

Como Funciona na Prática a Preservação de Sítios Web

Webinar promovido pelo grupo GT-BES-BAD

Ricardo Basílio, curador digital no Arquivo.pt

Objectivos

Identificar casos de utilidade da preservação da Web

Preservar um sítio Web de forma normalizada

Reutilizar o conteúdo preservado

Agenda

Introdução e abordagem ao tema

Exercício demonstrativo de preservação de um sítio Web

Sugestões de reutilização de conteúdos preservados

1. Introdução e abordagem do tema

Participantes do Webinar

Bem vindos!

Pessoas, Organizações, Projectos, Tecnologia

Uma pequena viagem



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

Pessoas, Organizações, Projectos, Tecnologia



Webrecorder.io

Webrecorder Player (de Ilya Kreymer) – aplicação local

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

Pessoas, Organizações, Projectos, Tecnologia



#Webarchiving

#curadoriadigital

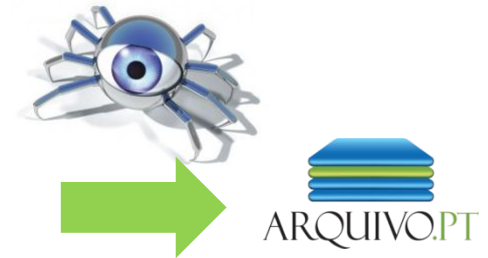
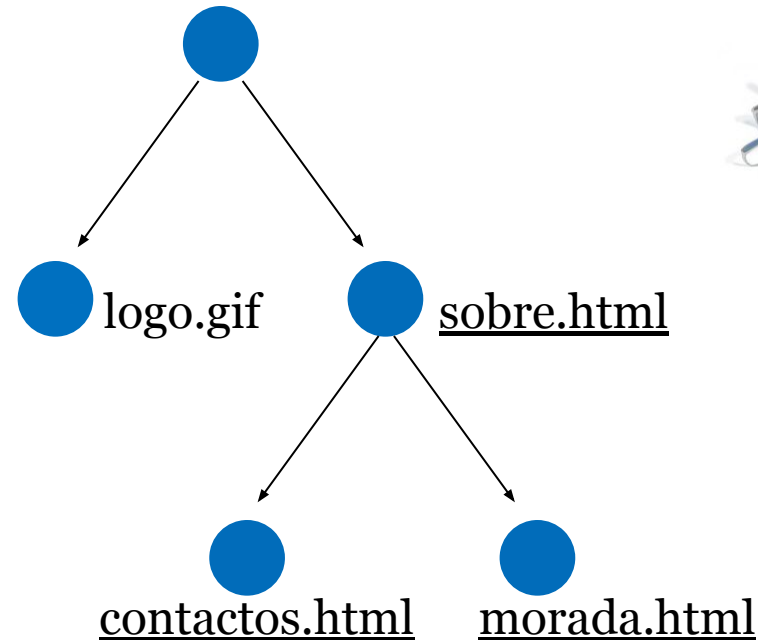
#memoriaweb



Web Crawler

recolhe

www.site.pt/index.html

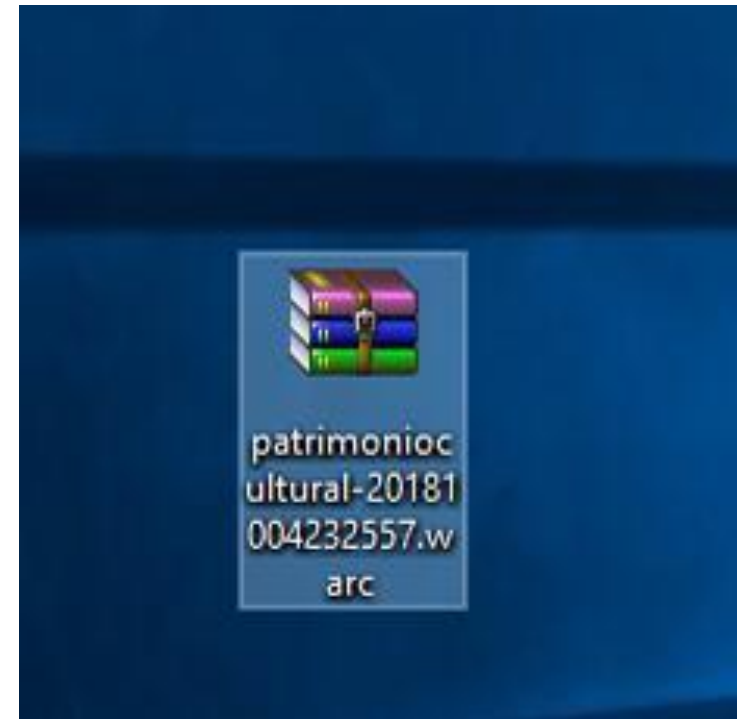




WARC

ISO 28500:2017

armazena



Pessoas, Organizações, Projectos, Tecnologia



pywb

indexa e faz o *replay*

wayback





Timestamp

ISO 8601

ex:20181004232557





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/>



Website – Sítio Web

- É um objecto composto
- É um contentor cujo conteúdo muda
- Tem nome e uma localização
- É propriedade de uma entidade
- É um canal de comunicação

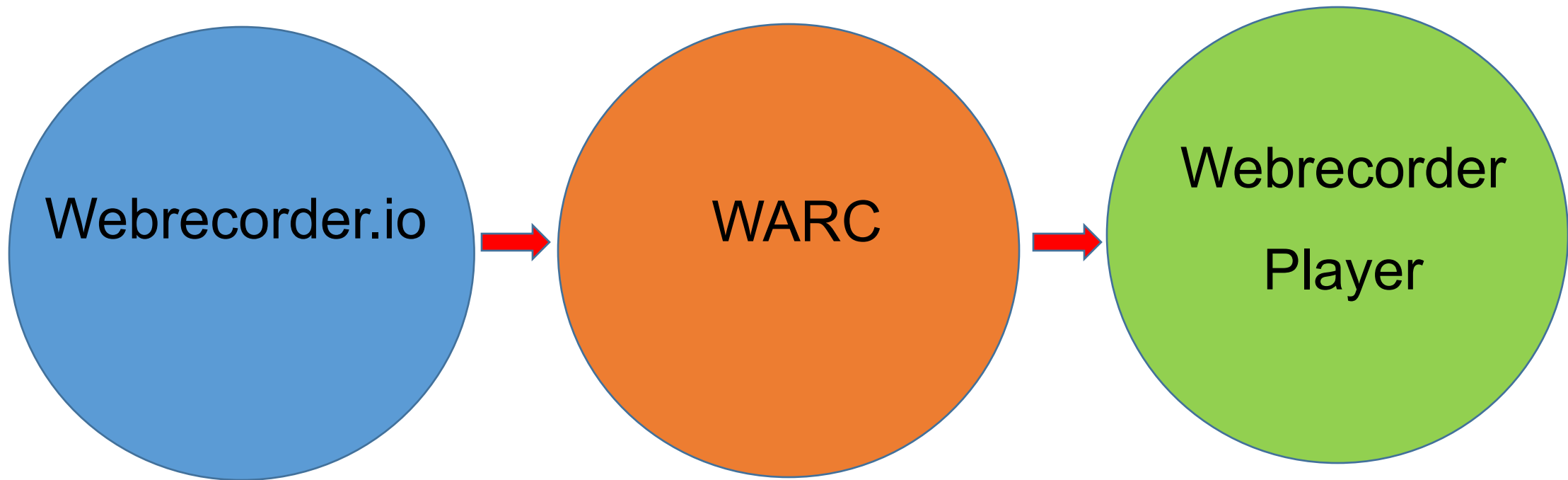


2. Exercício de preservação de um sítio Web

Vamos preservar um sítio Web



Vamos preservar um sítio Web



Vamos preservar um património comum



3. Reutilização de conteúdos preservados

Reutilização dos conteúdos preservados

- 1. Elaboração e partilha de colecções**
- 2. Re-publicação de conteúdos antigos**
- 3. Aplicações e uso das APIs**
- 4. Temas de investigação**
- 5. Lista de sites “arquivados”**
- 6. Exposição sobre a memória online de uma instituição**
- 7. Timelines sobre conteúdos passados da Web**
- 8. Políticas de recolha e curadoria dos sites institucionais**

Reutilização dos conteúdos preservados

Partilhe os seus WARCs

Veja a colecção online: <https://webrecorder.io/memoriaweb>

Faça o **upload** do seu WARC file para a pasta do GoogleDrive:

https://drive.google.com/drive/u/1/folders/1kF5hmvW-AmVG2fx8SrF8Jy18vSS_38pH

Reutilização dos conteúdos preservados

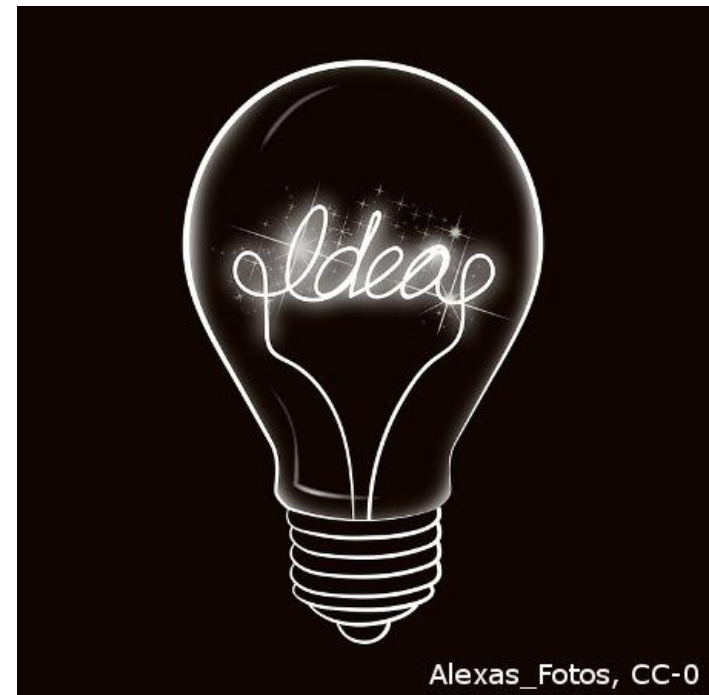
Post no Facebook

Pode publicar em <http://facebook.com/memoriawebparatodos>

Utilize hastags **#memoriaweb #webarchiving #curadoriadigital**

Reutilização dos conteúdos preservados

Outras Sugestões



Reutilização dos conteúdos preservados

Aplicações

Ex. “Conta-me histórias” - <http://contamehistorias.pt/arquivopt/>

Ao alcance de programadores

Reutilização dos conteúdos preservados

Temas de Investigação

Ao alcance de investigadores

Reutilização dos conteúdos preservados

Lista de sites “arquivados”

Saiba como apontar para versões preservadas

Ao nosso alcance

https://drive.google.com/drive/u/1/folders/1kF5hmvW-AmVG2fx8SrF8Jy18vSS_38pH

Reutilização dos conteúdos preservados

Exposição online

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

Ao nosso alcance

Reutilização dos conteúdos preservados

Políticas de recolha

e curadoria dos sítios web institucionais

Ao nosso alcance

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 dos profissionais 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

- TPDL, 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

Mais referências para a preservação de sítios Web

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

Convite

Associe-se ao Dia Mundial da Preservação Digital, tal como em 2017, e visite o Arquivo.pt na tarde de 29 de Novembro.

Esteja a par das novidades sobre preservação da memória Web e subscreva a mailinglist do Arquivo.pt

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)