

Preservação de sítios Web no contexto das Bibliotecas do Ensino Superior

Ricardo Basílio

Curador Digital - Arquivo.pt





Objectivo

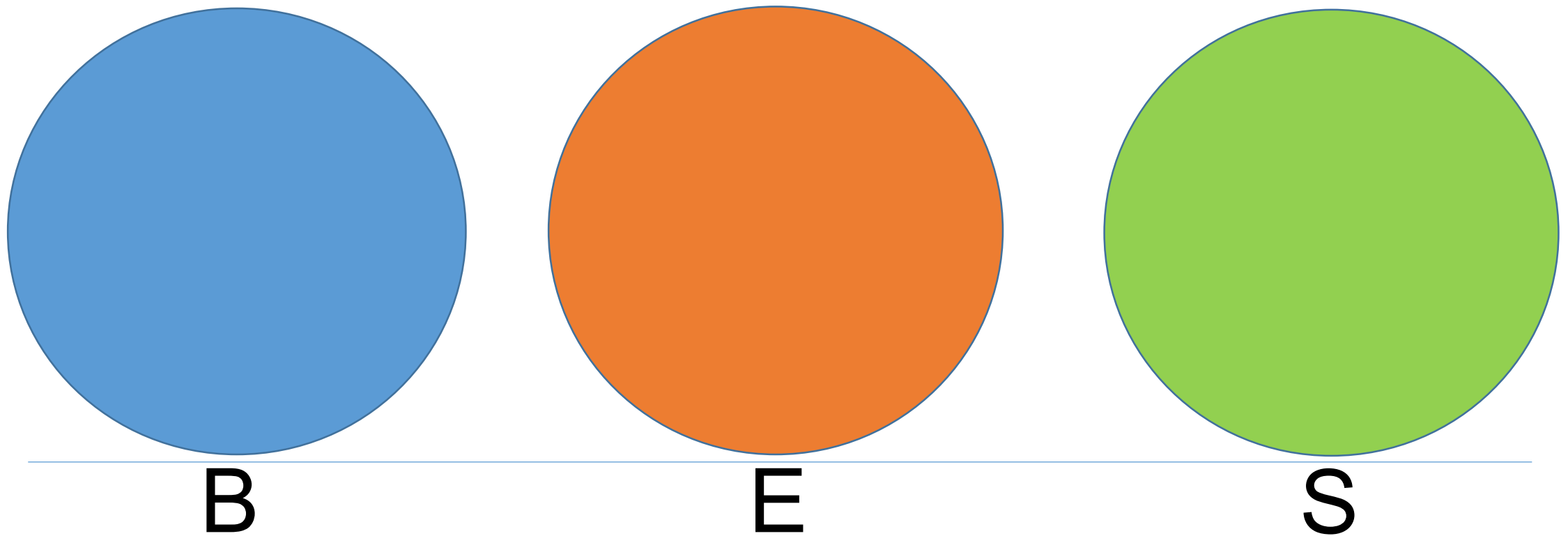
Aplicar processos de curadoria aos sítios Web institucionais

Agenda

- 1. As Bibliotecas do Ensino Superior como comunidade de suporte**
- 2. Como se faz para preservar os sítios Web institucionais**
- 3. Sugestões de reutilização de conteúdos preservados**

1. As Bibliotecas do Ensino Superior como comunidade de suporte

As Bibliotecas do Ensino Superior



Uma pequena viagem

- Pessoas
 - Organizações
 - Projectos
 - Tecnologia
- e vocabulário**



Tim Berners-Lee

CC 0. Créditos: Henry Tomas



Brewster Kahle

CC BY. Wikimedia

Pessoas, Organizações, Projectos, Tecnologia



IIPC

International Internet Preservation Consortium

URL: <http://netpreserve.org/>

Pessoas, Organizações, Projectos, Tecnologia



Rhizome

New Museum – New York

URL: <http://rhizome.org/>

Pessoas, Organizações, Projectos, Tecnologia



Folha Excel

Uma lista de URL's é a base para qualquer projecto



Webrecorder.io

Aplicação local para reprodução dos ficheiros WARC (funciona offline, é portátil, a rapidez depende do tamanho dos ficheiros. Descarregue última versão e execute:

URL: <https://github.com/webrecorder/webrecorderplayer-electron/releases/download>

Outros sistemas de recolha:

Web Curator Tool, Brozzler

Pessoas, Organizações, Projectos, Tecnologia

e vocabulário

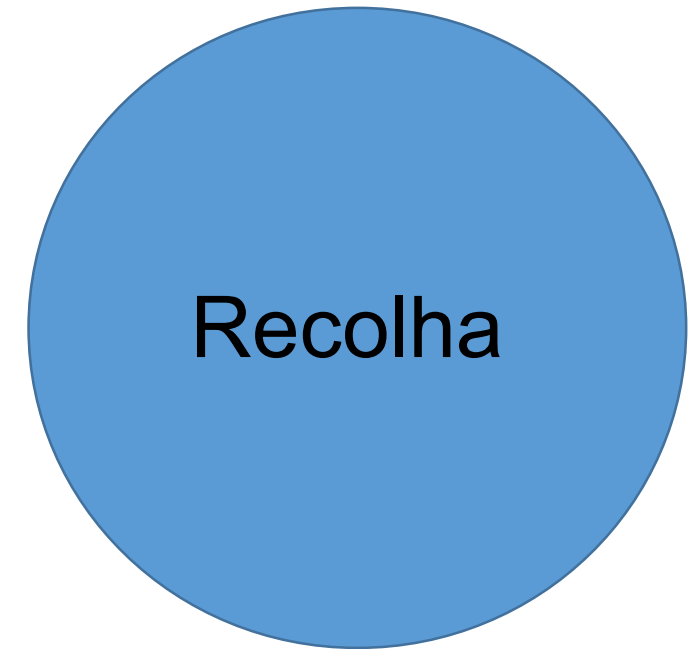
#Webarchiving

#digitalpreservation

#curadoriadigital

#memoriaweb

Web Crawler



Pessoas, Organizações, Projectos, Tecnologia

e vocabulário

WARC

ISO 28500:2017



Pessoas, Organizações, Projectos, Tecnologia

e vocabulário

wayback



Timestamp

14 dígitos

Exemplo (Natal à meia noite):

20181225000000

Pessoas, Organizações, Projectos, Tecnologia

e vocabulário

URLs

“seeds”

`http://ihc.fcsh.unl.pt/noticias/`



Protocol



Hostname + Domain



Path

URL com timestamp

<http://arquivo.pt/wayback/20030215185550/http://ihc.fcsh.unl.pt/>

Pessoas, Organizações, Projectos, Tecnologia

e vocabulário

Website

Sítio Web

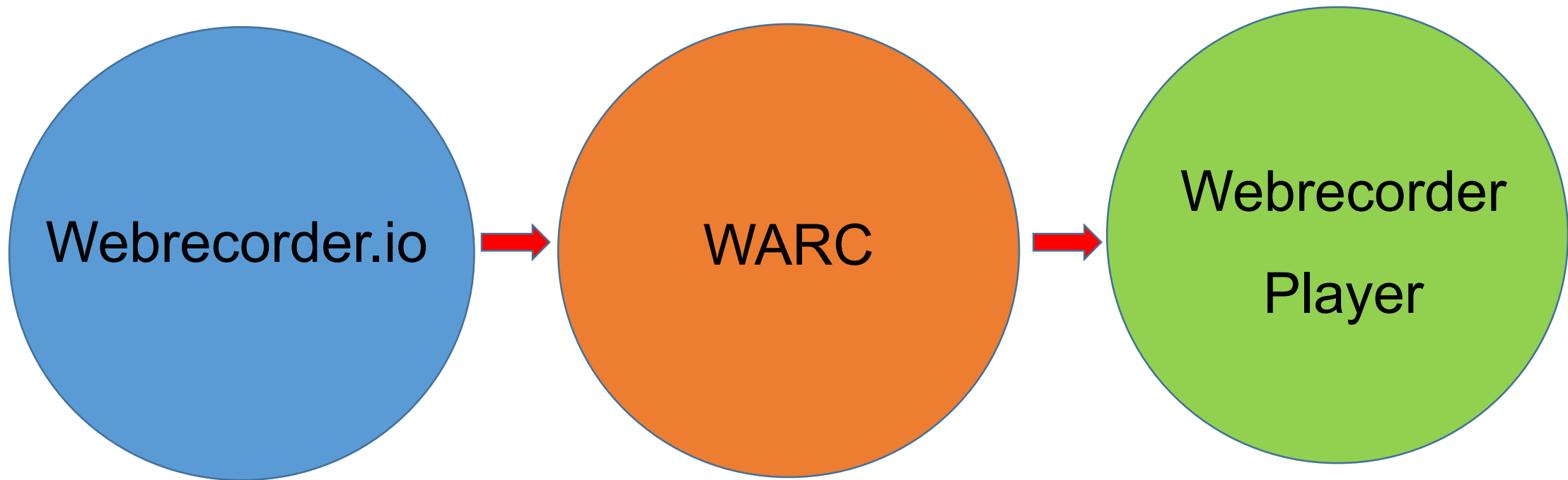


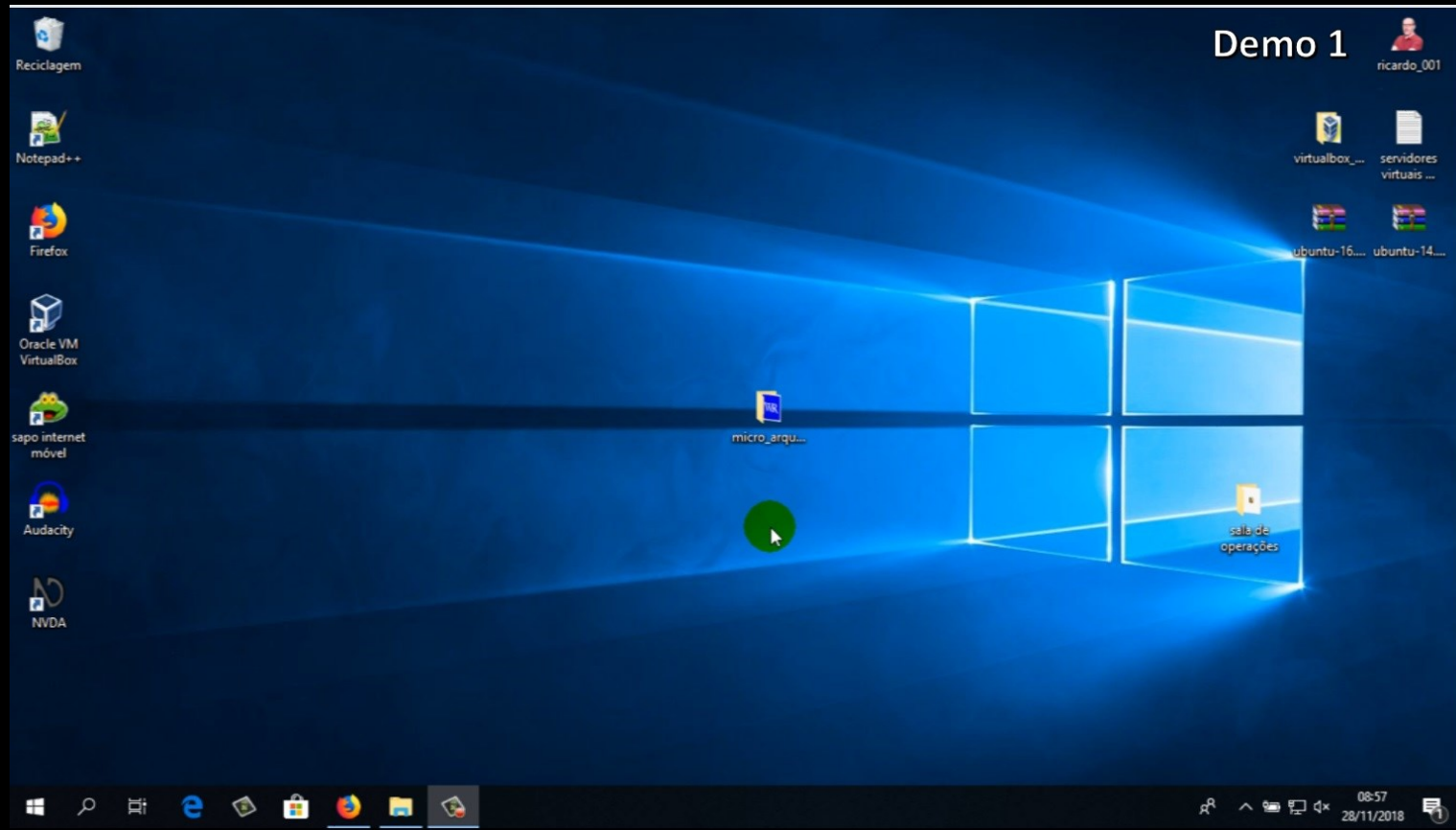
2. Como se faz para preservar os sítios Web institucionais

Como se faz para preservar

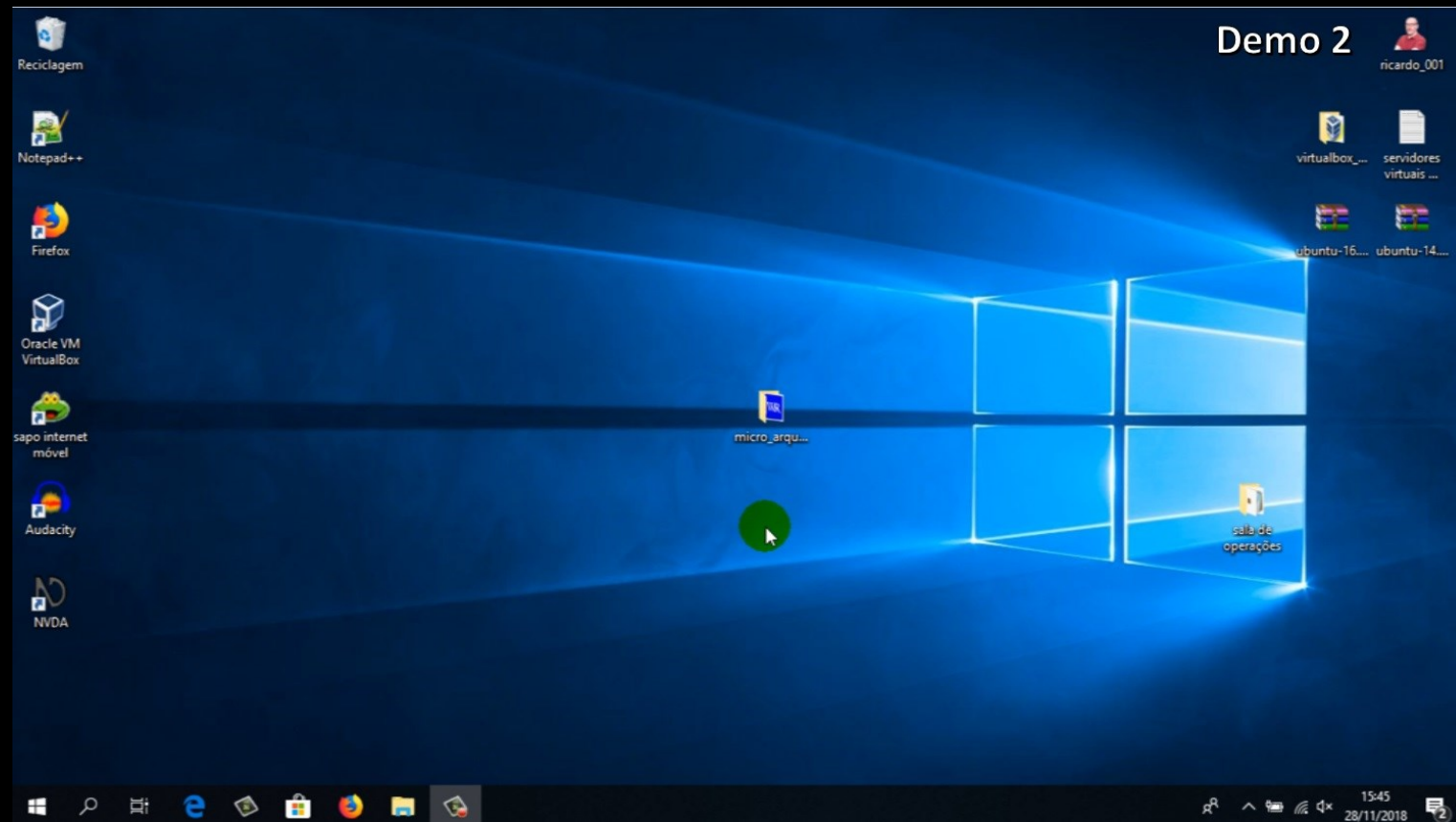


Como se faz para preservar

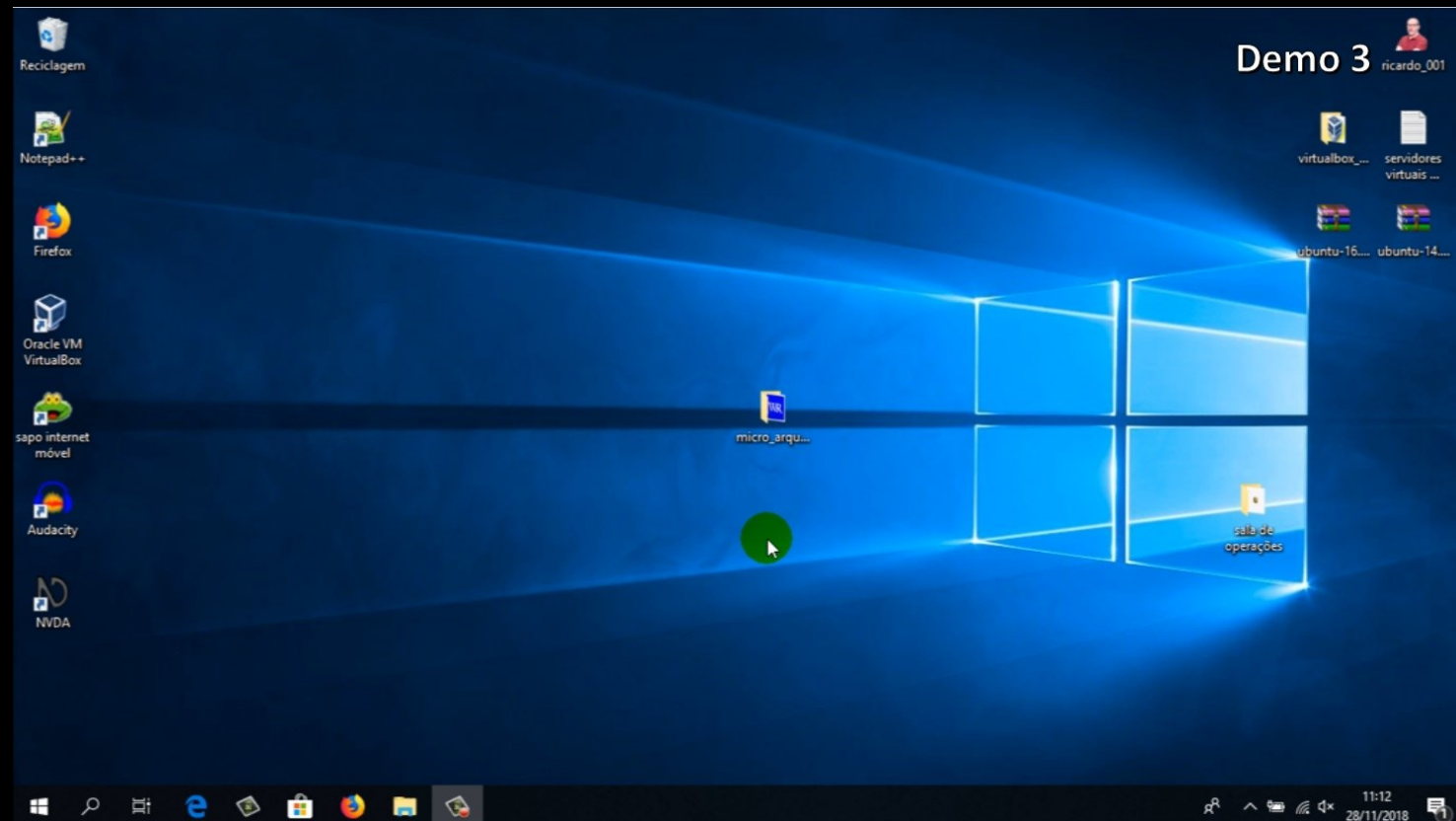




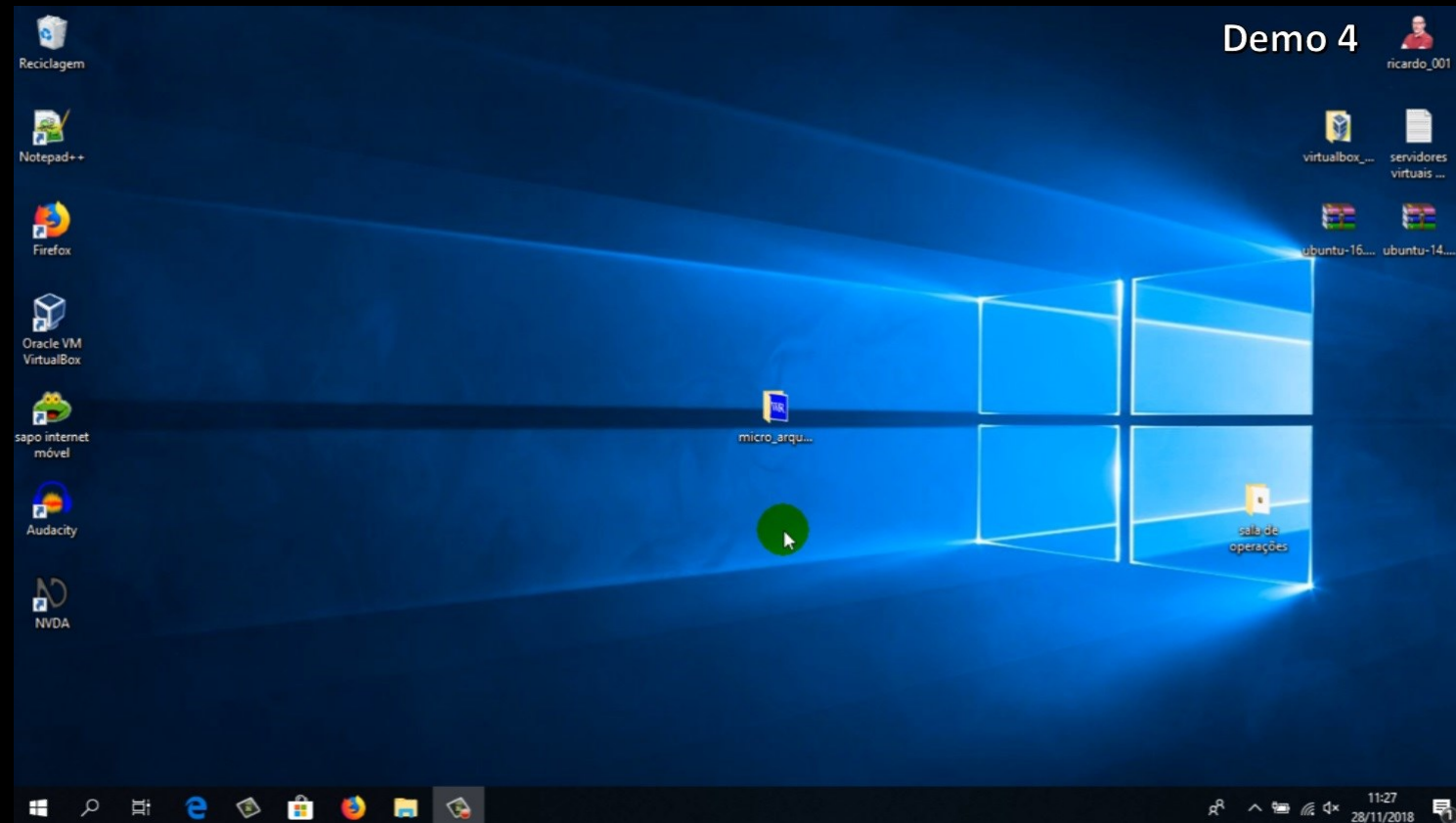
Preservar de forma personalizada
com alta qualidade



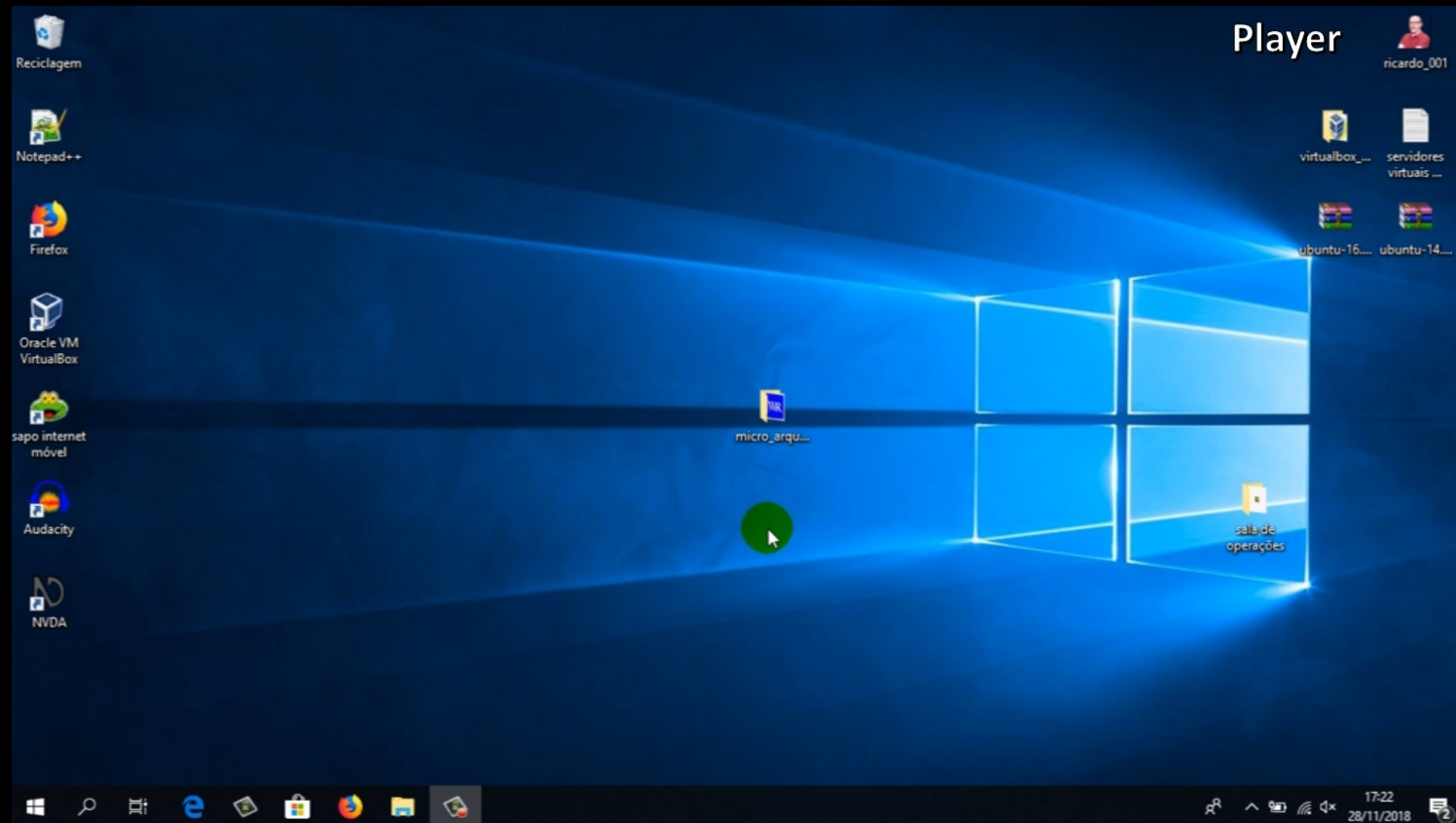
Melhorar as recolhas do passado



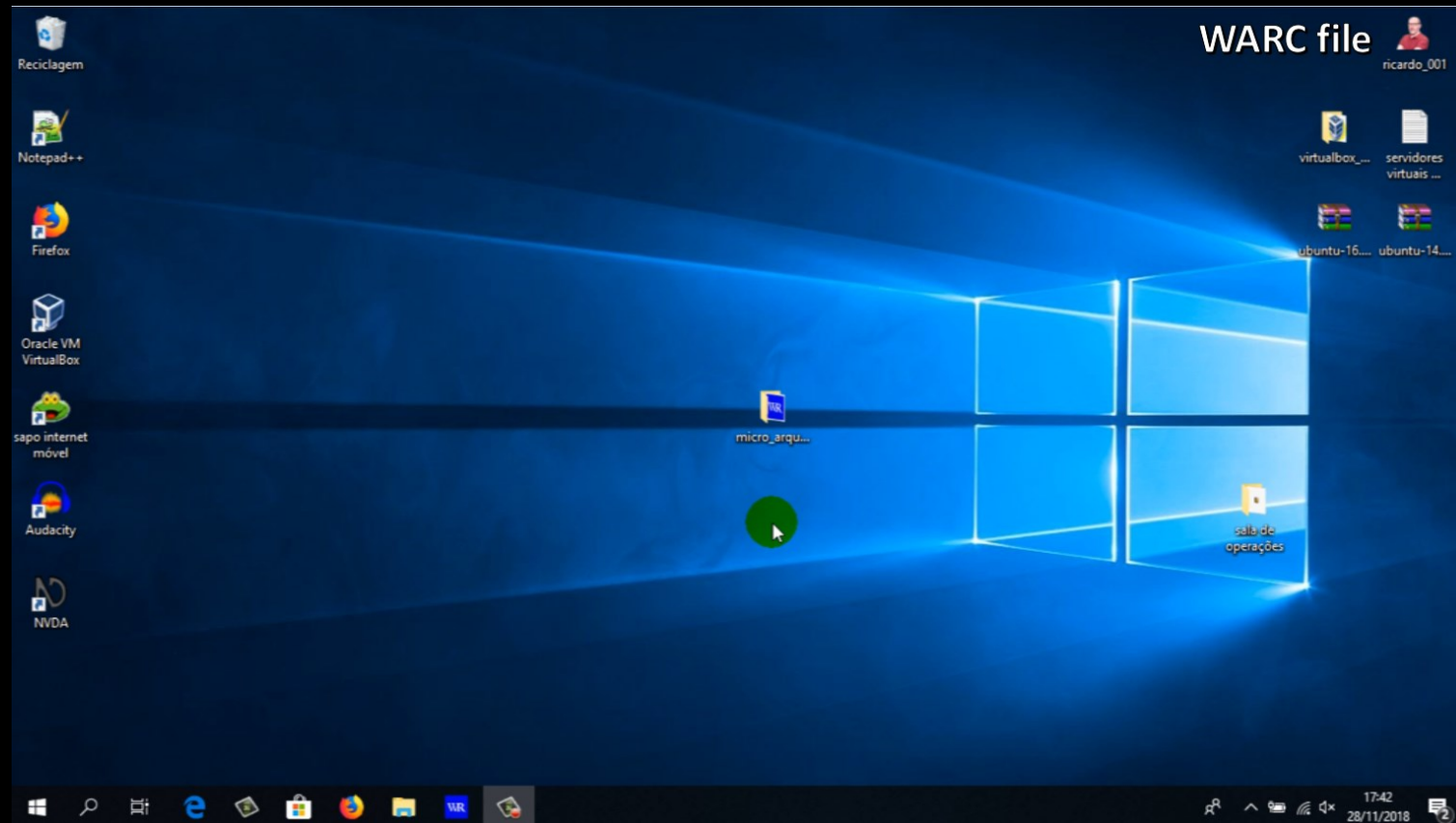
Gravar os canais sociais – Facebook e outros



Gravar os canais sociais – Facebook e outros



Gravar vídeos embebidos nos sites



Que informação está num ficheiro WARC?

Como se faz para preservar

Name	Date modified	Type	Size
 gpsiglas-20180919140621.warc	19/09/2018 15:11	File folder	
 gpsiglas-20180919140621.warc.gz	19/09/2018 15:07	gz Archive	53 094 KB
 memoriafcsh-20180919143445.warc.gz	19/09/2018 15:35	gz Archive	28 417 KB
 rec01-20180911164130.warc.gz	11/09/2018 17:43	gz Archive	289 124 KB
 rec02-20180910094031.warc.gz	10/09/2018 10:41	gz Archive	125 826 KB
 rec03-20180910120403.warc.gz	10/09/2018 13:04	gz Archive	66 921 KB
 rec04-20180910141253.warc.gz	10/09/2018 15:14	gz Archive	112 505 KB
 rec05-20180910151231.warc.gz	10/09/2018 16:14	gz Archive	173 719 KB
 rec06-20180910154740.warc.gz	10/09/2018 16:48	gz Archive	78 646 KB
 rec07-20180911101747.warc.gz	11/09/2018 11:19	gz Archive	283 249 KB
 rec08-20180911105343.warc.gz	11/09/2018 11:57	gz Archive	464 097 KB
 rec09-20180911111201.warc.gz	11/09/2018 12:12	gz Archive	30 441 KB
 rec10-20180911150115.warc.gz	11/09/2018 16:01	gz Archive	88 077 KB
 rec11-20180911155746.warc.gz	11/09/2018 16:58	gz Archive	116 408 KB
 rec12-20181116143714.warc	16/11/2018 14:39	WARC File	252 973 KB
 rec13-20181116143627.warc	16/11/2018 14:37	WARC File	70 974 KB
 rec15-20181116143540.warc	16/11/2018 14:35	WARC File	16 336 KB
 rec16-20181116142819.warc	16/11/2018 14:29	WARC File	383 894 KB
 rec17-20181116142623.warc	16/11/2018 14:27	WARC File	184 381 KB

3. Sugestões de reutilização de conteúdos preservados

Sugestões de reutilização

- **Publicação de conteúdos preservados sobre Faculdade**
- **Lista de sítios Web institucionais preservados**
- **Exposição sobre a memória online da Faculdade**

Reutilização dos conteúdos preservados

Redes Sociais

<http://facebook.com/memoriawebparatodos>



Reutilização dos conteúdos preservados

Lista de sites preservados

Nome do site (considerado como canal de comunicação de uma determinada entidade)	Endereço(s) (estes URLs tendem a ser longos podem ser escondidos em links, com aqueles que se apresentam na coluna 3)	Apresentação dos endereços em forma de link (<i>hostname</i> e <i>intervalo temporal</i> em que foram utilizados)
Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa (1997-2017)	• http://arquivo.pt/wayback/*/http://fcsh.unl.pt/	• fcsh.unl.pt, 1997-2017
Centro de Estudos Históricos (2000-2016)	• http://arquivo.pt/wayback/*/http://www2.fcsh.unl.pt/ceh/home.html/ • http://arquivo.pt/wayback/*/http://ceh-unl.tripod.com/ • http://arquivo.pt/wayback/*/http://www2.fcsh.unl.pt/ceh	• fcsh.unl.pt/ceh, 2000-2006 • ceh-unl.tripod.com, 2007-2008 • fcsh.unl.pt/ceh, 2009- 2016

Reutilização dos conteúdos preservados

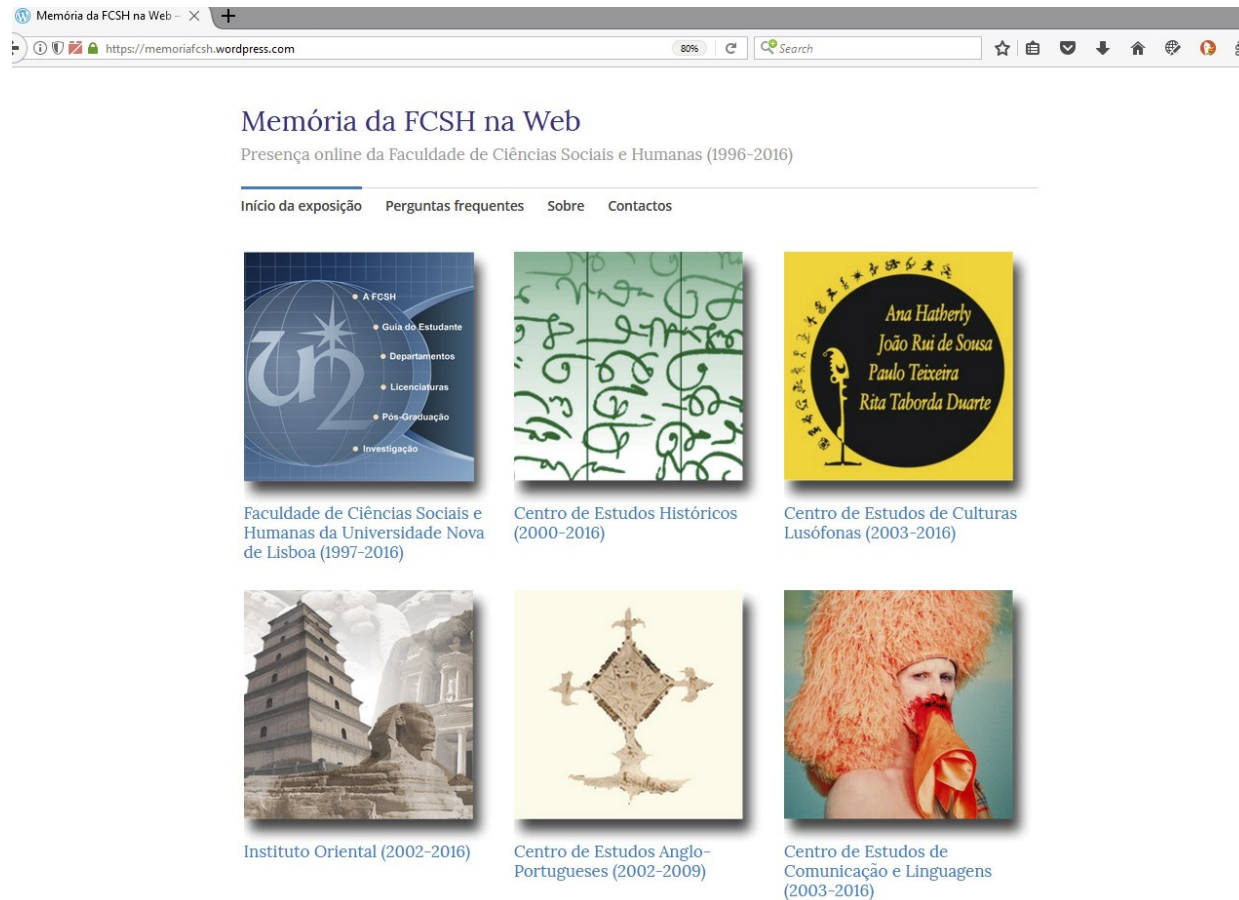
MemoriaFCSH.wordpress.com

Exposição online sobre a memória institucional

Tutorial: <https://youtu.be/yIOYZH9UErA>

Reutilização dos conteúdos preservados

Homepage



Reutilização dos conteúdos preservados

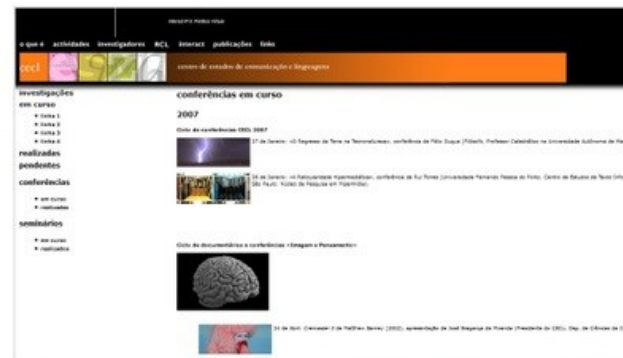
Item page

Memória da FCSH na Web

Presença online da Faculdade de Ciências Sociais e Humanas (1996-2016)

[Início da exposição](#) [Perguntas frequentes](#) [Sobre](#) [Contactos](#)

Centro de Estudos de Comunicação e Linguagens (2003-2016)



Pormenor da página "Actividades" (cecl.unl.com.pt, 2003)

Reutilização dos conteúdos preservados

Item page

Memória do sítio Web

A [primeira versão preservada é de 2003](#). Alguns elementos gráficos estão em Flash, o que limita a sua reprodução em vários dispositivos actuais. Em [2008](#), surge com um *design* ligeiramente diferente.

Em 2014, sítio Web, já com um *design* diferente, o sítio Web informa que o CECL [passou a integrar o centro inter-universitário CIC.DIGITAL](#), tornando-se seu pólo na FCSH.

O sítio Web do CECL foi um ponto de acesso para a Revista de Comunicação e Linguagens, a revista online Interact e o sítio Working papers, recursos que tiveram destaque na sua página de entrada. Foi também um suporte para registo de actividades correntes e passadas, projectos em curso ou finalizados.

Endereço

- [cecl.com.pt, 2003-2016](#)

Momentos para recordar

- [Revista Comunicação e Linguagens, 2006](#)
 - [Working papers, 2008](#)
 - [CD-ROM Cultura Digital numa notícia do site, 2008](#)
-

Reutilização dos conteúdos preservados

Lista de endereços

<http://cesem.fcsh.unl.pt/>

<http://fcsh.unl.pt/cesem>

<http://www.cetaps.com/>

<http://cham.fcsh.unl.pt>

<http://cicdigitalpolo.fcsh.unl.pt/pt/home/>

<http://www.cics.nova.fcsh.unl.pt/>

<https://institutodehistoriadaarte.wordpress.com/>

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

<http://www.clunl.edu.pt/>

<http://cria.org.pt/wp/>

<http://www.iap.fcsh.unl.pt/>

<http://negem.weebly.com/>

<http://cicdigitalpolo.fcsh.unl.pt/pt/home/>

<https://inovamedialab.org/>

<http://www.cecl.com.pt/pt/publicacoes/colecao-cultura-media-e-arte>

<http://cicdigitalpolo.fcsh.unl.pt/pt/cmj/>

<http://ojs.letras.up/index.php/prismacom/index>

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

<http://www.fcsh.unl.pt/inovtec/index.php/pt/projecto/>

<http://www.zas.gwz-berlin.de/cost/index.html>

<http://www.oneness.vu.lt>

<http://cria.org.pt/wp/nava/>

<https://necicria.wordpress.com/>

<https://laboh.jimdo.com/>

<http://etnografica.revues.org/>

<http://www.grandepequeno.com/>

<http://grandepequeno.blogspot.com/>

http://www.womenwriters.nl/index.php/Portuguese_Women_Writers

<http://litescape.ielt.fcsh.unl.pt/>

<http://hebrewilluminationinportugal.weebly.com/>

<http://www.cham.fcsh.unl.pt/ext/catedra/index.html>

<http://projecto-iem.wix.com/eicam>

<http://www.zotero.org/histmedport>

<http://heraldica.hypotheses.org/2606>

<https://seminarioquestoesdeestetica.wordpress.com/>

<http://morphology.squarespace.com/>

<http://linguagemeforcassquarespace.com/>

<http://wittgensteinpi.squarespace.com/>

<http://postnationaleu.squarespace.com/>

<http://filmphilosophy.squarespace.com/>

<http://arglab.squarespace.com/>

<http://cjpimi.ifilnova.pt/>

<http://photofilmiha.weebly.com/>

(continua...)

Reutilização dos conteúdos preservados

DADOS PARA DESCRIÇÃO DO SÍTIO WEB PRESERVADO	O que se pretende	Elementos utilizados
Nome	Nome ou título do sítio Web para apresentar na exposição	Sim
Contexto institucional	A quem pertence ou que entidade está nele representada. No caso dos <i>sites</i> de projectos e eventos um <i>site</i> pode pertencer a mais do que uma entidade.	Sim
<i>Hostname</i>	<i>hostname(s)</i> , que um sítio Web de uma entidade teve ao longo do tempo e indicação do intervalo temporal correspondente	Sim
Versões guardadas no Arquivo.pt	<i>Links</i> para as versões guardadas pelo Arquivo.pt, um por cada <i>hostname</i>	Sim
Imagem de destaque	<i>Printscreen</i> de uma página preservada para utilizar como imagem na exposição, legenda e URL da página	Sim
História do sítio Web	Nota descritiva das mudanças mais notórias	Sim

Reutilização dos conteúdos preservados

Memória Web da Imprensa Portuguesa
Exposição de páginas preservadas para recordar o passado e projectar o futuro

[Início](#) [No fundo do baú](#) [Sobre](#) [Contactos](#)

Tag: Aveiro



O Jomal de Estarreja (2009-2018)



Soberania do Povo (2002-2018)



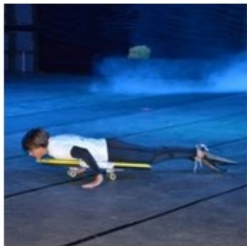
Correio da Feira (2013-2018)



Jornal da Bairrada (2000-2018)



Diário de Aveiro (2001-2018)



Defesa de Espinho (2014-2018)

Jornais estabelecidos por Distrito

Síntese

Síntese

- Há ferramentas disponíveis para recolhas de alta qualidade em pequena escala
- O Webrecorder permite experimentar localmente o processo de recolha, armazenamento e replay de conteúdos Web para efeitos de preservação
- As colecções constituídas não são inúteis. Bem pelo contrário, pela sua especialização, podem acrescentar algo às recolhas genéricas que já são feitas pelo Arquivo.pt
- O contributo das Bibliotecas do Ensino Superior para a preservação dos sítios Web é importante para toda a comunidade

Para ler, ver e ouvir

North Carolina State University Library Web Archiving. URL: <https://ncsu-libraries.github.io/web-archiving-docs/>

RISSE, Thomas - Web Archive Research and the role of digital academic libraries [Apresentação].

Workshop - TPD, 13 de Setembro de 2018, Porto. URL: <https://sobre.arquivo.pt/wp-content/uploads/Web-Archive-Research-and-the-role-of-digital-academic-libraries-by-Thomas-Risse.pdf>

RESNAIS, Alain - Toute la mémoire du monde. 1956. URL: https://youtu.be/i0RVSZ_yDjs

MATERIAL DE APOIO E OUTRAS REFERÊNCIAS

https://drive.google.com/drive/folders/1kF5hmvW-AmVG2fx8SrF8Jy18vSS_38pH?usp=sharing

Enquadramento legal do Arquivo.pt

[Decreto-Lei 55/2013](#)

“**promover** a preservação de conteúdos disponíveis na Internet nacional, garantindo a **disponibilização** deste à comunidade científica e ao público em geral” - Artº 3, al.2 n)

Sempre ao seu dispor!

ricardobasilio@fcsh.unl.pt e arquivo@contacto

No Arquivo.pt (directo): [+351] 300 005 204 Geral: [+351] 218 440 100

O meu twitter: @ricardobasilio_

