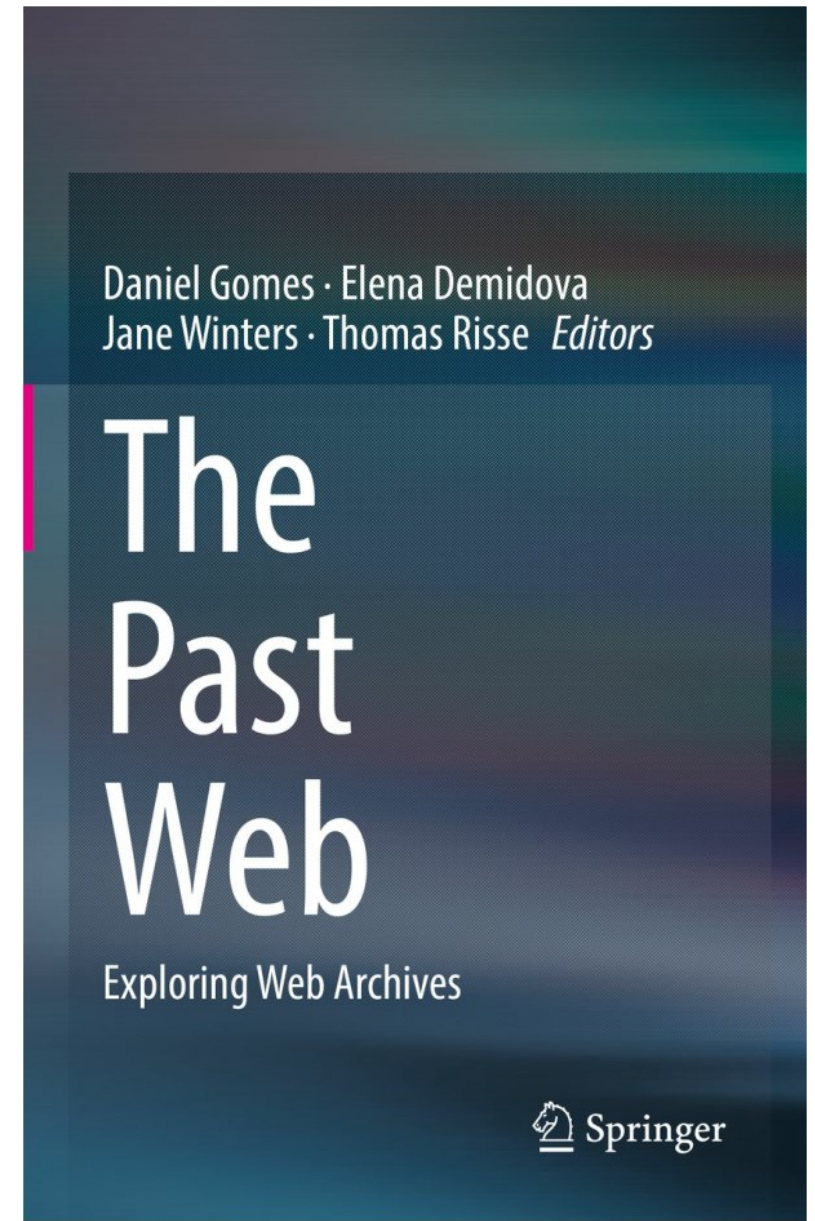


The Past Web: Exploring Web Archives

How a book is born?

Daniel.Gomes@fccn.Pt



Web archiving workshop 2003

3rd ECDL Workshop on Web Archives

August 21st, 2003
Trondheim, Norway

in conjunction with the [7th European Conference on Research and Advanced Technologies for Digital Libraries](#)



Objectives

Following the great success of the first two ECDL Workshops on Web Archiving in Darmstadt, Germany in 2001 ([WebArchiving 2001](#)), and Rome, Italy, in 2002 ([WebArchiving2002](#)), we are happy to invite you to the third Workshop in this series.

The workshop will provide a cross domain overview on active research and practice in the emergent domain of web archiving and studies on effective usage of this type of archives.

It is also intended to provide a forum for interaction among librarians, archivists, academic researchers and industrial researchers interested in establishing effective methods and developing improved solution for Web archiving.

Workshop Proceedings

Download [full Workshop Proceedings as a PDF file](#) (3,5 Mo).
See below for individual papers and presentations.

Afternoon session: 14:00 - 18:00 // Case Studies //

A Characterization of the Portuguese Web (download [paper](#))

Daniel Gomes, Mário J. Silva
Faculty of Sciences, University of Lisbon, Portugal

Building Thematic Web Collections: Challenges and Experiences from the September 11 Web Archive and the Election 2002 Web Archive (download [paper](#))

Steven M. Schneider, SUNY Institute of Technology & WebArchivist.org, USA
Kirsten Foot, Department of Communication, University of Washington & WebArchivist.org, USA
Michele Kimpton, Internet Archive, USA
Gina Jones, Library Services, Library of Congress, USA

Political Communications Web Archiving: Addressing Typology and Timing for Selection, Preservation and Access (download [paper](#))

Bernard Reilly, Gretchen Tuchel, James Simon, Center for Research Libraries, USA
Carolyn Palaima, Kent Norsworthy, Latin American Network Information Center - LANIC, USA
Leslie Myrick, Digital Library Group, New York University Libraries, USA


Learning by Doing: the Digital Archive for Chinese Studies (DACHS) (download [paper](#))


Jennifer Gross
University of Heidelberg, Germany

Archiving the Czech Web: Issues and Challenges (download [paper](#))

Petr Žabicka
Moravian Library in Brno, Czech Republic

Book “Web Archiving”, Springer 2006

 SpringerLink



© 2006

Web Archiving

Authors ([view affiliations](#))
Julien Masanès

Combines the librarian's application knowledge with the computer scientist's implementation knowledge

Introduces all aspects from website monitoring to deep Web preservation

Presents an unbiased view on current standardization and preservation projects

Book

96

11k

Citations Downloads

[Download book PDF](#)

[Table of contents \(10 ch...](#)[About this book](#)[Reviews](#)

Introduction

The public information available on the Web today is larger than information distributed on any other media. The raw nature of Web content,



Julien Masanès: the father of web archiving in Europe

Arquivo.pt, 2007

Menu

Tabela

2008

Março

16 Março às 04:49

16 Março às 16:46

Outubro

2009

2010

2011

2012

2013

2014

2015

2016

2017

2018

ARQUIVO.PT

arquivo-web.fccn.pt 16 Março às 04:49, 2008

Mapa do sítio Acessibilidade Contacto

☐ apenas na secção corrente

Entrada Como participar Fui arquivado? Funcionamento Ligações Perguntas frequentes Sobre o Arquivo

Você está aqui: Entrada

English Português

Arquivo da Web Portuguesa

O Arquivo da Web Portuguesa é um projecto da Fundação para a Computação Científica Nacional que tem como principal objectivo a preservação da informação publicada na web de Portugal.

Este projecto da [Fundação para a Computação Científica Nacional](#) (FCCN) visa a criação de um sistema de arquivo de conteúdos da web portuguesa, que terá como missão recolher periodicamente, armazenar e preservar a informação publicada.

A primeira fase do desenvolvimento do Arquivo teve início em Janeiro de 2008 e prevê-se que termine no prazo de 2 anos. Contudo, a manutenção de um sistema desta natureza e a preservação da informação arquivada é uma tarefa que deverá ser perpetuada posteriormente.

A Web possibilita que cada um de nós disponibilize informação e que esteja se torne acessível a todos, sem necessidade de recurso a editoras e meios de impressão tradicionais. Diariamente, são publicados milhões de conteúdos na web como por exemplo, textos, fotografias ou vídeos. A quantidade de informação que é publicada exclusivamente na web tem vindo a aumentar rapidamente nos últimos anos. No entanto, passado relativamente pouco tempo, a grande maioria desta informação deixa de estar acessível *online* e perde-se irremediavelmente.

Surge assim o interesse no arquivo e preservação da informação publicada na Web para que o conhecimento nela contido esteja acessível às gerações futuras.

O que é a Web portuguesa?

Entende-se por Web portuguesa, todos os conteúdos alojados sob o domínio .pt.

Numa primeira fase, pretende-se arquivar apenas conteúdos alojados sob este domínio nacional, embora posteriormente se possam vir a abranger todas as páginas escritas em língua portuguesa.

Para que serve o Arquivo da Web?

Os serviços a serem prestados pelo Arquivo da Web Portuguesa ultrapassam o âmbito histórico-cultural da preservação de informação digital. A existência de um Arquivo da Web de Portugal poderá:

- Contribuir para a expansão do uso do português enquanto língua para comunicação na Web;
- Disponibilizar conteúdos de interesse às diversas comunidades científicas, por exemplo, na área da História, Sociologia ou Processamento Computacional da Língua Portuguesa;
- Contribuir para o desenvolvimento da capacidade local de tratamento e prospecção de informação publicada na Web, reduzindo a dependência nacional de serviços estrangeiros;
- Fornecer provas em casos judiciais que tenham como base informação publicada na Web.

O Arquivo da Web Portuguesa e os outros arquivos da web

TPDL 2018@Porto:

Tutorial Research the Past Web using Web Archives



Writing “The Past Web”



Objectives of the book

- Raising awareness about the **importance of preserving** information published on the Web
- New up-to-date **pedagogical resource** about the state of the art in web archiving
- Disseminate **services offered** by web archives
- Share inspiring **research work** that has explored the Past Web

The Editors



Elena Demidova



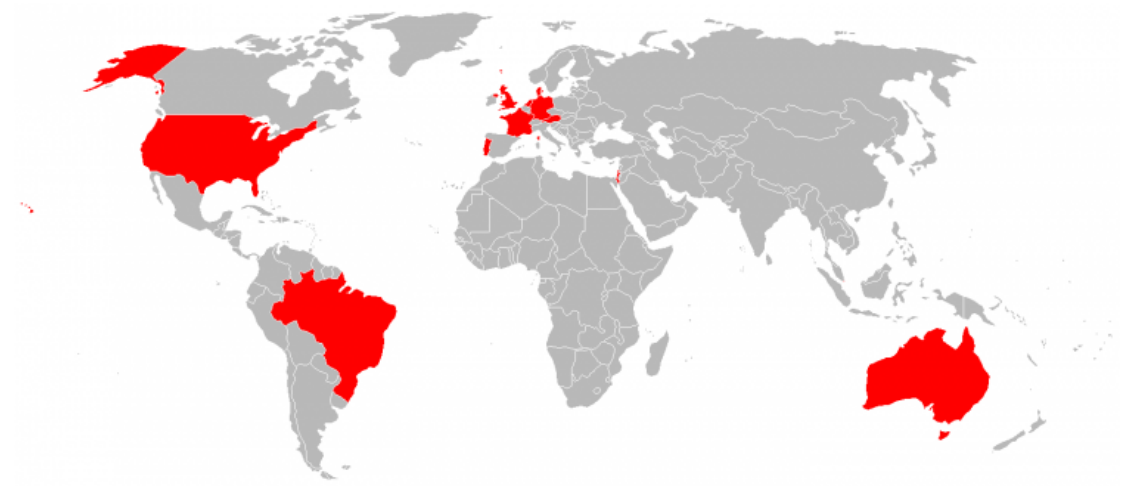
Jane Winters



Thomas Risse

40 authors from 12 countries of the World

Peter Webster
Shawn M. Jones
Ivy Huey Shin Lee
Saskia Huc-Hepher
Elena Demidova
Jane Winters
Paul Koerbin
Martin Klein
Ilya Kreymer
Miguel Won
Adam Jatowt
Naomi Wells
Dragan Espenschied
Herbert Van de Sompel
Shereen Tay
Michele C. Weigle
Sebastian Diering
Matthias Springstein
Janne Nielsen
Julien Masanès
Ralph Ewerth
Fernando van der Vlist
Eric Müller-Budack
Vitor Mangaravite
Anne Helmond
Anat Ben-David
Thomas Dugeon
Kader Pustu-Iren
Miguel Costa
Ricardo Campos
Helge Holzmann
Niels Brügger
Thomas Risse
Daniel Gomes
Arian Pasquali
Zeynep Pehlivan
Daniela Major
Michael L. Nelson



Who is this book for?

- **Professors**
to teach about web preservation
- **Researchers**
to study the past through the Web
- **Computer scientists**
to develop new applications to explore the past
- **Information professionals**
to also preserve online information
- **Citizens who use Internet (who doesn't?)**
to make the past as accessible as the present

Parte 1:

Why preserve the web?

The Era of Information Abundance and Memory Scarcity

Front Matter

Pages 1-3

The Problem of Web Ephemera

Daniela Major

Pages 5-10

Web Archives Preserve Our Digital Collective Memory

Daniela Major, Daniel Gomes

Pages 11-19

Parte 2:

How to collect information from the web?

Collecting before it vanishes

Front Matter

Pages 21-22

[National Web Archiving in Australia: Representing the Comprehensive](#)

Paul Koerbin

Pages 23-32

[Web Archiving in Singapore: The Realities of National Web Archiving](#)

Ivy Huey Shin Lee, Shereen Tay

Pages 33-42

[Archiving Social Media: The Case of Twitter](#)

Zeynep Pehlivan, Jérôme Thièvre, Thomas Durgeon

Pages 43-56

[Creating Event-Centric Collections from Web Archives](#)

Elena Demidova, Thomas Risse

Pages 57-67

Parte 3:

How to access historical web data?

Access methods to analyse the Past web

Front Matter

Pages 69-70

[Full-Text and URL Search Over Web Archives](#)

Miguel Costa

Pages 71-84

[A Holistic View on Web Archives](#)

Helge Holzmann, Wolfgang Nejdl

Pages 85-99

[Interoperability for Accessing Versions of Web Resources with the Memento Protocol](#)

Shawn M. Jones, Martin Klein, Herbert Van de Sompel, Michael L. Nelson, Michele C. Weigle

Pages 101-126

[Linking Twitter Archives with Television Archives](#)

Zeynep Pehlivan

Pages 127-139

[Image Analytics in Web Archives](#)

Eric Müller-Budack, Kader Pustu-Iren, Sebastian Diering, Matthias Springstein, Ralph Ewerth

Pages 141-151

Parte 4:

How to research the past web?

Researching the Past Web

Front Matter

Pages 153-154

[Digital Archaeology in the Web of Links: Reconstructing a Late-1990s Web Sphere](#)

Peter Webster

Pages 155-164

[Quantitative Approaches to the Danish Web Archive](#)

Janne Nielsen

Pages 165-179

[Critical Web Archive Research](#)

Anat Ben-David

Pages 181-188

[Exploring Online Diasporas: London's French and Latin American Communities in the UK Web Archive](#)

Saskia Huc-Hepher, Naomi Wells

Pages 189-201

[Platform and App Histories: Assessing Source Availability in Web Archives and App Repositories](#)

Anne Helmond, Fernando van der Vlist

Pages 203-214

Parte 5:

How to make web archives become infrastructures?

Web Archives as Infrastructures to Develop Innovative Services

Front Matter

Pages 215-216

[The Need for Research Infrastructures for the Study of Web Archives](#)

Niels Brügger

Pages 217-224

[Automatic Generation of Timelines for Past-Web Events](#)

Ricardo Campos, Arian Pasquali, Adam Jatowt, Vítor Mangaravite, Alípio Mário Jorge

Pages 225-242

[Political Opinions on the Past Web](#)

Miguel Won

Pages 243-252

[Oldweb.today: Browsing the Past Web with Browsers from the Past](#)

Dragan Espenschied, Ilya Kreymer

Pages 253-269

[Big Data Science Over the Past Web](#)

Miguel Costa, Julien Masanès

Pages 271-282

Parte 6:

The future of web archiving

A Look into the Future

Front Matter

Pages 283-283

The Past Web: A Look into the Future

Julien Masanès, Daniela Major, Daniel Gomes

Pages 285-291

Arquivo.pt preserved the cited links

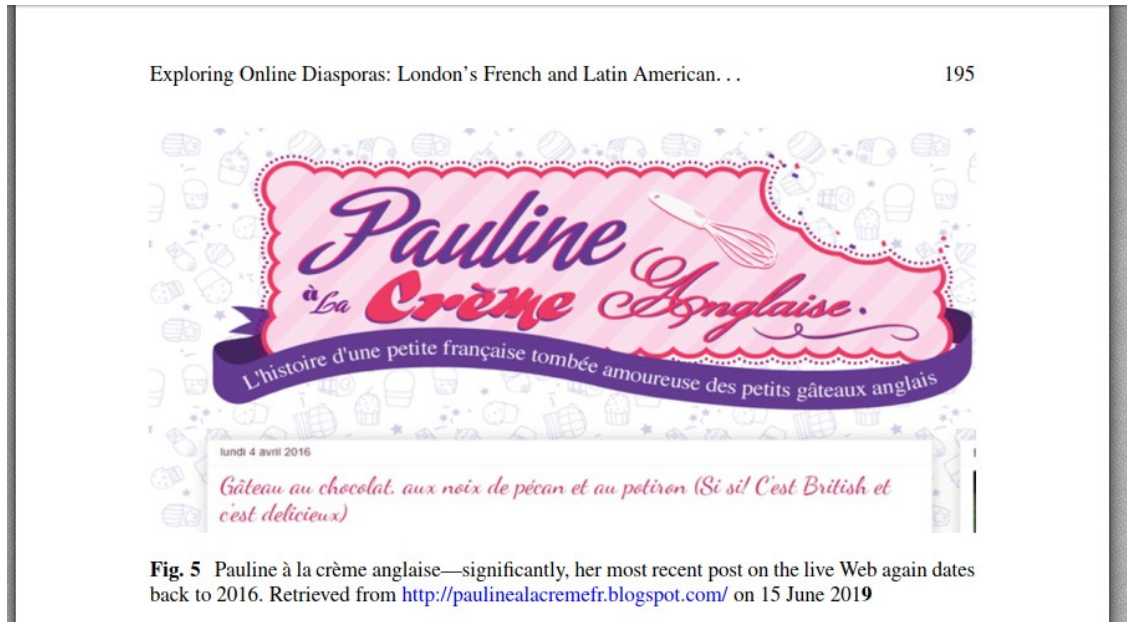
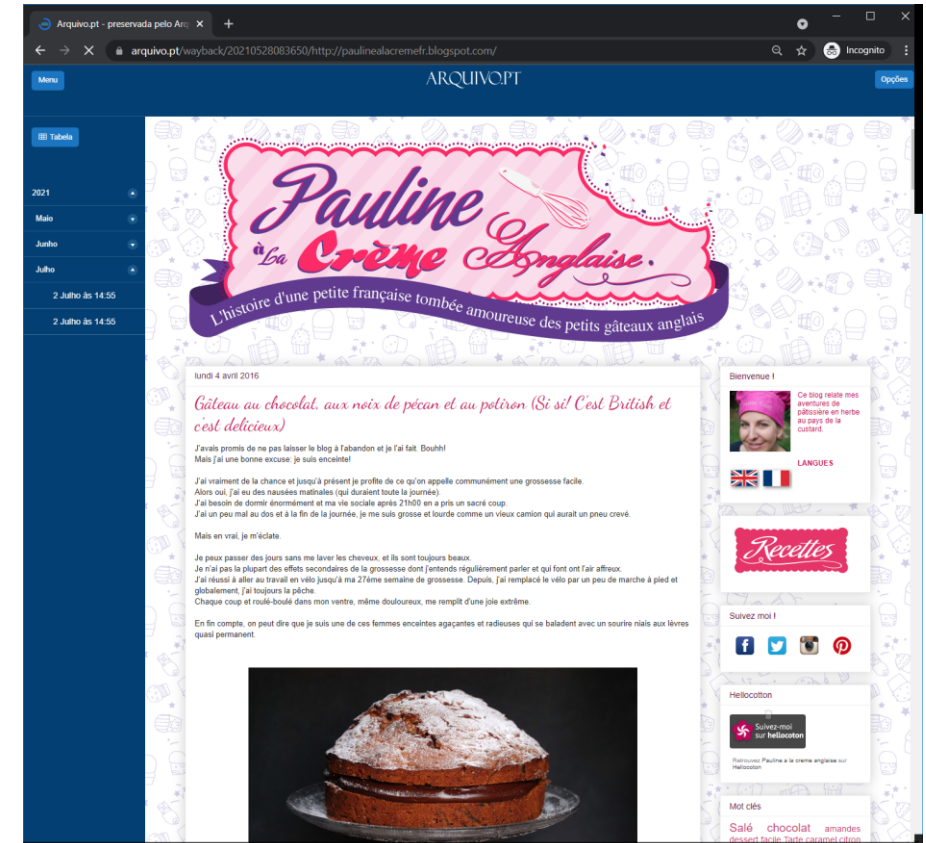
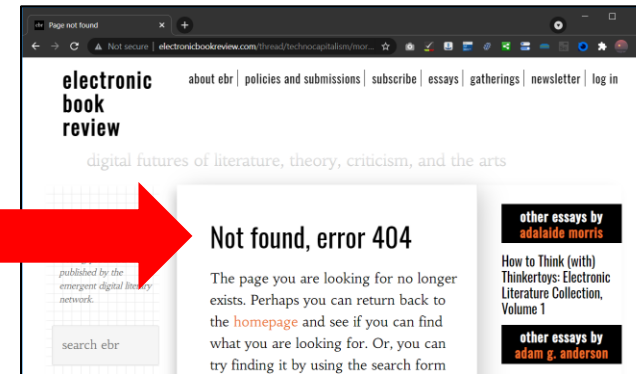


Image screenshot of part of a web page

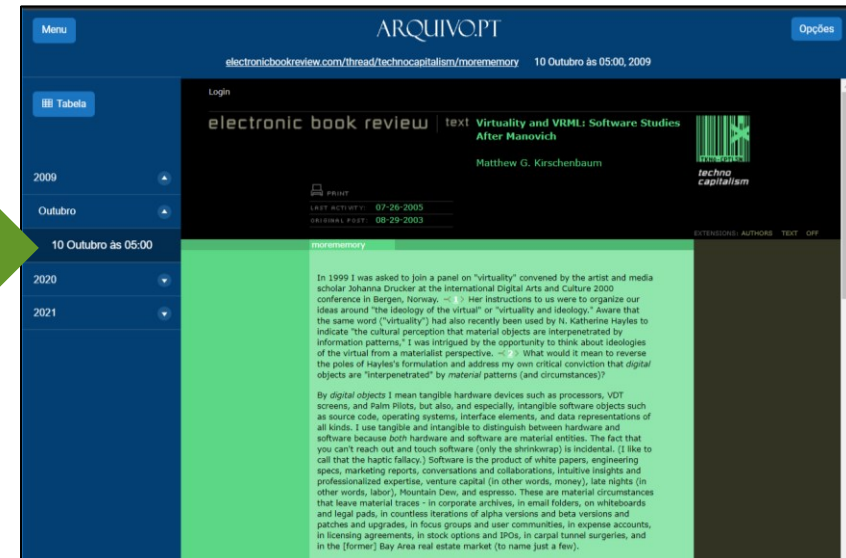


Archived web page, complete and browsable

Some citations of the book to online content are no longer available on the live-web!



But are available on the Past Web preserved by Arquivo.pt



Promising initial results

“The Past Web”:
8 400 downloads since july 2021



© 2021

The Past Web

Exploring Web Archives

Editors ([view affiliations](#))

Daniel Gomes, Elena Demidova, Jane Winters, Thomas Risse

Provides practical information about web archives, offers inspiring examples for web archivists and shares recent research results about access methods for exploring preserved information

Targets academics and advanced professionals in digital humanities, social sciences, history, media studies, and information or computer science

Serves as an initial reference for students in various areas of knowledge by introducing how to explore online history through web archives

Book

8.4k

Downloads



“Web archiving”:
11 000 downloads since 2006



© 2006

Web Archiving

Authors ([view affiliations](#))

Julien Masanés

Combines the librarian's application knowledge with the computer scientist's implementation knowledge

Introduces all aspects from website monitoring to deep Web preservation

Presents an unbiased view on current standardization and preservation projects

Book

96

11k

Citations Downloads

Read and disseminate it among your communities!

- Book available at:
www.springer.com/gp/book/9783030632908
- Preprint version available in open access at:
arquivo.pt/book

