

Welcome to WARCnet

Niels Brügger

WARCNET PAPERS



Welcome to WARCnet

Niels Brügger

Department of Media and Journalism Studies
School of Communication and Culture, Aarhus University
nb@cc.au.dk



WARCnet Papers
Aarhus, Denmark 2020

WARCnet Papers ISSN 2597-0615.

Niels Brügger: Welcome to WARCnet
© The author, 2020

Published by the research network WARCnet, Aarhus, 2020.
Editors of WARCnet Papers: Niels Brügger, Jane Winters, Valérie Schafer.

Cover design: Julie Brøndum
ISBN: 978-87-972198-0-5

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

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



DANMARKS FRIE
FORSKNINGSFOND
INDEPENDENT RESEARCH
FUND DENMARK

WARCnet Papers

Niels Brügger: *Welcome to WARCnet* (2020)

Welcome to WARCnet

Niels Brügger

Abstract: This publication starts by briefly introducing the prehistory of the researcher network WARCnet, 'Web ARChive studies network researching web domains and events'. Then follows an overview of the network, including a network analysis of the relations between network participants, and an outline of the planned network activities. The publication is rounded off by an epilogue.

Keywords: WARCnet, web archive, web history, national web, event

BEGINNINGS

Let us begin at the beginning. In the case of WARCnet the beginning holds promise of the future that the WARCnet research network is part of, since the project's two research topics — national web domains, and events on the web — were already present in the very early days of web archiving. In the Summer of 1997 the Royal Library of Sweden established the web archive collection Kulturarw3 with the aim of archiving the .se web domain, that is, an entire national web domain, just like the ones that are to be studied in the WARCnet network (cf. Brügger, 2011, p. 31). Also in the mid-1990s, in 1996, the American Internet Archive was established, and to demonstrate the value of preserving websites it started out by archiving web activity related to an important event, the US Presidential elections in 1996 (cf. Kimpton & Ubois, 2006, p. 202). Thus, one could argue that from the very beginning of web archiving the two lines of inquiry of the WARCnet network were already laid out.¹

However, the establishment of web collections does not necessarily lead to web archive research. To make that happen two things are needed: a research infrastructure to open up the web archives to researchers; and a critical mass of researchers who are interested in studying web archive collections. Two very early examples of researcher interest in events on the web exist: first, in 2001-02, the Danish pilot project 'netarkivet.dk' (the precursor to the national Danish web archive), in which three events on the web were studied: municipal elections, Parliamentary elections, and 9/11 (Christensen-Dalsgaard et al.,

1. For histories of the establishing of web archives see Brügger, 2011, p. 29-32; Laursen & Møldrup-Dalum, 2017; Koerbin, 2017; Webster, 2017, 2018.

2003, p. 8); and second, Kirsten Foot and Steven Schneider's pioneering book *Web Campaigning* (Foot & Schneider, 2006), about the U.S. elections of 2000, 2002, and 2004. However, these very early projects turned out to be isolated incidents, and their impact among researchers was limited because a wider research community which could support long-term interest was lacking.² It has taken some time to establish research infrastructures and researcher interest, but we are slowly getting there, and the WARCnet research network is the result of three previous attempts to establish a transnational research infrastructure and to raise awareness among scholars across countries.

In 2016–17 a two-step EU Horizon 2020 application for a transnational research infrastructure was submitted, 'A Research Infrastructure for the Study of the Archived Web' (RESAW). Unfortunately, the RESAW application was not successful, so in 2018 an application for a COST Action was drafted for the research network TRUEwarc, 'Transnational Researcher Use of Web Archives'. While the RESAW application was mainly about the development of a transnational research infrastructure for the study of the archived web, in combination with research projects, things were reversed with the TRUEwarc application where focus was more clearly on research and less on research infrastructure. But the TRUEwarc COST Action project suffered a similar fate to the RESAW application. It was then improved and resubmitted in 2019 with the acronym DIGICULTES, 'Digital Cultures in Europe through Web Archives', alas, still without success.

However, although none of these three applications was successful, their drafting has de facto established a sort of 'network in the making' on which the present network can build. One of the tangible forums where this informal proto-network has materialised is the RESAW Conferences that were initially established as one among other interim activities between the time when the RESAW application was initially planned to be submitted, in 2014, and the 2016 call where the application was actually submitted. The first RESAW Conference was organised in Aarhus, Denmark in 2015 as a one-off event, but since its great success reflected a growing interest in the field among web archivists as well as web archive scholars it was decided to make the conference the first in a series, of which the following two were organised in London (2017) and Amsterdam (2019), and the next two are planned to take place in Luxembourg (2021) and Marseille (2023).

Not surprisingly many of the participants in WARCnet have been part of this application journey from the very early beginnings in September 2012, when the first call for participants in what later became RESAW was sent out. It is therefore a great pleasure that so many have stayed on board to see that in 2019 the Danish Independent Research Fund Denmark | Humanities decided to support a transnational researcher network, with the national web archives as natural participants.

It is also worth noting that despite the fact that no transnational project as such has succeeded yet, a number of national research projects with participants from other countries have been launched, including the UK-based 'Big UK Domain Data for the Arts and

2. Winters (2017) identifies some of the main challenges for raising awareness among researchers about the potential inclusion of the archived web in their studies.

Humanities' (BUDDAH) project (2014, funded by the Arts and Humanities Research Council; see Cows, 2017), and the French 'Archives Sauvegarde Attentats Paris' (ASAP) (2016, funded by the Centre National de la Recherche Scientifique; see ASAP). In addition, other actors have played a pivotal role in contributing to the creation of transnational research infrastructures and raising awareness about web archives among researchers, most notably the International Internet Preservation Consortium (IIPC), whose open annual Web Archiving Conference (WAC) has for many years constituted a valuable forum where web archivists and researchers can meet and share ideas.

All in all, today the WARCnet network is part of a vibrant international web archive studies community, with a growing interest among web archives in collaborating with researchers, and where new researchers, in particular young researchers, continue to join in. The stage is set for taking the next steps in the development of the interface between web archives and researchers.

The WARCnet network is only made possible by the support of a number of organisations, institutions and individuals. I give thanks to the Danish Independent Research Fund Denmark | Humanities for having supported the idea for this network; to Aarhus University, and particularly the Department of Media and Journalism Studies in the School of Communication and Culture, for hosting the project; and to all the participants for their enthusiasm and dedication. A special thanks to the person who by coincidence started this journey. In September 2012, at the official opening of the Danish national research infrastructure Digital Humanities Lab, DIGHUMLAB, Sally Chambers, at the time Secretary General for the DARIAH-EU Coordination Office, suggested to me and a colleague that NetLab, the web archive section of DIGHUMLAB, should head a submission of an expression of interest to the EU Commission's consultation on 'Possible topics for future activities for integrating and opening national research infrastructures'. Sally Chambers, now at the Ghent Centre for Digital Humanities and a participant in WARCnet, is thanked for this somewhat casual suggestion many years ago that eventually became the RESAW application and later, through several detours, the WARCnet network.

WHAT WARCNET WILL BE DOING

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. WARCnet offers transnational interdisciplinary networking activities for researchers who study the archived web, and it reaches out to web archivists and IT developers. The networking activities are guided by an overarching research question: how have (trans)national web domains, and transnational events, developed on the European web?

WHO ARE 'WE'?

The WARCnet network is composed of researchers who have been pioneers within the emerging field of web archive studies and whose research interests and competencies within the social sciences and the humanities support research for the two case studies. To support the next generation of researchers and to strengthen local research environments the network also includes early career scholars who have all started to study the historical development of the web as it is found in web archives. Also, to facilitate close collaboration with relevant national web archives, the web archives from the researchers' host countries are also participating with their web collections and with their expertise in researcher engagement in relation to web archive studies. Finally, since it is a priority of WARCnet to be as inclusive as possible, a number of free seats are offered for participants who were not part of the initial network at the time of applying for funding (early career scholars are prioritised when allocating free seats).³

A network

Since WARCnet is a network it is relevant to have a closer look at what characterises the network: Are there any central nodes? Are there nodes with none or only few relations? Are there any clusters? Can one easily traverse the network? To answer these questions a network analysis of the WARCnet participants was performed, based on the relations between network members. When any two members know each other this is considered a relation, and 'knowing' is defined as 'having talked to'; if two participants have only heard of each other or been present at the same location (at a conference, for instance), but without talking to each other, this is not considered a relation. Obviously, the relations can only be defined in a tentative way since people may have talked to each other without my knowledge. Therefore the following must be taken with a pinch of salt.⁴ Nevertheless, the trends of the network graph are still revealing, as can be seen in the figures below.⁵

Before looking more closely at the visualisations let us see what network metrics can tell us about the overall characteristics of the WARCnet network. The network visualised below is composed of 41 nodes and 246 relations (edges), and it includes all the members of the initial network and the people who were invited to join the kickoff meeting as free seats (and who can later continue in the network). Given the relatively few nodes the network is not as well connected as one would think, since it has a network diameter of 4, meaning that it takes four steps to move through the graph between the most distant nodes. Also, the network has a density below medium (0.3) meaning that a third of the total number

3. A complete and updated list of network participants can be found at the project website warcnet.eu.

4. Another data source for analysing the network is references to publications between participants, but this would have limited the analysis by focusing on more formal relations, and thereby excluding informal communication and collaborations.

5. The networks have been created using the network analysis software Gephi (Force Atlas layout). Metrics were calculated in Gephi, graphs exported as a gexf file, imported in Polinode where the final visualisation was made.

of possible connections between all nodes is realised. However, the network is relatively effective to get around in, since it has an average path length of 1.78 which is not very close to the diameter of 4; the average path length measures the shortest possible route between all nodes in the network, and since it is not very close to the diameter this means that one does not have to go through too many nodes to get around in the network. The network could, of course, have been even more effective had the path length been closer to 0. Finally, regarding the centrality measures of the entire network's nodes the network has an average degree of 12, which means that on average each node has twelve connections to other nodes.

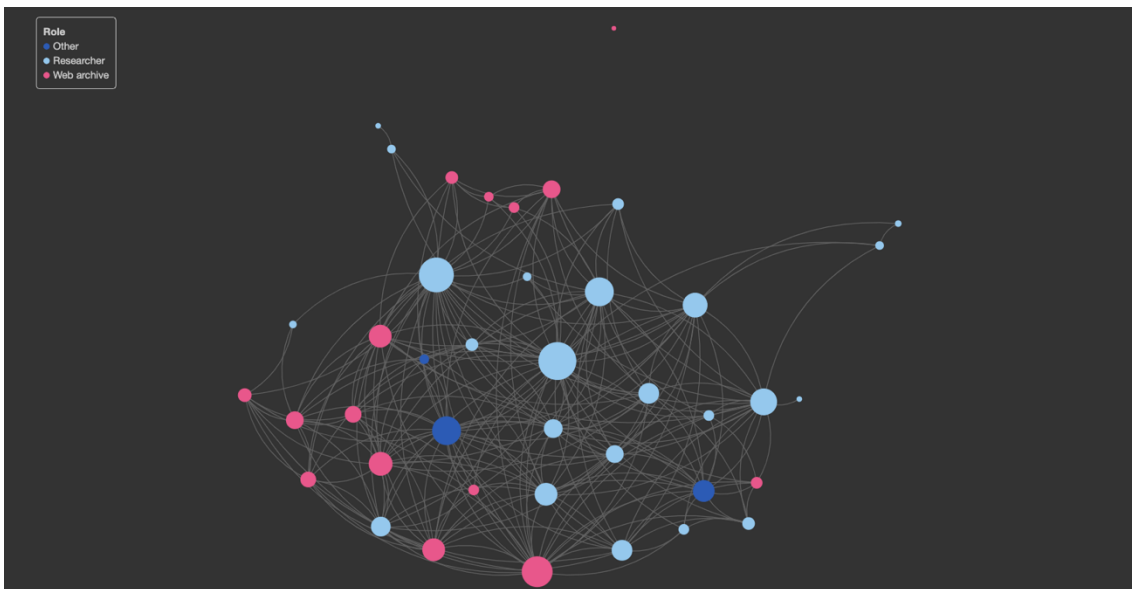


Figure 1: Degree of connectedness and role (a bigger visualisation can be found on page 14).

What do the visualisations tell us about the WARCnet network? First, let us look at the most connected nodes. In Figure 1 the size of a node mirrors how well connected it is, and, not surprisingly, the three Principal Investigators are central in the network. But this also goes for a number of other nodes — all in all approximately a quarter of all nodes. Then we have another quarter of the nodes that are medium well connected, and the rest, approximately half of all nodes, are connected to relatively few nodes. Second, it is relevant to look at the role that each node plays in the network, as either researcher, web archive, or 'other' (perhaps representing the IIPC, or an IT developer in a research infrastructure). In Figure 1 the role is identified by colour, and although this visualisation as an image does not show this very clearly, the interactive version in Gephi reveals that the majority of nodes are mostly connected to nodes of the same type — researchers tend to connect to researchers, web archives to web archives — and if cross-role connections exist it is usually to only 1 or 2 of the other type. This may not be all that surprising, but in each group there are a few exceptions, namely nodes that are also well connected to the other type of nodes — researchers connected to web archives, and vice versa — and in particular to more than a few nodes of the other type.



Figure 2: Degree of connectedness and status (a bigger visualisation can be found on page 15).

Finally, let us see what the visualisations can tell us about the status of network participants, in particular the people who have been invited in: how well connected are the invited free seats? The answer to this question can be found in Figure 2 where the size of nodes is identical to Figure 1, but colours now indicate if a node is an invited participant or has been a member since the network started. Again, the visualisation as an image does not show relations all that well, but an interactive version clearly highlights two tendencies. First, as expected, in general the free seats are only connected to a few other nodes, but differences also apply since a quarter are relatively well connected already. Second, and maybe more surprising, the majority of the invited free seats do not know each other, whereas, that well-connected quarter not only know other members already, they also tend to know each other.

This mapping of the starting point of WARCnet as a network makes it clear that there is room for improvement, and it also indicates some guidelines for this. Therefore, it is my goal that after a year and a half we will have created a network with a smaller diameter, a higher density, a smaller average path length, and a higher average degree: a network where the not-so-well-connected half has become more connected; where more participants from each type of role know more participants with another role, more researchers are connected to web archives and 'other', and vice versa; where more of the invited free seats (those already present, or new ones) will know more of the network members, including other free seats. If we succeed with this, all participants must try to connect as much as possible, and in particular to the most central nodes in the network, since they constitute the brokers who can establish a contact to other areas of the network. Hopefully, we will all be central as the network evolves.

Working Groups

The work with the research topics takes place in dedicated Working Groups. At the start of the project, the following four Working Groups have been established, whereas the fifth will be set up after a year.

Working Group 1: Comparing entire web domains

The networking activities of Working Group 1 will support research that aims to compare entire national and transnational European web domains, with a particular focus on a) the character of the web domains (e.g. number and size of websites, use of images, video and sound, hyperlink networks); b) the users (e.g. who the users are, and which websites they use); and c) the role of global social media websites such as Facebook and Twitter. Working Group 1 will be considering the following tasks:

- Plan and conduct coordinated joint research within the theme;
- Identify and assess the state of the art by reviewing existing literature in relation to the theme, with a view to mapping the gaps and prioritising a coordinated research agenda;
- Establish a standardised code book for what to investigate in relation to web domains, including reflections on how to define a national web domain, as well as standards for describing web domains, for mapping the available archived web content in different web archives, for establishing a corpus that can be said to represent a national web domain, and for the possible tools to be used.

Working Group 2: Analysing transnational events

The networking activities of Working Group 2 will focus on selected events from the following broad categories:

- 1) Unforeseen events: a) Terrorist attacks (train bombings in Spain (2004) and London (2005), the Charlie Hebdo, Copenhagen and Paris shootings (2015), Brussels and Nice attacks (2016)); b) the refugee crisis in Europe from 2015; c) the COVID-19 pandemic in 2020;
- 2) Predictable events: a) EU Parliamentary elections (2004, 2009, 2014, 2019); b) the London Olympics of 2012.

To create synergy, Working Group 2 will also investigate the relationship between the events and the different national web domains analysed in Working Group 1. Working Group 2 will be working with the following tasks:

- Plan and conduct coordinated joint research within the theme;
- Identify and assess the state of the art by reviewing existing literature in relation to the theme, with a view to mapping the gaps and prioritising a coordinated research agenda;
- Identify and map existing web material related to the selected events, and gather information to build diachronic transnational corpora.

Working Group 3: Digital research methods and tools

Material in web archives is fundamentally different from other digital source types, and therefore developing digital methods and tools to be used in web archive research is a key component of the network. At present the development of relevant methods and tools is

only an emerging field, and therefore a strong network is needed if the few existing initiatives are to grow. With a view to avoiding duplication of effort and to promoting existing initiatives the network offers a forum for the exchange of ideas. Thus, the possible tasks of Working Group 3 are:

- Analyse, test and customise existing methods and tools by assessing their possible usefulness in relation to the investigated themes;
- Experiment with new methods and tools by undertaking prototyping activities to underpin the potential development of new methods, tools and standards;
- Set up procedures to ensure all digital tools developed are made openly available (e.g. at GitHub or similar);
- Organise workshops, training schools, and/or hackathons/datasprints.

Working Group 4: Research Data Management across borders

As previously mentioned research infrastructures constitute the interface between web collections and researchers, and they therefore are an important element in the broader field of Research Data Management (RDM), that is, the question of how research data is handled. At a national level RDM within web archive studies is still in its infancy, let alone transnational RDM. Thus, the aim of Working Group 4 is to investigate the possibilities of sharing web archive-related data across borders, and the possible tasks of the group are to:

- Make an initial mapping of the legal and technical challenges for sharing web archive data between researchers from different countries (with a main focus only on web archives participating in the network);
- Develop a pilot project with sharing of derived data from web archives.

Working Group 5: Post WARCnet

The aim of Working Group 5 is to investigate the possibilities of making the network continue as the basis for national and/or transnational research applications. Therefore, the task of this Working Group is possibly to prepare joint research applications to national and international funding schemes.

ACTIVITIES, EXPECTED OUTCOMES AND COMMUNICATION

Since WARCnet is a network, the overall aim of all the activities is to support continued network building, which will be done through a number of activities, including meetings, conferences, network stays and publications.

Activities

The main activities of the network are to:

- Organise three network meetings, to take place in Aarhus (4–6 May 2020), Luxembourg (4–6 November 2020) and London (20–22 April 2021);
- Organise an international conference, to take place in Aarhus 3–5 November 2021, including a one-day meeting for network members only;

- Support network members' Short-term Network Stays (STNS);
- Engage in associated activities.

Network meetings constitute the core activity of the network. Here network members will get together and present the results of their work, coordinate cross-group activities, and discuss next steps. Also, each meeting will invite keynotes who are not network members, and workshops and hackathons/datasprints may also be organised. Between network meetings Working Group participants will meet (online or in person) to coordinate their work.

The closing conference will be open to anyone outside the network, and presentations will be considered for inclusion based on a call for papers. The day before the conference will be for network members only, and here the network activities will be rounded off and evaluated.

With a view to supporting the exchange of staff, WARCnet will accommodate a number of Short-Term Network Stays (between 1 and 5 days) where researchers are hosted by a participating research institution or a web archive (or both). This activity will foster closer collaboration between network members, and to ensure that this activity feeds back into the network, a member receiving an STNS will give a short presentation about the stay at the next coming project meeting.

Finally, WARCnet will support activities that are not organised by the network itself, but that are considered relevant for the topics being investigated. Such associated activities may be organised by research institutions, web archives, research projects, etc.

Expected outcomes

The expected outcomes of the network activities are, first, a network of researchers and web archives; and second, high-quality national and transnational research manifested in the following tangible outputs:

- 1–2 edited volumes with international publishers;
- 1–2 special sections/issues in international journals;
- 10-15 issues of the network publication series WARCnet Papers;
- 2–4 conference panels/round tables;
- Reading lists with relevant literature, free and publicly available;
- Experience with (trans)national Research Data Management.

In addition, network members may publish monographs, journal articles, book chapters, and conference papers within the network's research field, either individually or jointly.

WARCnet Papers is a series of papers related to the activities of the 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. Therefore, papers are primarily reviewed by the editorial board, but will be peer-reviewed if an author requires this; then a process of single-blind review will be organised.

The reading lists are considered a joint task, with network members invited to contribute with relevant literature within the project's fields of study. These reading lists are made publicly available on the project website warcnet.eu.

Communication

To support communication among network members, as well as communication to potential members and the wider public, WARCnet uses the following means of communication:

- The project website at warcnet.eu and warcnet.dk (redirects to cc.au.dk/warcnet where the website is located);
- A Facebook page (WARCnet), a Twitter account (WARC_net), a YouTube Channel (WARCnet Web Archive Studies) and a Slideshare account (WARCnetWebArchiveStu),
- A Newsletter (one can subscribe on the front page of warcnet.eu).

EPILOGUES?

In 2010, I ended the edited volume *Web History* with an epilogue entitled "The Future of Web History" (Brügger, 2010). Here it was noted that "there is still work to be done in establishing this emerging field [web history]" (p. 349), and I identified three possible paths that this work could follow: a) Broadening the conception of web archiving; b) Web history beyond national boundaries; c) Research infrastructure. The first point was about broadening web archiving beyond web crawling to include screen filming and the like, but this idea has never taken off. The second and third suggestions, however, have had a much more fruitful afterlife.

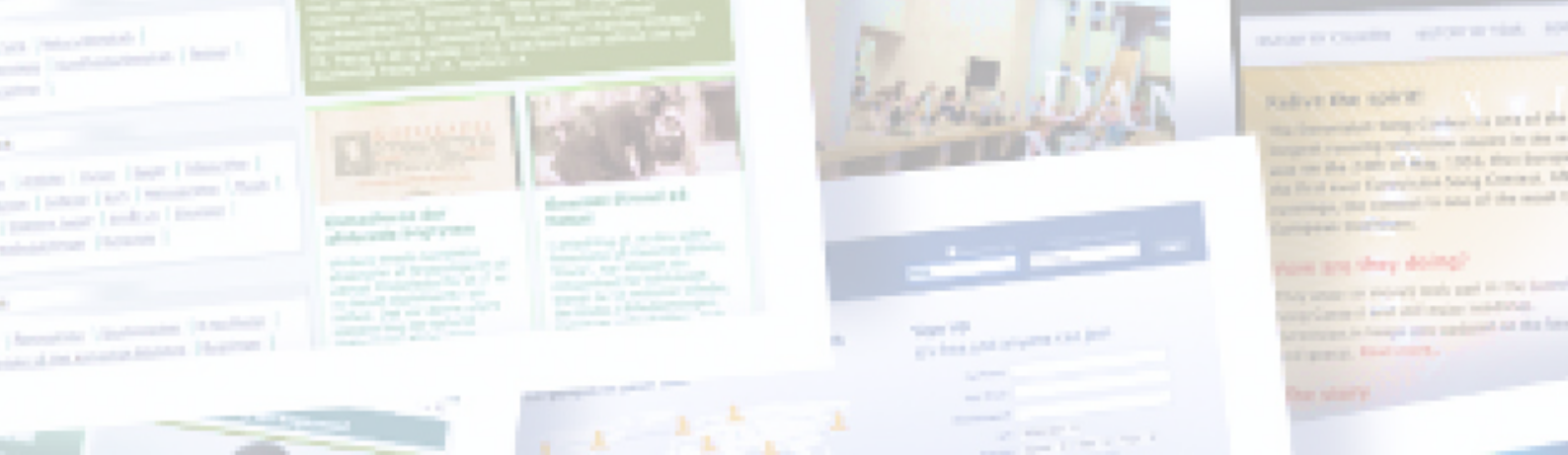
I made a plea that national web studies should be embedded in transnational research projects, in particular regarding events on the web. And I called for "research fora which can persist over time" in the form of "conferences, seminars, roundtables, email lists, international peer-reviewed journals", as well as encouraging researchers to "initiate collaboration with the national and transnational archiving institutions and organizations" (pp. 351–352). Conferences like the IIPC's open annual Web Archiving Conference (WAC) and the RESAW Conferences have been established, a number of roundtables have been organised, email lists have been created (and closed again), and the international peer-reviewed journal *Internet Histories: Digital Technology, Culture and Society* was launched in 2017. Hopefully, the WARCnet network can lay the foundations for fruitful transnational collaborations among web archives, among researchers, and between web archives and researchers in the future. When WARCnet ends let us decide if there still is a need for expressing more wishes in new epilogues.

REFERENCES

- ASAP — Archives sauvegarde attentats Paris, <https://asap.hypotheses.org>
- Brügger, N. (2010). Epilogue: The Future of Web History. In N. Brügger (Ed.). *Web History* (pp. 349–352). New York, NY: Peter Lang Publishing.
- Brügger, N. (2011). Web archiving — between past, present, and future. In M. Consalvo & C. Ess (Eds.). *The Handbook of Internet Studies* (pp. 24–42). Oxford: Wiley-Blackwell.
- Christensen-Dalsgaard, B., Fønss-Jørgensen, E., von Hielmcrone, Harald, Finnemann, N.O., Brügger, N., Henriksen, B., & Carlsen, S.V. (2003). *Experiences and Conclusions from a Pilot Study: Web Archiving of the District and County Elections 2001. Final Report for The Pilot Project 'netarkivet.dk'*. Copenhagen: netarkivet.dk. <http://netarkivet.dk/publikationer/webark-final-rapport-2003.pdf> (retrieved April 2020)
- Cowls, J. (2017). Cultures of the UK web. N. Brügger & R. Schroeder (Eds.). *The Web as history: Using web archives to understand the past and the present* (pp. 220–237). London: UCL Press.
- Foot, K.A., & Schneider, S.M. (2006). *Web Campaigning*. Cambridge, Mass.: The MIT Press.
- Kimpton, M., & Ubois, J. (2006). Year-by-year: From an archive of the Internet to an archive on the Internet. In J. Masanès (ed.), *Web Archiving* (pp. 201–12). Berlin: Springer.
- Koerbin, P. (2017). Revisiting the World Wide Web as artefact: Case studies in archiving small data for the National Library of Australia's PANDORA Archive. In N. Brügger (Ed.). *Web 25: Histories from the first 25 years of the World Wide Web* (pp. 191–206). New York: Peter Lang Publishing.
- Laursen, D., & Møldrup-Dalum, P. (2017). Looking back, looking forward: 10 years of development to collect, preserve, and access the Danish web. In N. Brügger (Ed.). *Web 25: Histories from the first 25 years of the World Wide Web* (pp. 207–227). New York: Peter Lang Publishing.
- Webster, P. (2017) Users, technologies, organisations: Towards a cultural history of world web archiving. In N. Brügger (Ed.). *Web 25: Histories from the first 25 years of the World Wide Web* (pp. 175–190). New York: Peter Lang Publishing.
- Webster, P. (2018). Existing web archives. In N. Brügger & I. Milligan (Eds.). *The SAGE Handbook of web history* (pp. 30–41). London: SAGE.
- Winters, J. (2017). Breaking in to the mainstream. *Internet Histories*, 1(1–2), 173–179.







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 and Valérie Schafer. In cases where a 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 2020-21, 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

