

# Exploring special web archives collections related to COVID-19: The case of the National Library of Australia

Olga Holownia, Friedel Geeraert,  
and Paul Koerbin

WARCNET PAPERS

WARCnet  
web archive studies

# Exploring special web archives collections related to COVID-19: The case of the National Library of Australia

*An interview with Paul Koerbin (National Library of Australia)  
conducted by Olga Holownia (IIPC) and Friedel Geeraert (KBR)*

[olga@netpreserve.org](mailto:olga@netpreserve.org)

[friedel.geeraert@kbr.be](mailto:friedel.geeraert@kbr.be)



WARCnet Papers ISSN 2597-0615.

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

© The author, 2022

Published by the research network  
WARCnet, Aarhus, 2022.

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: 978-87-94108-11-9

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)

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)

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)

# Exploring special web archives collections related to COVID-19: The case of the National Library of Australia

*An interview with Paul Koerbin (National Library of Australia) conducted by Olga Holownia (IIPC) and Friedel Geeraert (KBR)*

*Abstract: This WARCnet paper is part of a series of interviews with web archivists who have been involved in special collections related to COVID-19. The aim of the series is to provide a general overview of COVID-19 web archives. This interview with Paul Koerbin at the National Library of Australia was conducted in collaboration with the International Internet Preservation Consortium (IIPC).*

*Keywords: web archives, social networks, COVID-19, special collections, National Library of Australia, NLA*

This WARCnet paper is part of a series of interviews with web archivists who have been involved in special collections related to COVID-19. The interview was conducted on 19 August 2022 with Paul Koerbin at the National Library of Australia (NLA). Paul Koerbin is Assistant-Director Web Archiving, one of a team of three along with an Assistant-Director Web Archive Systems and a Senior Web Archivist. His role has responsibility for the business aspects of the NLA's web archiving program which includes strategy, policy and reporting duties, scoping and processing selective, thematic and bulk harvesting and managing relationships including those with the Internet Archive (for the .au TLD whole domain harvests) and the IIPC.

Web archiving formally commenced at the NLA in 1996. Initially this involved selection: scoping what would be collected with practical, operational collecting commencing in early 1997. Online publication, which specifically includes websites, web pages and web files, has been included in national legal deposit legislation since February 2016. In Australia national (Commonwealth) legal deposit provisions are included within the Commonwealth Copyright Act (1968), ss.195CA-195CJ. That is, legal deposit exists as a National Library specific exception within copyright legislation rather than in the National Library Act (1960). This gives a wide coverage (i.e. everything covered by the Copyright Act) as well as a high

level of legislative force. The material specifically covered by legal deposit is found in s.195CE of the Act.

The NLA uses a range of strategies to collect web content. This includes:

- 1) selective collecting targeting specific websites, each of which is collected according to an individually specified harvester (Heritrix, Browsertrix or HTTrack), scope (i.e. harvester profile), and harvesting schedule;
- 2) thematic selective collecting focused on events (e.g. elections, natural disasters) and topics (e.g. iconic Australian brands) using the same method as for selective collecting; and,
- 3) bulk seed-list and domain harvesting which includes thematic bulk harvests (i.e. bulk seed-list based harvests run in-house using Heritrix, currently mostly focused on collecting government websites) and whole domain harvests of the .au ccTLD harvests conducted annually in collaboration with the Internet Archive (using Heritrix and Brozzler).

At the end of August 2022, total content of the archive (figures rounded) comprised of data (WARC files): ~680 TBs and URL snapshots: ~15,000,000,000 records. The NLA's entire web archive corpus, known as the Australian Web Archive (AWA) — consisting of the 'PANDORA' selective web archive collection, the Australian Government Web Archive collection and the whole .au domain collection — is accessible through the NLA's discovery service called Trove. It is openly accessible online and can be searched by URL and text. Trove also includes browse paths to thematic collections. In addition, the older interface to the PANDORA selective web archive is also maintained (for the time being) as that provides browse paths to title lists and subject lists that are not available in Trove.

The NLA continues to maintain partner arrangements for the selective component (formerly known as PANDORA) of the larger web archiving program. Partners are not involved in domain or bulk harvesting. They provide curatorial resources while all infrastructure is developed and maintained by the NLA. The partner organisations include: the state libraries of Victoria, New South Wales, Queensland, South Australia and Western Australia; the Library and Archives of the Northern Territory; the Australian War Memorial and, the National Gallery of Australia.

The web archiving program within the NLA was subject to a significant restructure two years ago. The staffing was reduced from around six staff to three. The team now manages the web archiving operation as a 'product' with ongoing development of infrastructure, processes and procedures focused on the NLA's strategic priority to comprehensively collect Australia's online resources.

## THE REASONS FOR THE SPECIAL COLLECTION

*Why did you create a special COVID-19 collection?*

Paul Koerbin: Creating thematic collections around events including natural disasters has been a part of the NLA's web archiving approach for more than 20 years. So it was nothing novel, nor particularly challenging, to initiate a collection around COVID-19 in February

2020. By that time, it was evident that this pandemic was going to have a significant impact on Australian society and online responses to it should be documented. As I recall, one particular event, the major spread of the virus as the result of passengers arriving from a cruise ship, put the seriousness of the pandemic into focus. At the time of the emergence of COVID-19 we were also in the 'mopping up' stages of yet another natural disaster collection around devastating bushfires in the Australian summer of 2019-2020. So it was really just moving our collecting focus to COVID. It was, dare one say, and sadly, just business as usual.

## **THE SCOPE OF THE COVID-19 COLLECTION**

*What exactly did you collect? Websites, social media? Which specific platforms, hashtags, profiles or languages?*

Paul Koerbin: The simplistic answer is that, being a web archive, we collected online web resources (websites, or parts of websites, web pages etc). In saying that, I mean we did not approach the collection of content from the perspective of platforms or formats or forms, in other words, the sort of material we collect in the normal day-to-day operation of our web archiving activity. Formats, profiles, languages, platforms etc. are not overt selection considerations. The target is online content. Of course, platforms and formats certainly do become considerations in determining how successful (or not) we are in collecting material, but not in the fact of what we aim to collect. As Australia is a multicultural society, it was certainly the intent to collect material in community languages and to collect material relating to Indigenous Australians.

Generally speaking, social media is tricky since we do not have, at the NLA, a social media archiving infrastructure or program. Social media is not distinguished in our web archiving but this does mean that what we collect is limited and does not tackle the more difficult social media platforms. That said, we have collected numerous Twitter accounts. This focused on accounts by notable epidemiologists, statisticians and government services amongst others. The website approach to Twitter means we focused on accounts and we were not able to collect hashtag conversations. While the Australian Web Archive does not generally cover video content (YouTube) and material that is substantively audio-visual is explicitly outside the coverage of legal deposit, one of the partner organisations contributing to the collection did collect many YouTube videos of government spokespersons. This collecting was based on a government directive permitting the collecting of government content within that state.

The collection aimed for a broad coverage of material. Obviously, there is a lot of government material, from the federal government and from state and local government sources. As COVID unfolded, many cultural and sporting events collapsed and so we collected many online statements documenting this. Similarly with business and industry statements. Of course, news sites have been a critical source and that extends to opinion and commentary sites. As vaccines were developed, we also documented local medical industry sites as well as those publishing and analysing epidemiological data. Travel,



tourism and the higher education sector were hard hit and these too were documented. We also collected material reflecting opposition to government measures such as opposition to lockdowns and mandated mask wearing. Much of material relating to protests and the views of conspiracists and like would more likely be picked up in our general web archiving of news and opinion; however, one of the subcollections titled 'News, media, Twitter and opinion and protest sites (COVID-19)' does bring some of this material together.

These points of focus are reflected in the COVID-19 collection's 18 subcollections. Those subcollections also include a jurisdictional focus (i.e. Commonwealth and state jurisdictions) as the response to the pandemic was often different within the various state jurisdictions (sometimes these differences were quite at odds with one another). It is also worth noting that, in the early period when we were establishing the collection, these subcollection labels were refined numerous times. This probably accounts for the way some of the subcollections have layers of terms. So, the subcollections and their labels were determined by the content as it was collected rather than established 'a priori'. In this way, the collections reflect what we identified and collected as we embarked on the collecting; and as such also reflect the highly reactive, real-time approach to web archiving at the NLA more generally. It is certainly fair to say we spend more time in doing the collecting than pre-planning it. In the dynamic world of the web, this seems an appropriate response much of the time. Anyway, the objective was to find a balance between enough but not too much specificity in the breakdown of subcollections.

The COVID-19 collection and its subcollections can be browsed (and searched) in Trove and cover the following topics:

- ACT (Canberra) and Northern Territory government, local and regional organisations, services and information
- Business and industry statements, services and economic assessments
- Commonwealth Government services, advice and information
- Cultural institutions, arts bodies, artists, performers, venues and events — statements, notices and activities
- Medical industry bodies, health services, research, law and epidemiological data
- New South Wales government, local and regional organisations, services and information
- News, media, Twitter and opinion and protest sites
- Overseas Embassies and Consulates
- Queensland government, local and regional organisations, services and information
- South Australia government, local and regional organisations, services and information
- Sports organisations and events — statements and information
- Support services, charities, childhood and aged care organisations
- Tasmania government, local and regional organisations, services and information
- Travel information and tourism impact
- Universities and education sector statements and information
- Various bodies including unions, associations, religious organisations, conservation groups

- Victoria government, local and regional organisations, services and information
- Western Australia government, local and regional organisations, services and information.

*Could you provide more information with regards to the amount of data collected and the nature of the collected data?*

Paul Koerbin: The COVID-19 collection continues to grow as scheduled harvesting of many sites still continues. The current size is in the order of 8.2 TBs or 64 million files from around 35,000 harvested instances of 2,100 targeted sites.

I think it is also important to note that the COVID-19 collection is really only an entry point into the web archive and much more content related to the pandemic. Clearly, over the period since the pandemic started just about every website will refer to COVID-19 to a greater or lesser degree. It would be meaningless (and not feasible) to include everything in the COVID-19 collection. Therefore, there are some parameters set. We only included websites or parts of websites that were specifically and entirely (or mostly) related to issues in regard to the pandemic. So, in respect to news sites for example, only those parts of the websites clearly headlined by COVID news were included. There may be many more mentions or references to the pandemic in other content such as general news coverage. We would, most likely, pick this content up through other harvests (regular collecting of news sites or bulk harvests of gov.au and .au content) and, when browsing within the AWA delivery (the Trove delivery infrastructure for AWA which allows you to browse within the entire web archive corpus), it is possible to follow links to this other content. This is what I mean by the COVID-19 collection serving functionally as an entry point to pandemic-related content. In general, I think our thematic collections are a lens to the larger web archive, not the endpoint nor a definitive collection of material.

For the reasons touched on above, the collected material documenting the pandemic would quite obviously be much larger than that specifically identified within the formed collection.

*How did you go about archiving on a national level about an event that is fundamentally global?*

Paul Koerbin: I don't think this is particularly problematic — if you take social media out of the archiving equation. The focus of the collection is how the pandemic affects Australia and Australians and what the reactions and responses by Australians (government and otherwise) are. That is pretty straightforward. Legal deposit in Australia is not constrained by the ccTLD, so any material the Director-General considers relevant to Australians — and which would be covered by the Copyright Act — could be selected. In the case of this collection, that was not a significant issue. Most content was unambiguously within the coverage of legal deposit. The social media that we did collect, specific Twitter accounts, were primarily government information and epidemiological reporting and medical advice.

In respect to archiving nationally, the Commonwealth of Australia is a federation of states, so what was important was to document the different responses to the pandemic by

the respective states and by the Commonwealth. This was certainly made easier by the fact that 6 of the 8 Australian states and territories are participants in the PANDORA selective web archiving collaboration. So, responsibility for collecting was shared among participants with responsibility of their respective jurisdictions; and, of course, curators in the states had a closer view of issues pertaining their jurisdiction.

*Could you elaborate on this collaboration between the states and territories? Do the representatives have access to a seed nomination tool or can they launch crawls themselves? Are there any regular meetings to discuss starting new thematic collections or collections related to events?*

Paul Koerbin: In respect to process and procedure, all partner organisations use the web archiving workflow tool that is developed and maintained by the NLA, called PANDAS. This system allows agencies to manage their own contributions. Ultimately content ends up in the one central archive but selection, scoping, scheduling, QA etc. is all managed by each individual agency. All users of PANDAS can search the tool and see if a seed has already been selected. In practice, because of the division of responsibilities by jurisdiction, partners can be left to do their own selection without concern about encroaching on the responsibilities of other agencies. The PANDAS tool retains the organisation of work by contributing agency and by individual users within the agency. In other words, seed selection is not coordinated centrally by the NLA in a restrictive or proscriptive way.

In respect to the intellectual organisation of thematic collections, those collections which are national in scope, like the COVID-19 collection, are established by the NLA and partners are advised of the theme and structure. As appropriate, this may be done in discussion with representatives from partner agencies. Some thematic collections are established by partner agencies when they are specific to their jurisdiction, e.g. state election campaign collections. Generally, there are not regular meetings but the NLA does maintain an internal PANDORA mail-list to update partners on activities.

The NLA does not direct partners on what to collect in any specific detail. This means that collections may be inconsistent across jurisdiction in respect to the type and extent of content collected. For example, the State Library of Victoria tends to collect government video content extensively — based on specific permissions and their internal collection strategies — while this is not done by other agencies.

## THE FRAME OF THIS SPECIAL COLLECTION

*When did you start collecting and when did/do you plan to stop? What was the capture frequency?*

Paul Koerbin: We started in February 2020. At this stage we do not have a set end date although there is certainly some winding back of the harvesting currently underway. Some websites have already disappeared or 'moved on'; others are presenting less frequent updates. We will probably continue to leave the collection open, at least for the rest of 2022, perhaps longer. Possibly once the common social discourse moves to 'post-pandemic' — let us hope sooner rather than later — or perhaps even 'living with COVID' we may consider establishing a new collection to document that, drawing a line under the COVID pandemic collection. But the situation still seems too volatile to make that call just yet.

The question of capture frequency is not quite relevant to the NLA's collection. The collection is not a seed list-based harvest that is run in its entirety to a given schedule. Each archived website/document/resource has its own schedule. Some, many, were archived only once. Many more have been on schedules ranging from daily to quarterly (and many points in-between). Schedules also get modified as the collecting progresses, more or less frequent as the updating of content dictates.

*Could you elaborate on how you monitor the frequency with which content is updated to adapt the individual capture schedules? Is this an automated process?*

Paul Koerbin: The scheduling of harvesting frequency is manually applied to each title (i.e. seed). Schedules are established based on the nature of the publication. For example, news sites would most likely be done daily, while other sites might be done on any schedule from daily to weekly, monthly etc. There is no automated process to determine schedules. The schedule is reviewed by the curator when performing the quality checking process on harvests. If there is little change detected, the schedule may be lengthened, and vice versa. The workflow tool provides thumbnails of previous harvests to help with this process. Scheduling is more an art than a science, but the objective is to ensure the efficient capture of all the content. For example, a balance might be made between frequency and depth of harvest.

*How did you carry out quality control on the collection (if applicable)?*

Paul Koerbin: Quality control methods were the same as used for our everyday selective web archiving. The PANDAS workflow system allows the curator to view the content — i.e. a visual quality check — which can be cursory or in depth. The visual control displays thumbnails of the new harvest, the live site and previous harvest as well as presenting statistics for data size and number of files collected. Content that is, prima facie, harvested to an acceptable extent is clicked through to the production archive. Recent enhancements to the workflow system have made this process much more efficient, so that quality

checking and archiving can be done either in bulk or individually. For example, sites known to harvest well can be flagged to archive with little checking, while others that display problems — or display notable change — can be checked more thoroughly. Quality checking can result in changing the harvester and/or harvest profile to improve the harvest outcome.

*Did you encounter any issues, challenges, or limits related to the collecting activity?*

Paul Koerbin: There were of course the typical challenges and limits with regard to collecting web content, such as the technical challenges we always confront with certain websites. And there are always resource limitations with a very small team committed to the work.

The main distinctive challenge with the COVID collection is really the duration of the collecting. Most event collections are fast and furious, as is the norm with election collecting. Even if we are doing topical thematic collections a lot of the work will be front loaded and taper to the occasional addition of content. With the COVID collection there was, at least for the first year, and perhaps the second too, the need for a sustained effort as the issues around COVID and social and government responses evolved and documenting the pandemic required frequent (daily) harvesting of many resources.

Certainly, the COVID collection has occupied a very large part of our collecting time and focus over this period. That said, while continuing to build the COVID-19 collection we have also completed a very large election collection (along with consequent significant changes to collecting of government material arising from machinery-of-government changes); established a collection for the Russian invasion of Ukraine; and continued the daily processing of the many ongoing scheduled harvests we do.

The other main challenge I have already alluded to — and that is where to draw the boundaries around the collection given the issue of COVID having so thoroughly permeated online discourse. The challenge is really to make a collection that will usefully lead researchers (and the interested public) to pertinent resources without too much to wade through. It is a fine balance between making a collection genuinely useful and one that risks subverting itself by too much content. I'm not sure how well we have made that balance: time may tell. As with all web archiving endeavours, the urgent task is to capture content in a timely manner. The refinement in how we manage and organise that content can be left to later, although such work always runs the risk of languishing due to lack of resources and new priorities that emerge.

## **ACCESSIBILITY AND SEARCHABILITY**

*How can users access and search in this collection?*

Paul Koerbin: The collection with its 18 sub-collections can be browsed from the [Trove discovery service](#). The browse path is somewhat less than intuitive, it has to be said, since the COVID-19 collection is itself a subcollection three layers deep within the AWA. The user

needs to follow the path from the home page via links to 'Health' to 'Medical Conditions & Diseases' and then to 'Coronavirus (COVID-19) Pandemic and Australia, 2020-2022'.

Alternatively, users can use the simple or advanced search options in the Trove 'websites' zone to search by URL or text. It is possible under advanced search to limit searches to various government domains (e.g. gov.au or nsw.gov.au) or specific domains. Searching cannot however be limited to, or for filtered by, collection. When viewing a page within the AWA in Trove it is possible to open a side panel facet to see if it is part of a collection and link through to that collection, although, in all frankness, this is not very intuitive in the way this is currently presented.

*Have researchers already expressed interest in using the COVID-19 collection?*

Paul Koerbin: The collection is fully accessible, at least through browsing and through URL and text searching, so there is no need for researchers to make contact with us in order to access the content and so we have little idea how it is being used. We do not have an API available for programmatic analysis of the content, but we have not received any requests for this either.

While we do have *Google Analytics*, these do not provide details from which much can be interpreted. Since February 2020, the collection is only around the 11th most browsed and the *health.gov.au* domain the 3rd most viewed. Like most web archiving collections, its value needs to be considered over a much longer time as research develops around issues that will be supported by the web content collected.

*How do you communicate about this special collection?*

Paul Koerbin: We have not undertaken any special or specific communication in regard to this collection. The web archive does have a Twitter account which has provided very occasional updates but otherwise we have not specifically publicised the collection.

We did participate in the Documenting COVID-19 panel at the virtual 2021 IIPC Web Archiving Conference. This is the only public forum in which we have presented on our COVID collection.

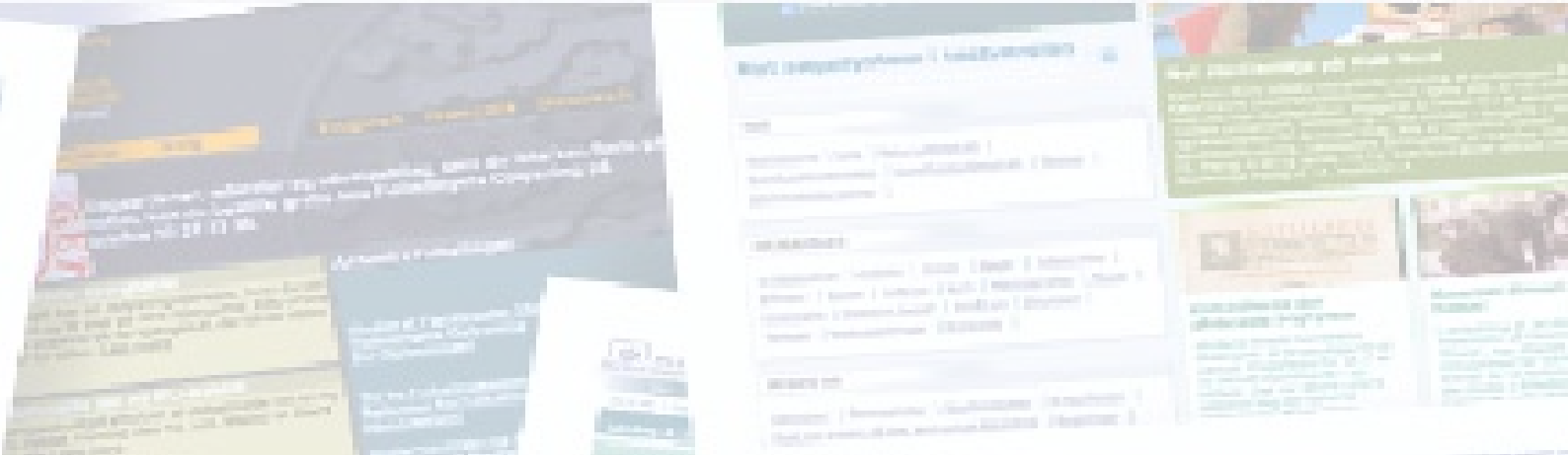
*Did you have any partnerships with local stakeholders (Archive-It, IIPC, etc.) during the collection process?*

Paul Koerbin: I am not sure of the meaning of 'local stakeholders' in this question. We did not use Archive-It, the Internet Archive or other service providers, nor did we partner specifically with the IIPC. The only local stakeholders were the PANDORA partner agencies mentioned previously who contributed to the building of the collection. We did make a small and quite limited contribution to the IIPC COVID collection, however our limited resources were focused on our national COVID collection.

## REFERENCES

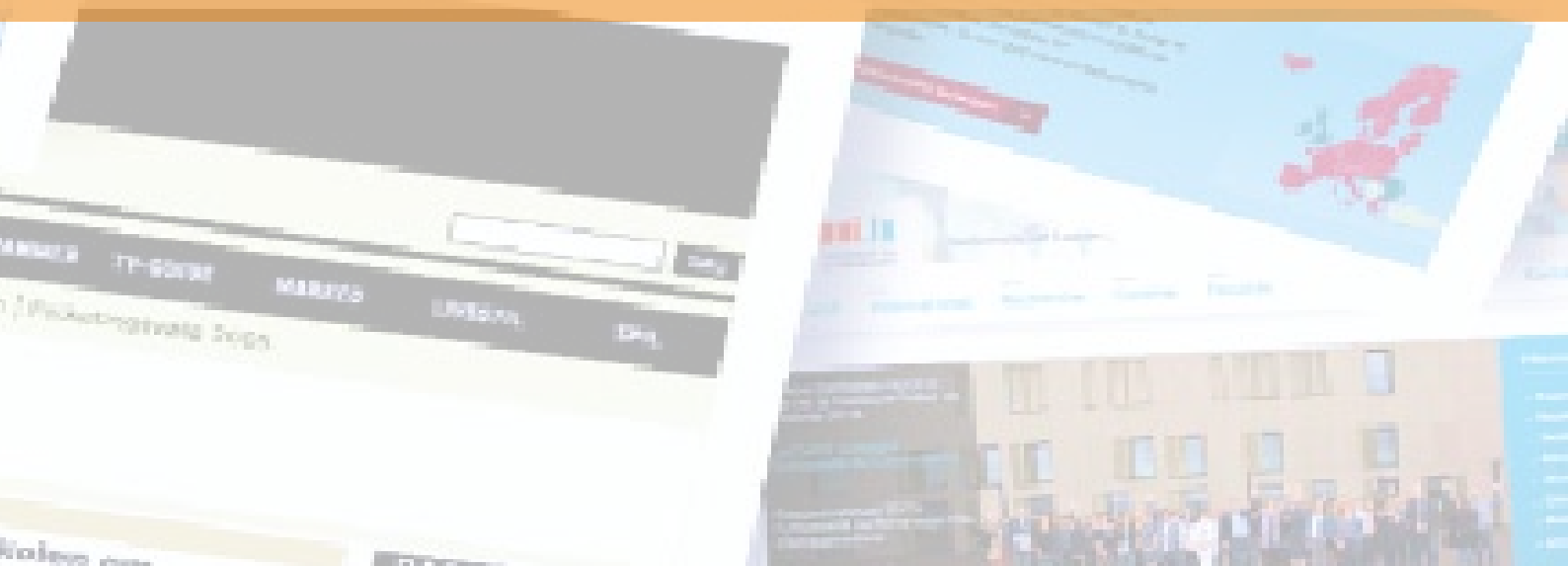
- National Library of Australia. TROVE (the Australian Web Archive). Retrieved from <https://webarchive.nla.gov.au/collection>.
- Commonwealth Consolidated Acts. *Copyright Act 1968 — SECT 195CE*. Retrieved from [http://www5.austlii.edu.au/au/legis/cth/consol\\_act/ca1968133/s195ce.html](http://www5.austlii.edu.au/au/legis/cth/consol_act/ca1968133/s195ce.html).
- National Library of Australia. *Bushfires, Australia (2019-2020) Collection*. Retrieved from <https://webarchive.nla.gov.au/collection/17777>.
- National Library of Australia. The COVID-19 collection and its subcollections. Retrieved from <https://webarchive.nla.gov.au/collection/17818>.
- National Library of Australia and Partners. *Pandora Australia's Web Archive*. Retrieved from <https://pandora.nla.gov.au/>.

*We would like to thank Paul Koerbin and Kelsey Socha for their help in proofreading this interview.*



**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