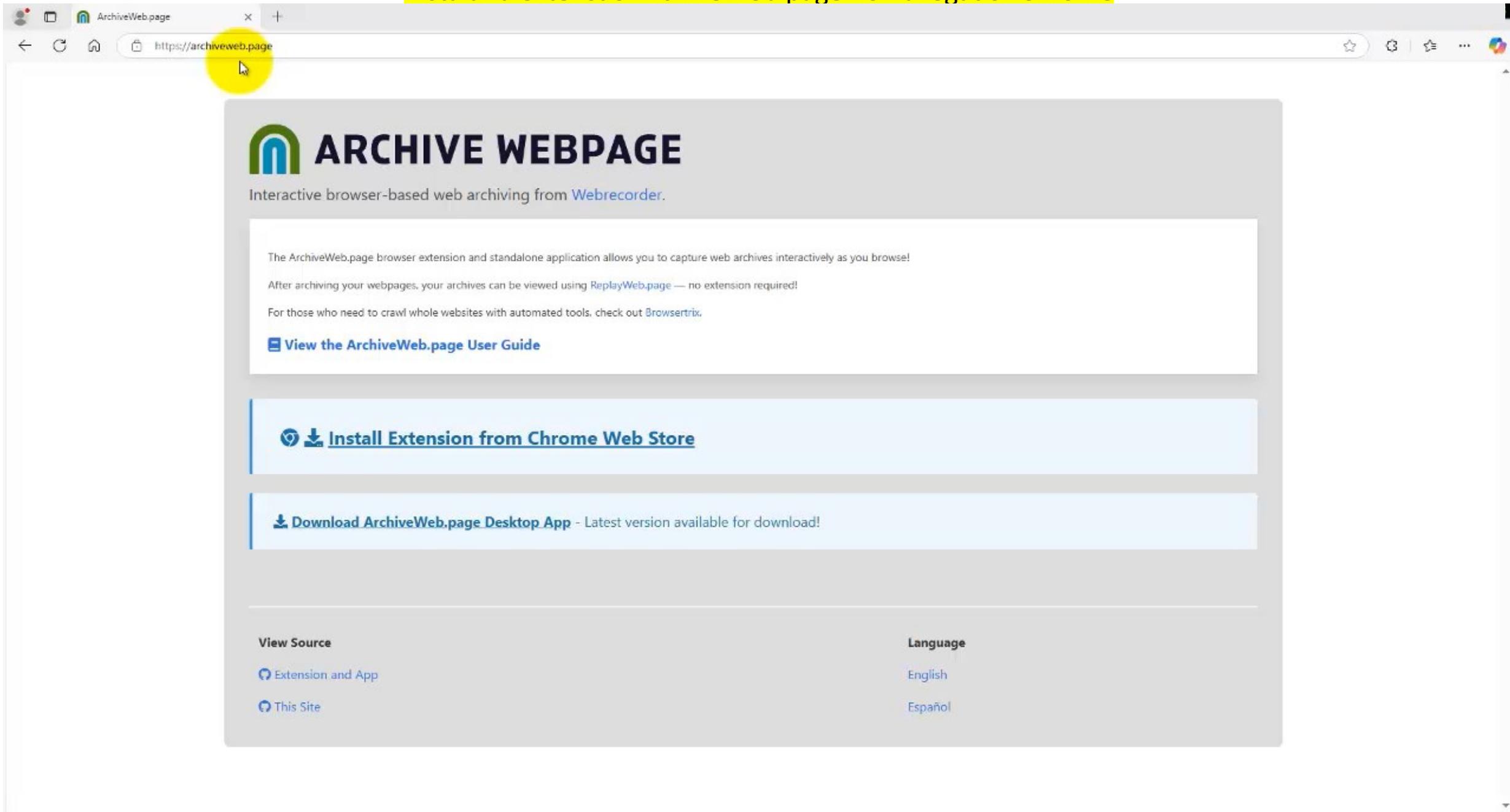
2. Demonstração de gravação utilizando o ArchiveWeb.page





Instalar a extensão ArchiveWeb.page

Instalar a extensão ArchiveWeb.page no navegador Chrome



ArchiveWeb.page

https://archiveweb.page

ARCHIVE WEBPAGE

Interactive browser-based web archiving from [Webrecorder](#).

The ArchiveWeb.page browser extension and standalone application allows you to capture web archives interactively as you browse!

After archiving your webpages, your archives can be viewed using [ReplayWeb.page](#) — no extension required!

For those who need to crawl whole websites with automated tools, check out [Browsertrix](#).

[View the ArchiveWeb.page User Guide](#)

[Install Extension from Chrome Web Store](#)

[Download ArchiveWeb.page Desktop App](#) - Latest version available for download!

View Source

- [Extension and App](#)
- [This Site](#)

Language

- [English](#)
- [Español](#)



Gravar uma página única

Gravar uma página única

The screenshot shows the website of the Câmara Municipal de Coimbra. The browser address bar displays <https://www.cm-coimbra.pt>. The page features a large video player with a scene of a coffee table in a room with stained glass windows. The text "à volta da mesa de um café." is overlaid on the video. To the left is a sidebar with navigation links: Notícias, Serviços Online, Balcão Virtual, Agenda, Viver, Turismo, Estudar, Investir, Obras, Transparência, and Comunicação. Below the video is a section titled "Atividade municipal" with navigation arrows and a "ver +" link. This section contains three promotional cards: "Encontro Nacional de Provedores dos Municípios" (6 MAIO 25), "viagem ao mundo da escrita: em tempo de camões" (2 DE MAIO A 29 DE AGOSTO), and "Save the date" for "coimbra 19'25" (2 e 3 julho, Convento São Francisco). The footer contains various logos, including "coimbra+ 170 años", "CENTRO 2020", and "CÂMARA MUNICIPAL COIMBRA", along with social media icons and the copyright notice "© 2025 . Câmara Municipal de Coimbra . Política de privacidade".



Gravar várias páginas

Gravar várias páginas

The screenshot shows the website of the Câmara Municipal de Coimbra. The browser address bar displays <https://www.cm-coimbra.pt>. The main content area features a large aerial photograph of the city's ruins, with the text "Ouvi-la falar de histórias" overlaid. Below the main image is a section titled "Atividade municipal" with a navigation bar containing left and right arrows and the text "ver +".

On the left side, there is a vertical menu with the following items:

- Notícias
- Serviços Online
- Balcão Virtual
- Agenda
- Viver
- Turismo
- Estudar
- Investir
- Obras
- Transparência
- Comunicação

At the bottom of the page, there is a footer with several news items and logos:

- Comissão de Gestão Integrada de Fogos Rurais de Coimbra reuniu ontem nos Paços do Concelho
- Casa da Cidadania da Língua inaugura no sábado mostra evocativa do Dia da Língua Portuguesa
- Nova edição do Coimbra BD já confirmada para os dias 24, 25 e 26 de abril de 2026
- MMC par...

Logos for "coimbra+ integrado 2020" and "CÂMARA MUNICIPAL COIMBRA" are also present.



Gravar muitas páginas

Gravar muitas páginas

Câmara Municipal de Coimbra x extension://fpeodllldobpkbkabp x + 4:50:54 05/02/2025

← → ↻ 🏠 🔒 https://www.cm-coimbra.pt

🏠 ☰ 🔍 PT

- Notícias
- Serviços Online
- Balcão Virtual
- Agenda
- Viver
- Turismo
- Estudar
- Investir
- Obras
- Transparência
- Comunicação

ouço-o nos seus risos!

Atividade municipal < > ver +

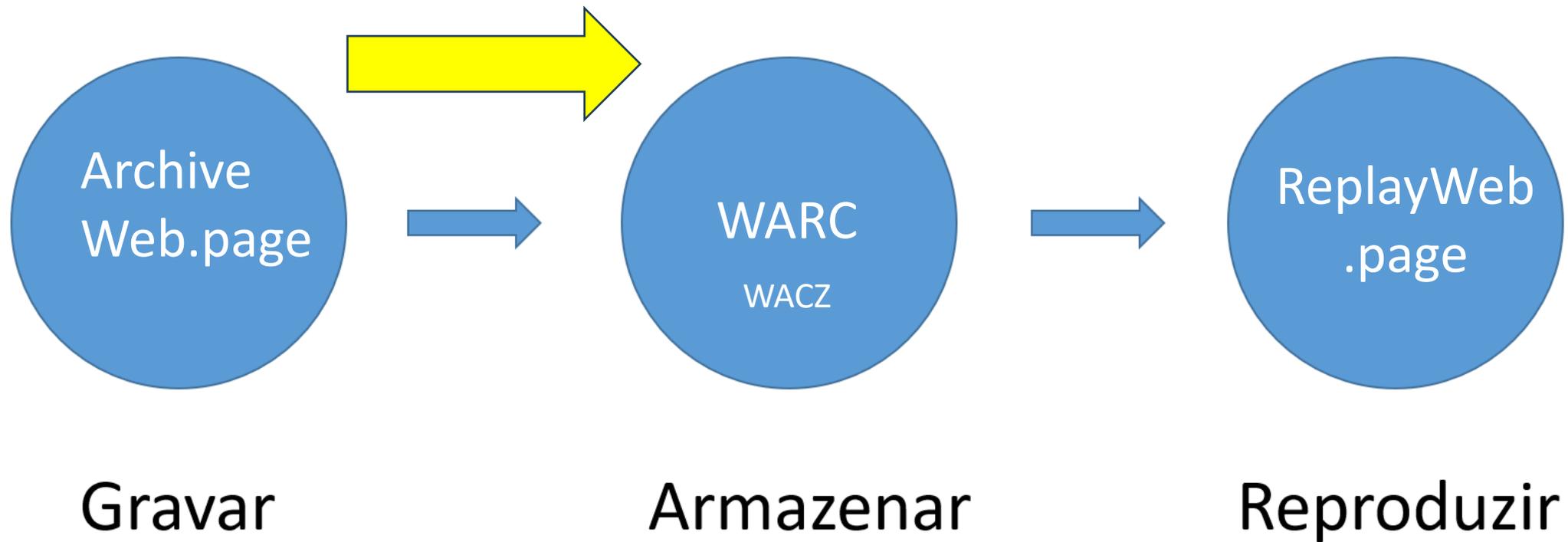
Seleção de crianças e jovens Comissão de Gestão Integrada de Fogos Rurais de Coimbra reuniu ontem nos Paços do Concelho Casa da Cidadania da Língua inaugura no sábado mostra evocativa do Dia da Língua Portuguesa Nova edição do Coimbra BD já confirmada para os dias 24, 25 e 26 de ab

coimbra+ CENTRO 2020 CAMARA MUNICIPAL COIMBRA

© 2025 . Câmara Municipal de Coimbra . Política de privacidade

Armazenar em formato
normalizado: WARC

Arquivar a Web



Formato WARC

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

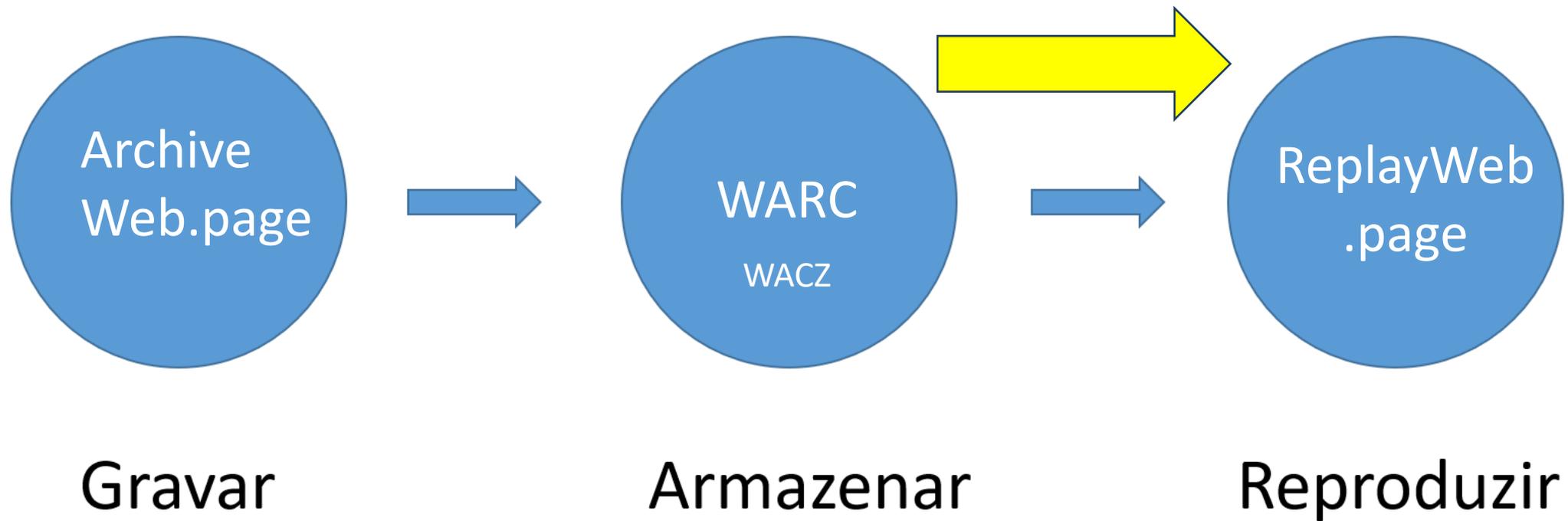
WACZ ficheiro comprimido e processado para facilitar a reproduzidos com o ReplayWeb.page

Download no formato normalizado WARC ou WACZ

Reproduzir os conteúdos em qualquer tempo e lugar



Arquivar a Web



Reproduzir localmente os conteúdos gravados

The image shows a browser window with the URL `replayweb.page`. The page title is "REPLAY WEBPAGE" and it includes navigation links for "User Docs" and "About". The main heading reads "Explore and Replay Interactive Archived Webpages Directly in your Browser. (Learn More)". Below this is a "Load Web Archive" section with a "Choose File..." button and a text input field for a URL. A green "Load" button is positioned to the right of the input field. An "Abrir" (Open) file explorer window is overlaid on the browser, showing the directory structure: `utilizador > Ambiente de trabalho > MeuArquivoDaWeb`. The file explorer lists several folders and files, with `01--gravar-página-única` selected. The file name field at the bottom of the explorer shows `01--gravar-página-única` and the file type is set to "WARC, WACZ, and HAR Files".

Arquivar a Web



O Arquivo Municipal de Sines, desde 2020, grava conteúdos Web de interesse para a memória e a história do Município. Usa o Webrecorder ArchiveWeb.page e o serviço Conifer que funciona de forma semelhante. Envia uma cópia dos ficheiros WARC e WACZ obtidos para o Arquivo.pt.

Faça como o Arquivo Municipal de Sines. Seja ativo na preservação dos conteúdos importantes.

Envie as suas gravações

Short link para voucher do filesender.fccn.pt

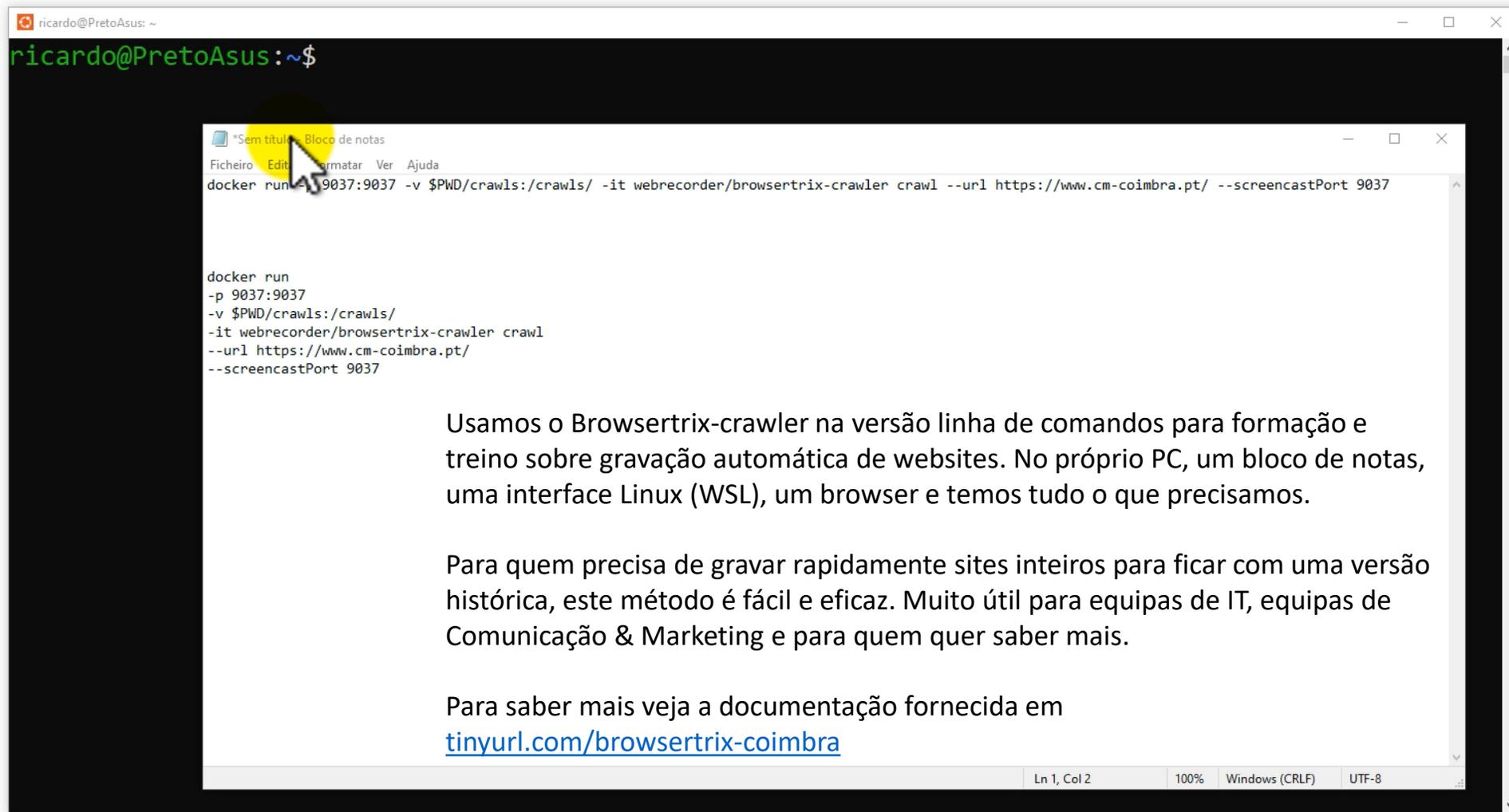
tinyurl.com/enviar2025

Para uma experiência mais completa, depois de gravar, exporte o seu ficheiro WARC ou WACZ e envie para o curador digital do Arquivo.pt. Poderá ver a sua gravação integrada no Arquivo.pt.

3. Gravação automática com o Browsertrix-crawler

Gravar sites inteiros com o Browsertrix-crawler

10:55:37 05/05/2025



The image shows a terminal window with a terminal prompt `ricardo@PretoAsus: ~$`. A Notepad window is open over the terminal, displaying the following command:

```
docker run -p 9037:9037 -v $PWD/crawls:/crawls/ -it webrecorder/browsertrix-crawler crawl --url https://www.cm-coimbra.pt/ --screencastPort 9037
```

Below the terminal and Notepad windows, there is explanatory text in Portuguese:

Usamos o Browsertrix-crawler na versão linha de comandos para formação e treino sobre gravação automática de websites. No próprio PC, um bloco de notas, uma interface Linux (WSL), um browser e temos tudo o que precisamos.

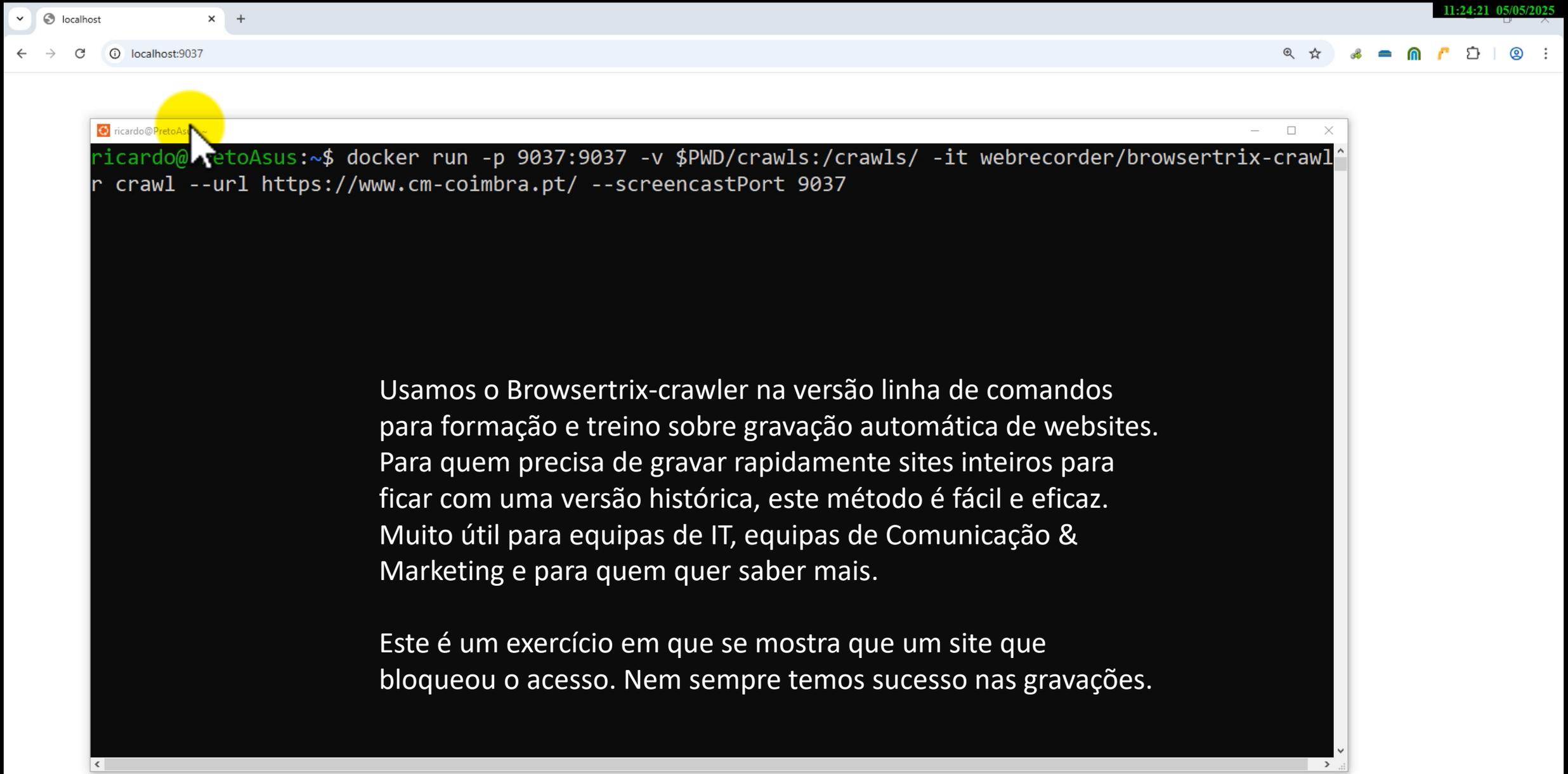
Para quem precisa de gravar rapidamente sites inteiros para ficar com uma versão histórica, este método é fácil e eficaz. Muito útil para equipas de IT, equipas de Comunicação & Marketing e para quem quer saber mais.

Para saber mais veja a documentação fornecida em tinyurl.com/browsertrix-coimbra

```
ricardo@PretoAsus:~$ docker run -p 9037:9037 -v $PWD/crawls:/crawls/ -it webrecorder/browsertrix-crawler crawl --url https://www.cm-coimbra.pt/ --screencastPort 9037
```

Usamos o Browsertrix-crawler na versão linha de comandos para formação e treino sobre gravação automática de websites. Para quem precisa de gravar rapidamente sites inteiros para ficar com uma versão histórica, este método é fácil e eficaz. Muito útil para equipas de IT, equipas de Comunicação & Marketing e para quem quer saber mais.

Este é um exercício em que se limita o âmbito (*scope*) para gravar apenas um subdomínio ou uma parte de um website. Ver documentação em tinyurl.com/browsertrix-coimbra.



The image shows a web browser window with the address bar set to localhost:9037. A terminal window is overlaid on the browser, displaying a Docker command. The terminal text is as follows:

```
ricardo@PretoAsus:~$ docker run -p 9037:9037 -v $PWD/crawls:/crawls/ -it webrecorder/browsertrix-crawler crawl --url https://www.cm-coimbra.pt/ --screencastPort 9037
```

Below the terminal window, there is a text block explaining the use of Browsertrix-crawler. The text is as follows:

Usamos o Browsertrix-crawler na versão linha de comandos para formação e treino sobre gravação automática de websites. Para quem precisa de gravar rapidamente sites inteiros para ficar com uma versão histórica, este método é fácil e eficaz. Muito útil para equipas de IT, equipas de Comunicação & Marketing e para quem quer saber mais.

Este é um exercício em que se mostra que um site que bloqueou o acesso. Nem sempre temos sucesso nas gravações.

tinyurl.com/browsertrix-coimbra

Conclusão

- Grave conteúdos em formato normalizado com o ArchiveWeb.page
- Crie um arquivo da Web útil para si e para a sua organização
- Use o Arquivo.pt para perpetuar conteúdos da Web importantes

Mail

contacto@arquivo.pt

Inscreva-se na mailing list (enviamos 1 vez por mês)

arquivo.pt/inscrever

FCCN serviços
digitais
fct

fct Fundação
para a Ciência
e a Tecnologia



UNIVERSIDADE D
COIMBRA