

# CitationSaver

## Preserva as referências Web dos artigos científicos

Ricardo Basílio, curador digital, FCT-FCCN. Email: ricardo.basilio@fccn.pt

Pedro Gomes, *developer*, FCT-FCCN. Email: pedro.gomes@fccn.pt

[arquivo.pt/citationsaver](http://arquivo.pt/citationsaver)

### Problema

- Os artigos científicos fazem referência a conteúdos da Web.
- Com o passar do tempo os conteúdos deixam de estar online ou mudam de endereço.
- As publicações perdem parte da sua consistência e do seu valor científico.
- Como fazer para que os conteúdos referenciados possam ser consultados no futuro?

### Metodologia

A criação do CitationSaver para uso da comunidade deriva da forma como o Arquivo.pt quer situar-se no ecossistema da gestão da Ciência. A missão do Arquivo.pt é preservar conteúdos Web e dar-lhes acesso. Não é desenvolver *software*. Para criar este serviço começou por identificar o problema das ligações quebradas nas publicações científicas. Comparou e testou o *software* disponível. Definiu um *workflow* em que o processamento é automatizado e invisível para o utilizador. Documentou o desenvolvimento no Github. Criou uma interface de utilizador no site do Arquivo.pt. Lançou o serviço, em fase experimental, para obter *feedback*, detetar erros e fazer melhorias.

### Resultados

A interface do CitationSaver está acessível no site do Arquivo.pt.

O funcionamento do serviço pode representar-se em **quatro passos**:

- 1- Utilizador **insere** artigo no CitationSaver
- 2- CitationSaver **extraí** todos endereços (URLs, tudo o que começa por http)
- 3- CitationSaver valida e **grava** automaticamente lista de endereços
- 4- Arquivo.pt **disponibiliza** conteúdos gravados

Há **três opções** para processar uma publicação no CitationSaver:

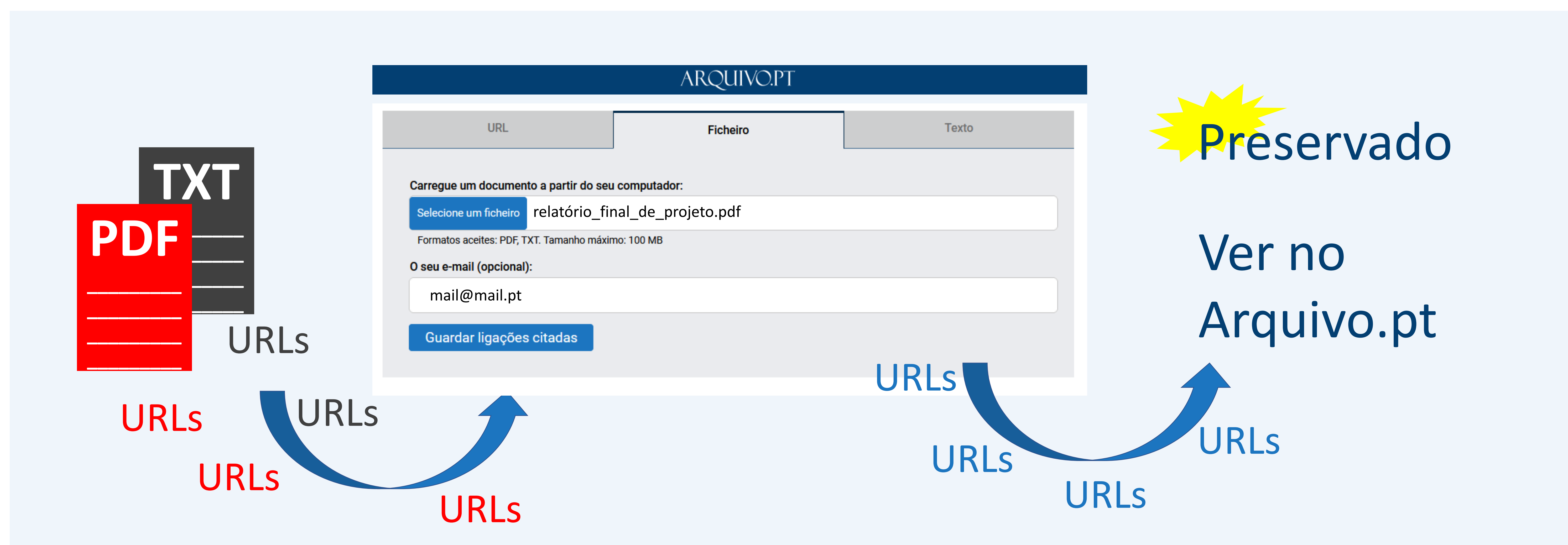
- a) **URL** - indicar o URL ou endereço Web onde o artigo se encontra publicado;
- b) **Ficheiro** – fazer *upload* do ficheiro PDF ou Word;
- c) **Texto** – copiar e colar texto onde estão referências a conteúdos Web.

Os conteúdos gravados ficam disponíveis no Arquivo.pt em algumas semanas. Ligações quebradas numa publicação podem ser recuperadas no Arquivo.pt.

Para testar a capacidade de processamento automático do serviço, foram submetidos 30 mil ficheiros PDF de publicações científicas, dos quais resultaram 670 mil URLs.

O *software* utilizado, a resolução de problemas e demais detalhes técnicos estão documentados no canal Github do Arquivo.pt

(<https://github.com/arquivo/CitationSaver>).



### Conclusão

O CitationSaver, tal como está disponibilizado pelo Arquivo.pt, permite aos profissionais incluir esta ferramenta na agenda da preservação digital institucional.

A **comunidade BAD** tem um papel fundamental para criar uma mentalidade de preservação, incluindo os conteúdos da Web.

O contributo do CitationSaver para a preservação de conteúdos citados nas publicações científicas é tanto mais significativo se o considerarmos em grande escala, aplicado a milhares de documentos.

O CitationSaver enquadra-se na missão do Arquivo de “promover a preservação de conteúdos disponíveis na Internet nacional, mas a sua **eficácia é maior com o apoio** da comunidade.