

O Arquivo.pt e a preservação da memória digital da Internet

Ricardo Basílio, curador digital

contacto@arquivo.pt



O que é o Arquivo.pt

- Serviço público da FCCN Serviços digitais 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)



Q

1991 6 Ago ————— 2025 24 Fev

Cerca de 1.183.988 resultados desde 1991 até 2025

cacilheirodotejo.blogspot.com

Cacilheiro do Tejo

20 Abril 2019

Cacilheiro do Tejo **Cacilheiro** do Tejo Diário de Bordo de uma histórica embarcação, ancorada na ... dezembro de 2011 Academia do CrepeCake é novo projecto **Cacilheiro** do Tejo A Tripulação do **Cacilheiro** ... boca e inspira um futuro mais doce. A Academia do CrepeCake reforça atributos da oferta **Cacilheiro** ...

almada-virtual-museum.blogspot.com/2019/03/ze-cacilheiro.html

Almada Virtual Museum: Zé Cacilheiro

12 Abril 2019















Almada Virtual Museum: Zé **Cacilheiro** Almada Virtual Museum ... de Almada conheço todas as pedras e ... Bibliografia Biblioteca em linha O

Q Pesquisar

1991 6 Ago 2025 24 Feb

Páginas Imagens Narrativa Pesquisa avançada

Cerca de 3.309 resultados desde 1991 até 2025

 → http://www.m-almada.pt/... 8 Maio 2021	 → http://liquidimages.com.p... 12 Novembro 2013	 → http://olhares.sapo.pt/asc... 14 Novembro 2016	 → http://www.stimuli.pt/cms... 3 Agosto 2010	 → http://joaofarinha.com.pt/... 21 Outubro 2008	 → http://miguel-moreira.pt/o... 18 Abril 2015	 → http://www.jgarraio.pt/ind... 12 Marco 2008
 → https://restosdecoleccion.... 27 Dezembro 2019	 → http://www.paulo-santos.... 20 Junho 2016	 → https://apc-instruments.c... 4 Setembro 2019	 → http://joaofarinha.com.pt/... 21 Outubro 2008	 → http://miguel-moreira.pt/o... 3 Junho 2020	 → http://www.reflexosonline... 23 Maio 2011	 → http://www.vart.pt/neves_... 8 Outubro 2014

Histórico de uma página web

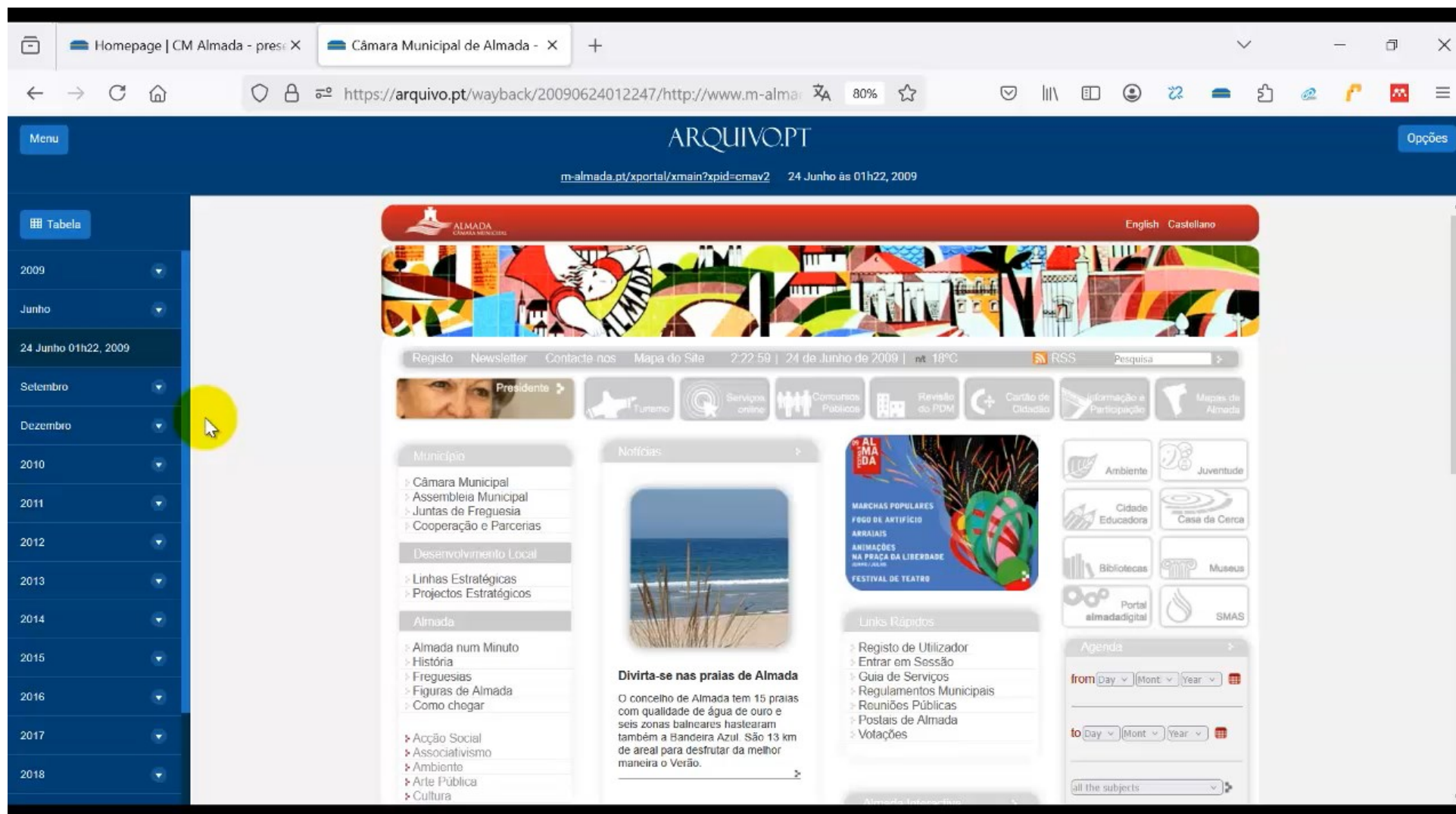


<https://arquivo.pt/wayback/20210421193054/http://www.m-almada.pt/>

Serviços para a comunidade

- **Recolhas de alta qualidade (a pedido)**
- Memorial do Arquivo.pt
- Formação

Serviços do Arquivo.pt para a comunidade: Recolhas de alta qualidade a pedido



O Arquivo.pt grava os conteúdos interativos do Website.

Serviços para a comunidade

- Recolhas de alta qualidade a pedido
- **Memorial do Arquivo.pt**
- Formação

Memorial do Arquivo.pt

Mantém o domínio

Redireciona para o
Memorial



Desliga o servidor

arquivo.pt/memorial

Memorial do Arquivo.pt



The screenshot displays the Arquivo.pt website interface. At the top, the header includes the 'ARQUIVO.PT' logo, a 'Menu' button, and a date stamp '30 Abril às 17h13, 2018'. A navigation sidebar on the left lists years from 2005 to 2015. The main content area features a banner for 'Cidades e regiões Digitais' with statistics: '25 projectos' and '235 autarquias'. Below the banner, there is a search bar and a list of project categories under the heading 'PROJECTOS', including 'Açores Digital', 'Algarve Digital', 'Almada Digital', 'ALO Digital', 'Aveiro Digital', 'Beira Baixa Digital', 'Beja Digital', 'Braga Digital', 'Entre Douro e Vouga Digital', and 'Évora Distrito Digital'. The main article is titled 'Algarve Digital é Case Study da CISCO' and includes a sub-header 'Algarve Digital' with a logo. The text of the article describes Cisco's role in the project and provides a URL for more information.

Exemplo do site Cidades Digitais acessível no Memorial do Arquivo.pt - <https://cidadesdigitais.pt>

Memorial do Arquivo.pt



Exemplo do site antigo da DGPC acessível no Memorial do Arquivo.pt - <https://siteantigo.dgpc.pt/>

Memorial do Arquivo.pt

Adequado para

- Websites com informação útil
- Websites que deixaram de ser mantidos
- Websites com tecnologia obsoleta

Vantagens

- Reduz a pegada ecológica
- Reduz os custos de manutenção
- Reduz ameaças de segurança à instituição

arquivo.pt/memorial

Serviços para a comunidade

- Recolhas de alta qualidade a pedido
- Memorial
- **Formação**

Formação

- Arquivo.pt: uma nova ferramenta para pesquisar o passado (módulo A)
- Bem publicar, para bem preservar (módulo B)
- Acesso e processamento automático de informação preservada da Web através de APIs (módulo C)
- Arquivar a Web: faça-você-mesmo! (módulo D)

arquivo.pt/forma

Formação



Valorização da memória Institucional arquivando conteúdos de interesse local – o caso do Arquivo Municipal de Sines

Resumindo

- Recolhas de alta qualidade (a pedido)
- Memorial
- Formação

arquivo.pt/catalogo

contacto@arquivo.pt

Para discussão

Após a apresentação

Devem ou não as instituições preservar os conteúdos públicos que colocaram nas redes sociais tendencialmente fechadas? Como fazer isso?

The screenshot shows a web browser displaying a Facebook page for 'Arquivo.pt'. The browser's address bar shows the URL: <https://arquivo.pt/wayback/20241218175147/https://www.facebook.com/Arquivo.pt/>. The page header includes the Facebook logo, a search bar, and navigation icons. The page content is divided into several sections:

- Apresentação:** A section describing the service, stating that Arquivo.pt allows users to search and access pages from the past since 1996. It also provides contact information: 'Página - Fornecedor de Internet', 'Av. Brasil, 101, Lisbon, Portugal', 'contacto@arquivo.pt', and 'arquivo.pt'. It mentions that 96% of users recommend it (14 reviews).
- Publicações:** A section showing a post from 'Arquivo.pt' dated '13 de dezembro'. The post text reads: 'Servir a investigação científica é missão do Arquivo.pt e aqui está um resultado. Diego Alves, visitou o Arquivo.pt, FCN - FCT, em agosto de 2022 para desenvolver o seu trabalho, no âmbito do Projeto CLEOPATRA, Rede de Formação Inovadora Marie Skłodowska-Curie. O Capítulo 3 do livro, "Robustness of Corpus-Based Typological Strategies for Dependency Parsing", em co-autoria com Daniel Gomes, é dedicado a este caso de uso do Arquivo.pt no domínio do Processamento de Linguagem Natural. "Event Analytics across Languages and Communities" foi publicado em acesso aberto, pela Springer, e está disponível em <https://link.springer.com/book/10.1007/978-3-031-64451-1>. Saber mais em <https://sobre.arquivo.pt/.../dados-abertos-para-/>'. Below the text is a photo of a man in a white shirt holding a blue book titled 'Event Analytics across Languages and Communities'. The post has 14 likes, 1 comment, and 4 shares.
- Fotos:** A section with a grid of photo thumbnails and a link to 'Ver todas as fotos'.

<https://arquivo.pt/wayback/20241218175147/https://www.facebook.com/Arquivo.pt/>

Caso de uso: A Direção Regional do Arquivo e Biblioteca da Madeira disponibiliza acesso a histórico da web madeirense. Para isso criou uma [descrição no catálogo](#).

The screenshot shows the website interface for the Direção Regional do Arquivo e Biblioteca da Madeira. The header includes the logo, the name of the institution, and navigation links: Início, Explorador, Coleções ABM, and Ajuda. There are also icons for mobile, refresh, and user profile.

Resultados da pesquisa avançada

Foram encontrados 10 resultados.

Ordenar ▾ Completo ▾ 🔔

Filtros

Bibliotecas ▲
 Arquivo e Bibliot... (10)

Coleções ABM ▲
Não existem filtros.

Tipo de Documento ▲
 Documento elec... (10)

dnoticias.pt
Autor: Empresa do Diário de Notícias
Formato: Documento Electrónico
Publicação: Funchal : Rua Dr. Fernão de Ornelas, 56/3.º, [2000]
Descrição física: 1 sítio Web : il. color.
Cota: RL[RECURSO DIGITAL]
Cont. Digital: [Consultar site](#)

🔍 Detalhes 📄 Reservar 🔗 Link ✉ E-Mail 🔄 Partilhar 📌 Guardar

<https://mailchi.mp/d63f5cc3412b/newsletter-do-abm-julho-24>

Deve ou não preservar-se os vídeos institucionais publicados no Youtube?
Em 2025 não é possível fazê-lo no formato arquivo da web. Como fazer?

Caso: Projeto de tradição oral do Politécnico de Lisboa, AMOPC



The screenshot shows the ARQUIVO.PT website interface. At the top, there is a dark blue header with the logo 'ARQUIVO.PT' in the center, a 'Menu' button on the left, and an 'Opções' button on the right. Below the header, the URL 'amopc.org/joaquim-manuel-prudencio-vieira' and the date '11 Abril às 13h59, 2022' are displayed.

The main content area is divided into two columns. The left column is a dark blue sidebar with a 'Tabela' button at the top. It contains a list of dates and months with dropdown arrows: 2022, Abril, 11 Abril 13h22, 2022, 11 Abril 13h59, 2022, 11 Abril 21h46, 2022, Maio, Julho, Agosto, Outubro, and 2023.

The right column features a video player with a black background and white text: 'An error occurred. Please try again later. (Playback ID: bbBBuoJw3cMZms5R) Learn More'. To the right of the video player is the title 'Joaquim Manuel Prudêncio Vieira' in bold black font, followed by 'parte 3 de 4'. Below the title is a description: 'Entrevista conduzida por Júlia Leitão Barros. Registado na ESCS em 25 de Março de 2021 por Miguel Carrapiço'.

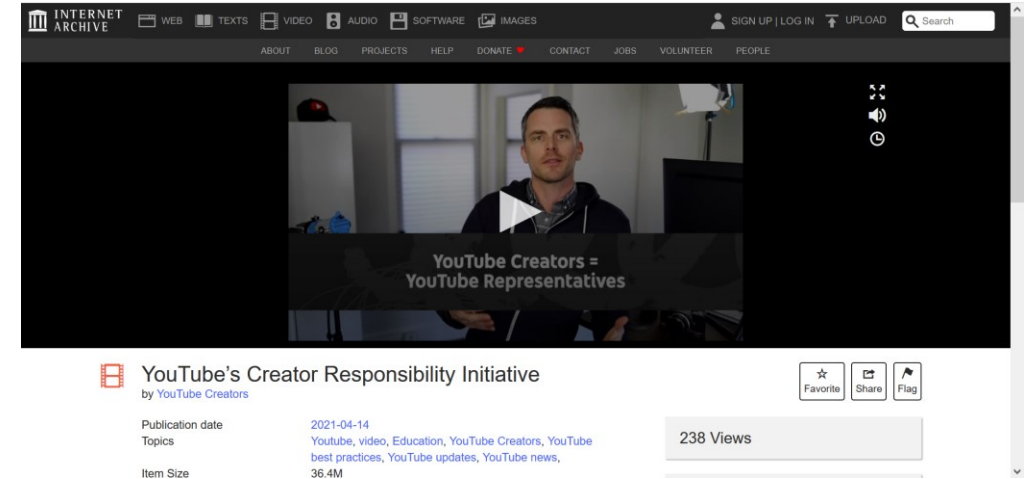
Below the video player and title is a list of video entries, each with a blue link and a timestamp: 'Joaquim Vieira parte 3 de 4 aos 1M2S', 'Joaquim Vieira parte 3 de 4 aos 5M15S', and 'Joaquim Vieira parte 3 de 4 aos 17M55S'. The text describes the content of each entry, such as 'diretor adjunto do Expresso, crescimento contínuo de tiragens, de 1981 a 1993, importância da Revista; António José Saraiva, 1º caderno, importância primeira página, notícia de internacional; Francisco Pinto Balsemão relação com o Expresso, não interferência, grupo Bilderberg; Vicente Jorge Silva, José António Saraiva; demissão do Expresso, notícia sobre Joe Berardo, SIC; Diário Popular; diretor gráfico, Luís Ribeiro, Mestre.' and 'entrada na Visão, convite Joaquim Furtado para diretor-adjunto para os programas da RTP, Cesário Borga diretor de informação; demissão da RTP, reportagem congresso da UNITA; Joaquim Vieira parte 3 de 4 aos 21M18S - queda da audiência da RTP, concorrência televisão privadas, tutela'.

<https://arquivo.pt/wayback/20220411135959/https://amopc.org/joaquim-manuel-prudencio-vieira/>

Deve ou não preservar-se os vídeos institucionais publicados no Youtube?
Em 2025, não é possível fazê-lo no formato arquivo da web. Como fazer?

O que faz o IA para ter uma coleção de vídeos do Youtube?

- Serviço para vídeos diferente da wayback
- Interface com metadados: <https://archive.org/details/YTV-20170220061354-crawl821>
- Interface para a reprodução de vídeos em vários formatos <https://archive.org/details/youtube-PO5z1bqW5yY>
- Coleção de vídeos <https://archive.org/details/youtubecrawl>
- Código utilizado: <https://github.com/yt-dlp/yt-dlp>



[Exemplo de vídeo do Youtube preservado](#)

A explorar:

Aplicação para extração e migração, TubeUp Video Stream Mirroring Application: <https://pypi.org/project/tubeup/>

Aplicação para capturar e reproduzir Youtube localmente: <https://github.com/tubearchivist/tubearchivist>