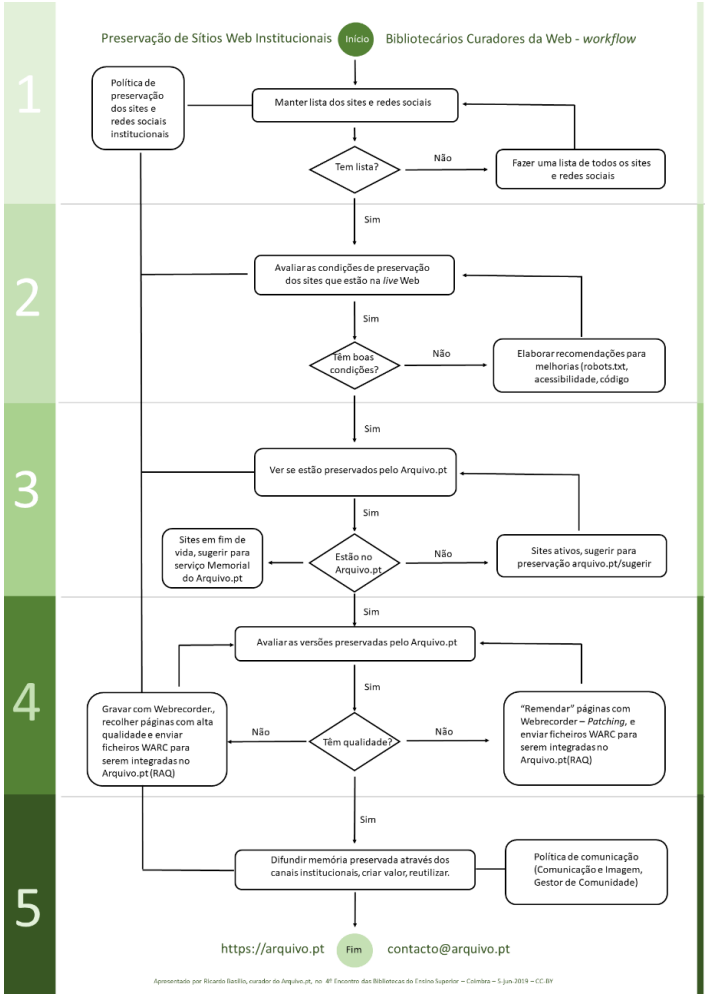
4º Encontro das Bibliotecas do Ensino Superior

Coimbra, 4-5 de junho de 2019

Workflow para a Preservação de Sítios Web Institucionais

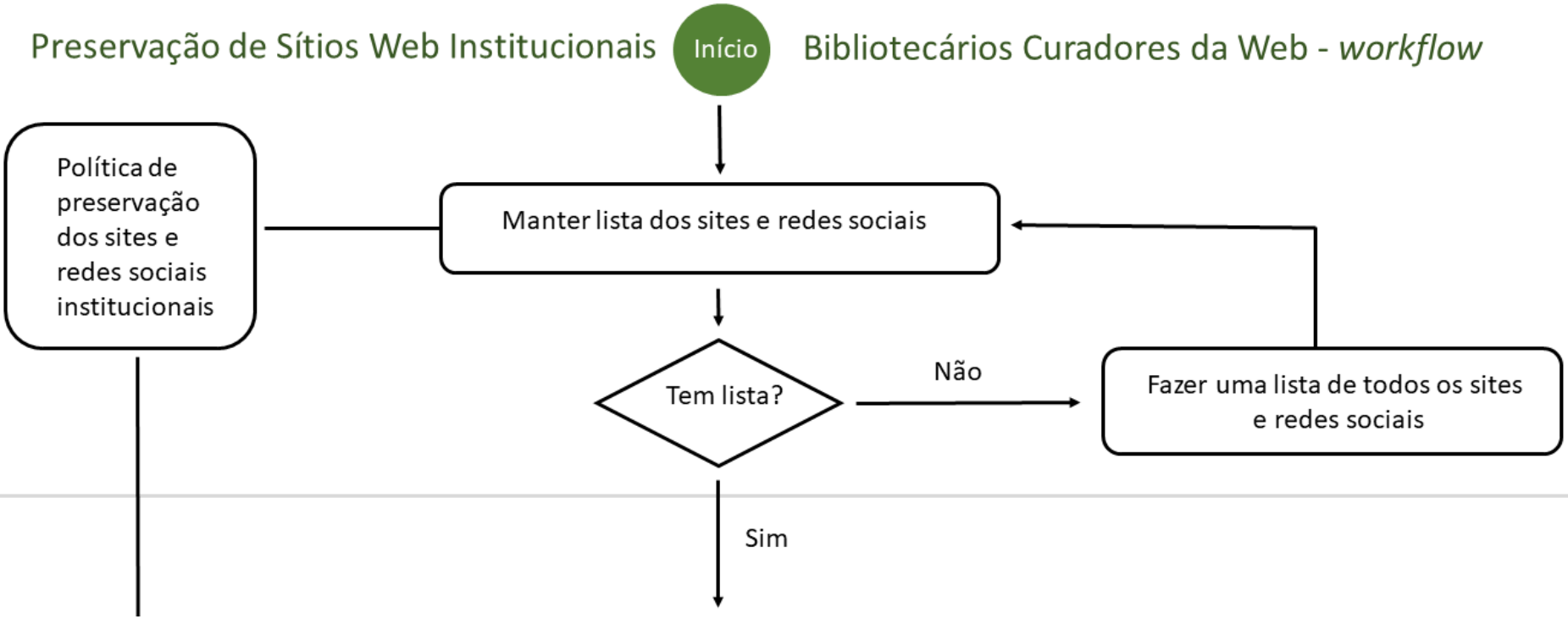
Ricardo Basílio, curador do Arquivo.pt



Download: https://is.gd/fluxogramacuradoria

1

Preservação de Sítios Web Institucionais



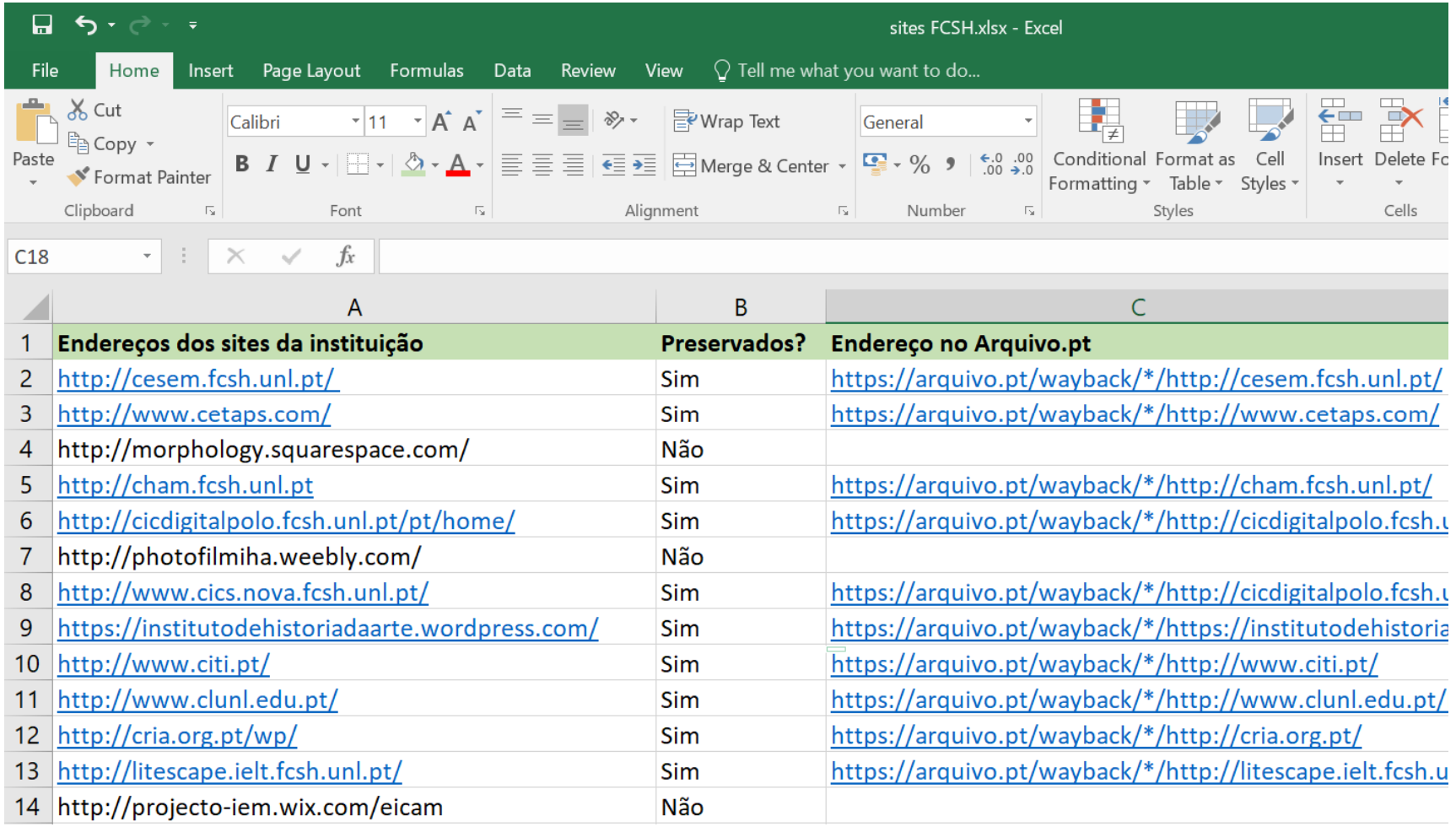
1

Tarefas de curadoria:

- **Manter uma lista de sites e redes sociais da sua instituição**
- **Saber se há referência a esta matéria nas políticas da instituição**

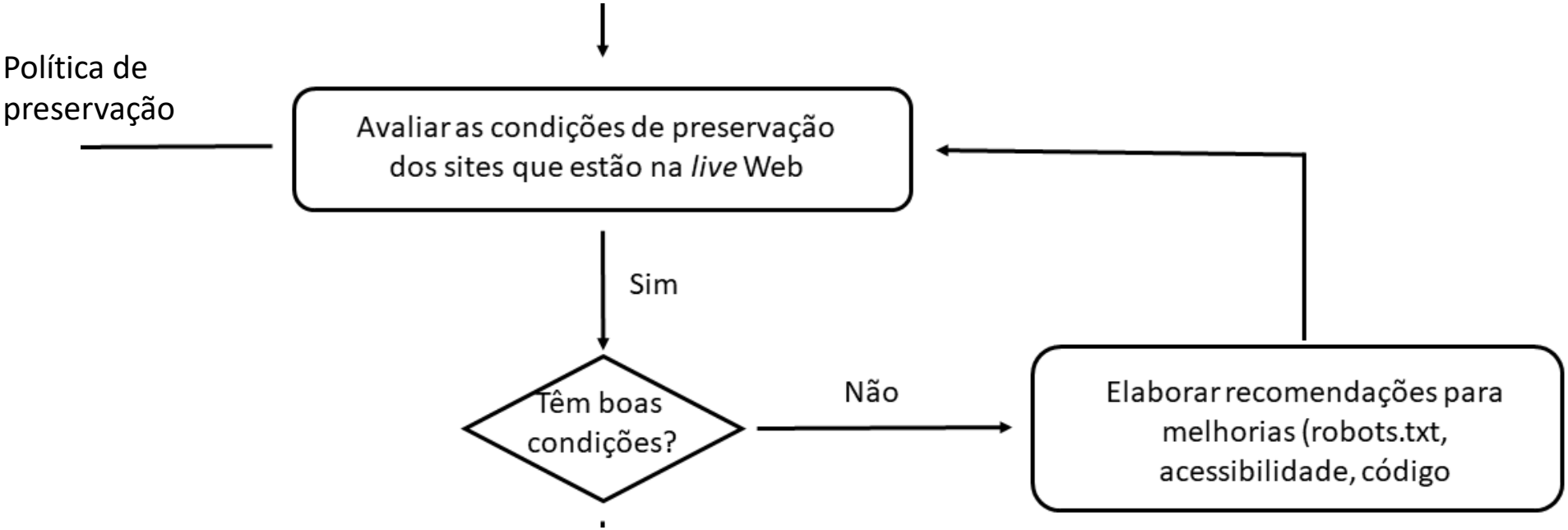
1

Exemplo: lista de sites da NOVA-FCSH



	A	B	C
1	**Endereços dos sites da instituição**	**Preservados?**	**Endereço no Arquivo.pt**
2	http://cesem.fcsh.unl.pt/	Sim	https://arquivo.pt/wayback/*/http://cesem.fcsh.unl.pt/
3	http://www.cetaps.com/	Sim	https://arquivo.pt/wayback/*/http://www.cetaps.com/
4	http://morphology.squarespace.com/	Não	
5	http://cham.fcsh.unl.pt	Sim	https://arquivo.pt/wayback/*/http://cham.fcsh.unl.pt/
6	http://cicdigitalpolo.fcsh.unl.pt/pt/home/	Sim	https://arquivo.pt/wayback/*/http://cicdigitalpolo.fcsh.u
7	http://photofilmiha.weebly.com/	Não	
8	http://www.cics.nova.fcsh.unl.pt/	Sim	https://arquivo.pt/wayback/*/http://cicdigitalpolo.fcsh.u
9	https://institutodehistoriadaarte.wordpress.com/	Sim	https://arquivo.pt/wayback/*/https://institutodehistoria
10	http://www.citi.pt/	Sim	https://arquivo.pt/wayback/*/http://www.citi.pt/
11	http://www.clunl.edu.pt/	Sim	https://arquivo.pt/wayback/*/http://www.clunl.edu.pt/
12	http://cria.org.pt/wp/	Sim	https://arquivo.pt/wayback/*/http://cria.org.pt/
13	http://litescape.ielt.fcsh.unl.pt/	Sim	https://arquivo.pt/wayback/*/http://litescape.ielt.fcsh.u
14	http://projecto-iem.wix.com/eicam	Não	

2



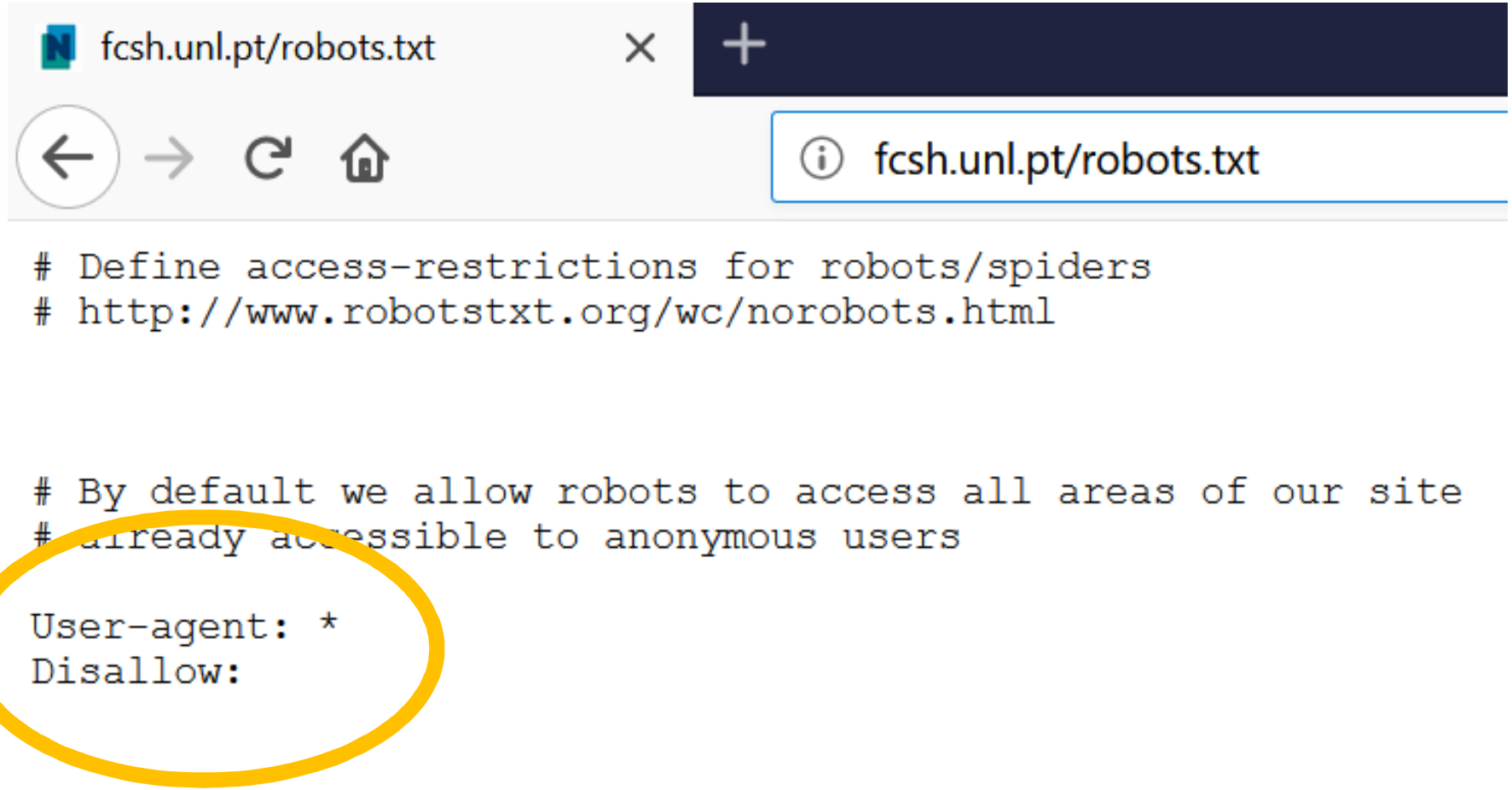
Ex. Avaliar com ArchiveReady.com

2

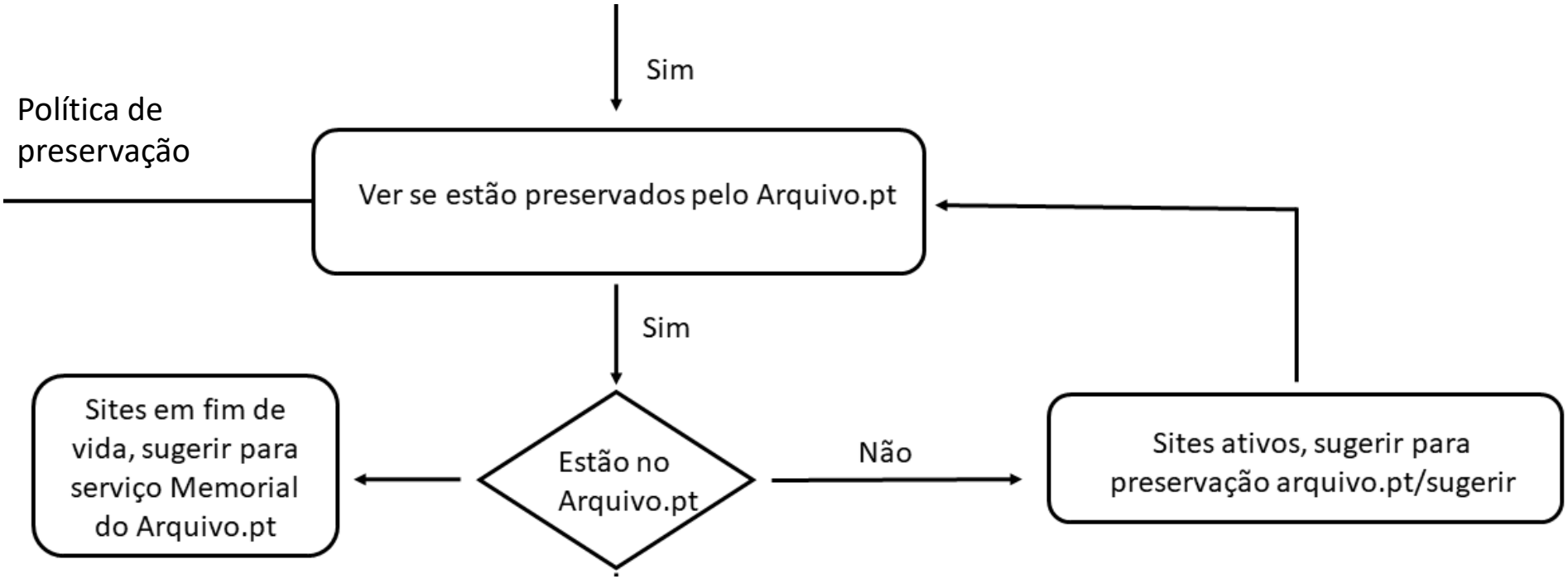
Tarefas de curadoria:

- **Avaliar as condições de preservação dos sítios Web (ArchiveReady.com)**
- **Elaborar recomendações de melhorias**

2


fcsh.unl.pt/robots.txt
fcsh.unl.pt/robots.txt
Define access-restrictions for robots/spiders
http://www.robotstxt.org/wc/norobots.html
By default we allow robots to access all areas of our site
already accessible to anonymous users
User-agent: *
Disallow:

3

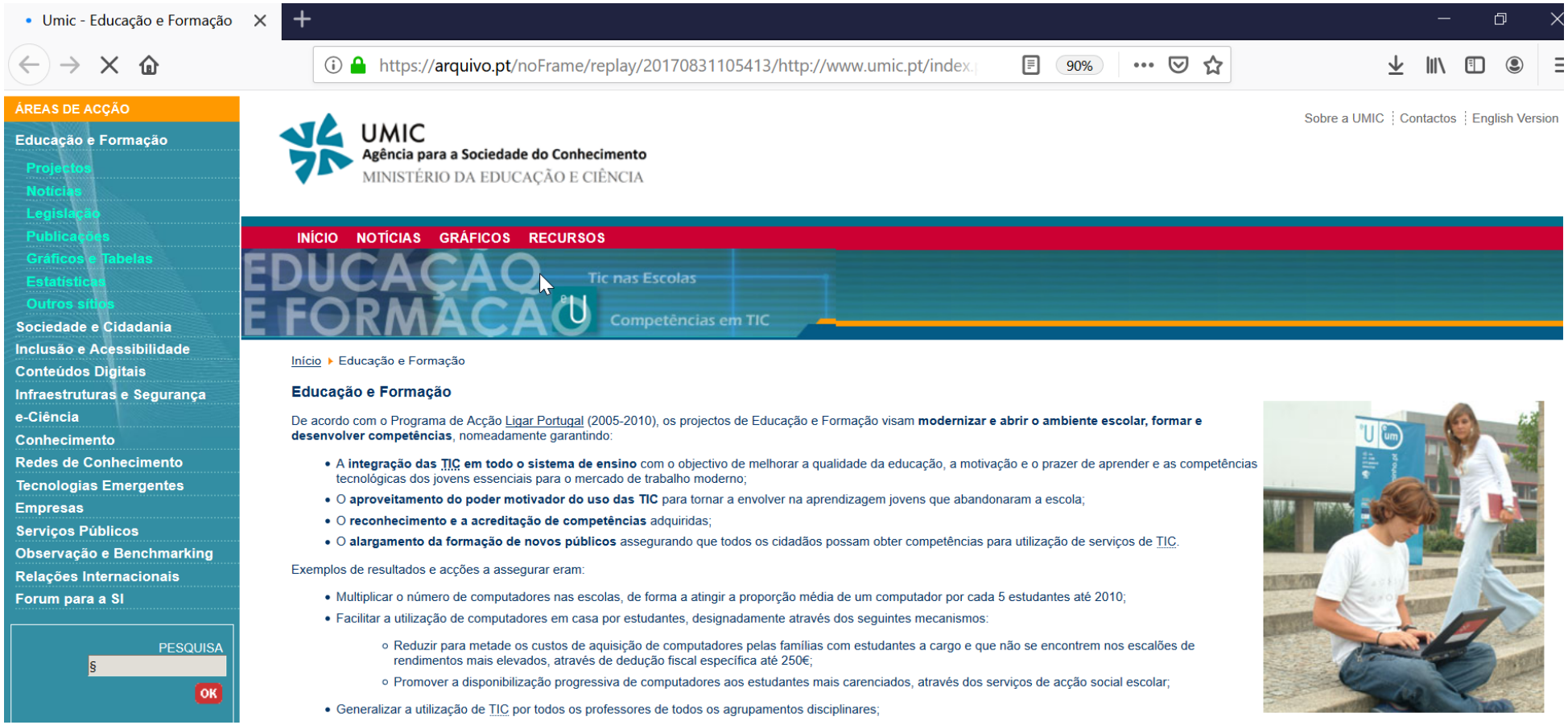

Sim
Política de preservação
Ver se estão preservados pelo Arquivo.pt
Sim
Sites em fim de vida, sugerir para serviço Memorial do Arquivo.pt
Estão no Arquivo.pt
Não
Sites ativos, sugerir para preservação arquivo.pt/sugerir

3

Tarefas de curadoria:

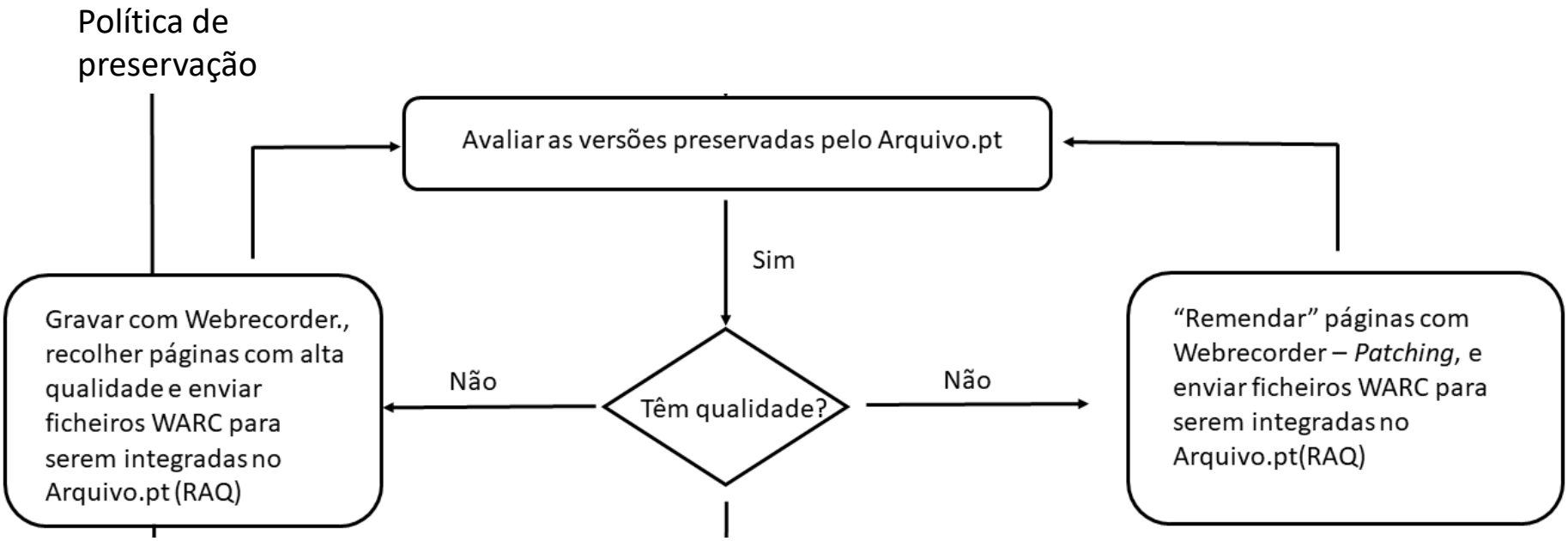
- **Ver se estão preservados no Arquivo.pt**
- **Sites ativos, sugerir para preservação arquivo.pt/sugerir**
- **Sites em fim de vida, usar o novo serviço Memorial do Arquivo.pt – entre em contacto**

3



Site da UMIC preservado no Memorial do Arquivo.pt

4


Política de preservação
Avaliar as versões preservadas pelo Arquivo.pt
Sim
Gravar com Webrecorder., recolher páginas com alta qualidade e enviar ficheiros WARC para serem integradas no Arquivo.pt (RAQ)
Não
Têm qualidade?
Não
"Remendar" páginas com Webrecorder – *Patching*, e enviar ficheiros WARC para serem integradas no Arquivo.pt(RAQ)

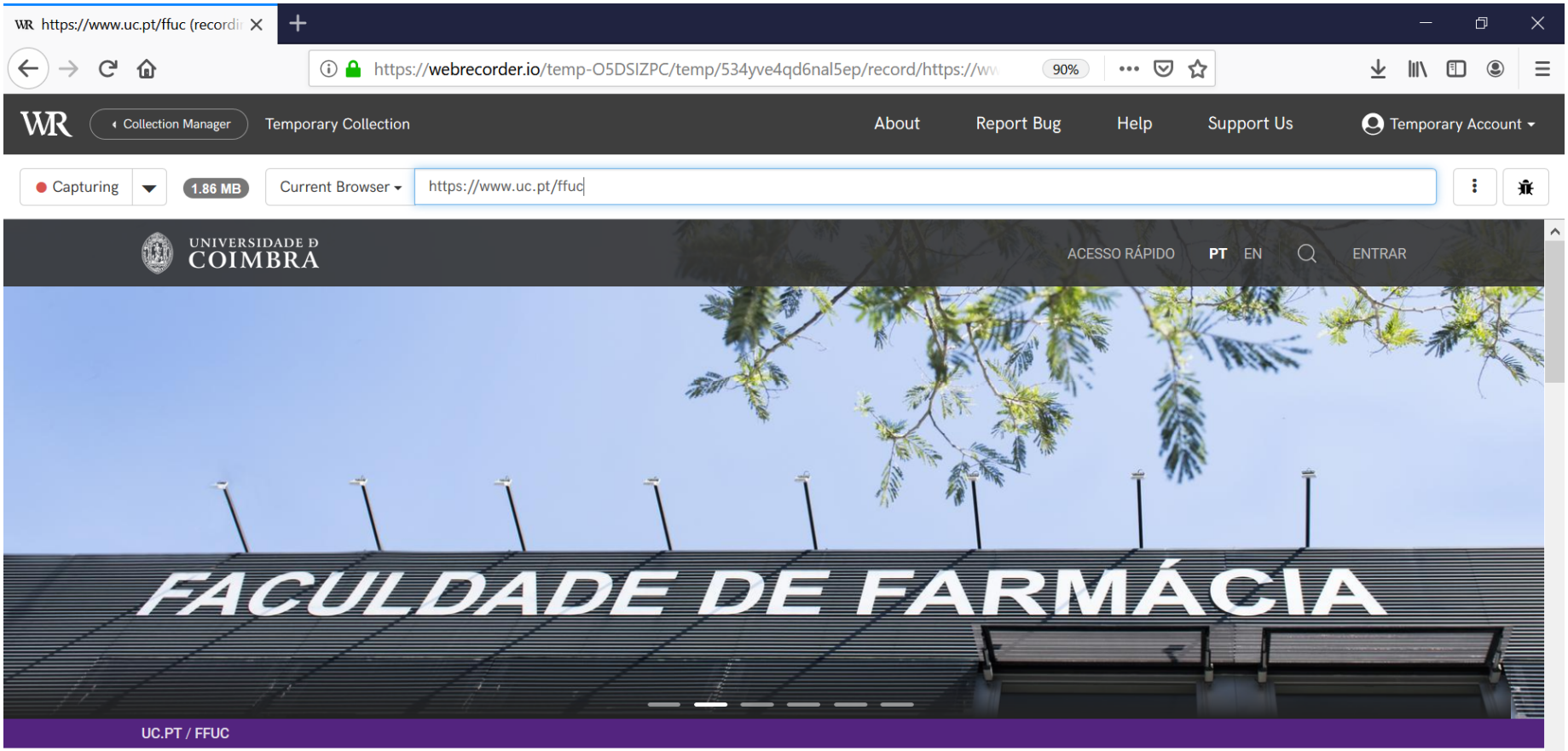
4

Tarefas de curadoria:

- **Avaliar as versões preservadas pelo Arquivo.pt**
- **Gravar manualmente páginas relevantes**

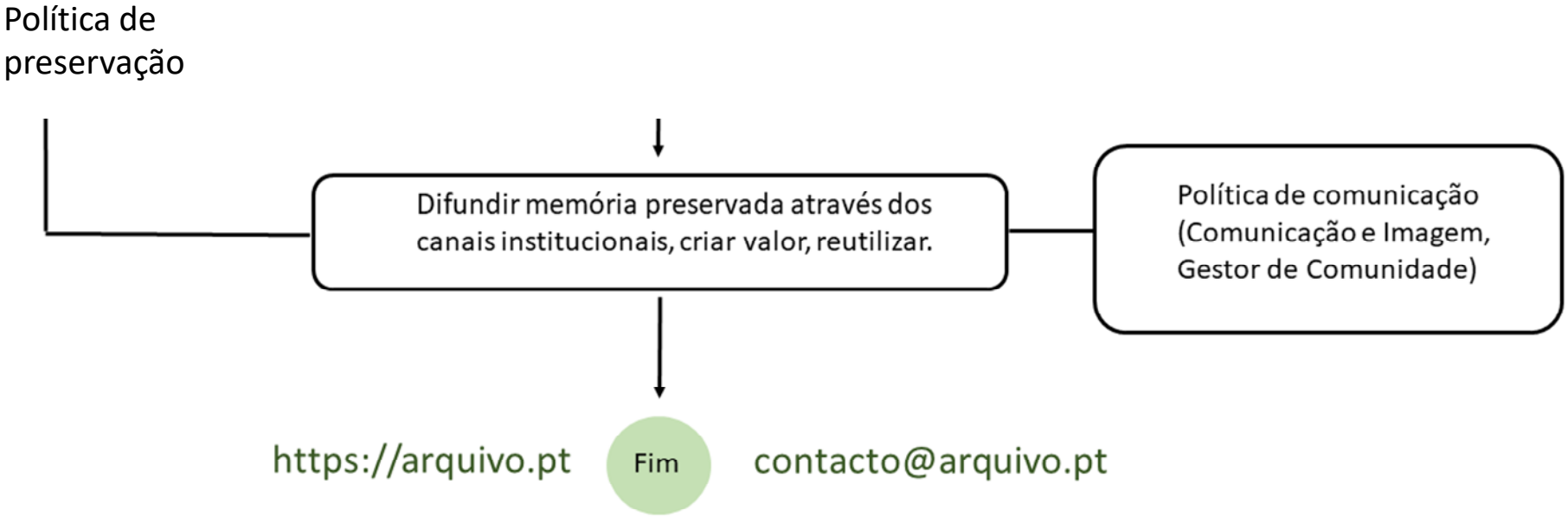
  (RAQ, i.e., recolhas de alta qualidade)
- **Remendar páginas antigas - *patching***

4



Gravação manual com o Webrecorder.io

5

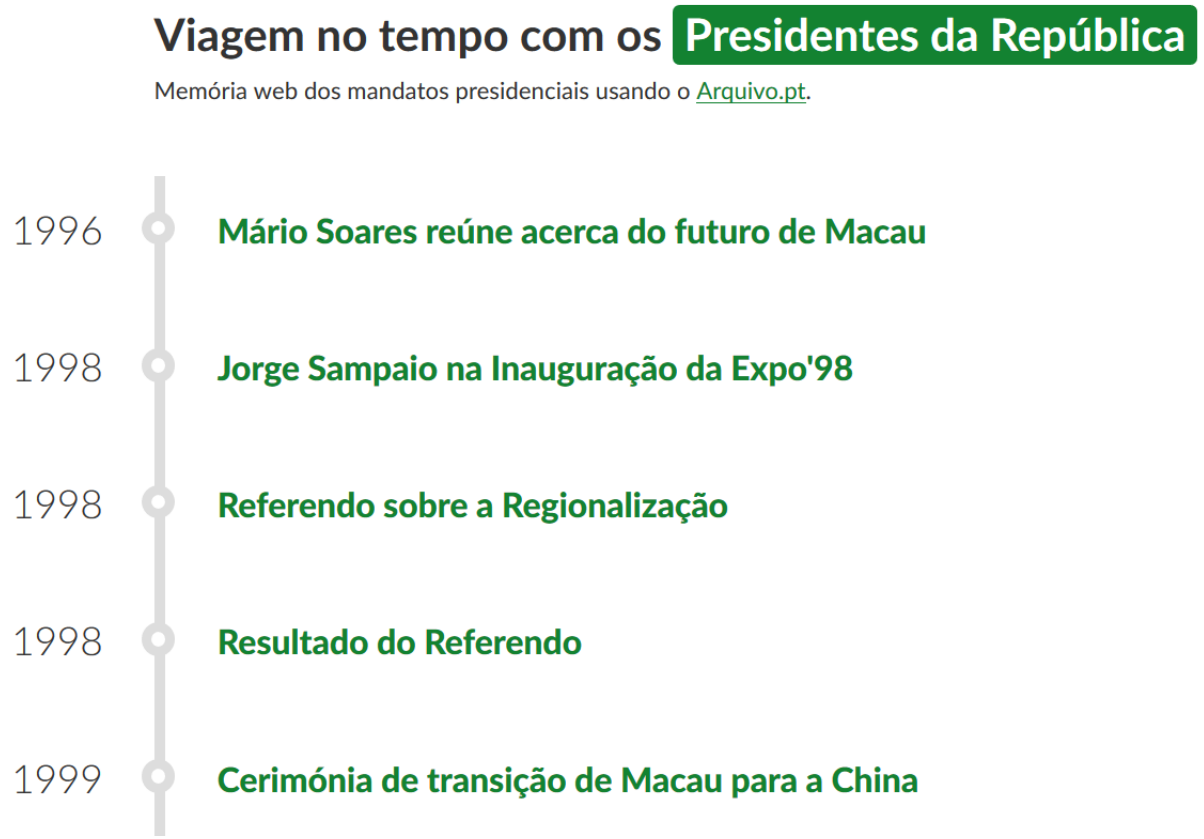

Política de preservação
Difundir memória preservada através dos canais institucionais, criar valor, reutilizar.
Política de comunicação (Comunicação e Imagem, Gestor de Comunidade)
https://arquivo.pt
Fim
contacto@arquivo.pt

5

Tarefas de curadoria:

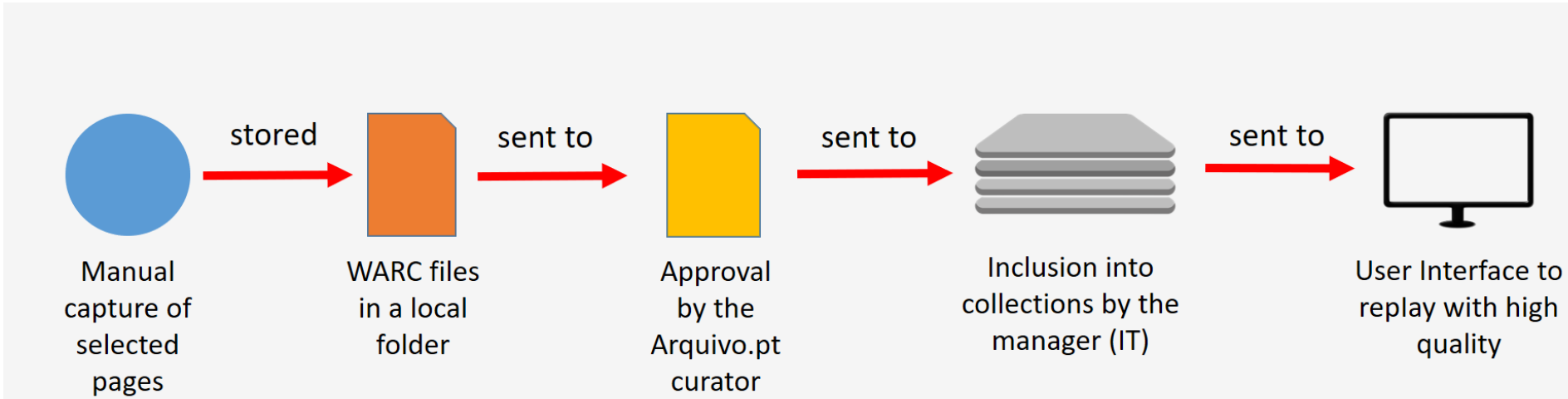
- **Difundir memória preservada através dos canais institucionais, criar valor, reutilizar**
- **Verificar as políticas de comunicação (Comunicação e Imagem, Gestor de Comunidade)**

5



Colaboração entre o Arquivo.pt e o Museu da Presidência da República

arquivo.pt/presidentes


stored
sent to
sent to
sent to
Manual capture of selected pages
WARC files in a local folder
Approval by the Arquivo.pt curator
Inclusion into collections by the manager (IT)
User Interface to replay with high quality

Librarians as Web Curators in Universities – Web Archiving Conference 2019 - Zagreb

	Training & impact *Chair: Inge Rudomino*	European Union & GDPR *Chair: Renata Petrušić*
16:10	**Daniel Gomes**: Arquivo.pt – memorial and other goodies	**Alejandra Michel**: The legal framework for web archiving: focussing on GDPR
16:30	**Maria Praetzellis, Sharon McMeekin & Abbie Grotke:** Building the IIPC training program	**Tom Storrar & Chris Doyle**: Creating an archive of European Union law for Brexit
16:50	**Julie Fukuyama & Simon Tanner**: Developing impact assessment indicators – making a proposal for the UK Web Archive	**Els Breedstraet**: Setting up an EU web preservation service for the long-term – tales of a (sometimes) bumpy road
17:10	**Ricardo Basilio & Daniel Bicho**: Librarians as web curators in universities	**Marinos Papadopoulos, Charalampos Bratsas, Michalis Gerolimos, Konstantinos Vavoussis, Eliza Makridou & Dimitra Hioti:** Text and data mining for the national library of Greece in consideration of GDPR
17:30	**Q&A**	**Q&A**
17:40	**Closing remarks (Panorama)**	
18:00	**Finish**	

http://netpreserve.org/ga2019/timetable/event/training-impact/

Dia Mundial da Preservação Digital,

28 de novembro de 2019

Download desta apresentação:
https://is.gd/pechakucharicardo