

# The first steps towards a Belgian web archive: a federal strategy

WAC – Zagreb  
June 6<sup>th</sup> 2019

Friedel Geeraert  
Sébastien Soyez

Royal Library (KBR) and State Archives of Belgium  
State Archives of Belgium

# Overview







# I The PROMISE project



# I The PROMISE project

---

Identify best practices in the field of web archiving

- Literature review
- Interviews with representatives of web archiving initiatives





## II The strategy

---

# Selection

**.be**  
**.vlaanderen**  
**.brussels**  
**.gent**

**gTLDs**  
**.org**  
**.com**  
...

**ccTLDs**  
**.fr**  
**.nl**  
...

**Web content  
created,  
produced,  
published, ...  
on the Belgian  
territory**

**IF**

**Websites are registered by  
Belgians**

**Content concerns Belgium or  
its general affairs**



# Selection

## BROAD CRAWL

- Sampling of the Belgian web
- Superficial capture
- Collected once a year
- Problem: no access to full list of Belgian domain names



## SELECTIVE CRAWLS

- Complete capture of web content
- More frequent captures



601 websites

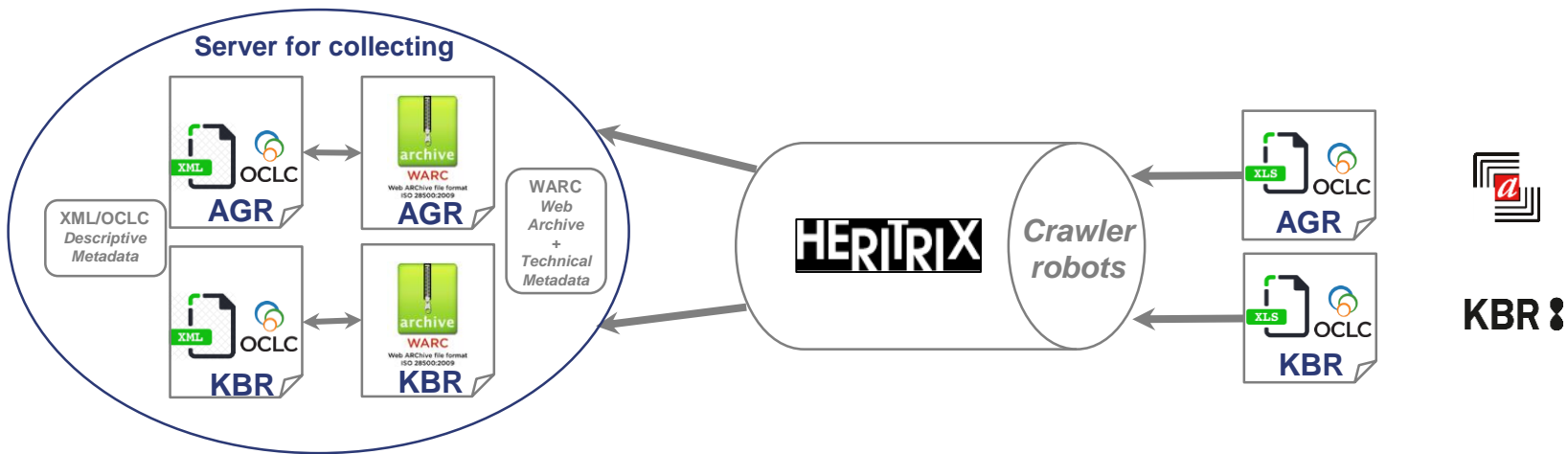
- Federal institutions
- Ministerial cabinets, ministers/secretaries of state
- Public organisms with link to federal level
- Provinces, the regions and the communities
- Projects funded by BELSPO

# KBR

928 websites  
1416 web pages  
37 sections of websites

- Special heritage collections
- Contemporary collections
- Spanish, Italian and Portuguese communities

# Collecting



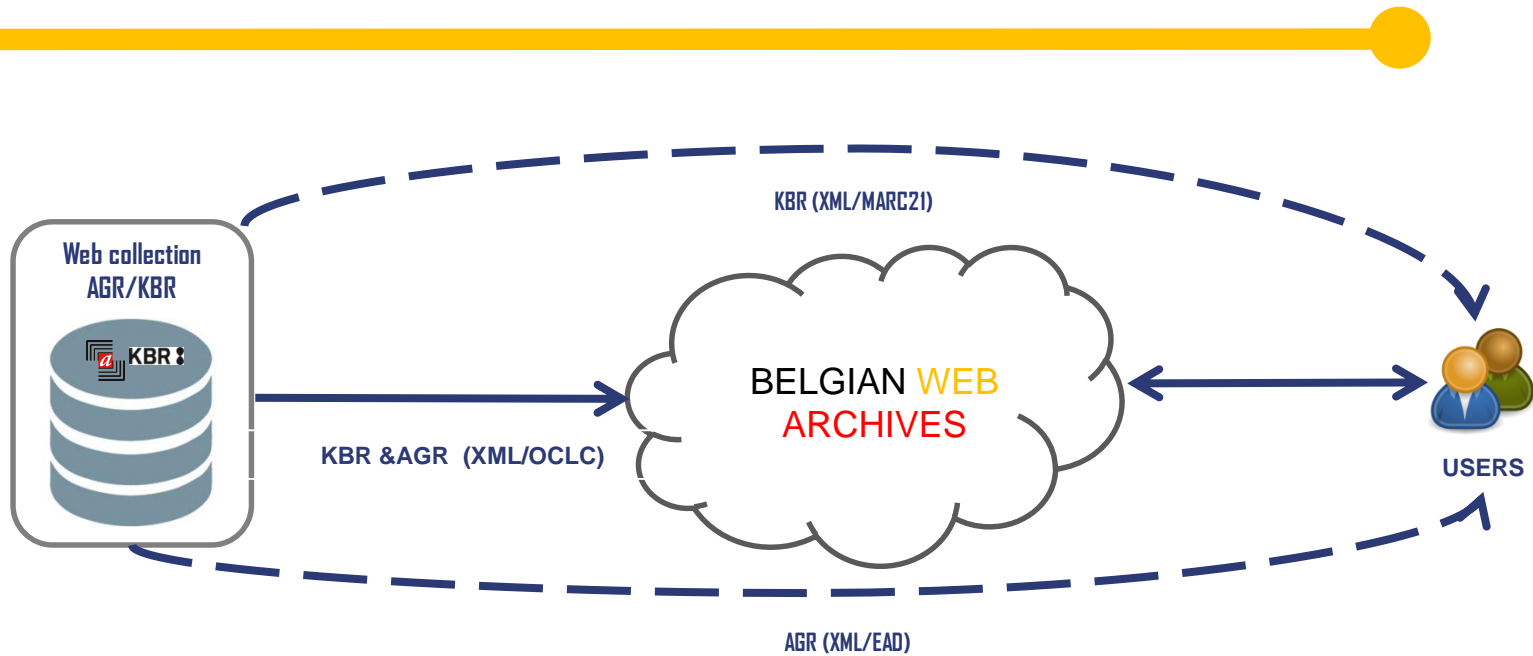
# Quality control



Lack of automated tools

Prototype visual correspondence

# Access



# Access



Law on  
Archives

Legal Deposit  
Law



Copyright  
legislation

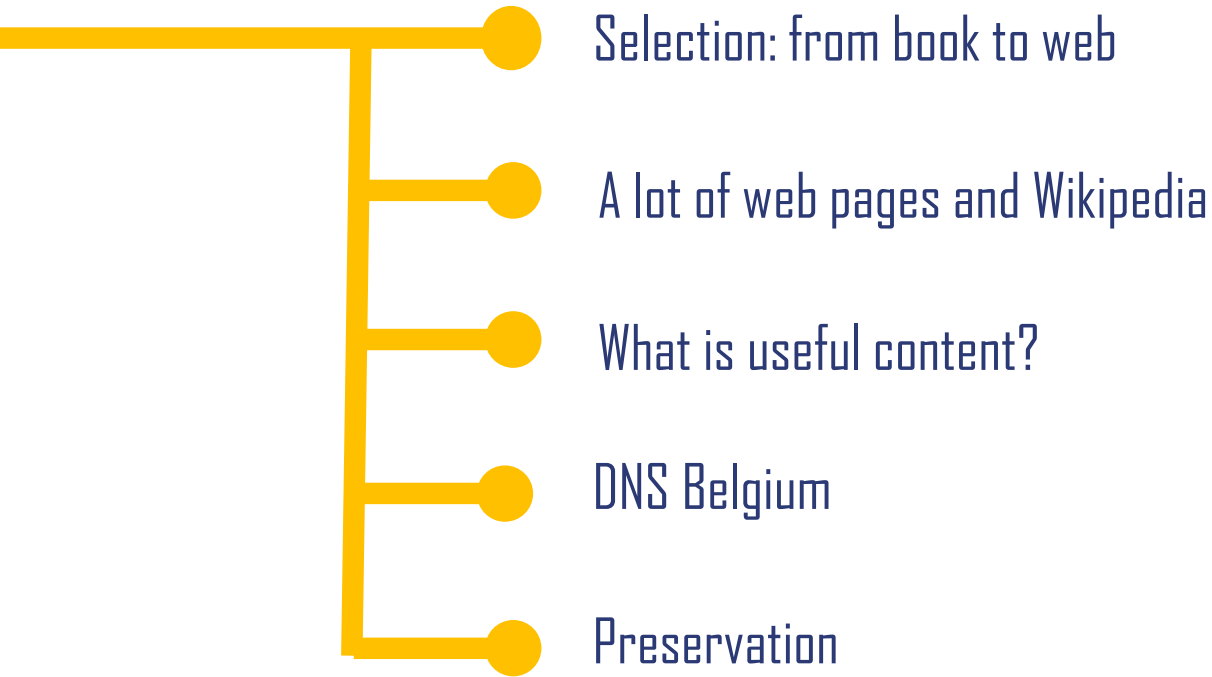


Privacy  
Legislation  
(GDPR)



Illegal  
content

# III Lessons learnt

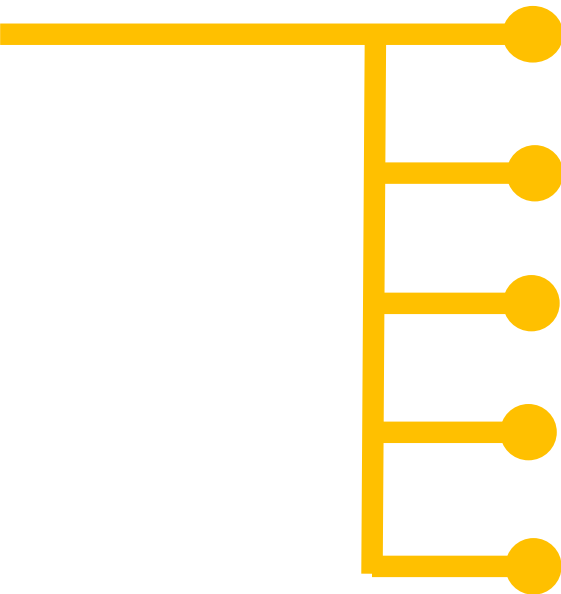


# III Lessons learnt

- Web crawling takes time
- Setting crawl parameters = trial and error
- Quality control = pain point
- Estimations: time and cost
- Too ambitious?



# IV Next steps



Validation of shared strategy



Access to archived web content



Recommendations and procedures



Web archiving = part of KBR strategy 2019-2021



October 18<sup>th</sup> - Colloquium 'Saving the web: the promise of a Belgian web archive'



# Thank you!



Friedel Geeraert

Researcher on PROMISE project

[friedel.geeraert@kbr.be](mailto:friedel.geeraert@kbr.be)