

Perspectives on web archive studies: Taking stock, new ideas, next steps

Niels Brügger, Valérie Schafer
and Jane Winters (Eds.)

WARCNET PAPERS

WARCnet
web archive studies

Perspectives on web archive studies: Taking stock, new ideas, next steps

Edited by Niels Brügger, Valérie Schafer, Jane Winters

With the participation of Alexandre Chautemps, Fabienne Greffet, Sharon Healy, Anne Helmond, Michael Kurzmeier, Janne Nielsen, Caroline Nyvang, Valérie Schafer, Katharina Schmid, Kees Teszelszky, Karin de Wild, Derren Wilson



WARCnet Papers
Aarhus, Denmark 2020

WARCnet Papers ISSN 2597-0615.

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

Published by the research network WARCnet, Aarhus, 2020.
Editors of WARCnet Papers: Niels Brügger, Jane Winters, Valérie Schafer, Kees
Teszelszky, Peter Webster, Michael Kurzmeier.

Cover design: Julie Brøndum
ISBN: 978-87-972198-4-3

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).



WARCnet Papers

Niels Brügger: *Welcome to WARCnet* (2020)

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

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

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

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

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

Perspectives on web archive studies: Taking stock, new ideas, next steps

Edited by Niels Brügger, Valérie Schafer, Jane Winters

With the participation of Alexandre Chautemps, Fabienne Greffet, Sharon Healy, Anne Helmond, Michael Kurzmeier, Janne Nielsen, Caroline Nyvang, Valérie Schafer, Katharina Schmid, Kees Teszelszky, Karin de Wild, Derren Wilson

Abstract: A round doc discussion about studying web archives to understand the history of (trans)national web domains and of transnational events on the web. What do we know, are there any new ideas, and what could be the next steps?

Keywords: web, web history, transnational web domains, transnational events

INTRODUCTION

Instead of physically bringing people together to participate in a roundtable discussion and discuss topics related to the activities of the WARCnet network, we decided to turn to a format that has previously been used by, among others, the journal *Internet Histories*, that is, a 'Round Doc' discussion. Participants in the WARCnet kickoff meeting 4-6 June 2020 were thus invited to participate in this roundtable conversation, taking 'place' not in a physical room but in a shared Google document. The questions below were asked by the editors, and participants were free to answer whichever questions they wished to address, or even to add questions themselves.

WEB ARCHIVES

Web archives constitute one of the main source types when studying the history of (trans)national web domains and of transnational events on the web. But European (and other) web archives are very diverse — what do you consider to be the major challenges when performing cross-national web archive studies (from a web archive as well from a researcher perspective)?

Kees Teszelszky: As Niels Brügger and others have already stated, web archives are messy. National web domains have no clear boundaries and often overlap state borders. I like to use the analogy between the borders of the Holy Roman Empire and the borders of the modern European states within the European Union. We cannot understand current digital culture without taking cross-national web domains into account.

Caroline Nyvang: To me, establishing benchmarks and baselines in order to be able to compare results from various web archives is an important step in facilitating comparative research.

Fabienne Greffet : The major challenge is actually to get comparable data when it comes to websites/social networks. And this is only possible if the sources are more or less collected in the same way. This is the main challenge that both national archives and WARCnet have to face. Otherwise, web sources will always be incomplete/partial for comparative research; and consequently, some countries will be over-represented in human and social science research, while others will be underrepresented.

Janne Nielsen: Some challenges will probably be recurring, no matter how we approach the study (e.g. issues relating to the characteristics of the archived web and how archives' scope, strategies, harvesting settings etc. influence what is archived), while others will depend on what we are trying to study. For instance, we might be able to study a transnational event by studying the event in different archives and then compiling and discussing the findings across archives, but if we aim to do comparative studies as such, the issue of how the data is affected by the choices made by the different archives becomes much more significant. So we need to figure out how we can ensure comparable data across web archives and on what levels. One way forward might be to try to map (in as much detail as possible) the differences 1) between the archives, and 2) within each archive over time, to create a background for assessing how these characteristics might affect the findings.

How and why should/could the European scale be a relevant one for web archives studies?

Fabienne Greffet : Europe is not only a continent, it is also a political construction. A number of legal frames/constraints, such as the GDPR apply specifically to the internet in the EU and the European continent. A number of European events, such as EU elections or sport competitions, also make a claim for some kind of “memory”. In the 1970s, the European

Commission created the Eurobarometers in order to both monitor and promote the construction of European values and a European identity. Nowadays, opinion polls are no longer the only way to monitor this. Digital traces, in their diversity, are also at stake in the monitoring and promotion of European values.

Valérie Schafer: As noted by Julia Pohle in an interview we conducted for a special issue of the academic journal *Internet histories* on the internet and the EU market: “The internet is a global network of networks, which affects all parts of the world; however, much like globalisation, the digital transformation does not have the same effect in each country and for each society or community. This is particularly true at the global level, where digital connectedness has led to new modes of inclusion and exclusion. But also within the European Union, the digital transformation takes different forms in its highly developed Member States. Specific uses, developments and policies are strongly influenced by traditions, norms, values and institutions, which vary from country to country” (Schafer et al., 2020: 217). What Julia Pohle underlined for the internet is also true for web cultures. Working on this issue of inclusion and exclusion, whether through the study of entire domain names or European events, as envisioned by the WARCnet network, may give us a better understanding of past and current European digitisation, while also raising the question of the emergence of a “digital European public sphere” and its realities and limits. Studying web archives on a European scale also provides us with an excellent laboratory to help overcome issues of multilingualism and the difficulties they represent for researchers, and in this respect it invites us to explore the potentialities of distant reading of metadata, visual studies and other sometimes underestimated aspects.

What lessons, if any, can be learned from national web domain studies to inform transnational research?

Kees Teszelszky: One lesson that we have learned is that a definition of the borders of a national domain is like a Procrustean bed. It will never fit properly and you will lose valuable data by using too narrow a definition.

Janne Nielsen: The first lesson that comes to mind is the difficulty of creating corpora — a challenge, which will only increase as we try to make corpora that allow for comparisons across archives. Related to this are several technical aspects, which should be addressed, and where discussions of standards, best practice etc. might benefit from the lessons learned in existing studies. Other factors that can inform transnational research include the value of documentation and the necessity for close collaboration between researchers, curators and IT-developers to uncover the workings of the archives and how this influences what we can study, and how we can study it. It is an obvious advantage of WARCnet that our network includes people who occupy these different positions.

COMPARING ENTIRE WEB DOMAINS

There is no doubt that the archived web is an important source type when studying a nation's — or nations' — web domain(s). But should it be the only source type? And if not: which other source types could/should be included to better embed the web in its historical context(s)?

Alexandre Chautemps: The domain registries (Afnic for France, DK Hostmaster for Denmark, RESTENA DNS LU for Luxembourg, and so on; cf. the list of the members of CENR: <https://centr.org/about/members.html>) could be a very interesting source allowing for better knowledge of the live web and its evolution over time. They publish statistical data such as number of domains, regional distribution, proportion of domains with a foreign owner, etc. They also publish, each year, lists of registered domains, which helps to assess the representativeness of the web archive. It might also be interesting to study the evolution of the population of internet users for the country/region considered. In France, ARCEP (the French regulatory authority for telecommunications) has published every year since 2003 a report which includes numerous data in that respect: the proportion of citizens of the country having access to the internet, the level of computer and smartphone ownership per household etc. We might also be interested in studying practical guides for internet users (numerous documents in printed form since 1992, and in BnF's web archive (see chapter "Le web mode d'emploi" in *Le web des années 90: [guided tour]*, Paris, BnF, 2016, <https://www.bnf.fr/fr/centre-d-aide/parcours-guides-dans-les-archives-de-linternet#step06>) and website directories in printed form (since 1995 in BnF's catalog). We might also think of printed sources which allow the study of economic aspects of the internet, such as market studies about internet advertising (present in the BnF's collection since 1998), and market studies about e-marketing and cyber-marketing (present in the BnF's collection since 1996).

Derren Wilson: Printed sources for web design and development are sometimes translated from English into other languages (e.g. <https://abookapart.com/products/responsive-web-design>) but it is difficult to think of an influential printed source for web development that has been translated from a European language into English. Some software used internationally to create and test websites is designed and built in Europe (<https://www.vanamco.com/ghostlab/>, <https://www.mamp.info/en/windows/>, <https://www.drupal.org/>). It would be interesting to see if familiarity with English was considered important for the recruitment of web employees to keep up-to-date with the evolution of web practices and if the types of job roles were the same (does the role of webmaster last longer in some countries? Is there a Europe-wide split between 'web designer' and 'front-end developer' as there is in the United Kingdom?).

Janne Nielsen: I agree that it would be valuable to consult sources with statistical, financial and technical data over time, as well as printed sources about practices, both for users and web designers.

How can we engage stakeholders and the public at large with such studies? What might be their economic as well as technical and social impact?

Valérie Schafer: The study of national web domains allows us to highlight national trends but also to compare their development between one country to the next, and in this respect it provides valuable information on both the archived web and developments in the living web. This general background and overview may be very helpful for students, who are starting to analyse web archives, but it is also extremely instructive for web archive professionals. However, in order to make this research attractive to a wider audience I would recommend targeting two groups in particular: 1) The community that deals with domain names in Europe. I am a member of the Scientific Council of Afnic, that Alexandre previously mentioned, which manages various domains in France including .fr. I have no doubt that Afnic, like other domain name managers in Europe, would be interested in finding out about this research and also in providing us with feedback, for example on the communication, economic and legal policies surrounding domain names, which may in turn also enable us to contextualise our research more effectively; 2) In a more entertaining way, this research may also be of interest to the general public. The latter may be interested, for example, in discovering which were the five most visited websites in their country at the end of the 1990s, the moment when Facebook took off, or the personalities or GIFs which have been most popular on the web since the mid-1990s. Including national web discovery quizzes and a grand European quiz on the WARCnet project website would seem to me to be an excellent idea. Ready, set, play!

Sharon Healy: I agree Valérie. We should consider novel ways to engage with the general public, which in turn will increase awareness of not only the current research project, but of the availability of web archives across Europe. By way of an example, I came across a great visualisation by Data is Beautiful, “Most Popular Websites 1996–2019” (<https://www.youtube.com/watch?v=2Uj1A9AguFs>). I am not too sure where the data came from but the visual provides for a good discussion in the YouTube comments. I then used the visualisation as part of a class discussion, on the rise and fall of popular websites. It was a nice way for students to think about the historical side of web development, as well as the politics.

Derren Wilson: I would like to see more use of web archives to encourage debate about the histories and definitions of web design and user experience. The web has been around for long enough to have helped shape how we interact with technology now, for better or for worse.

Anne Helmond: Following on from Derren, such studies into changing web design practices could provide more detailed insights into the evolution of online trackers as well as ‘dark pattern’ design practices. These may be used to inform the wider public about which actors on the web have come to collect data about our web activities (and how they do it) and which design features are nudging users in particular directions.

Michael Kurzmeier: As Anne and Derren stated, the study of web histories can be a great starting point for a public to develop concepts to describe and critique the web. Historical understanding of events and conditions can help to look past an assumed uniformity and permanence of technology.

How can researchers be enabled to work with hybrid collections, combining paper, digitised and born-digital materials?

Janne Nielsen: That is a really good (and difficult) question and maybe this is an area where it would be valuable for Working Group 1 (Comparing entire web domains) and Working Group 4 (Research Data Management across borders) to join forces to discuss ways of creating, preserving and possibly sharing collections with different types of materials from different archives.

Karin de Wild: Data from various sources (from metadata, images and transcribed text to entire collection records, born-digital material, etc.) can be made available in machine-readable format under an open licence (also known as cultural open data). To support connections between the data, one can make use of linked data (a set of best practices for publishing structured data on the web). When it is not possible to work with WARC files within linked data (when it is not possible to make them openly available online), there might be opportunities to work with the metadata or derivative data. For example, Rhizome's ArtBase makes use of Wikidata, the data management platform for Wikipedia, which is encoded in RDF (linked data). Here, they store the metadata of their collection of digital art. Together with Lozana Rossenova and Dragan Espenschied, I recently worked on applying the W3C PROV standard so that the provenance of these online artwork can be described, even when the artwork is part of various museums and archives (Rossenova et al., 2019).

ANALYSING TRANSNATIONAL EVENTS

What could web archives change about our understanding of transnational events and our narratives of European events?

Valérie Schafer: We may expect to read new narratives, to have access to a European history that is more bottom-up, less institutional but also closer to citizens' debates that do not necessarily take place in traditional arenas (press, political circles). I will cite just one example: the research project that Frédéric Clavert has just launched, entitled *Unsupervised Europe: European citizens and online debates about European integration*. Frédéric will try "to figure out how European citizens relate to EU democratic institutions at a time when political debate has frequently hinged on the democratic deficit of such bodies", primarily by focusing on French-speaking newsgroups such as fr.soc.economie, the main Usenet newsgroup for discussion of economic issues, including those related to mainstream economic policies (<https://unsure.hypotheses.org/author/clavert>). The project can

be expected to yield new insights into critical discourses on the EU, and web archives are an obvious extension of this research, from Euroscepticism online to the collections documenting Brexit, for example in the UK Web Archive special collections.

How can we identify and address gaps and silences in web archives relating to transnational events?

Kees Teszelszky: The Whole Earth Web Archive (WEWA) of the Internet Archive is a good starting point (see <https://blog.archive.org/2019/10/30/the-whole-earth-web-archive/>). It is a sample set of 50 small nations whose archived web content was extracted from the Internet Archive's collections, in order to build special search and access features on top of this subcollection, and create a dedicated discovery portal for searching and browsing archived websites of underrepresented nations around the world.

How can we decide which events should be the target of focused collecting activities on the web, when resources are limited?

Anne Helmond: One entry point could be Wikipedia which has a section on "Current Events". It is an excellent resource for monitoring ongoing events worldwide and for collecting event-related keywords for querying for archivable sources in search engines. In particular, one could monitor the subsection entitled "Ongoing events", collect the linked Wikipedia pages of these events, and scrape and archive the reference lists of these pages. In an ideal scenario these steps could be automated so as to have an ongoing "live" collection of archived current events.

Karin de Wild: The web is used and studied across disciplines, in various cultural and social contexts. It seems important to ask experts from various disciplinary fields and socio-cultural backgrounds and we might like to consider democratizing the decision-making process. It is also important to keep in mind that: "The collecting problem that needs to be resolved is insoluble not simply because of the problems of deciding what to collect, but by the unrealistic belief that when something is collected it will be kept in perpetuity" (Knell, 1999: 16).

DIGITAL RESEARCH METHODS AND TOOLS

Digital methods and tools can be considered a key component in web archive studies. Should these tools be plug-and-play style or tools based on some level of scripting? And should the tools be placed on the researcher side or on the web archive's side?

Anne Helmond: Ideally there should be tools available for various levels of (technical) skills, and their setup and workings, and the workflows they enable, should be clearly documented. In order to gain further traction as a method there should be various easy-to-use tools available for beginning and aspiring web archiving researchers. These tools should

have a web interface that is available and accessible from anywhere, run on the researchers' side in their browser, and their outputs should at least include a standard CSV format. That is, we should not only consider how the tools look but also what kind of input or output they require.

Michael Kurzmeier: The defining aspect for me is what can be achieved with these tools. If something can be done in a quick and easy way, I see no need to complicate tools. If it is a question of functionality that a GUI tool cannot provide, ideally there should be a learning curve which takes the user to the point where they can understand why functionality was chosen over accessibility.

Katharina Schmid: Regarding the question of where tools should be placed, I think web archives are well equipped to use tools that are able to handle large amounts of data and special file formats like WARC files. Specific tools like the Archives Unleashed Toolkit (<https://archivesunleashed.org/aut/>) can be used to generate derivative datasets in formats researchers are more familiar with. Generally, the derivatives are smaller than the original WARC files and easier to process outside of specialized infrastructures. Derivative data can be shared openly via repositories or custom APIs, where this is legally possible, or on the basis of extraction agreements, following the example of Netarkivet (<http://netarkivet.dk/in-english/>).

Should tools be general and answer basic common needs and/or custom-made tools adapting to precise research needs?

Derren Wilson: I think there is space for both. It is important when working with a new medium or technology that the user can quickly feel they are accomplishing something — the Danish web archive's visualisation of tags for example (<http://labs.statsbiblioteket.dk/tags/>). As users grow in confidence and are willing to spend more time setting up and customising tools — because they can see the value of doing this for their research — more sophisticated tools will help describe the complexities and limitations of web archiving as users configure tools for their specific requirements.

Karin de Wild: I also think we need both general and more custom-made tools. To open the web archives to a broader (research) community, should we also consider programs that are widely used, like Gephi or analytics platforms like Tableau or Alteryx? This is actually related to what Anne mentions above, that it is important to also consider the outputs.

Is the openness of tools a core value for you or something secondary?

Valérie Schafer: Ideally I would of course like not only tools but also web archives themselves to be open. I dream of the possibility of sharing data, publishing screenshots of web archives, ensuring that research can be easily replicated and moving across national borders without hindrance — all of which would also facilitate collective and collaborative work.

There are obviously many obstacles, particularly with regard to copyright. But to come back to the tools, of course, their openness guarantees the accountability, transparency and readability of our research. While the notion of FAIR data is increasingly valued (we mentioned it during Ian Milligan's keynote and it was at the core of Sally Chambers' presentation too), the question of FAIR tools and analysis must also be raised. Obviously this should not be at the expense of the user-friendly nature of the tools and their accessibility. Ian Milligan expressed this very well in his keynote "You shouldn't need to be a web historian to use web archives: lowering barriers to access through community and infrastructure" that he presented at this WARCnet kickoff meeting, and the Archives Unleashed toolkit and the datathons he organises embody this principle.

Derren Wilson: I think it is very important that tools are open — a good tool should provide readable code methods, useful documentation and be customisable. It might even be that a currently available tool serves as the basis for a different tool written in a different language because the code gives examples of how to use particular APIs or other practices that aren't documented anywhere else.

Michael Kurzmeier: I agree with Derren that tools should be open, and the same is true for file formats. Openness in both helps accessibility and customization. Open file formats, in particular, greatly help to keep data usable and prevent data from being locked into a proprietary system.

What kinds of training would be of most benefit for researchers and web archivists, and how should this be delivered?

Anne Helmond: There could be training sessions at web archiving institutions for researchers wishing to work with their data. These could be half-day sessions explaining the basic idea of how their web archive works and is constructed, how one can search or request particular data, and how these datasets may be saved on-site or on-disk for further analysis. To better match the questions and needs of researchers with the options available at web archiving institutions, researchers could submit their research question in advance. The final hour of the session could be dedicated to discuss the feasibility and operationalization of the submitted research questions. On the other hand there could be two-day training sessions by researchers who have developed tools at their universities. The first half-day session could be dedicated to a tool demonstration where researchers follow the demonstration step-by-step. Then the second half could be dedicated to streamlining the research questions researchers have submitted in advance and to assisting with technical questions. The first half of the second day would have researchers working with the tool independently, with the aim of collecting their data. Things often work quite well if you follow clear instructions during a workshop, but once you aim to use the tool on your own many new issues and questions arise. This is why a few hours where researchers work independently with expert users available to answer questions would be ideal. The final hour and a half of the day could be dedicated to people briefly presenting their preliminary results and to

addressing what went well and what did not go well. This could then also serve as a form of user feedback on the tool and its instructions.

Michael Kurzmeier: Speaking about the situation in Irish HLE — I would like to host sessions bringing together researchers from related disciplines who have an interest in web archives but no real starting point. The training would then introduce what web archives are available to research, how to find and engage with material, how to cite, etc. A lot of groundwork needs to be done here, but I think that by addressing students and early career researchers across disciplines the topic of web archives could be brought into a wider public.

RESEARCH DATA MANAGEMENT ACROSS BORDERS

When it comes to management of web research data, in many countries it still remains to be negotiated where to strike the balance between the responsibilities of the two main actors, the web archives and the research institutions. What do you consider the main challenge when the issues related to research data management are moved to a transnational scale?

Alexandre Chautemps: Dealing with the diversity of the institutions, of the legal frameworks, and with the heterogeneity of the various collections.

Sharon Healy: I propose two types of challenges here related to screenshots. 1) Screenshots as datasets: various researchers have used screenshots to create a dataset for the study of web histories (Brügger, 2018: 80-81). What happens to the dataset, is it accessible and reusable, can it be shared across borders? 2) Screenshots in publications: on a national level, there is a lack of clarity on the legalities of reproducing a screenshot of a webpage (live or archived) in a publication. While the doctrine of Fair Use may be applied in some instances to reproduce a screenshot of a webpage, there are several instances where this is problematic. For instance, images on a webpage may be the copyright of other parties, and while a screenshot may just be concerned with the design of a webpage, it could amount to a copyright infringement. In Working Group 4 'Research Data Management across borders', should we also be looking at the use of screenshots as part of the "mapping of the legal and technical challenges for sharing web archive data between researchers from different countries"?

Is sharing metadata the new frontier? Or what do you see as the new frontier to conquer?

Alexandre Chautemps: Sharing metadata is indeed a good way to (partially) make our collections accessible, but there are other possibilities such as sharing derivative data. Not all the derivative data can be legally shareable (full text and even short extracts from the text of the archived websites cannot in France, since the right of quotation is not relevant for data resulting from an automated process) but some of them can, for example host and

domain names, lists of URLs, links between web pages in the corpus, keywords, MIME types, crawl logs, and so on.

How can we ensure common values, ethical frameworks and web research data management at the European level?

Alexandre Chautemps: Something like a code of practice... There is such a code established by the WHO (<https://www.who.int/about/ethics/code-of-conduct-for-responsible-research>). I do not know about such codes of practice in the digital humanities field.

How can we balance the rights of the individual to privacy with the rights of researchers to access data?

Alexandre Chautemps: The individual right to privacy is not unlimited, particularly in the framework of academic use of data, since the European directive GDPR includes a provision about processing data for the purposes of academic, artistic or literary expression. This provision is diversely implemented in the various national laws and needs to be explored to identify further possibilities for data reuse. Additionally, possibilities of pseudonymisation and anonymization must also be investigated. However, such treatments are complex, heavy, and can cause loss of intelligibility of the data.

How can we make web archives and derived datasets Findable, Accessible, Interoperable and Reusable (FAIR), within current constraints?

Alexandre Chautemps:

- Findable: Collections of web archiving institutions are findable for researchers who have already at least a vague knowledge of their web archiving activities. But there is not really a global tool such as a union catalogue of all the web archiving institutions' collections yet. Additionally, I do not think that a standard set of metadata such as OCLC's *Descriptive metadata for web archiving* would be used by all our institutions. So our collections are findable for those who already know them or who are willing to search a lot (on our websites, on the IIPC's website, in the web archiving literature, by email exchanges with the web archiving curators...).
- Accessible: our collections are accessible onsite in each web archiving institution, and, for some European web archiving institutions, accessible remotely under certain conditions. Metadata and derivative data can be given to the researchers with an agreement concluded between the research organisation and the web archiving institution.
- Interoperable: Interoperability is facilitated by the large (but not unanimous) use of formats like W/ARC for archive storage, so scripts and programs written to process archived data can theoretically be used for processing data coming from different

institutions, with some minor adjustments. There is also a certain homogeneity in the crawling metadata, since numerous web archiving institutions use Heritrix.

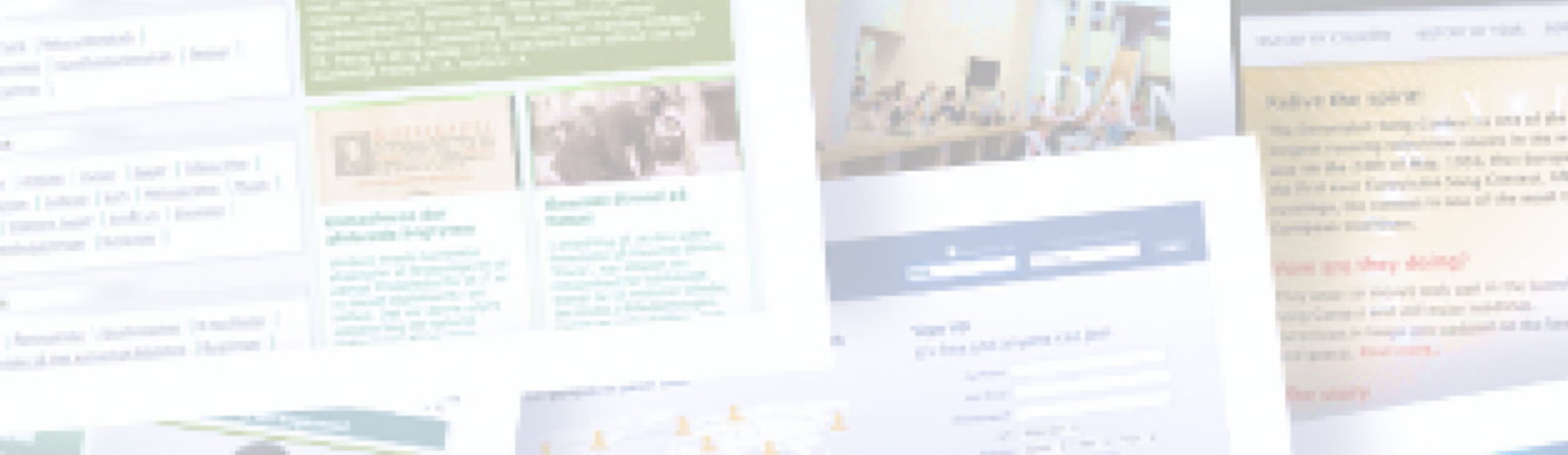
- Reusable: for reuse, issues are firstly legal. There is no problem for the dissemination of research results, but the raw data are protected for a long time by copyright and potentially contain personal data, so cannot be disseminated without permission. Another issue is the preservation of the research works themselves. I am talking about the results but also about tools (programs and scripts written to process the data), diaries of the research works, intermediate states of the processed data and the research papers, and so on. Institutions such as national libraries have no infrastructure dedicated to this type of preservation, and I'm not sure that all the research institutions would treat research data preservation issue as a priority.

If you could summarize the WARCnet network in one sentence, quotation, what would it be?

Valérie Schafer: I cannot resist quoting the founder of the World Wide Web, Tim Berners-Lee, but I am finding it difficult to choose between two of his quotes: "The Web does not just connect machines, it connects people", which is a very simple sentence but also a very true one, and "Innovation is serendipity, so you don't know what people will make", as it opens new horizons and invites us to embrace creativity, stay humble but also think big!

REFERENCES

- Brügger, N. (2018). *The Archived Web: Doing History in the Digital Age*. The MIT Press.
- Knell, S.J. (1999). *Museums and the Future of Collecting*. Aldershot: Ashgate.
- Rossenova, L., Wild, K. de, Espenschied, D. (2019). *Provenance for Internet Art: Applying the W3C PROV model*. Amsterdam: iPres.
- Schafer, V., Fickers, A., Howarth, D., Musiani, F., Pohle, J., & Winseck, D. (2020). "The internet and the European market" from a multidisciplinary perspective: a "round-doc" discussion, *Internet Histories*, 4(2), 217-227, DOI: 10.1080/24701475.2020.1739360.



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



warcnet.eu

warcnet@cc.au.dk

twitter: @WARC_net

facebook: WARCnet

youtube: WARCnet Web Archive Studies

slideshare: WARCnetWebArchiveStu