

# Catalog of tools for digital preservation

servatio

202



# The Web is the largest and most widely used source of digital objects



# Which keep on getting lost...



 RESEARCH TOPICS
 PUBLICATIONS
 OUR METHODS
 SHORT READS

 Home
 > Research Topics
 > Internet & Technology

REPORT | MAY 17, 2024

SHARE 🟦

### When Online Content Disappears

38% of webpages that existed in 2013 are no longer accessible a decade later

BY ATHENA CHAPEKIS, SAMUEL BESTVATER, EMMA REMY AND GONZALO RIVERO

# **Ready-to-use** tools to safeguard digital objects published online are required



## Arquivo.pt Catalog



sobre.arquivo.pt > About > Services catalog of the Arquivo.pt web archive

#### Services catalog of the Arquivo.pt web archive

Last updated on October 26th, 2023 at 05:47 pm

#### Public services

- Search and access web-archived data since the 1990s
- Application Programming Interfaces (APIs)
- · Suggest websites to be preserved
- SavePageNow: immediately archive web pages
- · Integration of historical web data collections
- Training on web preservation
- Open data listing archived web information on various topics
- CitationSaver: extracts links from documents and archives the correspondent web pages
- Arquivo404: presents web-archived pages instead of "pages not found"

#### For partners

- Memorial preserves your old website information before deactivating it
- High-quality archive of websites (on-demand)
- Creation of collections and thematic exhibitions
- · Itinerant exhibition of posters at your facilities

Toolkit of 13 **free** tools to preserve digital objects published online.

Anyone can contribute to safeguard Humanity's Digital Legacy.

arquivo.pt/catalog

## Tools for **acquisition**: collect before it vanishes!

- Suggest websites
- SavePageNow
- Integration of historical web data collections
- CitationSaver
- High-quality archive
- Memorial

### Save page now: immediately archive a webpage

### ARQUIVO.PT

### SavePageNow Record webpages on Arquivo.pt

### https://www.dpconline.org/events

e.g. www.fccn.pt

The SavePageNow service allows a page to be archived at the exact moment the user makes the request.

The archived contents are later integrated in the Arquivo.pt collection.

Learn more

arquivo.pt/savepagenow

Record

### **CitationSaver** preserves links cited in documents



### arquivo.pt/citationsaver

# Arquivo.pt Memorial: preserves the content of deactivated websites



## Tools for **access:** if you cannot access, it is not preserved

- Search and access
- Application programming interfaces
- Arquivo404

### Search **texts** from past pages



#### www.nla.gov.au/padi/topics/18.html

#### PADI - Digital preservation strategies

#### 5 October 2009

PADI - Digital preservation strategies Home About PADI Search Browse Topics Feedback Contributions What's New? PADI Partners About ICADS Digital preservation strategies Limit to: Emulation ... number of researchers as a useful strategy in conjunction with other digital preservation methods ...

### Browse the **history** of an address

ARQUIVO.PT	
Q dpconline.org	Search
1991 <sup>6</sup> Aug Pages Images Narrative	Advanced search
About 40 results from 1991 to 2024	
☐ Table III List	
2014	1 version
2016	3 versions 🗸
2017	1 version 🔹
2018	2 versions
2019	7 versions
2020	2 versions
2021	9 versions
2022	12 versions
2023	3 versions



## Replay options: complete the page!



### Search for images from the past!



# Application Programming Interfaces to develop new applications



### Your new app



### Arquivo.pt APIs

#### **APIs**

Daniel Gomes edited this page 11 days ago · 33 revisions

### APIs specific to Arquivo.pt that enable the full exploration of its functions

- Arquivo.pt API (Full-text & URL search)
- Image Search API v1.1 (beta version)

APIs based on international standards to enable interoperability among web archives and code reuse

- CDX-server API (URL search)
- Memento API (URL search)

### Tools for **dissemination**: preserved digital objects are valuable assets!

- Online thematic collections
- Itinerant exhibition of posters
- Open data sets
- Training

## Exhibitions of web-archived documents

#### Memória dos Museus na Web

Websites dos museus portugueses (1996-2021)

Home Lista de museus em Portugal Sobre



Museu Nacional de Arte Antiga



Museu Nacional de

Ciência

História Natural e da



Museu Nacional de Arte Contemporânea – Museu do Chiado



Museu Nacional dos Coches



Museu Nacional Soares dos Reis



Museu Nacional Machado de Castro





27 datasets





arquivo.pt/expos

# Training courses on web preservation and research

- New ways of searching the past (module A)
- Publish preservable information on the web (B)
- Automatic processing of information preserved from the Web (C)
- Do-it-yourself web archiving (D)



arquivo.pt/training

## Arquivo.pt Catalog: final remarks

Free toolkit to safeguard digital legacy published online

Applied to preserve at-risk digital objects

- 1.7 million users in 2023
- 470 TB of preserved digital objects served
- Software is free **open source**
- **Open access** technical and scientific publications

Safeguards and **enable others** to safeguard online digital objects