

Using web archives for researchers

15:15 - 16:30

Sally Chambers – GhentCDH and DARIAH

Eveline Vlassenroot – imec-mict-UGent

Jesse de Vos – Beeld & Geluid

Valérie Schafer – C2DH

Patricia Blanco – MSc in Digital Humanities (KU Leuven)



Introduction

Sally Chambers - GhentCDH and DARIAH



User requirements for a web archive

Eveline Vlassenroot – imec-mict-UGent



Research question:

What are the requirements and needs of potential users of web archives?

*How do they, or do they want to, **access, use and consult** web archives?*



KBR



Methodology

- Survey with a local, regional, national and international focus
- Ran 2 months (April 2018 - end of May 2018)
- 3 languages: Dutch, French and English
- Total: 587 respondents (after cleaning: 399)

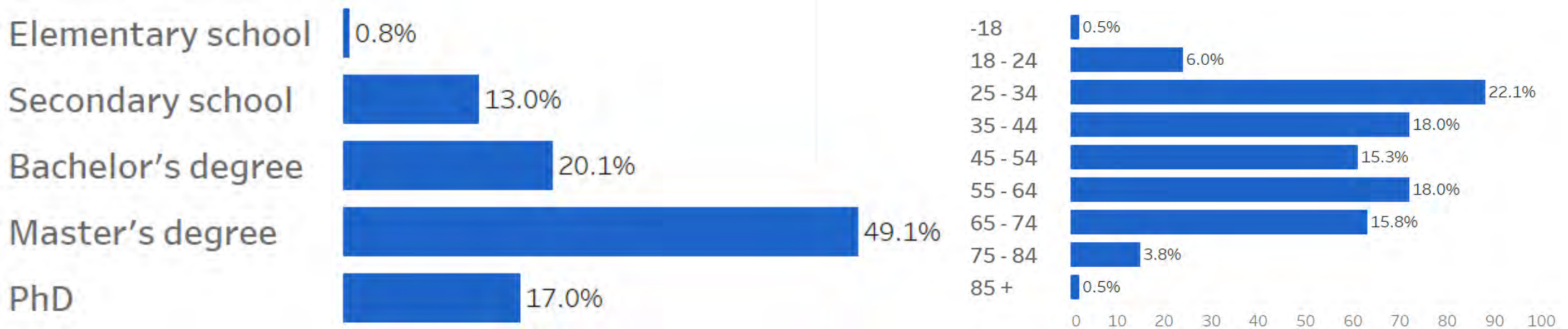
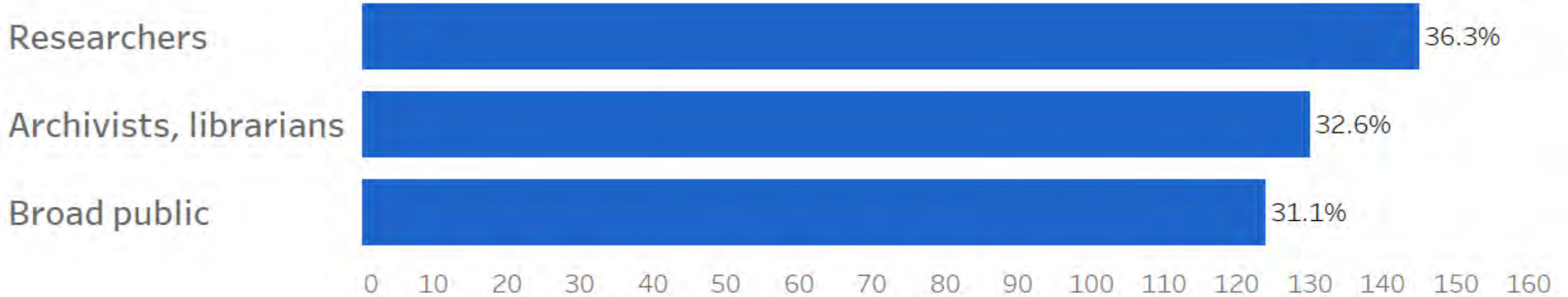


Target groups

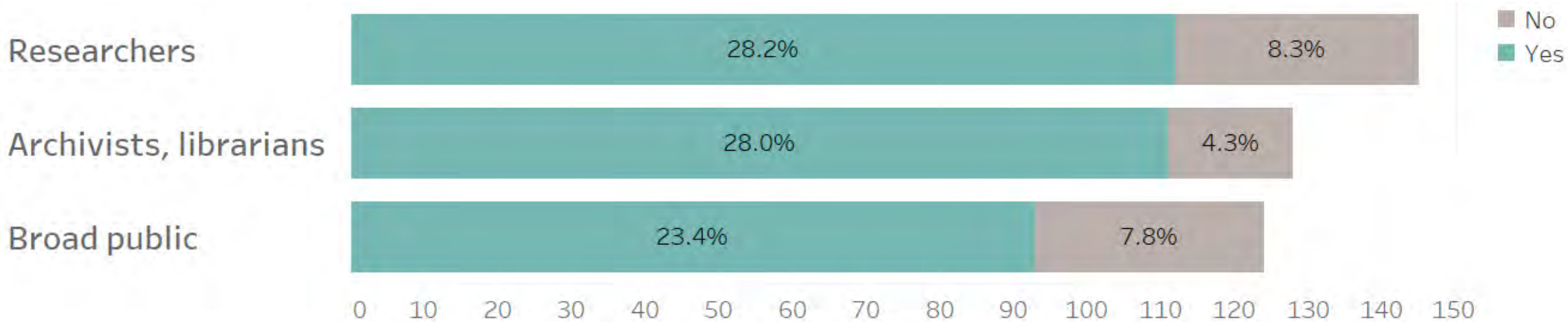
1. **Research:** students, academics, anyone involved in research more broadly
2. **Archives, libraries and governmental institutions**
3. **General public**



Socio demographics (n=399)



Are you aware of the existence of web archives? n=397



Can you provide an example of a web archive?



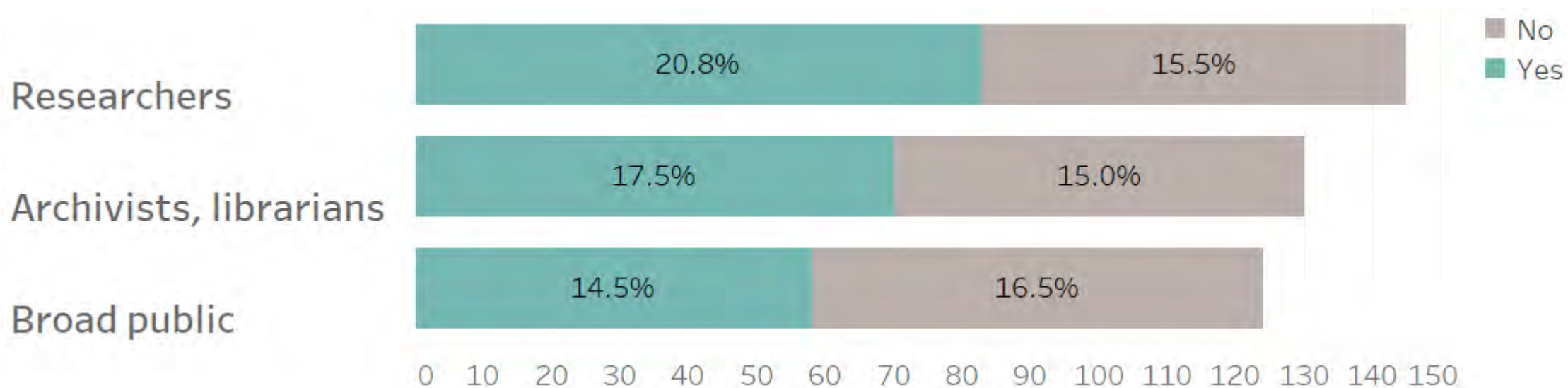
KBR



HE2B
HAUTE ÉCOLE
BRUXELLES-BRABANT



Have you ever used a web archive? n=399



Why are you not using web archives? n=180

I didn't know of their existence.

45.6%

No relevant data available in the web archive for my topic.

18.3%

Access is (too) restricted.

11.1%

Too difficult to work with (lack of research infrastructure).

5.0%

Incomplete or inaccurate archived materials.

1.7%

Other

18.3%

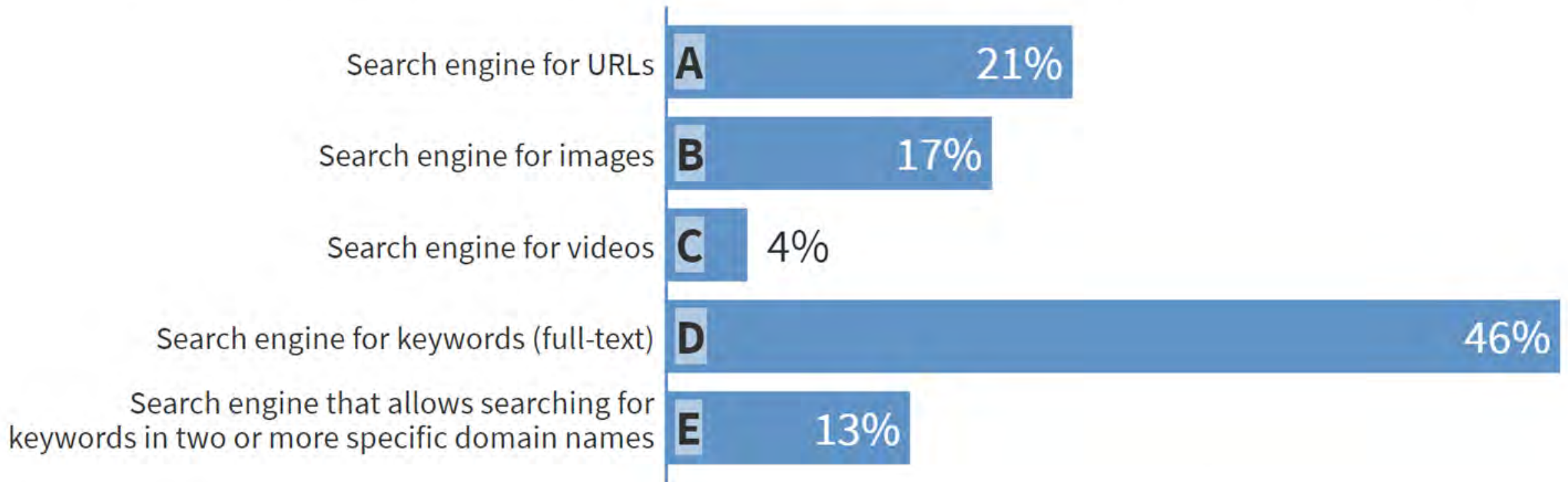
0 10 20 30 40 50 60 70 80 90



Respond at PollEv.com/friedelgeera923

Text **FRIEDELGEERA923** to **22333** once to join, then **A, B, C, D, or E**

Which search functionalities do you deem most important for the Belgian web archive? (max. 2 options)



Select the three most important web archive functionalities in your opinion.

n=290

(1)

Search engine for keywords

Search engine for images

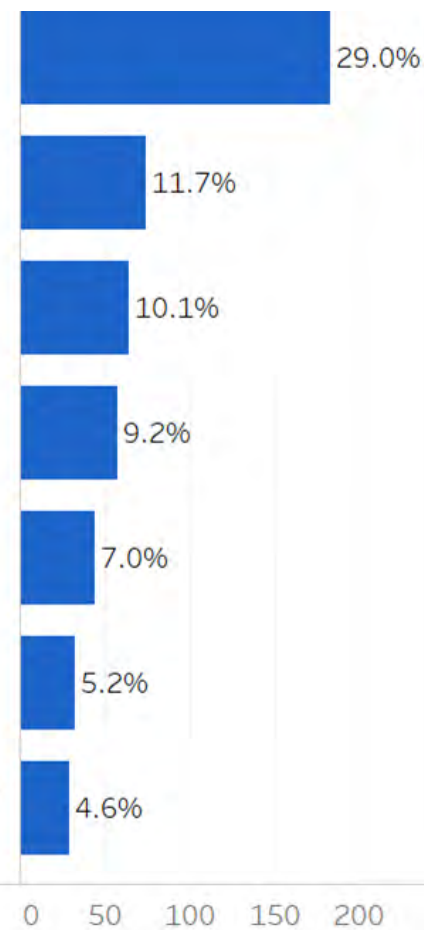
Collections of archived websites specific to my research area

The possibility to download data in multiple data formats

Search engine for URLs

The possibility to download data in multiple ways

Tools that help with the discovery of patterns, trends or relationships in data



KBR



Select the three most important web archive functionalities in your opinion.

n=290

(2)

A tool that enables the creation and extraction of a corpus from a larger collection

4.4%

Overview of personal search history

4.4%

Tools that help me analyse the data

4.0%

Side-by-side visualization of a website throughout the years

3.8%

Search engine that allows searching for keywords in two or more specific URLs

2.7%

Search engine for videos

2.2%

Auto-completion of search fields in search engine

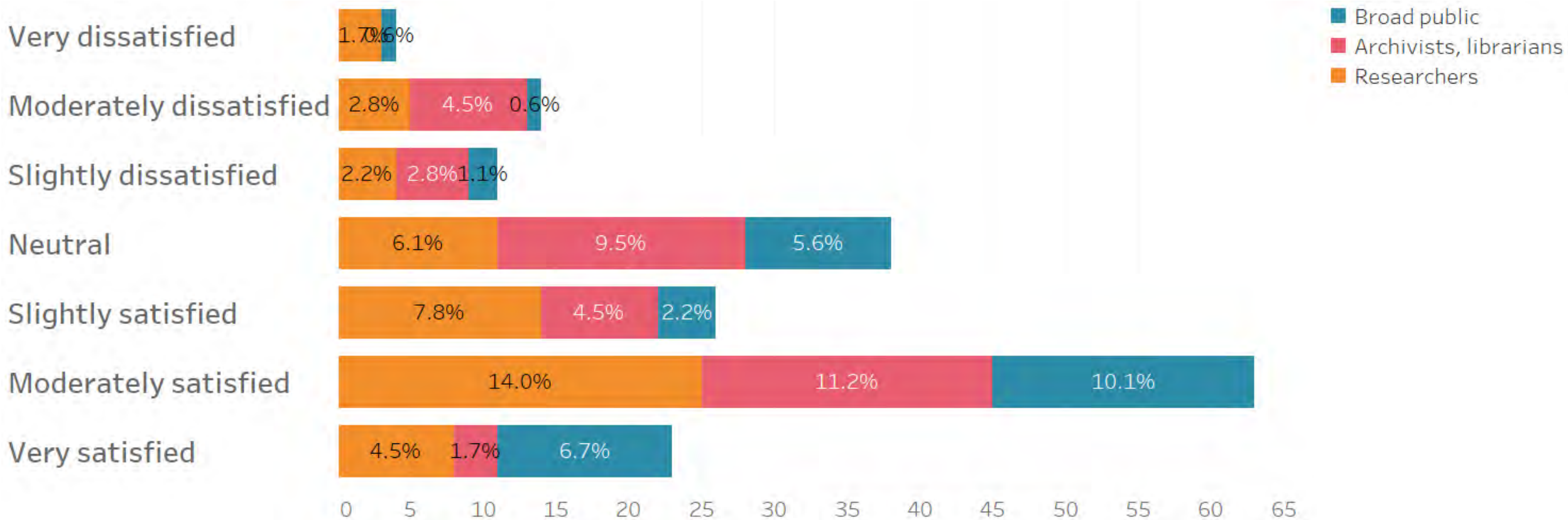
1.6%

0 50 100 150 200

**KBR**

How satisfied are you with this or these webarchive(s)?

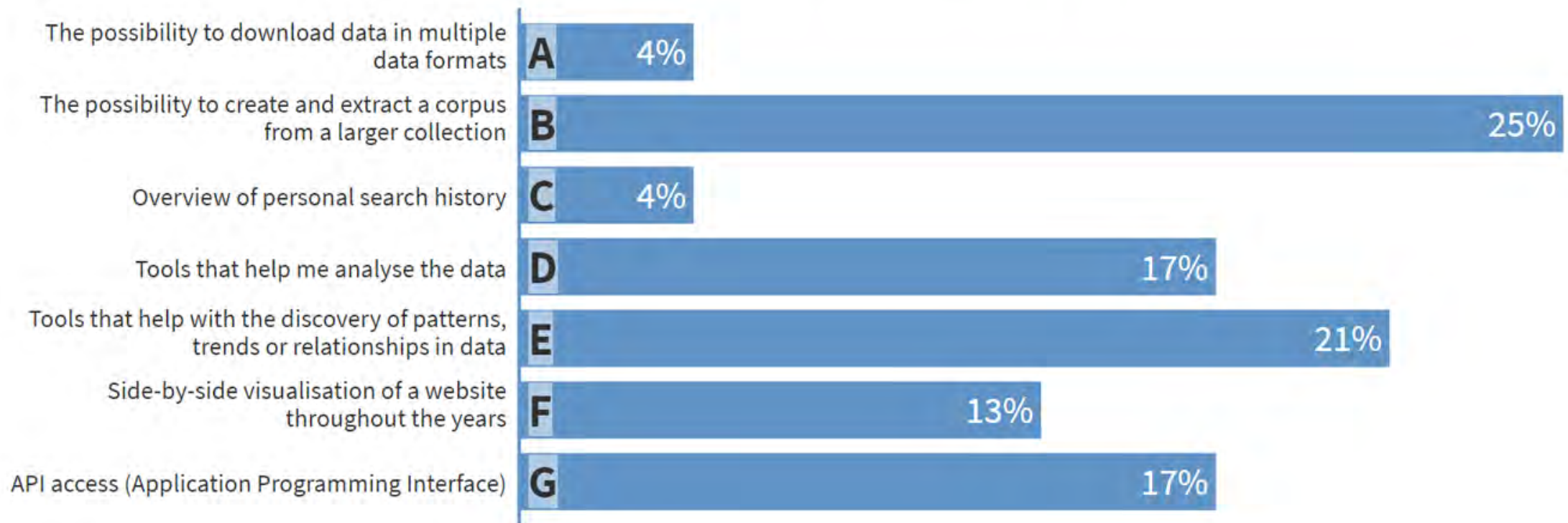
n=179



Respond at PollEv.com/friedelgeera923

Text **FRIEDELGEERA923** to **22333** once to join, then **A, B, C, D, E...**

Which other functionalities would you like to see in the Belgian web archive? (max. 3 options)



In working with data from web archives, what challenges have you faced? n=292

(1)

Challenges in accessing the data (e.g. web archive data was only available in the reading room).

17.6%

A lack of appropriate descriptive information about the web archives (e.g. lack of metadata).

16.4%

Difficulty in finding relevant materials (e.g. limited search options).

14.9%

Lack of content / collections in my research area.

9.7%

0 20 40 60 80 100 120 140 160 180



KBR



In working with data from web archives, what challenges have you faced? n=292

(2)

Data is in such a form that it is hard to process (messiness).



Lack of transparency and clarity regarding intellectual property rights.



The size of the data is too large.



Challenges conducting data analysis (e.g. lack of resources or tools to analyze the data).



0 20 40 60 80 100 120 140 160 180



KBR



Other interesting topics:

- Use of data from web archives
- Most important methods of analysis
- Most helpful metadata
- Subjects or topics of most interest
- Contribution to the selection of material
- Format to download the data
- Receiving guidance
- Collecting own web archives
- Quality control



Online dashboard:



“Clearly Google is the norm for the participants. They expected a search engine to work on full text, just like Google.”

Ras, M., & Van Bussel, S. (2007). Web archiving user survey. Online at: https://www.kb.nl/sites/default/files/docs/kb_usersurvey_webarchive_en.pdf.

“Access and use (...) remains a perceived area of need. Likewise, metadata is identified as an area that would benefit from ongoing knowledge-sharing around best practices. Social media and quality assurance continue to be recognized as areas for which better and more accessible tools are needed.”

Bailey, J., Grotke, A., McCain, E., Moffatt, C., & Taylor, N. (2017). Web Archiving in the United States: A 2016 Survey. National Digital Stewardship Alliance. Online at: https://ndsa.org/documents/WebArchivingintheUnitedStates_A2016Survey.pdf.

“(...) better discoverability options for the archived content, data selection and management, as well as better access to more ways of analysing the data is needed.”

Costea, M.-D. (2018). Report on the Scholarly Use of Web Archives. NetLab. Online at: http://netlab.dk/wp-content/uploads/2018/02/Costea_Report_on_the_Scholarly_Use_of_Web_Archives.pdf.





Hockx-Yu, H. (2013). Web Archiving and Scholarly Use of Web Archives. Online at: <http://docplayer.net/10376122-Scholarly-use-of-web-archives.html>.



KBR



Coordination of Webarchiving in the Netherlands

Jesse de Vos – Beeld & Geluid



Using Web archives for researchers: experience, challenges, issues

Valérie Schafer – C2DH



Saving the Belgian Web

Web archiving practices, research opportunities and limitations

Patricia Blanco – MSc in Digital Humanities (KU Leuven)

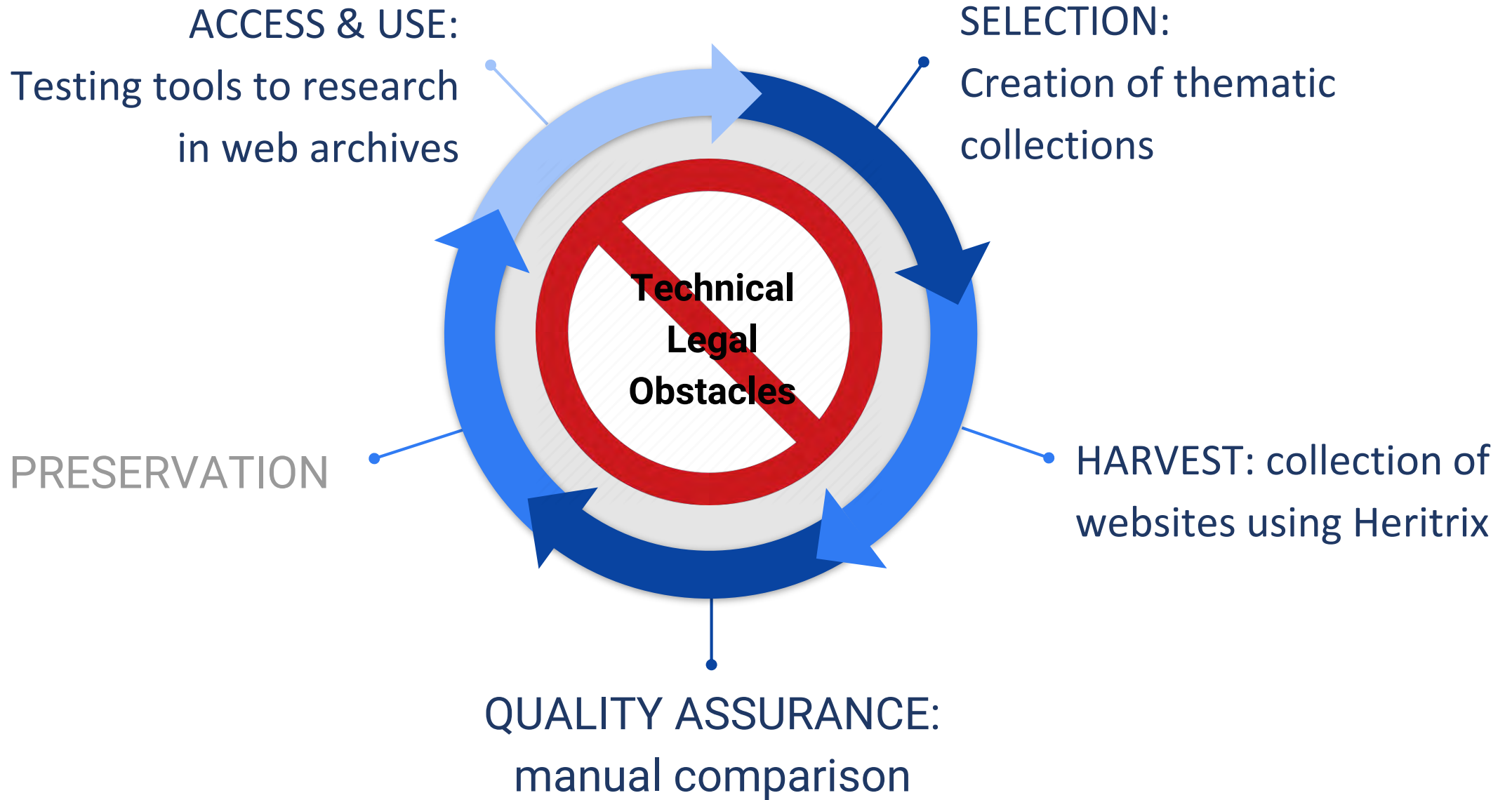


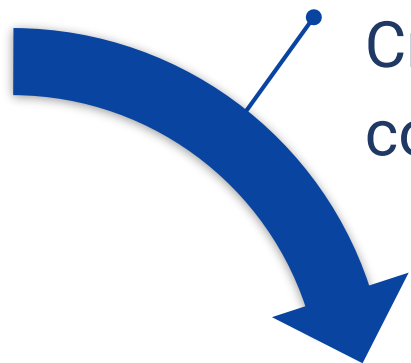
Web Archives?



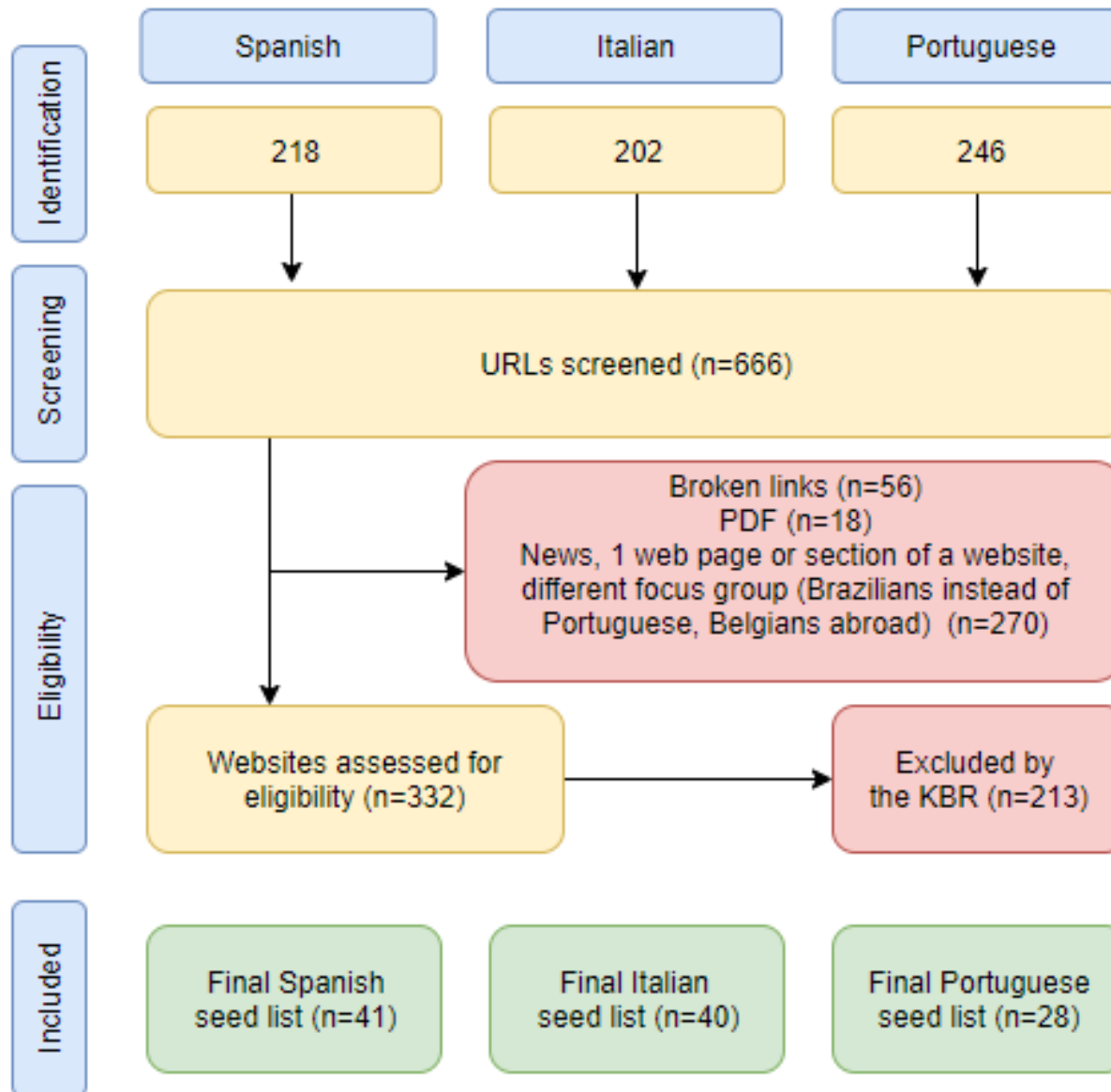
<http://bit.do/webarchiving>

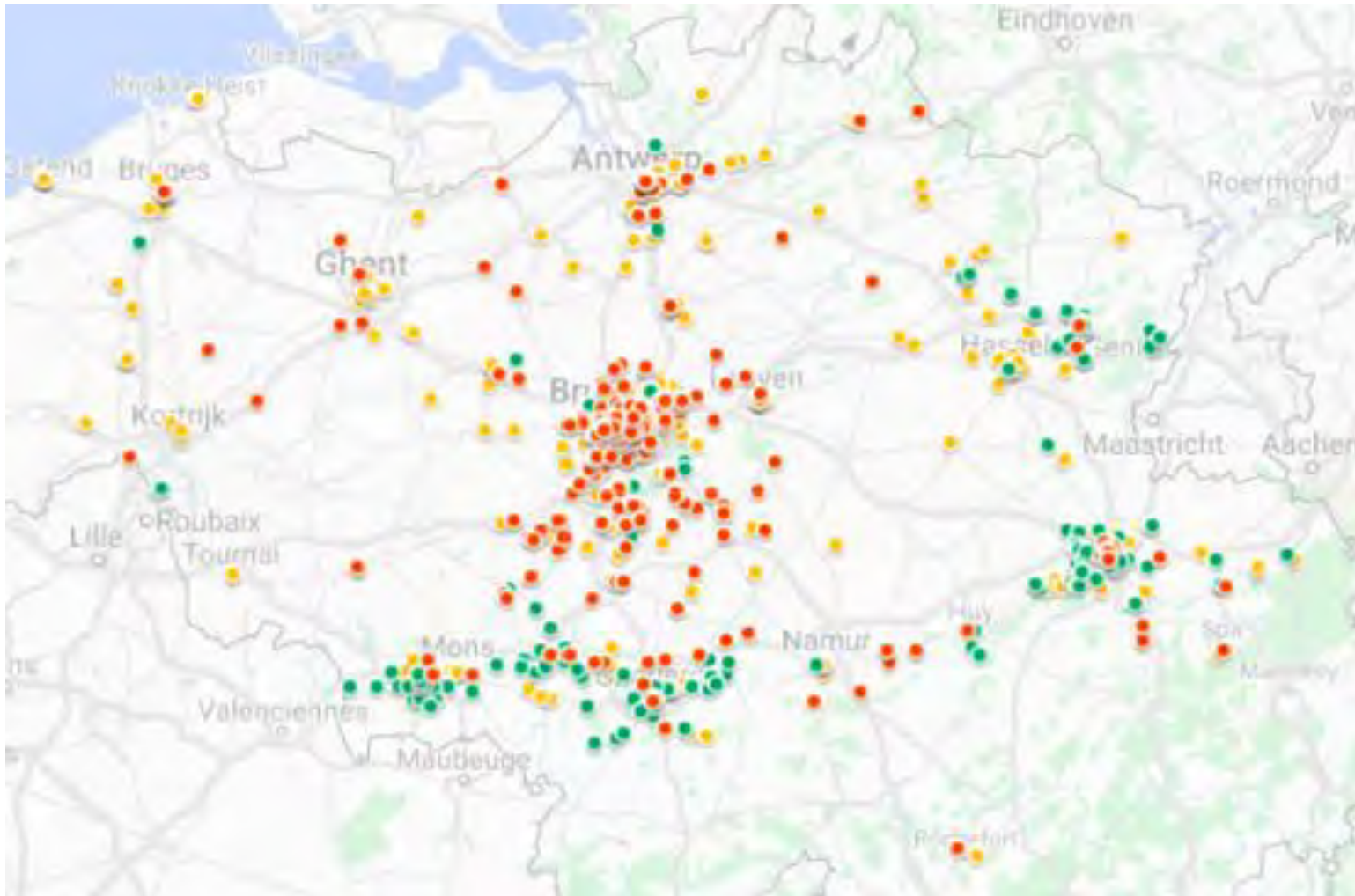






SELECTION: Creating a thematic collection





- Portugalnet
- Hispagenda
- Italiansnet

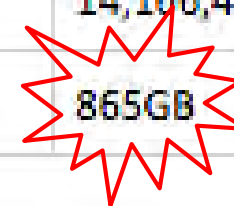
<http://bit.do/promise3>



KBR



	Spanish	Italian	Portuguese
Initial seed list	41	40	28
URLs downloaded	1,772,554	2,610,369	2,604,372
URLs queued	87,438	7,938,266	14,166,405
Size	81 GB	163 GB	865GB



	Spanish	Italian	Portuguese
Number of WARC files	22	11	36
Total size ⁷²	~ 70 GB		



The Archives Unleashed Toolkit

Archives Unleashed The Archives Unleashed Project

Home

About the Project

Getting Started

Archives Unleashed Toolkit

Introduction

Archives Unleashed Cloud

Archives Unleashed Notebooks

Warclight

Get Involved

Events

Publications

News (on Medium.com)

The Archives Unleashed Toolkit

<https://archivesunleashed.org/>

Where's the documentation?

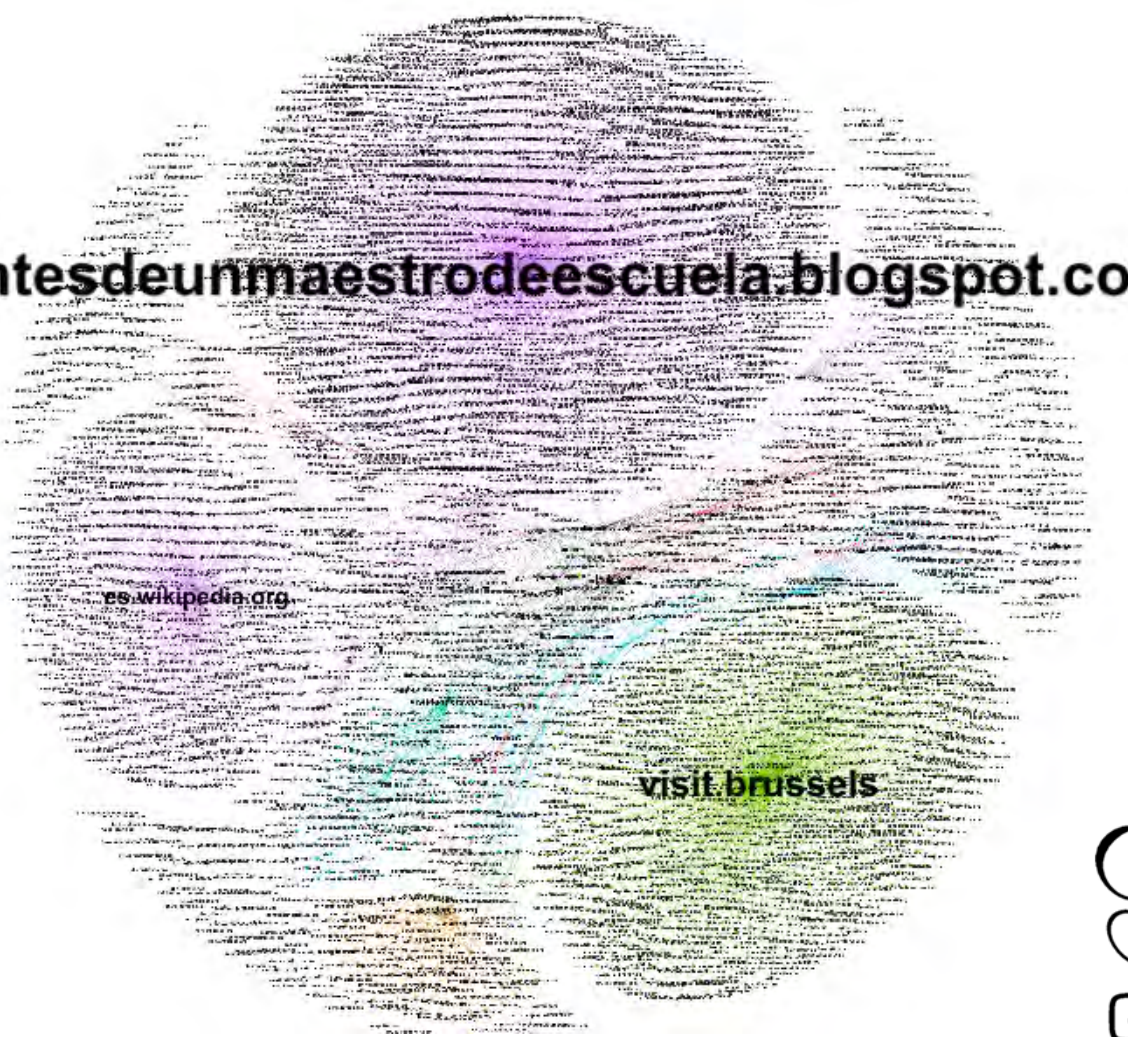
We've reworked our website, and the documentation now lives [here](#).

Introduction

```
Welcome to
Scala version 2.3.0
Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_161)
Type in expressions to have them evaluated.
Type :help for more information.
scala> :paste
// Entering paste mode (ctrl-D to finish)
import io.archivesunleashed.
```



apuntesdeunmaestrodeescuela.blogspot.com



KBR



(20190313, www.scopribruxelles.com, https://www.scopribruxelles.com/attivita/brussels-card/util/mixin, HTTP/1.1 404 Not Found Date: Wed, 13 Mar 2019 22:23:21 GMT Server: Apache Upgrade: h2,h2c Connection: Upgrade, close Set-Cookie: currency=EUR; expires=Fri, 13-Mar-2020 04:12:07 GMT; Max-Age=31556926; path=/ Set-Cookie: _client_id=29252316-1869-E18F-6E95-ECC393E68834; expires=Sun, 30-May-2027 22:23:21 GMT; Max-Age=259200000; path=/ Cache-Control: no-cache, max-age=0, must-revalidate, no-store Access-Control-Allow-Origin: https://ws-paylands.paynopain.com Transfer-Encoding: chunked Content-Type: text/html; charset=UTF-8 114ad Pagina non trovata - Civitatis.com Civitatis Bruxelles Guida di viaggio Info generali Info Close Info generali Organizza il viaggio Organizza il viaggio Documentazione necessaria Prezzi Lingua Tempo Orario commerciale Giorni festivi Vedi tutto Storia Brussels Card Foto Vedi tutto Cosa vedere Cosa vedere Close Cosa vedere Top 10 Top 10 Grand Place Bruges Manneken Pis Museo degli Strumenti Musicali Gand Atomium Palazzo di Giustizia Palazzo del Cinquantenario Gallerie Saint Hubert Mini Europe Vedi tutto Monumenti e attrazioni turistiche Monumenti e attrazioni turistiche Grand Place Manneken Pis Atomium Palazzo di Giustizia Palazzo Reale Palazzo del Cinquantenario Cattedrale di Bruxelles Basilica del Sacro Cuore Notre Dame du Sablon Parco di Bruxelles Gallerie Saint Hubert Mini Europe Quartiere Europeo Jeanneke Pis Palazzo di Coudenberg Vedi tutto Musei Musei Musei Reali delle Belle Arti del Belgio Museo degli Strumenti Musicali Museo Magritte Museo di Storia Militare Autoworld Museo della Città Museo del Cinquantenario Istituto Reale delle Scienze Naturali Museo Belvue Museo Horta Museo Charlier Museo del Fumetto Vedi tutto Visite vicine Visite vicine Bruges Gand Anversa Amsterdam Vedi tutto Vedi tutto Trasporto Trasporto Close Trasporto Metro Tram Autobus Taxi Biglietti e abbonamenti Vedi tutto Come arrivare Come arrivare Close Come arrivare Aeroporto di Charleroi Aeroporto di Bruxelles Vedi tutto Dove mangiare Dove mangiare Dove dormire Dove dormire Shopping Shopping Bruxelles in 48 ore Bruxelles in 48 ore Mappa Mappa Escursioni e visite guidate Transfer aeroporto Hotel in offerta Hotel Appartamenti Autonoleggio Voli + No tiene elementos EUR Euro Dollaro USA Sterlina Real brasiliano Peso messicano Peso argentino Le mie prenotazioni Aiuto FAQ Contattaci Carrello € Concludere la prenotazione Il tuo carrello è vuoto Condizioni generali Avviso legale Privacy Come sei arrivato qui? Ops, la pagina che stai cercando non esiste Info generali Cosa vedere Trasporto Come arrivare Dove mangiare Dove dormire Shopping Bruxelles in 48 ore Mappa Attività Ricevi le ultime offerte Bruxelles Guida di viaggio Escursioni e visite guidate Transfer aeroporto Hotel Appartamenti Autonoleggio Voli Civitatis Destinazioni Guide di viaggio di Civitatis Chi siamo? Civitatis Magazine Supporto Contatto FAQ Condizioni generali Avviso legale Politica di Privacy Cookie Lavora con noi Fornitori Affiliati Agenzie Come ci valutano 9,4 / 10 ★★★★★ ★★★★★ Più di 135.000 opinioni su civitatis.com Scarica la nostra APP Pagamento sicuro garantito è una marca di Civitatis Tours SL. C.I.C.M.A 2998 Facebook Twitter Linkedin Instagram 0)



KBR





VOYANT
see through your text



KBR



INTERNET ARCHIVE WayBackMachine

Please try our Beta Wayback Machine - Now with Site Search!

[DONATE](#)

Explore more than 279 billion web pages saved over time



Tools

- [Wayback Machine Availability API](#)
Build your own tools.
- [WordPress Broken Link Checker](#)
Banish broken links from your blog.
- [404 Handler for Webmasters](#)
Help users get where they were going.

Subscription Service

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. Visit [Archive-It](#) to build and browse the collections.

Save Page Now

Capture a web page as it appears now for use as a trusted citation in the future.
Only available for sites that allow crawlers.



Thank you

Questions?

P.BLANCO.NUNEZ@GMAIL.COM



Questions, wrap-up and discussion

