THE STUDY OF ARCHIVED WEB: USER PERSPECTIVES

NetLab Forum | 27th February 2018

Maria Costea
MA student 4th semester, Aarhus University, School of Culture and Society, Sustainable Heritage Management

Niels Brügger (supervisor)
Professor MSO, Head of Centre for Internet Studies and NetLab, Aarhus University

Background

- Archiving institutions are switching focus from data collection to users

- Efforts are being made by web archiving communities to develop the necessary research infrastructure for the study of web archives

- Feedback is necessary in order to develop tools and policies in accordance with users' needs

Background

- User studies:
  - Web Archive of the Netherlands (2007);
  - the Portuguese Web Archive (2010),
  - the British Library (2012);
  - the French National Library (2012).

- Dougherty et al. (2010): interviews with web archive stakeholders, exploring some of the challenges of use

- Use cases: *Using Web Archives to Understand the Past and the Present* (Brügger & Schroeder 2017)*,* as well as on the IIPC website

- Empirical knowledge of web archive use is, thus, still limited

Research Questions

Exploratory study, based on the following questions:

- Who are the users and non-users of web archives?

- What are the reasons why some researchers do not use them?

- How and why are web archives used for research?

- What are the data and research infrastructure needs of researchers using archived web materials?

- To what extent does the existing research infrastructure meet user needs?

- In which ways could the functionalities of web archives be improved to meet these needs?
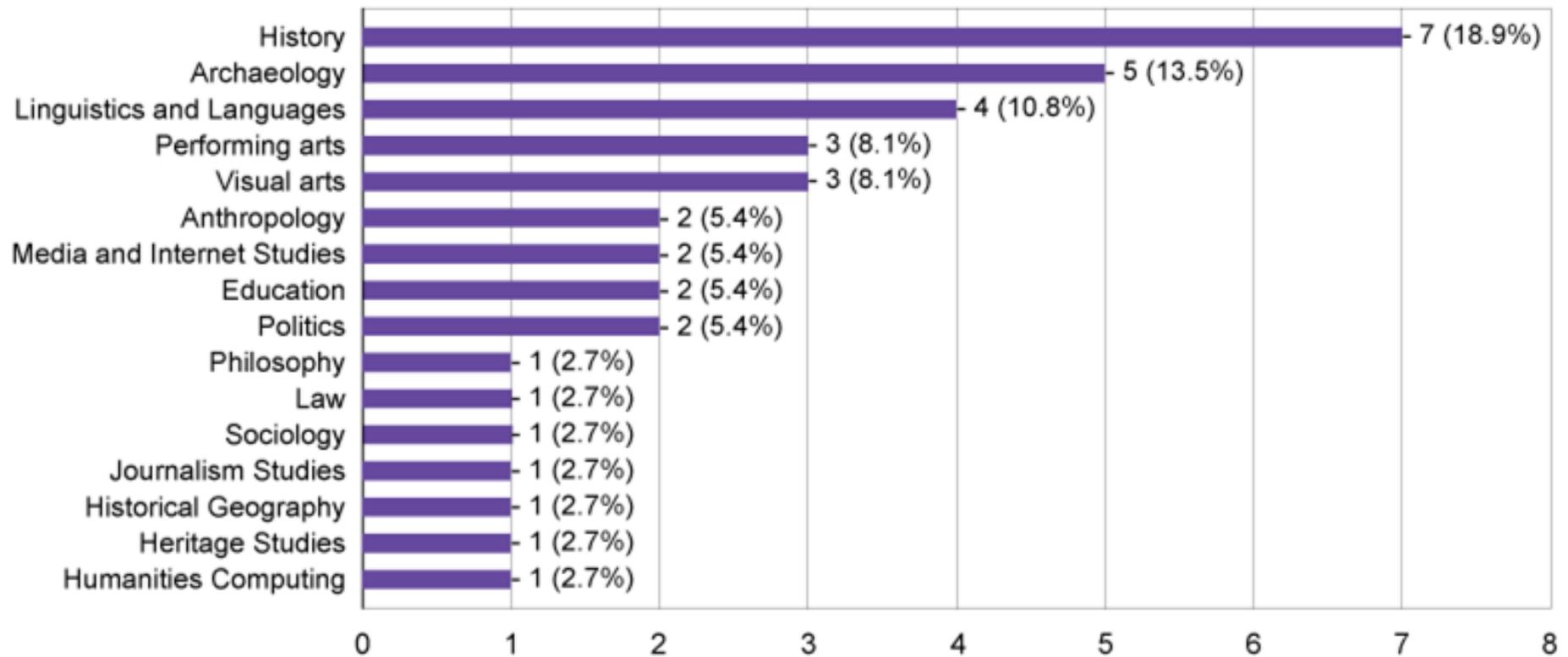
Data collection:

- Online survey
  - Invitations sent to academic staff of humanities and social sciences departments at Aarhus University and the University of Copenhagen
  - Sample: 88 individuals (36 users, 52 non-users)

- Interviews with users and non-users
  - 3 users were invited to detail their use method and describe the challenges they've experienced while using web archives
  - A test was performed with 2 non-users on the Internet Archive's Wayback Machine, providing feedback on usability
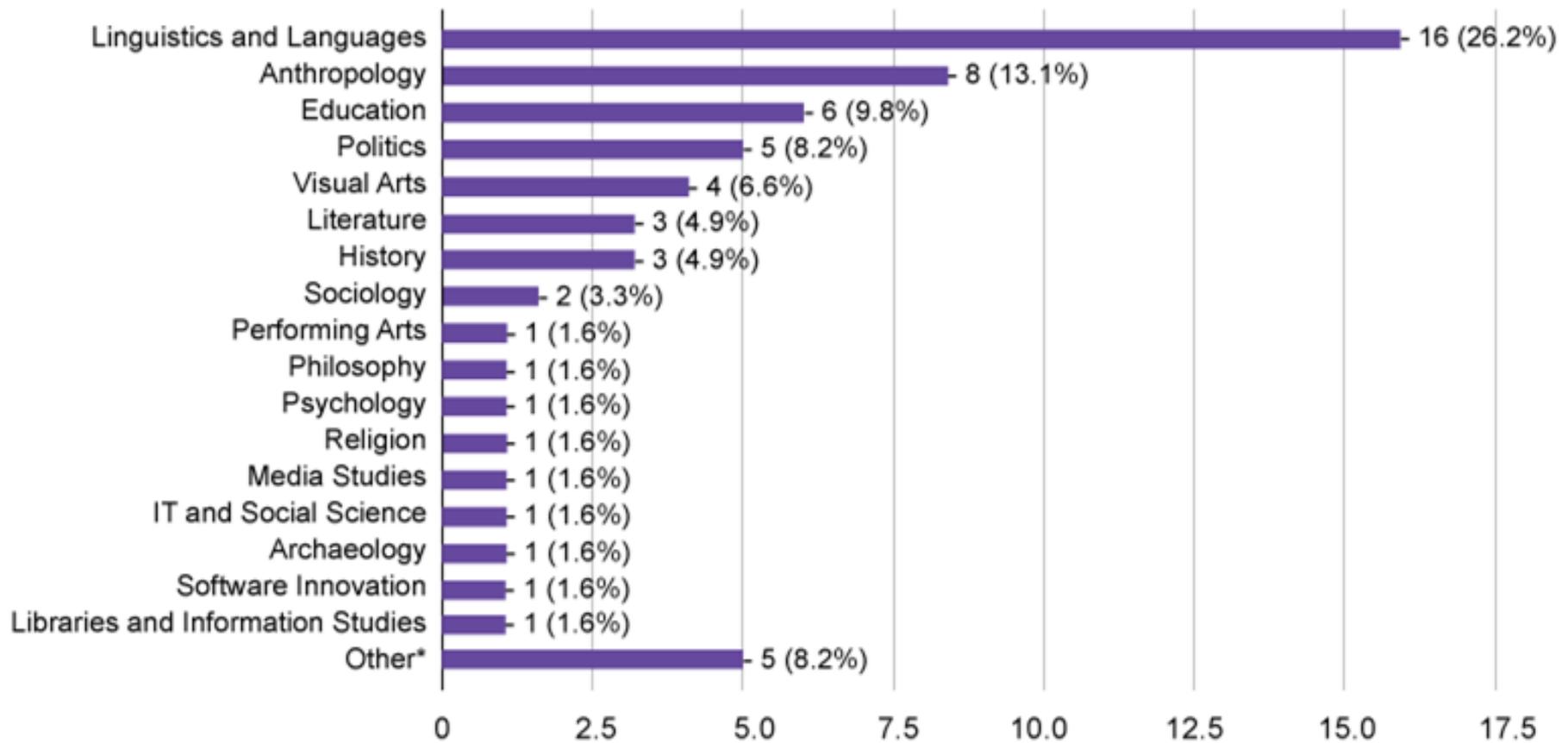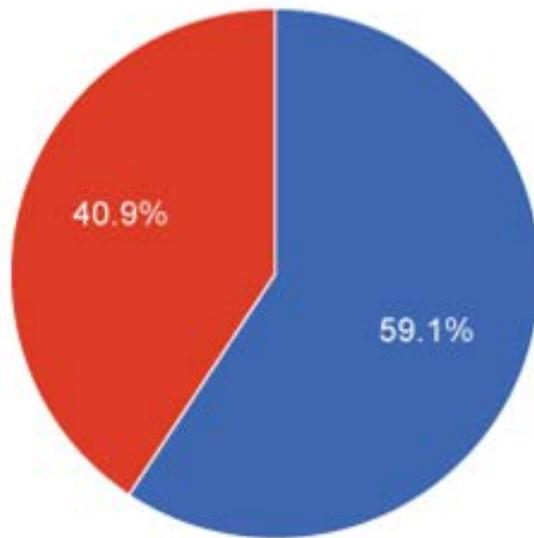
Online survey results

Fields of research: Users

Online survey results

Fields of research: Non-users

Online survey results
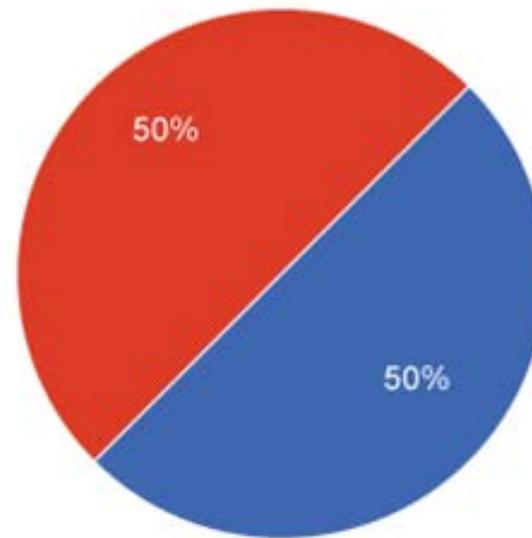
Were you aware that copies of Danish websites are archived by Netarkivet (the National Danish web archive) and made accessible to researchers?
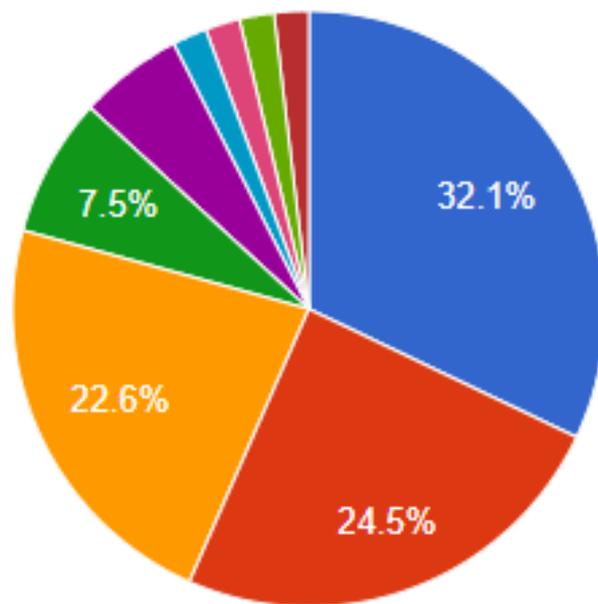

40.9%

59.1%

Overview for all participants

50%

50%

Web archive users

● No
● Yes

Reasons for not using web archives

- 🔵 I don't find the content relevant to my research — 32.1%
- 🔴 I didn't know of their existence — 24.5%
- 🟠 I do not know how to use them or what exactly they contain — 22.6%
- 🟢 I knew they existed, but they did not come to mind as a potential source for research — 7.5%
- 🟣 I do not know how to use them or what exactly they contain
- 🔵 I don't think they can provide me with accurate/sufficient information
- 🩷 Copyright and privacy law difficulties with offline use
- 🟢 I just recently started my research
- 🔴 I haven't had the opportunity to use web archives in research, but have used them to find materials for teaching

Online survey results

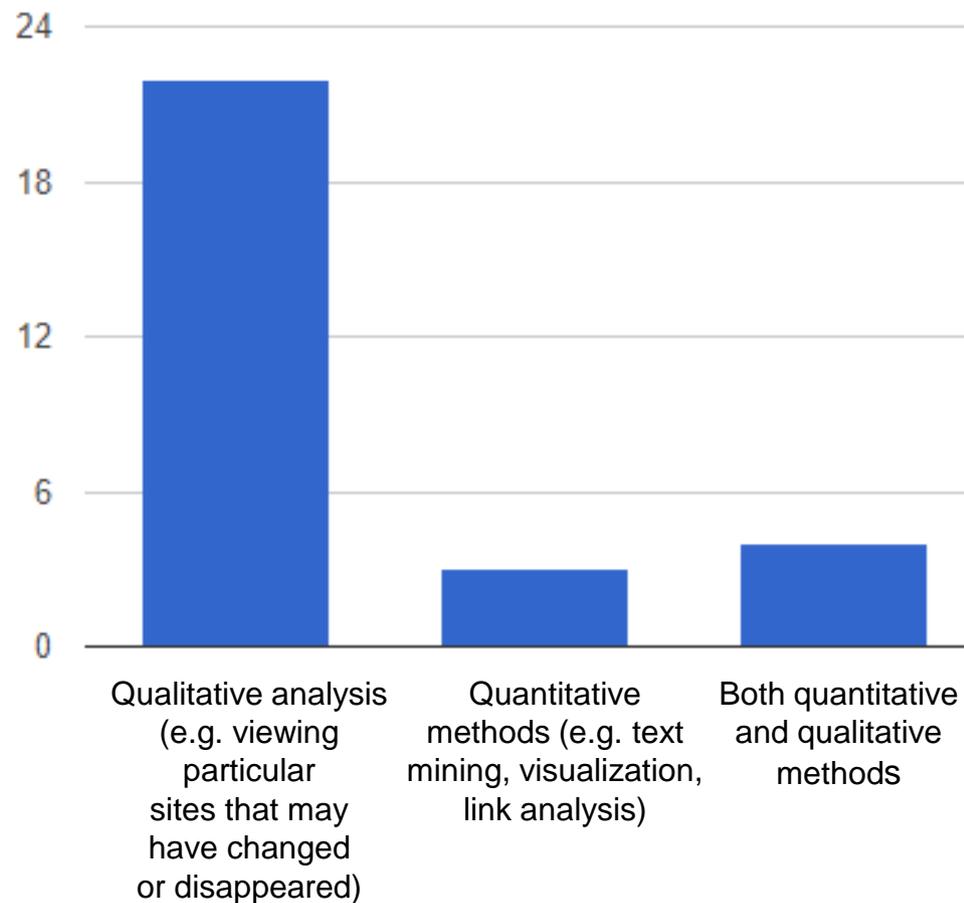## Value of web archives to research (users and non-users)

- 21.6% (n=11) of non-users and 56% (n=14) of users offered positive ratings, comments generally stating that this is due to their importance as a source for studying contemporary society

- A number of non-users (13.7% n=7) and users (44% n=11) provided a neutral rating, stating that that while the content of web archives is very valuable to research, there are a number of different issues which currently detract from their usefulness, related to:
  - Data reliability, completeness and integrity
  - Access and availability
  - Usability

- Lower scores were given by non-users who evaluated web archives in terms of personal relevance

Online survey results: Users

Which research methods do you use/have used with web archives?

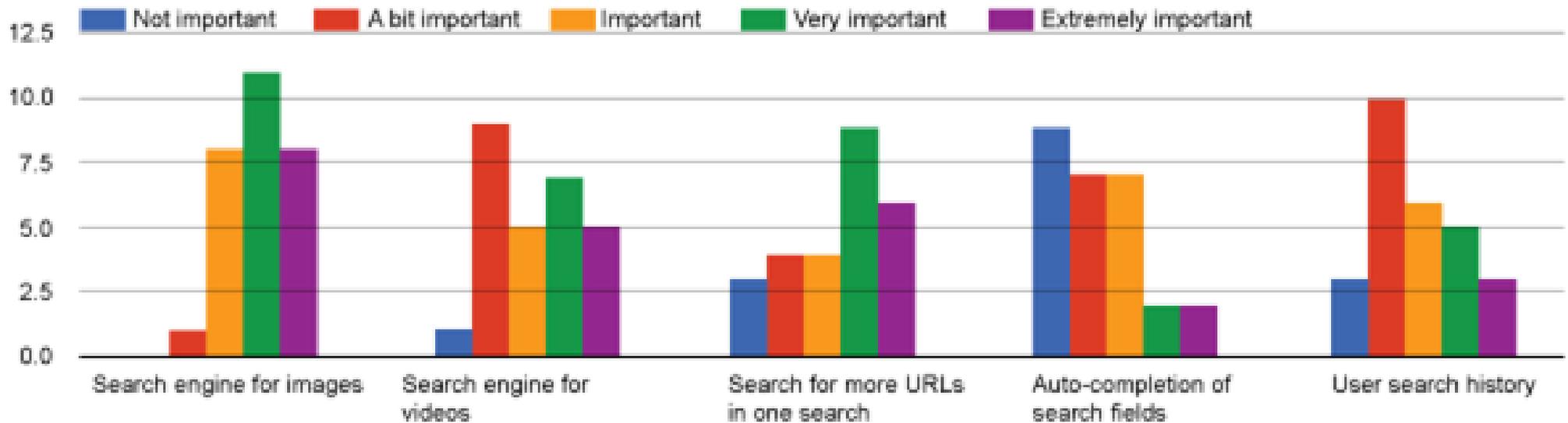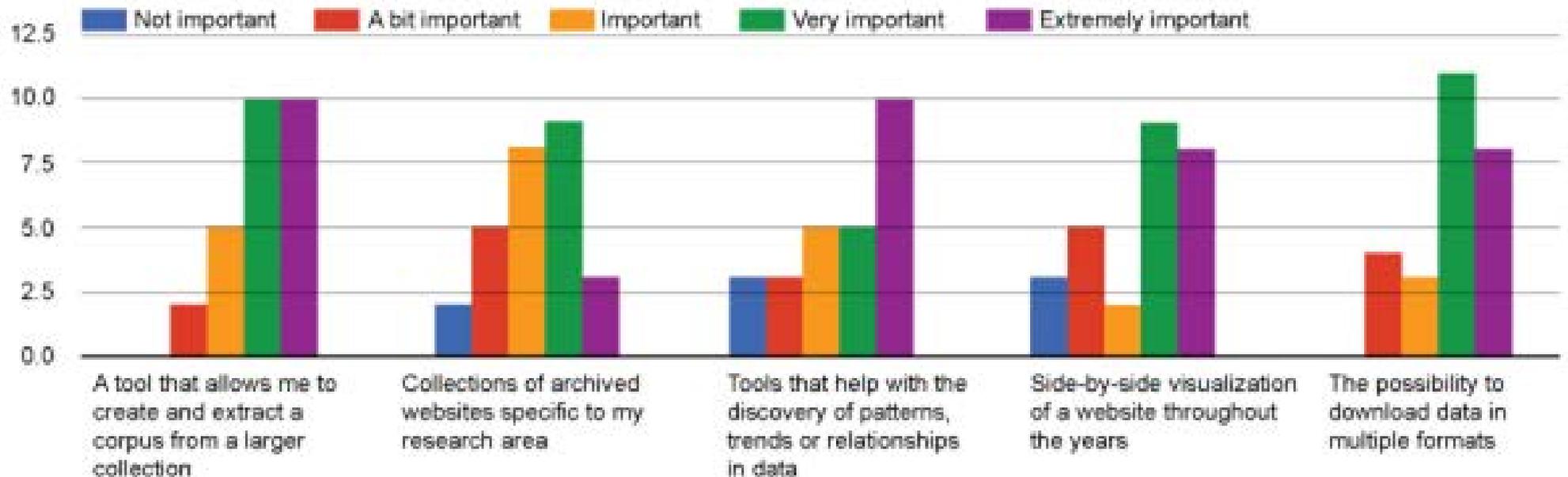**What researchers have already studied using web archives**

- Music history
- Organizational and personnel change in specific public offices
- Development of specific websites
- Development of style conventions
- Political discourse analysis
- Memory studies
- History of Danish radio broadcasting (DR web pages)
- How visual coverage of politics changes over time
- Church archaeology
- Social work research: finding websites of earlier municipal institutions
- Historical research
- Researching public comments pages
- Research on creative policy materials
- History of administration/legislation
- Children's media history – web pages for children
- Communication of design on social media

Online survey results: Users

How important do you think it is that the following search functionalities are included in web archives?

How important do you think it is that the following data selection, visualization and extraction functionalities are included in web archives?

Use cases

1. The historical evolution of the DR's role as an educator, examining how educational programs have been represented on DR's website

2. Analysis of PSB development when they have to switch from regular broadcasting (radio and television) to digital media, and the interactions that occur between the two types of platforms.

3. Multiple music history projects, such as researching the so-called "digital music revolution", for the period 2005-2012.

Challenges and suggestions

- Need for transparency concerning metadata and documentation in order to provide an accurate and reliable picture of the research object (e.g. which [if any] parts are missing from a web page)

- Users had to rely on other sources to find relevant case studies for research **>** overview of existing materials in the archive would improve this issue

- Difficult selection process due to no tools to bookmark or manage content. Users rely on self-made systems **>** wish for a "workspace" within the archive

- Being able to use the larger dataset as well as create smaller corpora for big data analysis

- Participants tested the Internet Archive's URL search, full text search and playback view

- Participant suggestions:
  - The possibility to search for different file types both within archives and within a selected domain would greatly help researchers seeking a specific type of content
  - Overview of archived content (also visual) in order to quickly see changes along time
  - Filters for results
  - Language too technical for beginners **>** provide at least an explanation of the jargon
  - User guides for casual and first-time users

Conclusions

- A significant segment of the research community in the humanities and social sciences is still unaware of the existence of web archives or how to use them.

- Researchers highly appreciate the value of the content, but they need better tools (e.g. discoverability, content management) and better policies to make use of this important resource

- Limited use of quantitative methods

- Further research needed on the use of web archives