

The WARCnet network: The first year

Niels Brügger

WARCNET PAPERS

WARCnet
web archive studies

The WARCnet network: The first year

Niels Brügger

`nb@cc.au.dk`



WARCnet Papers
Aarhus, Denmark 2021

WARCnet Papers ISSN 2597-0615.

Niels Brügger: The WARCnet network: The first year
© The author, 2021

Published by the research network
WARCnet, Aarhus, 2021.

Editors of WARCnet Papers: Niels
Brügger, Jane Winters, Valérie Schafer,
Kees Teszelszky, Peter Webster,
Michael Kurzmeier.

WARCnet
Department of Media and Journalism Studies
School of Communication and Culture
Aarhus University
Helsingforsgade 14
8200 Aarhus N
Denmark
warcnet.eu

Cover design: Julie Brøndum
ISBN: 978-87-94108-03-4

The WARCnet network is funded by the Independent Research Fund Denmark | Humanities
(grant no 9055-00005B).



WARCnet Papers

Niels Brügger: *Welcome to WARCnet* (May 2020)

Ian Milligan: *You shouldn't Need to be a Web Historian to Use Web Archives* (Aug 2020)

Valérie Schafer and Ben Els: *Exploring special web archive collections related to COVID-19: The case of the BnL* (Sep 2020)

Niels Brügger, Anders Klindt Myrvoll, Sabine Schostag, and Stephen Hunt: *Exploring special web archive collections related to COVID-19: The case Netarkivet* (Oct 2020)

Friedel Geeraert and Nicola Bingham: *Exploring special web archives collections related to COVID-19: The case of the UK Web Archive* (Nov 2020)

Friedel Geeraert and Barbara Signori: *Exploring special web archives collections related to COVID-19: The case of the Swiss National Library* (Nov 2020)

Matthew S. Weber: *Web Archives: A Critical Method for the Future of Digital Research* (Nov 2020)

Niels Brügger: *The WARCnet network: The first year* (Jan 2021)

Valérie Schafer, Jérôme Thièvre and Boris Blanckemane: *Exploring special web archives collections related to COVID-19: The case of INA* (Aug 2020)

Niels Brügger, Valérie Schafer, Jane Winters (Eds.): *Perspectives on web archive studies: Taking stock, new ideas, next steps* (Sep 2020)

Friedel Geeraert and Márton Németh: *Exploring special web archives collections related to COVID-19: The case of the National Széchényi Library in Hungary* (Oct 2020)

Friedel Geeraert and Nicola Bingham: *Exploring special web archives collections related to COVID-19: The case of the IIPC Collaborative collection* (Nov 2020)

Caroline Nyvang and Kristjana Mjöll Jónsdóttir Hjörvar: *Exploring special web archives collections related to COVID-19: The Case of the Icelandic web archive* (Nov 2020)

Sophie Gebeil, Valérie Schafer, David Benoist, Alexandre Faye and Pascal Tanesie: *Exploring special web archive collections related to COVID-19: The case of the French National Library (BnF)* (Dec 2020)

All WARCnet Papers can be downloaded for free from the project website warcnet.eu.

The WARCnet network: The first year

Niels Brügger

Abstract: This WARCnet Paper provides an overview of the WARCnet network's organisation and of the network's different activities in 2020, including dissemination of results.

Keywords: WARCnet, activities, 2020

IN THE SHADOW OF COVID-19: CHALLENGES AND OPPORTUNITIES

In the first year of the WARCnet network, 2020, the activities have taken place in the shadow of COVID-19. Two network meetings were planned, the kick-off meeting in Aarhus, Denmark in May, and the second meeting in Luxembourg in November. But the outbreak of the COVID-19 pandemic in Europe in March closed down most of the European countries, including travel and research activities, and therefore both planned meetings had to move online.

However, despite these extreme circumstances the network has managed to get started and to have more activities taking place than could be hoped for, had we even had the chance to meet in person. Nevertheless, the COVID-19 pandemic definitely has come with a number of challenges, but it has also opened up new opportunities.

On the one hand, the challenges. The funding was given to enable network building by bringing network members together at a series of meetings and allowing for going on Short-Term Network Stays. This has not been possible, participants have not been able to meet in person yet, which is definitely a challenge for a network. Online meetings are fine, but only online meetings is not fine in the long run. In particular this can be a challenge for people who are not familiar with other members of the network. As a consequence of this the network has (almost) not been able to spend any of the funding, despite the fact that a lot of the planned activities have been realised. Last, but definitely not least, many of the network participants have been challenged and affected in their daily lives to such an extent that it has been difficult and stressful to participate in the network activities at the level they would have liked to.

On the other hand, opportunities. One of the aims of the network was to study transnational events, and for better and worse the pandemic has provided a very concrete and current example of a transnational event. It has lent the network the possibility of

investiga-ting a large scale event, even while it is (unfortunately) unfolding, including the immediate collecting of testimonies about web archives' practices of collecting the web activity related to an event. Replacing physical meetings with online sessions has also made it possible to expand the network beyond the initial group by including a considerable number of new members, because funding resources were saved — apparently there was a need. On a more practical level the pandemic has made it much more straight-forward for network members to participate in online meetings in Working Groups between meetings — this would probably have been done anyway, but now it became more 'natural' to do it this way. Finally, with the pandemic the network got the opportunity to extend the project period by one year, thanks to the kindness and responsiveness of the funder, the Danish Independent Research Fund Denmark | Humanities.

The following paragraphs outline what the network has managed to do in the shadow of COVID-19.

NETWORK PARTICIPANTS, MEETINGS, AND WORKING GROUPS

The WARCnet network is a researcher network that is so fortunate as to also have web archives among its members. By the end of 2020 WARCnet had 38 participants, including 29 researchers, seven web archives, one IT developer, and a transnational international web archiving organisation, the International Internet Preservation Consortium (IIPC). When the network started there were 23 members, but in conjunction with the two first meetings 16 new members were included (and one member has left the network).

Network members have two venues for actually forming a network: Network meetings and Working Groups.

Network meetings

Initially, four three-day network meetings were planned, including a closing conference, but because of the extension of the project period the project will be hosting five three-day network meetings, including the closing conference. The kick-off meeting was held in May 2020 'in' Aarhus (Zoom, organised by Aarhus University, the School of Communication and Culture, Department of Media and Journalism Studies), the second 'in' Luxembourg (Zoom, organised by the University of Luxembourg, The Luxembourg Centre for Contemporary and Digital History (C²DH)), and the following meetings are planned to be in Aarhus (April 2021), London (November 2020), and Aarhus (June 2022 (closing conference)).

The kick-off meeting took place under the heading 'Taking stock, new ideas, next steps', and the aim was to get an overview of where we stood with web archive based research: What can web archives offer for the types of studies we would like to do? What do we have of studies (past, ongoing) about national webs, and events? Which tools are available? To what an extent can we manage the archived web material to prepare it for our studies, in each web archive and across web archives? And the kick-off meeting also invited participants to think about new ideas about the points above, and about what would be the next steps, in general and within this project. Therefore, within each of the sessions

there were two types of presentations: (a) Presentations from participants who had been within the field for a longer period of time, these presentations mainly focused on the first point above, (b) Brief presentations from participants who were new to the field, or external participants, and they primarily focused on the second point. The kick-off meeting included two keynotes, one by Matthew S. Weber, Associate Professor in the Hubbard School of Journalism and Mass Communication at the University of Minnesota, and one by Ian Milligan, Associate Professor, Department of History at University of Waterloo, and 35 pre-sentations by network members.

The heading for the second meeting in Luxembourg in November was 'Reporting progress, new ideas, next steps', and the aim was to get an overview of where we were after the kick-off meeting, and therefore some of the same questions were reiterated: What have web archives offered for the types of studies we would like to do? What approaches, methodologies, challenges and case studies have been selected and already started? What are the first results? What do we have of studies (past, ongoing) about national webs and events? To what an extent can we manage the archived web material for our studies, in each web archive and across web archives? Which tools are available and which tools are needed? What are the main challenges and achievements in each Working Group at this stage? Therefore, within the meeting there were three types of sessions: (a) Brainstorming sessions for each Working Group, (b) Working Groups' presentations to other Working Groups, gathering all WARCnet participants (research questions, methodologies, case studies, challenges, achievements, work plan and planned outreach), (c) Common sessions, gathering all WARCnet participants, to promote coordination between Working Groups and address key issues such as public engagement as well as outreach. The November meeting included two keynotes, one by Simone Natale, Associate Professor at the University of Turin, and one by Marta Severo, Professor of Information and Communication Sciences at Université Paris Nanterre, and 17 presentations by network members. Finally, a Twitter challenge aimed at the wider public was launched (described more in detail below).

Working Groups

Each network participant has been invited to join one of the four Working Groups (WGs), and joining more than one WG was also possible:

- WG1: Comparing entire web domains (chair: Janne Nielsen, PhD, Assistant Professor, School of Communication and Culture, Department of Media and Journalism Studies, Aarhus University), 10 participants,
- WG2: Analysing transnational events (chair: Valérie Schafer, PhD and HDR, Professor in Contemporary European History, The Luxembourg Centre for Contemporary and Digital History(C²DH), University of Luxembourg), 9 participants,
- WG3: Digital research methods and tools (chairs: Anat Ben-David, PhD, Associate Professor, The Open University of Israel, and Michael Kurzmeier, PhD Researcher, Centre for Digital Humanities, Maynooth University), 17 participants,

- WG4: Research Data Management across borders (chair: Sally Chambers, MA, Digital Humanities Research Coordinator, Ghent University, Ghent Centre for Digital Humanities), 8 participants.

To enable synergy the WARCnet Working Groups are strongly connected in a 'hashtag model' (Figure 1). WG1 and WG2 each focuses on one of the project's two topics: Entire web domains, and transnational events. But they are also connected to each other, since web domains constitute the backcloth on which events take place, and new web forms being used during events may affect the web domain. WG3 is about Digital Methods and tools which is a field in its own right, but also closely related to the two strands of research taking place in WGs 1 and 2. And the same goes for WG4 that focuses on Research Data Management, since the management of the project's research data is an indispensable topic to address in the research in WGs 1 and 2 as well as in relation to digital methods and tools.

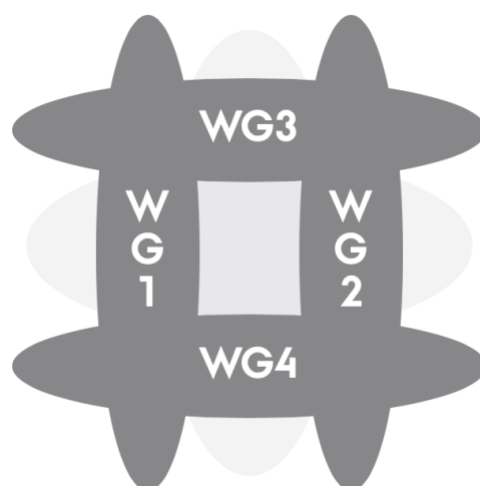


Figure 1: The 'hashtag model' that shall ensure synergy between WGs.

The WGs started working after the kick-off meeting in May 2020, and so far they have by and large been working independently with the overall aim to formulate a work plan for future work. This also includes reflections on how WGs can collaborate in concrete ways to ensure the expected synergies. In brief, the WGs first had to find their rationale and become WGs in their own right before starting to reach out to other WGs. At the time of writing all WGs have put forward ideas about what should be the first steps to be taken in the near future. The following points outline some of the achievements and plans for each WG:

- WG1: WG1 met online three times in 2020. To prepare for transnational research comparing national web domains WG1 started (a) mapping the characteristics of relevant web archives (scope, strategies, access etc.) building on the work done in the Belgian PROMISE project, and (b) discussing what characteristics/metrics would be relevant and possible to compare across national web domains. WG1 also discussed how to best approach the web archives with questions and made a plan for how to visualise web archive collections.

- WG2: The research group met three times online in 2020. It agreed on a timeline and a working process and on the case studies to be analysed (e.g. the COVID-19 crisis and the Euro-prides since 1996, as seen through web archives). The WG started a series of interviews with web archivists on the COVID-19 crisis as well as a survey targeting GLAM. It also developed a theoretical approach for studying transnational events through web archives and organised a datathon on COVID-19 metadata and derived data from several European web archives, that will take place in January 2021. The WG successfully submitted a proposal for the RESAW 2021 conference to be held in June 2021.
- WG3: WG3 has met online three times in 2020. Its goal was defined as developing digital methods to be used in web archive research as well as analyse, test and customise existing methods and tools by assessing their usefulness. Three projects are currently under development at WG3: (a) A scoping study: perform a systematic review of the types of research that was conducted using different types of archived web material, (b) a scoping study of skills: map the various skills and IT proficiency necessary to perform web archive research, and explore how to guide users of web archives to different datasets/research scenarios based on their available skills, (c) Sandboxes, which will explore the availability of existing datasets for data exploration or for testing skills/research scenarios. Additionally, the WG discussed holding a workshop at one of the upcoming general meetings.
- WG4: WG4's nine members have a rich range of expertise including researchers using web and social media archives, digital cultural heritage professionals and technical experts. WG4 met online three times during 2020. Following an iterative scoping exercise, WG4 decided to prepare a WARCnet Paper on 'Research Data Management Support for Research using Web Archives'. A key aspect of this work is to undertake semi-structured interviews with researchers to better understand their current research practices, with a particular emphasis on data. Additionally, WG4 successfully submitted a pre-conference workshop proposal for the RESAW 2021 conference, together with the colleagues from the BE-Social project, on User requirements for working with social media and web archives.

DISSEMINATION AND OUTREACH

From early 2020 the WARCnet network has started disseminating knowledge about what the network has been working with, and about planned activities. Communication about network activities has taken place on the project website at warcnet.eu, on social media platforms, and in a Newsletter, whereas results have been disseminated in the network's publication series WARCnet Papers, and network initiated conference presentations have been submitted and accepted. Finally, WARCnet's Twitter Challenge has constituted an example of outreach to a wider public.

warcnet.eu, social media, and Newsletter

A project website was established in January 2020 at the web address warcnet.eu (hosted by Aarhus University). On the website one can find relevant information about the project's life, including presentations from the kick-off meeting in May and from the network meeting in November, either as a video or as a slideshow. In 2020 warcnet.eu had 3,087 visitors. A quarter of the visitors come from Denmark, and then follows the United Kingdom (13%), the United States (13%), France (9%), and the Netherlands (9%). Most users access the website from a stationary computer (85%), and 15% use either a mobile phone or a tablet. Not surprisingly the landing page is the most popular, and then follows the web page with the WARCnet Papers (6% of the visitors go there), the WARCnet team page (4%), the web page about the kickoff meeting in Aarhus (4%), and the About page (4%). In addition, the landing page and the web page with WARCnet Papers are also the most used entry points to the website, 45% of the visitors start on the front page, 9% starts on the WARCnet Papers page. If we then look at how users first came to the website, the vast majority simply typed in the web address in their browser (75%), whereas 14% came from social media, 6% from external referrers, and 5% from search engines. Among those coming from social media almost all of them came from Twitter. Finally, when looking at the distribution of visitors over the year it is clear that the main visitor activity is related to network meetings (May and November), but the interest in the website already starts to grow 2-3 months before the events take place.

WARCnet has established a presence on Facebook, Twitter, YouTube, and Slide-share of which the latter two are mainly used for disseminating presentations from network meetings. The Facebook page has 13 followers, the Twitter profile has 194 followers, the YouTube channel has 9 subscribers and has been viewed 888 times, and the Slideshare page has been viewed 7,262 times, including 6 downloads.

Finally, WARCnet publishes a 3-4 pages Newsletter with information about the network. Four issues have been sent out in 2020 to 70 subscribers from all over the world.

WARCnet Papers

Before the kick-off meeting in May it was decided to launch a publication series, WARCnet Papers, with the aim of offering a home for network related material that for different reasons does not fit well into the more traditional publication outlets such as academic journals or edited books. To ensure the relevance of the publications, WARCnet Papers strives to publish with a rapid turnover, all publications are publicly available at the project website, and they cover content types such as keynotes, interviews, round table discussions, presentations, extended minutes, reports, white papers, status reports, and similar.

All WARCnet Papers go through an editorial review, and in all other cases than interviews, round table discussions, and similar WARCnet Papers go through a process of single blind review. The editorial process of the WARCnet Papers is managed by an editorial group with six members.

Since the series was launched in May 13 papers have been published in 2020, including a welcome paper about the WARCnet project, two keynotes, and a so-called 'round doc' discussion (see the complete list at the end).

One special WARCnet Papers project deserves to be singled out, namely the COVID-19 interviews launched by WG2. The idea was to collect empirical information about how web archives select and collect web material related to the COVID-19 pandemic. All these interview publications — by and large — follow the same interview guide with the same questions which makes it possible to make comparisons across web archives. These publications constitute valuable information in their own right about how cultural heritage institutions handle this transnational event, but they will also serve as the empirical material on which future academic papers and publications initiated by the WG2 members can build, including qualitative and well as quantitative approaches, such as analyses of word frequencies across all interviews. As of today nine WARCnet Papers have been published about the COVID-19 pandemic in web archives, and more are in the pipeline.

It is interesting to know if the WARCnet Papers are actually read, and although download figures are only indicative of an interest they show if there is any interest at all. The 13 WARCnet Papers published during 2020 have been visited 307 times on warcnet.eu in 2020.

Conference papers and publications

Although it is still early in the network's project period a number of network initiated presentations and round tables have already been submitted and accepted for the 4th RESAW Conference 'Mainstream vs marginal content in Web history and Web archives' to be held in Luxembourg June 2021 (cf. the list at the end).

One of the challenges for web archive researchers is to have material in national web archives extracted for further study. In 2018 the national Danish web archive Netarkivet hosted at the Royal Danish Library allowed for the extraction of material for research purposes, based on a delivery agreement between the library and the researcher's institution. Since this agreement may serve as an inspiration for other national web archives the Royal Danish Library has had the agreement translated into English, initiated by an interest formulated in WG4.

Outreach: WARCnet Twitter Challenge, and video reporting

With the aim of reaching out to a wider community outside academia and to raise awareness about the archived web and web history in general WARCnet launched the 'WARCnet Twitter Challenge' in relation to the network meeting in November. The 'WARCnet Twitter Challenge' invited anyone to (re)discover the old web and to participate in the challenge by tweeting their reply using the hashtag #WarcnetChallenge, and by adhering to the format of the challenge:

Your web archiving discovery challenge entry must consist of three parts: 1. Your discovery (what have you found?), 2. Your method (how did you find it or discover it?), 3. Your story (why is your find so special?). (warcnet.eu)

17 entries were received, 5 were nominated, and one was voted the winner by network members (all entries and more related information can still be found on twitter.com, #WarcnetChallenge). The challenge is supposed to continue during the entire project period, with different themes, and with a final winner elected at the closing conference in 2022.

WARCnet has also used other means of reaching out, namely a two-minute video presentation about the first year of WARCnet activities, to be published in relation to the C²DH Annual Report at the University of Luxembourg.



MANAGEMENT

The WARCnet network is headed by three PIs, and together with three network members (one Early Career Researcher, one web archive representative, and one IT developer) they form the Steering Group. The purpose of the Steering Group is to oversee the management of the WARCnet network, including, among others, defining, planning and approving network activities, establishing and implementing a procedure for the addition of new members to the network, and monitoring the project budget.

Also, to ensure the coordination between the WGs a Coordination group has been formed, consisting of all Steering Group members and the four WG chairs. Finally, a Short-Term Network Stay Committee has been set up to manage the allocation of funded network stays. Last but not least the administration and management of the project is supported by two student helpers, both from Aarhus University.

COMING YEARS

Because of the COVID-19 pandemic the WARCnet network has been able to extend its activities by one year. Two network meetings are planned in 2021 (Aarhus in April, London in November), and in addition to presenting and discussing the network's research activities it is also considered to include a training workshop about analytical software to support the researcher use of web archives. The closing conference is scheduled to take place in June 2022, in Aarhus, and it will be open to participants from outside the network.

Hopefully, it will also be possible to travel again in 2021, thus enabling the planned Short-Term Network Stays where network members can visit one of the institutions of another network member. Finally, the four individual WGs continue to meet and plan activities, as indicated above. Hopefully, 2021 will be the year when the WARCnet network is getting well out of the shadow of COVID-19.

BRIEF OVERVIEW OF DISSEMINATION RELATED TO NETWORK ACTIVITIES

Accepted conference presentations

- Brügger, N. Link structures between small, medium size, and big websites on the Danish web in 2006 and 2015. The 4th RESAW Conference 'Mainstream vs marginal content in Web history and Web archives', to be held in Luxembourg June 2021.
- Geeraert, F., Bingham, N., Brügger, N., Gebeil, S., Nanni, F., Nyvang, C., Schafer, V., Strandgaard Jensen, H., de Wild, K., Winters, J. Representation, participation and inclusivity: European web archives collecting the digital traces of COVID-19. The 4th RESAW Conference 'Mainstream vs marginal content in Web history and Web archives', to be held in Luxembourg June 2021.
- Helmond, A., Brügger, N., Winters, J., Schafer, V., Chambers, S., Nielsen, J., Ben-David, A. Roundtable on the WARCnet project. The 4th RESAW Conference 'Mainstream vs marginal content in Web history and Web archives', to be held in Luxembourg June 2021.
- Mechant, P., Vlassenroot, E., Chambers, S., Brügger, N., Aasman, S., Geeraert, F., Pranger, J. User requirements for working with social media web archives, Pre-conference workshop. The 4th RESAW Conference 'Mainstream vs marginal content in Web history and Web archives', to be held in Luxembourg June 2021.
- Winters, J. Web archives as sites of collaboration. 'Engaging with Web Archives: Opportunities, Challenges, Potentialities' conference, Maynooth, Republic of Ireland, September 2020.

Publications

- Royal Danish Library (2020). Delivery agreement for digital copy delivery from the Netarchive (online archive) of the Royal Danish Library. Aarhus/Copenhagen: Royal Danish Library. kb.dk/webarchiveaccess.


WARCnet Papers

- Brügger, N. (2020). *Welcome to WARCnet*. Aarhus: WARCnet.
- Brügger, N., Myrvoll, A., Schostag, S., Hunt, S. (2020). *Exploring special web archive collections related to COVID-19: The case of Netarkivet*. Aarhus: WARCnet.
- Brügger, N., Schafer, V., Winters, J. (2020). *Perspectives on web archive studies: Taking stock, new ideas, next steps*. Aarhus: WARCnet.
- Gebeil, S., Schafer, V., Benoist, D., Faye, A., Tanesie, P. (2020). *Exploring special web archive collections related to COVID-19: The case of the French National Library (BnF)*. Aarhus: WARCnet.
- Geeraert, F., Bingham, N. (2020). *Exploring special web archives collections related to COVID-19: The case of the IIPC Collaborative collection*. Aarhus: WARCnet.
- Geeraert, F., Bingham, N. (2020). *Exploring special web archives collections related to COVID-19: The case of the UK Web Archive*. Aarhus: WARCnet.

- Geeraert, F., Németh, M. (2020). *Exploring special web archives collections related to COVID-19: The case of the National Széchényi Library in Hungary*. Aarhus: WARCnet.
- Geeraert, F., Signori, B. (2020). *Exploring special web archives collections related to COVID-19: The case of the Swiss National Library*. Aarhus: WARCnet.
- Milligan, I. (2020). *You shouldn't Need to be a Web Historian to Use Web Archives*. Aarhus: WARCnet.
- Nyvang, C., Hjörvar, K.M.J. (2020). *Exploring special web archives collections related to COVID-19: The Case of the Icelandic web archive*. Aarhus: WARCnet.
- Schafer, V., Els, B. (2020). *Exploring special web archives collections related to COVID-19: The case of BnL*. Aarhus: WARCnet.
- Schafer, V., Thiévre, J., Blackemane, B. (2020). *Exploring special web archives collections related to COVID-19: The case of INA*. Aarhus: WARCnet.
- Weber, M.S. (2020). *Web Archives: A Critical Method for the Future of Digital Research*. Aarhus: WARCnet.

Outreach

- Schafer, V., Winters, J., Brügger, N. (2020). Video presentation of WARCnet for C²DH Annual Report, 3 minutes. Luxembourg: The University of Luxembourg.
- WARCnet Twitter Challenge (2020). WARCnet, <https://cc.au.dk/en/warcnet/warcnet-twitter-challenge/>



WARCnet Papers is a series of papers related to the activities of the WARCnet network. WARCnet Papers publishes keynotes, interviews, round table discussions, presentations, extended minutes, reports, white papers, status reports, and similar. To ensure the relevance of the publications, WARCnet Papers strives to publish with a rapid turnover. The WARCnet Papers series is edited by Niels Brügger, Jane Winters, Valérie Schafer, Kees Teszelszky, Peter Webster and Michael Kurzmeier. In cases where a WARCnet Paper has gone through a process of single blind review, this is mentioned in the individual publication.

The aim of the WARCnet network is to promote high-quality national and transnational research that will help us to understand the history of (trans)national web domains and of transnational events on the web, drawing on the increasingly important digital cultural heritage held in national web archives. The network activities run in 2020-22, hosted by the School of Communication and Culture at Aarhus University, and are funded by the Independent Research Fund Denmark | Humanities (grant no 9055-00005B).

WARCNET PAPERS

