

Reporting from two Short-term Network Stays at The Institut National de l'Audiovisuel and the Bibliothèque nationale de France, and The Royal Danish Library

Beatrice Cannelli

WARCNET PAPERS



Reporting from two Short-term Network Stays at The Institut National de l'Audiovisuel and the Bibliothèque nationale de France, and The Royal Danish Library

Beatrice Cannelli

beatrice.cannelli@postgrad.sas.ac.uk



WARCnet Papers
Aarhus, Denmark 2023

WARCnet Papers ISSN 2597-0615.

Beatrice Cannelli: *Reporting from two Short-term Network Stays at The Institut National de l'Audiovisuel and the Bibliothèque nationale de France, and The Royal Danish Library*
© The author, 2023

Published by the research network
WARCnet, Aarhus, 2023.

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

Cover design: Julie Brøndum
ISBN: 978-87-94108-21-8

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* (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)

Susan Aasman, Nicola Bingham, Niels Brügger, Karin de Wild, Sophie Gebeil and Valérie Schafer: *Chicken and Egg: Reporting from a Datathon Exploring Datasets of the COVID-19 Special Collections* (Dec 2021)

Olga Holownia, Friedel Geeraert, Abbie Grotke, Jennifer Harbster and Gulnar Nagashybayeva: *Exploring special web archives collections related to COVID-19: The case of the Library of Congress* (Feb 2022)

Niels Brügger: *The WARCnet network: The second year* (Dec 2022)

Michael Kurzmeier: *Using a national web archive for the study of web defacements? A case-study approach* (Aug 2023)

Helle Strandgaard Jensen: *Any Teletubbies Caught in the Web?* (Aug 2023)

Niels Brügger: *The WARCnet network: The third year* (Aug 2023)

Kees Teszelszky: *Understanding the history of national domain crawls: mapping and archiving the national web domain in Denmark and France* (Aug 2023)

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)

Karin de Wild, Ismini Kyritsis, Kees Teszelszky and Peter de Bode: *Exploring special web archive collections related to COVID-19: The Dutch Web archive (KB)* (Nov 2021)

Michael Kurzmeier, Joanna Finegan and Maria Ryan: *Exploring special web archives collections related to COVID-19: The National Library of Ireland* (Feb 2022)

Emily Maemura: *Towards an Infrastructural Description of Archived Web Data* (May 2022)

Olga Holownia, Friedel Geeraert and Paul Koerbin: *Exploring special web archives collections related to COVID-19: The case of the National Library of Australia* (Dec 2022)

Helena Byrne, Beatrice Cannelli, Carmen Noguera, Michael Kurzmeier, Karin de Wild: *Looking ahead: after web (archives)?* (Aug 2023)

Friedel Geeraert, Marie Haškovcová, Luboš Svoboda and Markéta Hrdličková: *Exploring special web archives collections related to COVID-19: The case of the National Library of the Czech Republic* (Aug 2023)

Niels Brügger: *Why onsite meetings are important: Reporting from five Short-Term Network Stays* (Aug 2023)

Beatrice Cannelli: *Reporting from two Short-term Network Stays at The Institut National de l'Audiovisuel and the Bibliothèque nationale de France, and The Royal Danish Library* (Aug 2023)

WARCnet Special Reports

Sharon Healy, Helena Byrne, Katharina Schmid, Nicola Bingham, Olga Holownia, Michael Kurzmeier, Robert Jansma: *Skills, Tools, and Knowledge Ecologies in Web Archive Research* (Aug 2022)

Sharon Healy, Helena Byrne, Katharina Schmid, Juan-José Boté-Vericad, Lanna Floody: *Towards a Glossary for Web Archive Research: Version 1.0* (Aug 2023)

Sharon Healy, Helena Byrne: *Scholarly Use of Web Archives Across Ireland: The Past, Present & Future(s)* (Aug 2023)

Reporting from two Short-term Network Stays at The Institut National de l'Audiovisuel and the Bibliothèque nationale de France, and The Royal Danish Library

Beatrice Cannelli

Abstract: This WARCnet paper reports from two Short-Term Network Stays (STNS) at web and social media archiving institutions in France and Denmark. The STNS were made to carry out interviews and fieldwork for PhD research focusing on obstacles and practices related to the development of social media archives.

Keywords: social media archives, archiving practices, web research

SOCIAL MEDIA ARCHIVING: CHALLENGES AND PRACTICES

This WARCnet paper reports from two Short-Term Network Stays (STNS) aimed at collecting information about the challenges and context in which social media are currently being archived for my ongoing PhD research titled *Archiving Social Media: A Comparative Study of the Practices, Obstacles, and Opportunities Related to the Development of Social Media Archives*. As part of my PhD study, I am conducting a comparative study of international cultural heritage institutions that are either already involved with or are planning to undertake the long-term preservation of social media content.

As a digital assemblage of data, objects and human interactions, social media holds a unique cultural value, offering privileged insights into history, social trends, politics, and economic matters (Brügger, 2017; Burkey, 2020; Fondren & Menard McCune, 2018). In the last decade, archivists have come a long way in the practice of archiving social media, but they are still facing significant challenges (Espley et al., 2014; Marshall & Shipman, 2012; Thomson, 2016). My research aims to contribute to the debate about the manifold issues faced by cultural heritage institutions concerned with the archiving of social media and the future of social media preservation by analysing and comparing the direct experience of information professionals and other stakeholders, aiming ultimately to define best practices and guidelines for upcoming projects.

In relation to my study, the Bibliothèque nationale de France (BnF), the Institut National de l'Audiovisuel (INA) and The Royal Danish Library (KB) were among the institutions that I planned to interview given the experience they have acquired in almost two decades of dealing with web, at first, and then social media archiving.

Thanks to WARCnet I had the opportunity to visit in-person, first, the two French institutions in Paris between 19 and 22 April 2022 during which I met members of the respective web archiving teams at both the Bibliothèque François Mitterand and at the INA's offices in Bry-sur-Marne, located just outside Paris; and then the Royal Danish Library's web archiving team located in Copenhagen and Aarhus between 25 and 30 September 2023.

When it comes to web and social media archiving in France the effort of archiving content under the digital legal deposit legislation has been equally divided between INA (that archives web content related to radio and TV) and the BnF (that preserves the rest of the web). As for Denmark, the *Netarkivet* at the Royal Danish Library has been archiving the web since 2005, collecting material published on the Internet in Danish, by Danes and about Danes, according to the Danish Legal Deposit Act. The Danish Web Archive also includes content from social media platforms that is publicly available on the internet.

THE BIBLIOTHÈQUE NATIONALE DE FRANCE (BNF)

At the BnF, the Digital Legal Deposit team (*Dépôt Légal numérique*) is in charge of developing and maintaining *l'Archives de l'Internet*, that is the French web and social media archive. My visit at the BNF began with a quick tour of the Bibliothèque François Mitterand — an impressive building composed by four towers, each one dedicated to specific subjects and where related books are preserved, and the reading rooms that are located at the base of the towers. The first part of my visit at the BNF mostly focused on a general introduction to the French legal framework in which the *Archive de l'Internet* project is embedded and how they came to develop the web archive. This was followed by an overview of the appraisal, selection, collection processes and the tools used for the collection of web content, with a special focus on social media as this is the main topic of my research.

The French web archive at BNF currently preserves nearly one and a half petabyte of data, and, in terms of social media archiving, BnF is currently archiving content from: Twitter (multiple harvests a day), Instagram (2-3 times a year), YouTube (2-3 times a year). At the time of the visit, the BNF was planning to start archiving content from TikTok about the 2022 general elections. As for Facebook, the harvest is currently suspended due to some issues with Facebook's API limits. Alongside annual broad crawls, the BNF has a few *Emergencies collections* related for example to the 2015 terrorist attacks, or those preserving content from blogs that have been shut down (e.g., *Le Monde* and *Libération*). They also contribute to the International Internet Preservation Coalition (IIPC) collaborative collections.

The second day of the STNS focused mostly on information related to access to the French web archive, research opportunities and the support they provide to researchers. The head of the web archiving team showed me the reading room where the on-site

terminals are located and from where users can access and consult the web content preserved by the BnF. The presentations given by members of the team about access and research support in regards to the web archive gave me the opportunity to learn more about the wide variety of content they are collecting, their workflows, as well as the rich network of BnF sites and other libraries spread across the French territory from which it is possible to access data that are part of the French web archive, facilitating researchers' access to the French Web archive. Moreover, through the interviews and informal talks with the BnF electronic legal deposit team members I was able to learn more about their experience and their efforts in trying to solve the issues they face when archiving social media, especially those posed by the constant changes implemented by social media on their platforms, making archiving this content a continuous attempt to quickly find solutions in order to capture content before it changes or disappear.

THE INSTITUT NATIONAL DE L'AUDIOVISUEL (INA)

In line with its activity of archiving radio and television programs for preservation and research purposes, INA is responsible for the legal deposit of the French web in connection with audiovisual communication. INA currently preserves around 16,069 websites, and more than 15 thousand Twitter accounts and 2,500 hashtags. INA also archives content from YouTube with videos published by more than 9 thousand French accounts and authors.

I spent the first day of my visit at INA at the Bibliothèque François Mitterrand where I met some of the members of the web archiving team. I was shown around the reading room where the computers to access and consult INA's web archive content are located. During the interview we discussed some of the issues encountered by INA when archiving content from social networks, such as data limits imposed by APIs, difficulties related to capturing an ever-changing platforms and data (e.g. likes, number of visualisations per video) and access. I was then given the opportunity to browse their social media collection from one of the computers available in the reading room. The collection is searchable through *WebMédia* which is their self-developed browser; on *WebMédia* is it also possible to generate some data visualisations about, for example, the top 15 most used emoji icons, mentions or links related to a hashtag or account, and other stats such as distribution of tweets and retweets, verified and non-verified accounts. Probably one of my favourite things available on *WebMédia* is the *SocialTV*: a section of the interface where users can find a selection of videos that, once played, shows a rolling feed of Tweets posted at the same time as the program was airing.

The second day of my visit at INA was at their offices and training centre located in Bry-sur-Marne, where I met INA's research and development team. Here, I had the chance to ask more questions about the technical issues related to the harvesting of Twitter and how such data is made searchable and accessible to users. I was also shown a preview of the new interface and filters that were going to be implemented on *WebMédia*, improving the way users can browse and filter the archived social media.

THE ROYAL DANISH LIBRARY

The *Netarkivet* at the Royal Danish Library (KB) is responsible for the collection and preservation of the Danish part of the web, including social media, in the fulfilment of their mandate. The KB archives social media sites such as Twitter, Facebook and YouTube on a selective basis with a frequency that varies from weekly to annual, aiming at providing a comprehensive and representative snapshot of what is published on the Danish websphere.

The first part of the STNS took place in Copenhagen where I spent two days shadowing the daily activities carried out by the *Netarkivet* web curator. Prior to the meeting, I was granted access to the *Netarkivet* so that I could explore the collections for the duration of my STNS. During the visit, I was given a tour of the Royal Danish Library main building — called *the black diamond* for its architectural profile — and then the curator illustrated all the main characteristics of the Danish Web archive interface and its collections, with a special focus on social media and the issues they have been facing. Moreover, as part of the STNS, I was given the opportunity to participate to one of their team meetings and a provisional videocall with a researcher who was seeking assistance from KB with the harvesting of content from Twitter for a study they were planning to conduct. While the interview was fundamental to gather information about the challenges related to the preservation of social media, the informal talks over lunch, observation carried out in the course of the visit and the discussions I had with the curator over many practical aspects concerning web and social media archiving was a source of invaluable information that no interview alone could have provided.

The second part of the STNS took place in Aarhus where I met the IT-specialist who worked on the development of the *Netarkivet*'s SolrWayback. After a brief tour of the University Library and its deposits, I had the chance to meet the team stationed in Aarhus and learn more about the SolrWayback and specific challenges related to the collection of and providing access to Twitter content.

CONCLUSIONS

The Short-term Network stays in France and Denmark have had a positive impact on my research allowing me to meet and talk in-person to web archivists and professionals working daily in the development of social media collections and dealing first-hand with the many issues these born digital items pose. The STNS also offered the unique opportunity to have interesting discussions with various members of both France and Denmark web and social media archiving teams, especially around the way content is selected, its value for society, the vital interactions with researchers and the difficult relationship with social media platforms or, at least, in the rare instances where a channel of communication was even managed to be established directly with service providers.

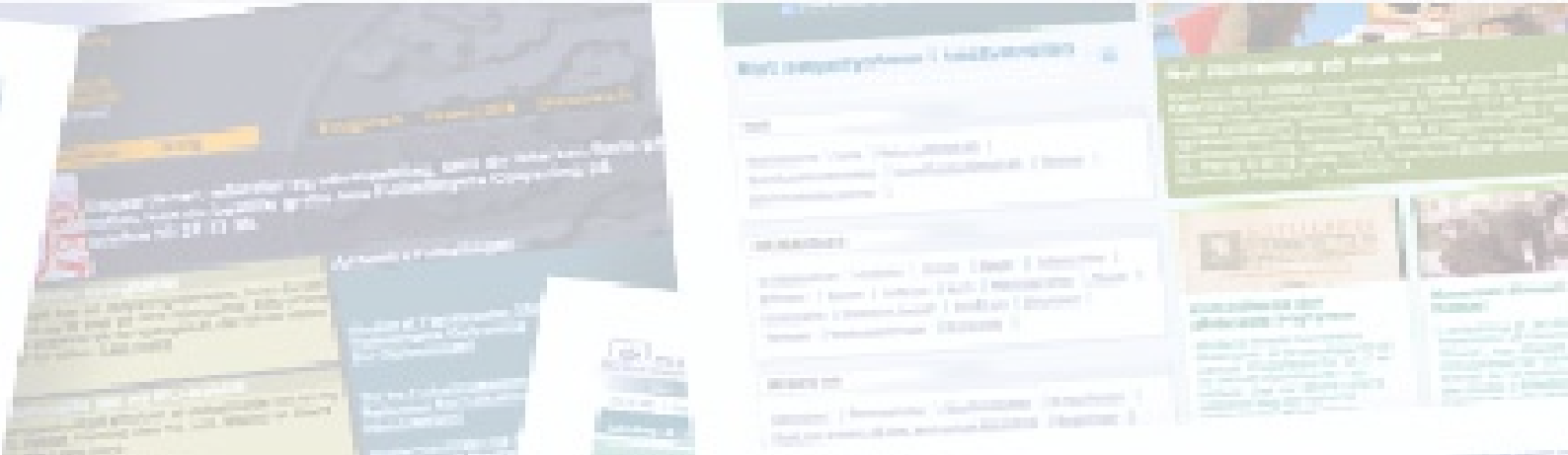
The quality and quantity of data collected throughout these two STNS was essential to my study and especially to the final goal of my research which is to identify a set of guidelines for the development of social media archives, that hopefully could support and offer food for thought to institutions interested in archiving social media content as well as

to those, such as BnF, INA and KB who already have a well-established and long-standing experience with this type of content.

Finally, I would like to reiterate that the value of these STNS did not just reside in the ability to conduct interviews in person, but especially in the chance to meet members of the respective web archiving teams developing personal networks, gather unique insights through observation or informal discussions, and also being able to browse web and social media archiving collections which would be otherwise impossible as access is often limited to on-site or at least bound to a location within the national borders. All of this would have never been possible without the WARCnet short-term network stays.

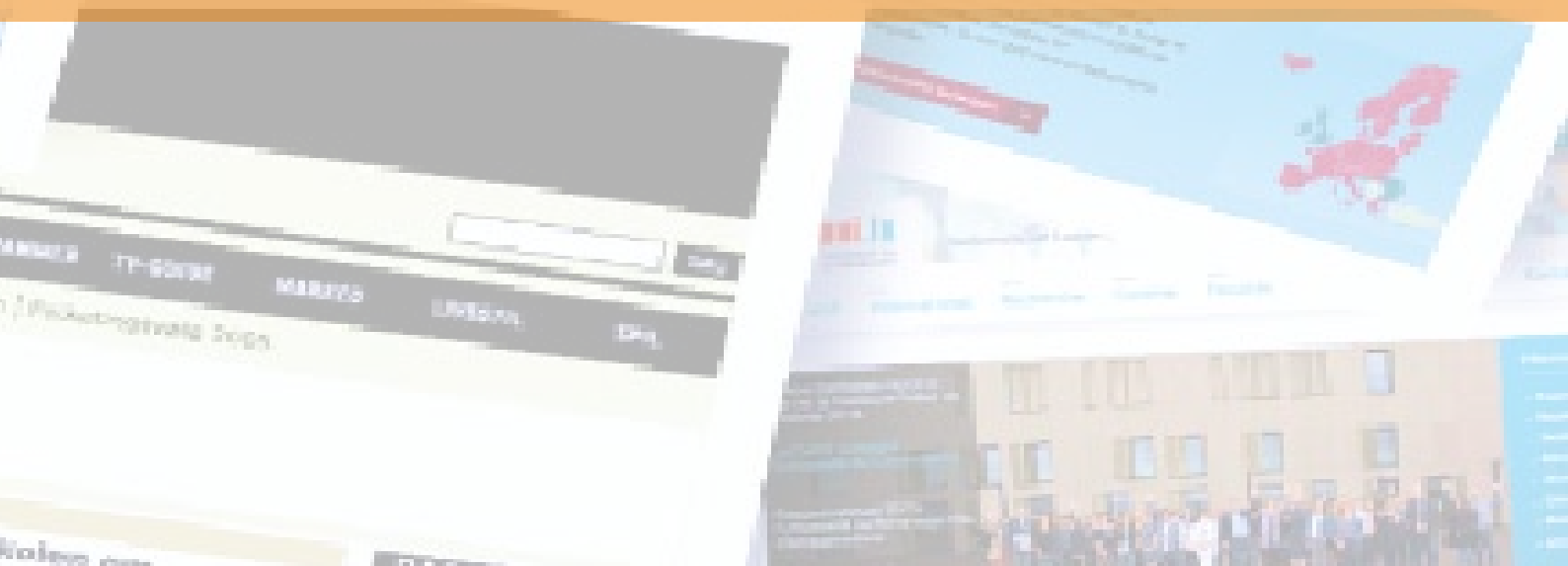
REFERENCES

- Brügger, N. (2017). Web History and Social media. In J. Burgess, A. E. Marwick, & T. Poell (Eds.), *The SAGE Handbook of Social Media* (pp. 196–212). SAGE Publications.
<http://ebookcentral.proquest.com/lib/ulondon/detail.action?docID=5151795>
- Burkey, B. (2020). Repertoires of Remembering: A Conceptual Approach for Studying Memory Practices in the Digital Ecosystem. *Journal of Communication Inquiry*, 44(2), 178–197. <https://doi.org/10.1177/0196859919852080>
- Espley, S., Carpentier, F., Pop, R., & Medjkoune, L. (2014). Collect, Preserve, Access: Applying the Governing Principles of the National Archives UK Government Web Archive to Social Media Content. *Alexandria*, 25(1–2), 31–50. <https://doi.org/10.7227/ALX.0019>
- Fondren, E., & Menard McCune, M. (2018). Archiving and Preserving Social Media at the Library of Congress: Institutional and Cultural Challenges to Build a Twitter Archive. *Preservation, Digital Technology & Culture*, 47(2), 33–44. <https://doi.org/10.1515/pdte-2018-0011>
- Marshall, C. C., & Shipman, F. M. (2012). On the institutional archiving of social media. *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries*, 1–10. <https://doi.org/10.1145/2232817.2232819>
- Thomson, S. D. (2016). *Preserving Social Media* (16–01; DPC Technology Watch Report). <https://www.dpconline.org/docs/technology-watch-reports/1486-twr16-01/file>



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