

Publish and Preserve



Recommendations to publish preservable information

How to preserve information collected from the Web

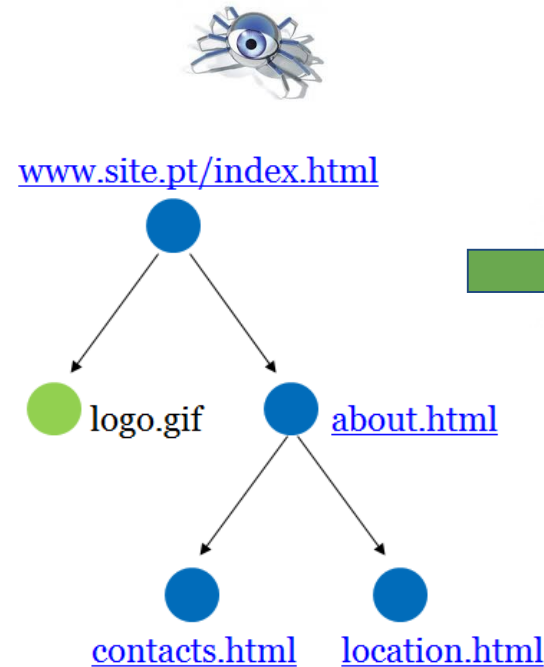
Robustify service

How do web archives work?

Web archiving workflow



Crawls are performed by Robots




INTERNET ARCHIVE



Web Archives

Replay of preserved pages



www.washingtonpost.com/

Icons: Print, Camera, Mobile, Facebook, Twitter, Email, PT

List versions

2008

2009

2010

2011

2012

2013

2014

2015

2016

January

February

March

May

June

October

November

December

1

2

3

Search

Sections

Politics Opinions Sports Local National World Business

The Washington Post

Dec. 13, 2016 Edition: U.S. & World | [Regional](#)


In the News Grindr murder Trump pledge Cindy Brady 'Daily Show' Opiates Good Samaritan Kanye West

ExxonMobil chief is Trump's choice for State despite concerns about Russia ties

- The nomination of ExxonMobil chief executive Rex Tillerson likely faces a tough fight in the Senate, where some Republicans have said they are worried about his relationship with Russian President Vladimir Putin.
- Like others in the new Trump Cabinet, Tillerson lacks any experience in government but will try to apply his experience in the business world to the realm of diplomacy.

By Steven Mufson, Philip Rucker and Karoun Demirjian · 42 minutes ago

- Wonkblog:** Trump's pick for secretary of state argued against one of the president-elect's biggest promises
- WorldViews:** If ExxonMobil were a country, it would be the



▶ Who is Rick Perry? 2:12

(Photo by Getty; video by The Post)

Perry picked for Energy Department, an agency he once vowed to abolish

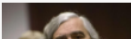
Former Texas Gov. Rick Perry is likely to shift the department away from renewable energy and toward fossil fuels, whose production he championed as Texas governor while serving for 14 years.

By Juliet Eilperin and Steven Mufson · 44 minutes ago


Joe Davidson Federal Insider

Energy Dept. rejects Trump's request to name workers

Scientists are unsure about what President-



Video




▶ Play Video 3:03

(Jason Aldag/The Washington Post)


What is the One China policy, and why is Beijing so infuriated with Trump?

UP NEXT



Politics in 'a galaxy far, far away'

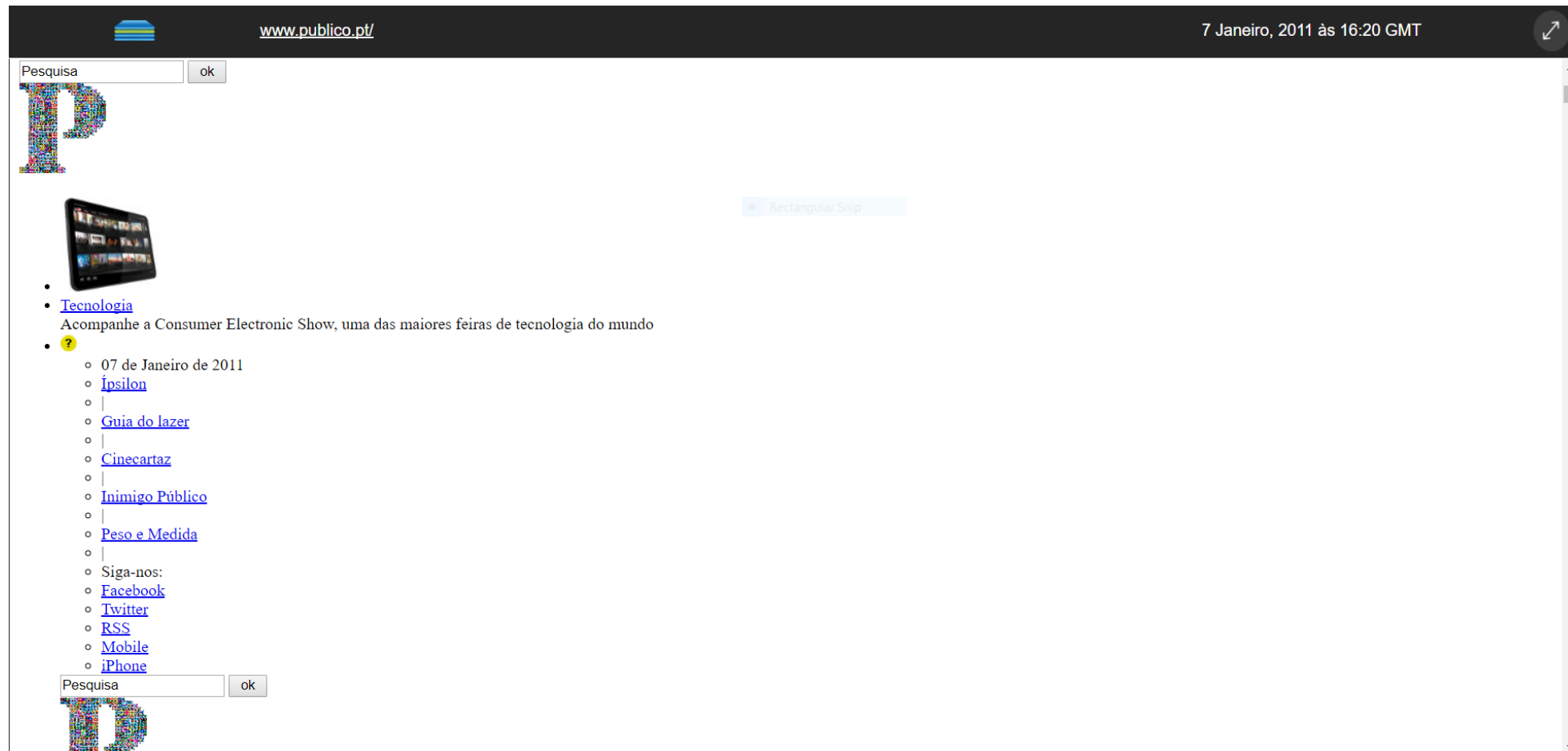
▶ Play Video 1:45



Make the side dish Obama ate at Harlem's Red Rooster restaurant

▶ Play Video 2:15

Sometimes, things go wrong. ☹️



6

**Recommendations to publish preservable
information**

Identify the publication date (#1)

Which is the publication date?



observador.pt/2015/01/07/e...no-militar/

Facebook Twitter Newsletters

2015

January 8 05:38

OBSERVADOR

INÍCIO POLÍTICA ECONOMIA PAÍS MUNDO DESPORTO CIÊNCIA CULTURA OPINIÃO VAIDADES VÍDEOS EXPLICADORES ESPECIAIS

ATAQUE EM PARIS

Charlie Hebdo: Ataque experiente com precisão "militar"

ATUALIZADO

Vários especialistas em segurança garantem que os terroristas que atacaram a redação do Charlie Hebdo tinham treino militar. Gesto que alude Estado Islâmico está a intrigar analistas.



ÚLTIMAS / MUNDO

- 23:16 O homicídio em nome de Alá que chocou a Holanda
- 22:11 "Quem teme tempestades passa a vida a rastejar"
- 19:39 Hackers derrubam sites do governo alemão
- 19:10 Rushdie: "Religião merece ser objeto de sátira"
- 19:00 Desenhadores inundam internet com homenagens
- 18:54 "Terrorismo é um dos resultados da austeridade"
- 18:45 Xeiقة Munir: "ficamos chocados com a violência"

FOTOGALERIA EM DESTAQUE



Os Cookies ajudam-nos a melhorar a sua experiência como utilizador. Ao utilizar os nossos serviços, está a aceitar o uso de cookies e a concordar com a nossa política de utilização.

Identify the publication date (#1)

ARQUIVO.PT www.diarioviseu.pt/  EN

Diário de Viseu 1ª Página da Edição de Hoje

Developed by **pictonio** www.pictonio.com

Terça, Setembro 1, 2015 - 17:30 **Business**

NOTÍCIAS SUPLEMENTOS ASSINATURAS CONTACTOS ESTATUTO EDITORIAL FICHA TÉCNICA

In Memoriam: **Adriano Lucas (1925-2011)** Director: **Adriano Callé Lucas** *Há 17 anos a informar..*

Siga-nos no Facebook

 **Diário de...**
13000 gostos

 Gostar da Página

Sê o primeiro dos teus amigos a gostar disto



FICTON é principal montra de afirmação do concelho
Edição de: Terça, Setembro 1, 2015



Mickael Carreira, Pedro Abrunhosa, Jorge Palma e Cuca Roseta são os nomes mais sonantes do cartaz deste ano da FICTON – Feira Industrial e Comercial do Concelho de Tondela, que este ano se realiza entre os dias 11 e 16 de Setembro. Serão seis dias de festa durante os quais a autarquia tondelense ... [Ler mais](#)

Vit. Setúbal empresta Forbes ao Académico
Edição de: Terça, Setembro 1, 2015

O Académico de Viseu chegou a acordo com o Vitória de Setúbal para o empréstimo do avançado Forbes até ao final da temporada. Fora das opções do técnico Quim Machado para esta época, o guineense chega a Viseu para jogar novamente na II

Publicidade

Projetos de Inovação

pictonio www.pictonio.com

Mercados

Sete conselhos de trading

As empresas americanas que mais ganharam no primeiro

Allow web archive robots (*robots.txt*) #2

DN.pt: how it was (2016)

 [Diário de Notícias](#) [Jornal de Notícias](#) [TSF](#) [O Jogo](#) [Dinheiro Vivo](#) [Volta ao Mundo](#) [Delas](#) [Classificados](#) [ASSINAR](#) [LOGIN QUIOSQUE](#) [▶ OUVIR RÁDIO](#)

Terça-Feira | 15 de março de 2016 | 11:28 | Fundado em 29 de dezembro de 1864



Diário de Notícias

 **ÚLTIMA HORA** EXPLOÇÃO DE CARRO EM BERLIM FAZ UM MORTO

Esquerda quer menos alunos por turma para reduzir indisciplina

Num grupo de 4200 alunos houve mais de 9 mil expulsões no ano passado. PS, BE e PCP de acordo: solução é ter menos crianças por...

- ↳ Manuais vão ser dados aos alunos. Mas é só um empréstimo por...
- ↳ Orçamento (real) da Educação ultrapassa os 6019 milhões

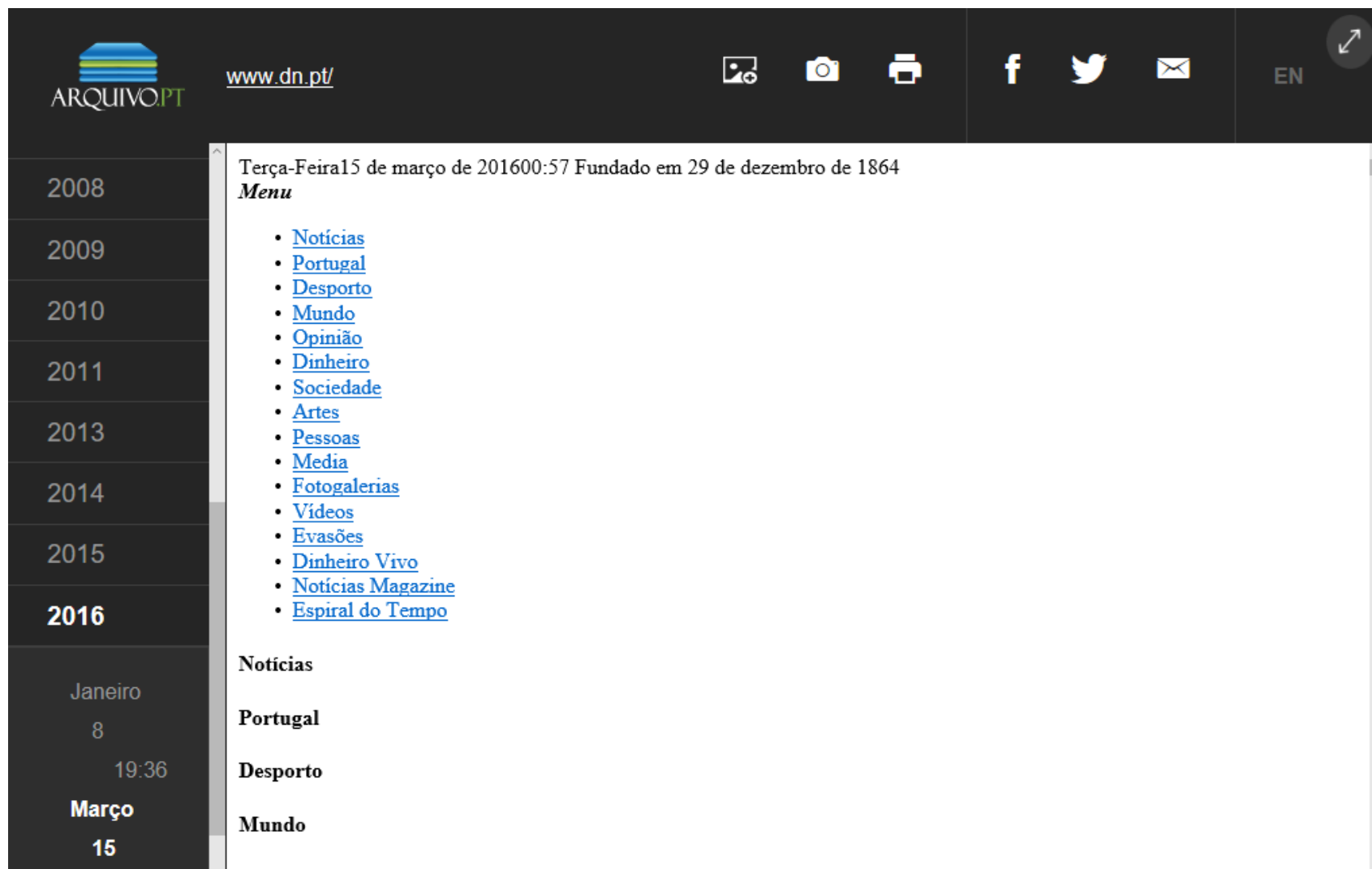


PUB

 [ASSINAR DN](#)

DN.pt: how it was preserved



The screenshot displays the DN.pt website interface. On the left is a dark sidebar with a year and date selector. The main content area has a dark header with the 'ARQUIVO.PT' logo, the URL 'www.dn.pt/', and various utility icons. Below the header, the main content area shows the date 'Terça-Feira 15 de março de 2016 00:57' and the founding date 'Fundado em 29 de dezembro de 1864'. A 'Menu' section lists various categories with blue links. Below the menu, the first few news items are visible, including 'Noticias', 'Portugal', 'Desporto', and 'Mundo'.

ARQUIVO.PT www.dn.pt/

Terça-Feira 15 de março de 2016 00:57 Fundado em 29 de dezembro de 1864

Menu

- [Notícias](#)
- [Portugal](#)
- [Desporto](#)
- [Mundo](#)
- [Opinião](#)
- [Dinheiro](#)
- [Sociedade](#)
- [Artes](#)
- [Pessoas](#)
- [Media](#)
- [Fotogalerias](#)
- [Vídeos](#)
- [Evasões](#)
- [Dinheiro Vivo](#)
- [Notícias Magazine](#)
- [Espiral do Tempo](#)

Noticias

Portugal

Desporto

Mundo

2016

Janeiro

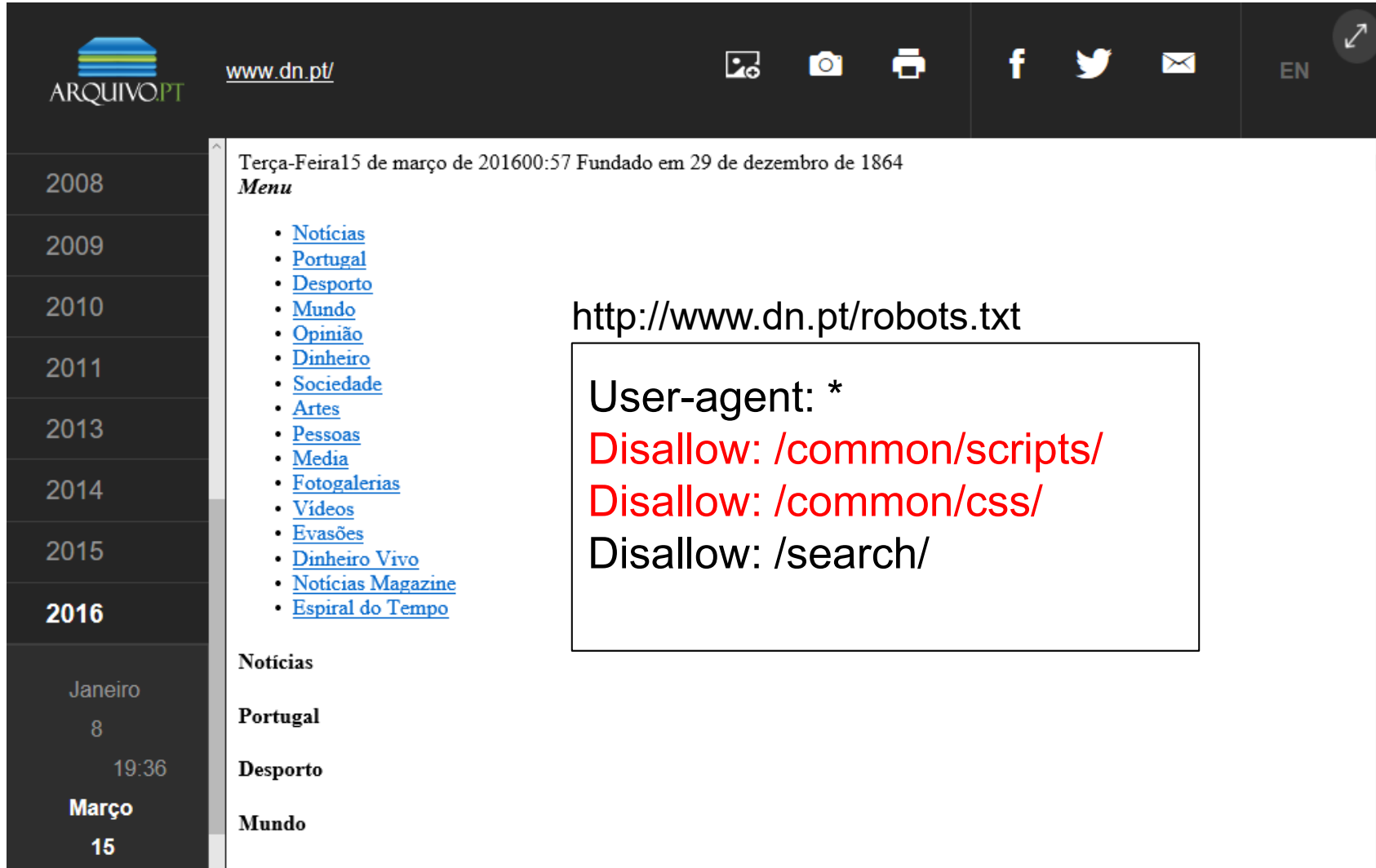
8

19:36

Março

15

Robots Exclusion Protocol was the source of the problem



The screenshot shows the DN.pt website interface. On the left is a sidebar with a date selector showing 'Março 15'. The main content area has a dark header with the 'ARQUIVO PT' logo, the URL 'www.dn.pt/', and social media icons. Below the header, the date 'Terça-Feira 15 de março de 2016 00:57' and the founding date 'Fundado em 29 de dezembro de 1864' are displayed. A 'Menu' section lists various categories like 'Notícias', 'Portugal', 'Desporto', etc. To the right of the menu, the URL 'http://www.dn.pt/robots.txt' is shown, followed by a box containing the robots.txt file content:

```
User-agent: *  
Disallow: /common/scripts/  
Disallow: /common/css/  
Disallow: /search/
```

"Disallowing crawling of Javascript or CSS files in your site's robots.txt directly harms how well our algorithms render and index **your content and can result in suboptimal rankings."**

source: **Google Webmaster Central Blog, [Updating our technical Webmaster Guidelines](#)**

B-on.pt: how it was preserved

The screenshot displays the B-on.pt website interface as it appeared in 2009. On the left, a dark sidebar features a calendar for June 2009, with the 22nd highlighted at 19:38. The main content area has a top navigation bar with the URL <http://www.b-on.pt/> and social media icons. Below this is a search bar with the placeholder text 'expressao'. A red bar with the word 'English' is positioned above the main content. The main content area includes a section titled 'O que é a b-on?' with links for 'Missão e objectivos', 'Regras de utilização', and 'Contacto'. Below these links is an RSS 2.0 feed icon. The central part of the page features a search interface with tabs for 'Pesquisa rápida', 'MetaPesquisa', 'Recursos', 'Periódicos', and 'Área Pessoal'. The 'Pesquisa rápida' tab is active, showing two search fields with the placeholder text 'introduza uma expressao' and a search button labeled 'pesqui:'. Below the search fields is a section titled 'Seleccione o conjunto onde deseja efectuar a pesquisa:' with radio buttons for 'Artes e Humanidades', 'Ciências Sociais', 'Ciências', 'Ciências da Saúde', and 'Eng./Tecnologia'. The 'Trial ebrary' section follows, featuring the Ebrary logo and a text block stating that the Ebrary platform of ebooks is available in trial for all B-on members until July 15, 2009. It provides a link to access the platform and mentions that it contains over 170,000 books. At the bottom of the page, there are four logos: 'Ministério da Ciência', 'UMIC', 'FCCN', and 'União Europeia'.

<http://www.b-on.pt/>

b-on Biblioteca do Conhecimento Online

expressao

English

O que é a b-on?

Missão e objectivos

Regras de utilização

Contacto

RSS 2.0

Pesquisa rápida | MetaPesquisa | Recursos | Periódicos | Área Pessoal

Campo: Todos os campos introduza uma expressao E pesqui:


Campo: Todos os campos introduza uma expressao (Exemplos)




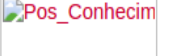

Seleccione o conjunto onde deseja efectuar a pesquisa:

☐ Artes e Humanidades ☐ Ciências Sociais ☐ Ciências

☐ Ciências da Saúde ☐ Eng./Tecnologia

Trial ebrary

 A plataforma de ebooks da ebrary está disponível, em trial para todos os membros b-on, até ao dia 15 de Julho 2009. Para aceder a esta plataforma de ebooks verifique por favor, [aqui](#), o endereço URL que foi atribuído à sua Instituição. Para que possa usufruir, de forma mais rápida e eficaz das potencialidades de pesquisa e leitura em mais de 170.000 livros, dispõe de um pequeno vídeo em inglês no endereço <http://www.ebrary.com/corp/collateral/flash/unityh/unityh.html>

Default exclusions by the Content Management Systems (CMS) are the source of the problem



User-agent: *
Disallow: /administrator/
Disallow: /cache/
Disallow: /components/
Disallow: /editor/
Disallow: /help/
Disallow: /images/
Disallow: /includes/
Disallow: /language/
Disallow: /mambots/
Disallow: /media/
Disallow: /modules/
Disallow: /templates/
Disallow: /installation/
Disallow: /dmdocuments/

B-on.pt: *robots.txt* correctly configured

The screenshot displays the B-on.pt website interface. On the left is a dark sidebar containing a calendar for June, July, August, and September, with the date 18th highlighted. The main content area has a dark header with the URL 'http://www.b-on.pt/' and social media icons. Below the header is a banner with the 'b-on' logo and the text 'biblioteca do conhecimento online'. The main content area is divided into sections: 'Início' (Home) with links to 'O que é a b-on?', 'b-on de A-Z', 'Projectos e Serviços', and 'Adesão'; 'Pesquisa Rápida' (Quick Search) with a search bar and a dropdown menu; 'Pesquisa Avançada' (Advanced Search) with a search bar and a dropdown menu; 'Ajuda?' (Help?) with links to 'Guia do Utilizador', 'FAQs', 'Formação b-on', 'Tutoriais', and 'e-learning'; and 'Como posso?' (How can I?) with a dropdown menu and a button 'Pergunte ao seu bibliotecário!'. The right sidebar features a section 'A b-on para...' (B-on for...) with categories: ESTUDANTES, PROFESSORES, INVESTIGADORES, PROFISSIONAIS DE SAÚDE, and BIBLIOTECÁRIOS. Below this is a section 'Queremos melhorar!' (We want to improve!) with a link 'Envie-nos a suas sugestões' and a section 'Comunidade b-on' (B-on community) with the text '35 Utilizadores'.

http://www.b-on.pt/

english Mapa Contactos Subscrição de Alertas Regras de utilização

b-on biblioteca do conhecimento online

À distância de um clique!

Home

Início
O que é a b-on?
b-on de A-Z
Projectos e Serviços
Adesão

Alertas b-on

Serviço de Pesquisa »

e-Learning b-on

Estamos no RCAAP

Pesquisa Rápida

Insira a expressão de pesquisa em Seleccione o conjunto de pesquisa OK

A pesquisa será efectuada em todos os campos (Assunto, Título, Autor, ISSN, ISBN e Ano)

Ajuda?

- Guia do Utilizador »
- FAQs »
- Formação b-on »
- Tutoriais »
- e-learning »

Como posso?

Seleccionar a opção

Pergunte ao seu bibliotecário!

Destaques

Taylor & Francis - aviso

Logo Taylor & Francis

Por questões de lentidão dos seus servidores o editor Taylor & Francis foi retirado dos conjuntos de pesquisa do portal b-on. Assim, todos aqueles que queiram pesquisar neste editor terão de o fazer directamente na interface do mesmo ou através do item Recursos do serviço de pesquisa b-on.

A b-on para...

ESTUDANTES

PROFESSORES

INVESTIGADORES

PROFISSIONAIS DE SAÚDE

BIBLIOTECÁRIOS

Queremos melhorar!
Envie-nos a suas sugestões »

Comunidade b-on
35 Utilizadores

Allow access to Web Archives

Arquivo.pt robot is correctly **identified**.

User-agent: Arquivo-web-crawler
Disallow:



**An Universal Resource Locator for each
resource (#3)**

Resources hidden behind forms

[CANDIDATOS](#) [ESTUDANTES](#) [ALUMNI](#) [EMPRESAS](#) 

 **INSTITUTO UNIVERSITÁRIO DE LISBOA**



[ESTUDAR ▾](#) [INVESTIGAR ▾](#) [INTERNACIONAL ▾](#) [AGENDA ▾](#) 

[O ISCTE-IUL ▾](#)

bolsas



[GERAL](#) [PESSOAS](#) [DOCUMENTOS](#)

3 RESULTADOS

 ISCTE-IUL

 [Formulário \(Bolsas de 3.º Ciclo\)](#)

 [Requerimento \(Bolsas de 3.º Ciclo\)](#)

 [Regulamento de Atribuição de Bolsas Internas a Estudantes do 3º Ciclo](#)

Resources hidden behind forms



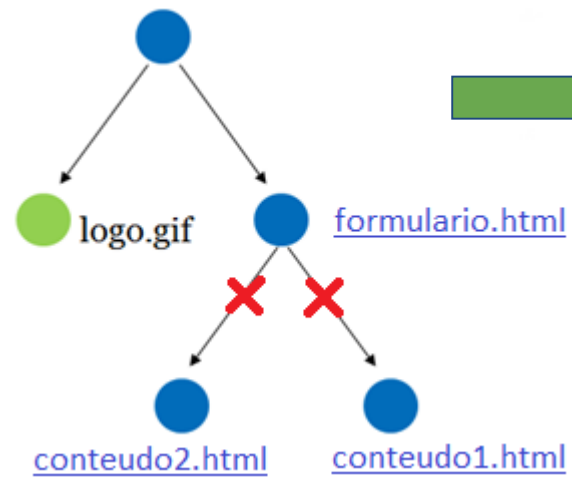
Não Arquivado

Pesquisar noutros arquivos

Resources hidden behind forms



www.site.pt/index.html



Site Maps help people (usability) and machines (SEO) to access the information



- ↳ Parlamento
- ↳ Presidente
- ↳ Deputados e Grupos Parlamentares
- ↳ Comissões Parlamentares
- ↳ Intervenções e Debates
- ↳ Fiscalização Política
- ↳ Actividade Parlamentar e Processo Legislativo
- ↳ Relações Internacionais
- ↳ Orçamento do Estado e Contas Públicas
- ↳ Revisões Constitucionais
- ↳ Diário da Assembleia da República
- ↳ Gestão do Parlamento
- ↳ Dossiers Temáticos
- ↳ Arquivo e Documentação
- ↳ Legislação
- ↳ Livraria Parlamentar

Mapa do Site

[Página Inicial](#)

- [Presidente](#)
- [Acessibilidade](#)
- [Administrador](#)
- [Contactos](#)
- [Correio](#)
- [Correio AHP](#)
- [Correio BIB](#)
- [Correio CICRP](#)
- [Correio DAC](#)
- [Correio DAPAT](#)
- [Correio DAPLEN](#)
- [Correio DE](#)
- [Correio DGF](#)
- [Correio DILP](#)
- [Correio DP](#)
- [Correio DRAA](#)
- [Correio DRHA](#)
- [Correio DRI](#)
- [Correio DSAF](#)
- [Correio DSATS](#)
- [Correio DSDIC](#)

Maintain the same URL along time (#4)

Problem: broken timeline

Versions list												
463 versions of iscte.pt												
http://iscte.pt												
2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
22 Jan	28 Jan	3 Feb	7 Jan	13 Jan	4 Jan	6 Jan	20 Mai					
25 Mar	2 Feb	21 Mar	12 Jan	2 Feb	9 Jan	18 Jan	20 Mai					

Versions list												
34 versions of http://iscte-iul.pt												
http://iscte-iul.pt												
2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
								6 Jun	21 Jan	21 Jan	5 Nov	26 Set
								10 Jun	22 Jan	22 Jan	7 Nov	30 Set
								4 Aug	20 Mai			27 Out
								5 Aug	21 Mai			
									2 Jul			

Single unbroken timeline

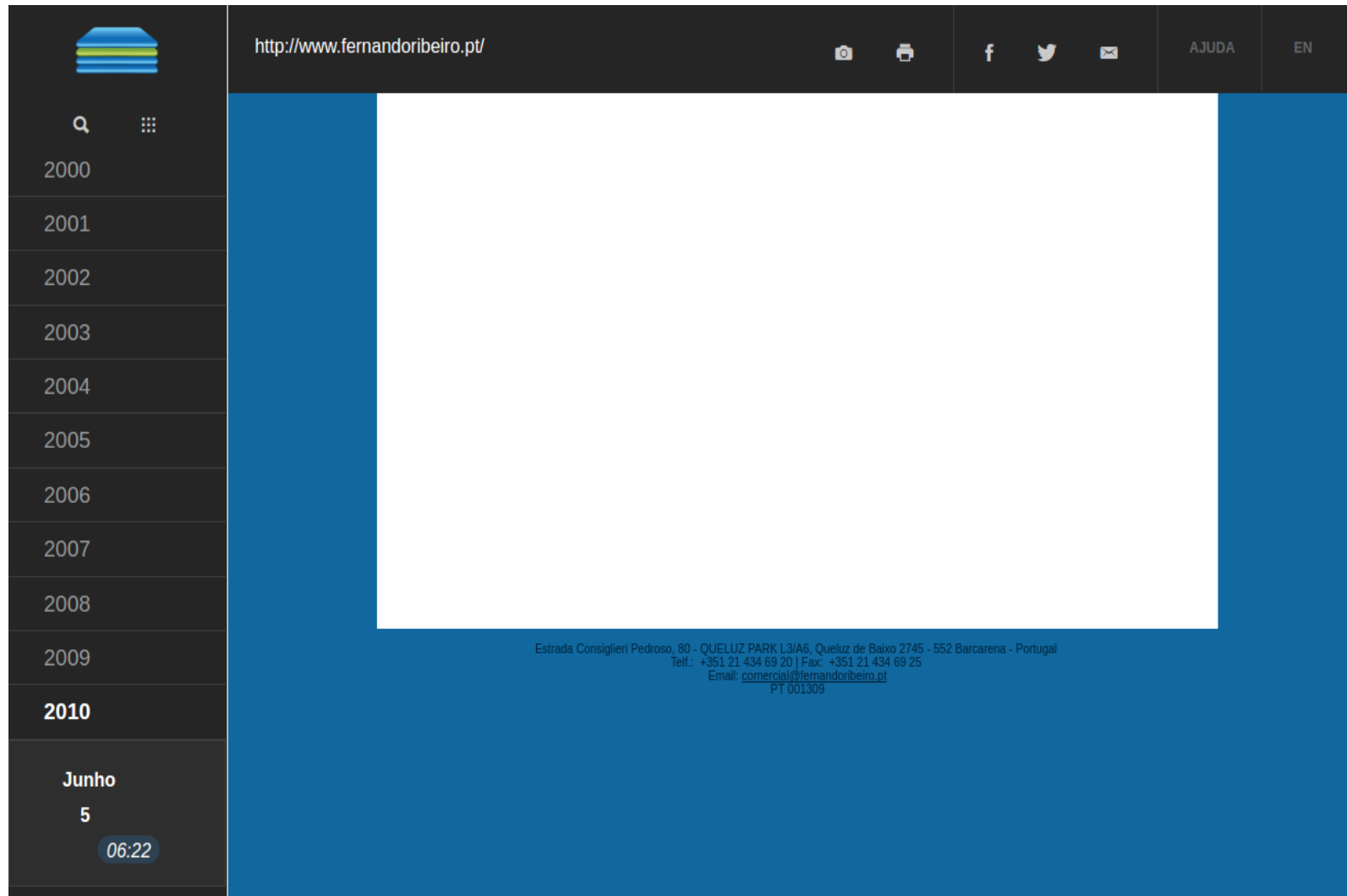
Versions list

463 versions of iscte.pt

2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
22 Jan	28 Jan	3 Feb	7 Jan	13 Jan	4 Jan	6 Jan	20 Mai	31 Mai	21 Jan	21 Jan	5 Nov	26 Set
25 Mar	2 Feb	21 Mar	12 Jan	2 Feb	9 Jan	18 Jan	20 Mai	6 Jun	20 Mai	22 Jan	7 Nov	30 Set
25 Mai	18 Feb	21 Abr	13 Jan	6 Feb	13 Jan	26 Jan	21 Mai	5 Aug	22 Mai			
27 Mai	24 Mar	5 Jun	1 Feb	7 Feb	19 Jan	28 Jan	23 Jun					
5 Jun	27 Mar	11 Jun	4 Feb	2 Mar	21 Jan	6 Feb	24 Set					
24 Set	9 Abr	11 Jun	4 Feb	2 Abr	24 Jan	15 Feb	26 Set					
27 Set	10 Abr	12 Jun	8 Feb	25 Abr	27 Jan	11 Mar	17 Dez					
13 Nov	10 Abr	14 Jun	8 Feb	27 Abr	5 Feb	14 Mar	18 Dez					
26 Nov	31 Mai	23 Jun	9 Feb	2 Jun	20 Feb	21 Out						

Use preservable data formats (#5)

Problem: Flash usage



“It **breaks** with the Web’s **fundamental** interaction principles”

NN/g Nielsen Norman Group

Evidence-Based User Experience Research, Training, and Consulting

Search

[HOME](#) [TRAINING](#) [CONSULTING](#) [REPORTS](#) [ARTICLES](#) [ABOUT NN/G](#)

Topics

[E-commerce](#)

[Intranets](#)

[Mobile & Tablet](#)

[User Testing](#)

[Web Usability](#)

[See all topics...](#)

Author

[Jakob Nielsen](#)

[Don Norman](#)

[Bruce "Tog" Tognazzini](#)

[See all authors...](#)

Recent Articles

Flash: 99% Bad

by [JAKOB NIELSEN](#) on October 29, 2000

Topics: [Web Usability](#)

Summary: Although multimedia has its role on the Web, current Flash technology tends to discourage usability for three reasons: it makes bad design more likely, it breaks with the Web's fundamental interaction style, and it consumes resources that would be better spent enhancing a site's core value.

About 99% of the time, the presence of Flash on a website constitutes a usability disease. Although there are rare occurrences of good Flash design (it even adds value on occasion), the use of Flash typically lowers usability. In most cases, we would be better off if these multimedia objects were removed.

Flash tends to degrade websites for three reasons: it encourages design abuse, [it breaks with the Web's fundamental interaction principles](#), and it distracts attention from the site's core value.

Mission Manifesto Join the Fight!



The movement to rid the world of the Flash Player plugin

<http://www.occupyflash.org/>

Choose a media format with:

Freely available, without legal rights that restrict its use.

A standard issued by an official organization (e.g. [W3C](#)).

Openly documented through a [public and free specification](#).
Widely used.

Read and written by several software platforms, including [open-source](#) technologies.

Not compressed or compressed without information loss.

Preservable media formats

Text

HTML, XHTML ou XML

Open Document Text (.odt)

PDF/A-1 ISO 19005-1 (.pdf)

Image

PNG (.png)

JPEG2000

Video

AVI without compression (.avi)

Not preservable media formats

Text

Microsoft Word (.doc)

Images

Macromedia Flash (*.swf)

PhotoShop (.psd)



Video

Windows Media Video (.wmv)

Metadata to describe content (#6)

Who is the author?

When was it published?

www.ciberevora.pt/ensinar/...exto02.htmlEN

Tabela

1997

Julho

12

03:58

1998

Internet: Um pouco de História...

Mas afinal... de onde é que a Internet surgiu?

Bem, a Internet "evoluiu" a partir de um sistema criado em 1969. Nessa altura o departamento de defesa do governo Americano chegou à conclusão de que necessitavam de desenvolver um sistema para permitir a troca de informação entre cientistas e fontes militares de todo o país.

Um sistema de 4 computadores foi então criado, de nome ARPANET.

O sistema evolui rapidamente, o seu nome mudou para ARPANET, e em 1972 já tinha crescido de 4 para 37 computadores. Ao mesmo tempo a sua utilização foi também mudando. Além de trocarem informação técnica com fins militares e científicos, os utilizadores da ARPANET começaram a trocar correio electrónico (e-mail), algumas vezes de conteúdo importante, outras vezes de conteúdo trivial...

As coisas continuaram, e em 1983 a ARPANET já tinha crescido de tal forma que foi decidido criar uma rede separada, chamada MILNET, para os fins militares que a ARPANET entretanto deixou de servir.

Todo o tipo de instituições e companhias começaram gradualmente a ganhar interesse na Internet... e em 1990 a Internet como nós a conhecemos começou...

Uma verdadeira explosão que teve a sua origem na invenção da W.W.W. (World Wide Web) - a forma mais popular de navegar na Internet. O crescimento da Internet tem sido fenomenal. Em 1985 tinha cerca de 5 mil utilizadores. Neste momento tem cerca de 45 milhões...

É um crescimento de mais de 6000% !!!! E isto é só o começo...

Metadata usage (Dublin Core)

```
<meta name="DC.Type" content="Text" />  
<meta name="DC.Creator" content="Daniel Gomes" />  
<meta name="DC.Date.Created" content="2009-08-21" />  
<meta name="DC.Date.Modified" content="2009-11-10" />
```

This information summarizes, **enrich** or completes the contents adding more value to the information.


Computers can use this information.

Recommendations to avoid preservation problems

1. Identify the **publication date**
2. **Allow** web archive robots
3. An **Universal Resource Locator** for each **resource**
4. Maintain the **same URL along time**
5. Use **preservable** data formats
6. Metadata to **describe** content


How preservable is a website?


<http://archiveready.com>


 **ArchiveReady**
website archivability evaluation tool


Checking website: <http://www.dn.pt/>


[Check again »](#)

 Summary

 **HTML and CSS** 13


 **HTTP** 4


 **Media** 13

 **Sitemaps** 3

Overall Rating

65%

 One page printable HTML

 Results in EARL XML format

Results

	Rating
Web Archivability Facet	
Accessibility	50%
Cohesion	65%
Metadata	80%
Standards Compliance	65%

About

Web ar
these p
informa
website
standar

It must
Archiva
of Web

Do yo

<http://archiveready.com>



ArchiveReady

website archivability evaluation tool

Checking website: <http://www.dn.pt/>

<http://www.dn.pt/>

Check again »

 Summary

 HTML and CSS 13

 HTTP 4

 Media 13

 Sitemaps 3

Evaluation complete!  1 Correct  2 Errors

01.  There are "Disallow:" commands in your robots.txt file.

Disallow: /common/scripts/

Disallow: /common/css/

Disallow: /admin/

Disallow: /search/

Disallow: commands are used to block web bots from specific website areas. They may

 Accessibility

Reference: <http://www.robotstxt.org/robotstxt.html>

Create your own Web Archive

Web Recorder

Webrecorder

Collect & Revisit the Web

URL to capture

Add to collection

Login to add to Collection...

Select browser

Use Current Browser ▼

▸ Session settings

Start

<https://webrecorder.io>

A Project By
RHIZOME

Capturing

2.58 MB

Current Browser

http://www.tpd1.eu/tpdl2018/



REGISTRATION

PROGRAM +

WORKSHOPS

CONFERENCE INFORMATION +

TRAVEL INFORMATION +

CONTACT

TPDL 2018



Porto **22nd International Conference**
Portugal **on Theory and Practice**
Sep. 10-13 2018 **of Digital Libraries**

FACULTY OF ENGINEERING, UNIVERSITY OF PORTO

Digital Libraries for Open Knowledge



When

The TPD1 2018 Conference
will take place from Monday,



How

Interested in participating in
the TPD1 2018 Conference?



Who

Want to know more about
the people behind the TPD1



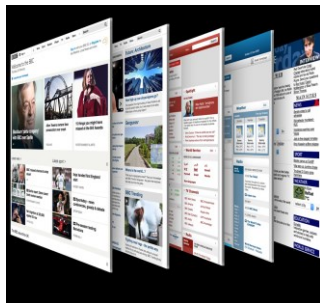
Where

The TPD1 2018 Conference
will be held at the Faculty of



Twitter

Tweets by @tpdl2018



1

Select

Webrecorder

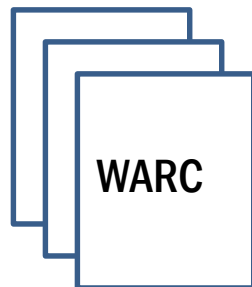
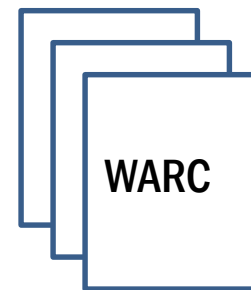


2

Capture

WARC (Web ARChive)

ISO 28500:2017



3

Replay



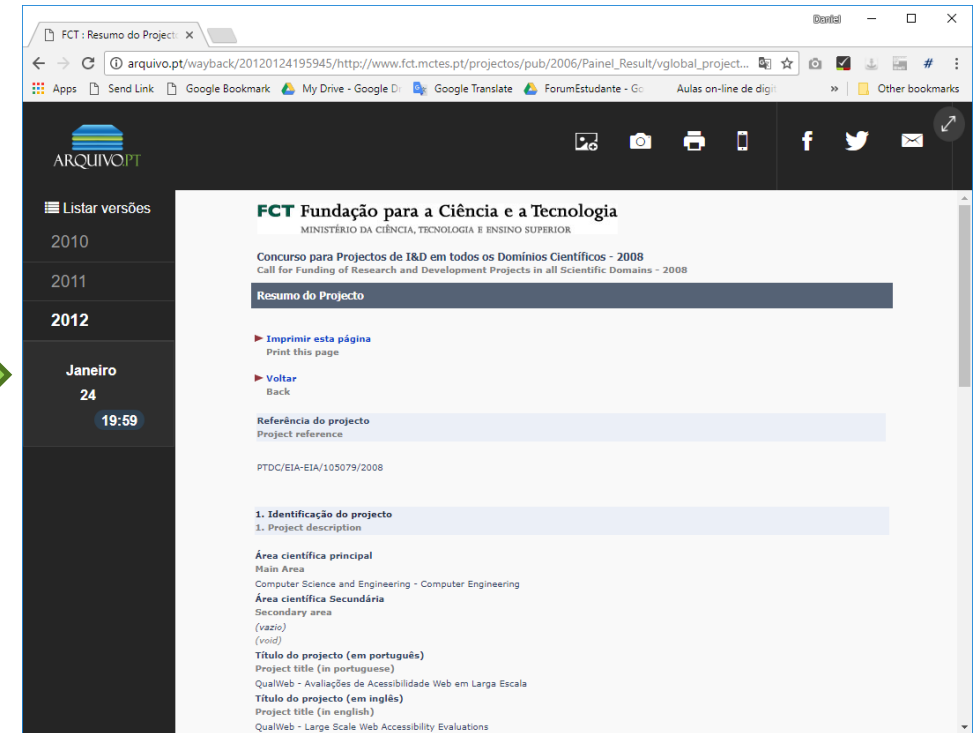
Robustify

Minimizes errors due to broken links, redirecting to a web-archived resources

robustify.arquivo.pt: broken links redirect to preserved pages @Arquivo.pt



Live-web page with broken link



Users follow link to preserved page
at Arquivo.pt

How to use robustify in your website

Insert this snippet of code in your website:

```
<script src="http://robustify.arquivo.pt/robustifyArquivoPT.js"></script>  
<script> robustify({}); </script>
```

Recommendations:
arquivo.pt/recommendations
daniel.bicho@fccn.pt

