

Exploring special web archives collections related to COVID-19: The National Library of Ireland

Michael Kurzmeier, Joanna Finegan and Maria Ryan

WARCNET PAPERS



Exploring special web archives collections related to COVID-19: The National Library of Ireland

Joanna Finegan and Maria Ryan (NLI) conducted by Michael Kurzmeier (Maynooth University)

Michael.Kurzmeier@mu.ie



WARCnet Papers
Aarhus, Denmark 2022

WARCnet Papers ISSN 2597-0615.

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

© The authors, 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: 978-87-94108-06-5

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)
- 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)

Exploring special web archives collections related to COVID-19: The National Library of Ireland

An interview with Joanna Finegan and Maria Ryan (NLI) conducted by Michael Kurzmeier (Maynooth University)

Abstract: This WARCnet paper is part of a series of interviews with European 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.

Keywords: web archives, social networks, COVID-19, special collections, Ireland, National Library

This interview is part of a series of WARCnet Papers about how Web Archives have collected material relating to the COVID-19 pandemic. This interview was conducted with the digital collection staff at the National Library of Ireland (NLI), Joanna Finegan and Maria Ryan, on November 17, 2021. As a national library, the NLI is responsible for selecting, preserving and making available publications issued in Ireland, including Websites. The NLI has a special focus on Irish language publications. Ireland does not have a legal deposit legislation, and thus the NLI has focused on selected sources to archive in-depth. The ongoing pandemic presented both challenge and opportunity to archive Irish voices in relation to COVID-19. The NLI uses Archive-it to manage and make its collection of archived Web material accessible.

THE REASONS OF THE SPECIAL COLLECTION

Michael Kurzmeier: Did you conduct a special COVID collection? If so, what was your motivation?

Maria Ryan: Yes we did, we did conduct one as part of one of the collections in our openly available selective web archive, because obviously, the pandemic has had such a massive impact on every country, not least Ireland. So, it's important to reflect its impact, as best we

could. It wasn't something that we couldn't do, and it fell within our remit. And I know, for other people that do web archiving, there would have been a special focus in a Covid collection. But for us it was the case that we needed to capture the broadest picture of the pandemic in Ireland, because it was very clear that it was going to affect all aspects of our life. And it was important to preserve the impact of the pandemic on our country through web archiving.

Joanna Finegan: This is our only shot at preserving the impact in terms of Ireland's life online, because we don't have a domain crawl. So that may be a different situation to some of the other countries and the preservation programs in that way. For us, it is selective web archiving only. Which is why we had to take that opportunity.

THE SCOPE OF THE COVID-19 COLLECTION

What did you collect? websites? social networks (in this case, which DSN, specific hashtags, profiles, languages)?

Maria Ryan: We collected websites, both in the English and Irish language, and we collected the broadest picture possible of our life in Ireland with COVID. So that was every industry, every sector, and obviously, especially focused on health. So, we would have started by looking at government guidance, for example the HSE [Health Service Executive] website, HSPC [Health Protection Surveillance Centre], and the Department of Health, and in general the government's response to Covid. And then also media and commentary on the online-only media and the commentary that we get through the National Library. The National Library is a legal deposit library, so we receive print or digital copies of print publications in Ireland. That covers all the newspapers. We knew we had that. And this is why we focused on online-only-commentary, things like Thejournal.ie and Tuairisc.ie, which is an online-only Irish language journal. And then obviously, the news online. We started getting a picture quite rapidly. The collection is still growing, and it is websites only and in both languages, because we have a remit to do that.

THE FRAME OF THIS SPECIAL COLLECTION

When did you start? When did/do you plan to stop? What was the capture frequency?

Maria Ryan : I suppose when Covid stops being front page news. There is no definitive timeline. I think we will be dealing with this for a long time. And after that I can envisage a *life after COVID* collection or something to that effect. We started in March 2020. If you remember, there was a general election in February of that year. And so you can see in that collection, through some of the news and commentary websites that we would have kept captured as part of the election, that COVID is coming to the forefront of the news. First it is reporting on the cases in China, and then it reached us totally, but focused

collection started in March. And from about mid-March, we were capturing certain websites. There were about six websites that we captured on a weekly basis, and they were the government's and HSE websites, because then the advice was changing constantly. At this point we were looking at the impact that Covid was having on the Health Service and the HSPC, and the government and the media sites. We would crawl them on a weekly basis from the start.

And as things started to settle throughout the year, we then looked to expand that to also include the sectors that it was affecting. So that was everything like tourism, hospitality, weddings, and in the middle of last year, there was so much talk about going to the hairdresser. And so we looked at all those different sectors, so official government responses and then also the non official response of a few campaigns that were ongoing. For example, *we can be zero*, we looked at those responses, and those campaigns that were rising, but also, some industries pivoted to making hand sanitizer, and things like that. And we looked at those sorts of examples, and also community responses. We wanted to understand what people are doing in response to COVID. And I suppose in light of yesterday's news [announcement to halt easing of restrictions made on the 16th], we will look again at the various sectors that have been affected and their response.

The COVID-19 collection is openly available. What we have collected is already available from the online selective web archive portal on our website. And as Joanna said earlier, this was our only shot at capturing any of that response online. And we are focused on capturing and selecting on a smaller scale, but obviously, a broader domain crawl would have picked up more content relating to COVID-19 In Ireland, but there's still no legislation to enable the National Library to carry out a domain crawl. So that's been a major, a major block for us.

Joanna Finegan: Nicola Bingham spoke about the challenge of keeping your selective collection focused. And in terms of usability for readers, obviously, that's a massive challenge with COVID, growing only as the situation continues. What would be interesting to look at, would be other national initiatives just to see what their response is. But I don't think we're at a stage where we're closing just yet. It's a bit like Brexit, which is another collection, which we haven't closed that we expected to have done so but we're not there on that either.

How was quality control done on the collection?

Maria Ryan: We have a thorough quality assurance process, and we carried that out on all of those websites. We have publicly available close to 200 websites that we've captured on various times. COVID became our primary focus last year. So, there were many things we did not get to do because of it. But all material has been fully worked through that workflow.

Could you give me a quantitative overview of the data?

Joanna Finegan: There is about 3.7 TB of material on the portal which comprises both sites which are regularly crawled and a broader set of sites, captured less frequently, which represent the most affected sectors of society. Part of this material stems from the March 2020 general election collection and overlaps with the emerging COVID situation.

Also there are a number of other discrete collections happening regularly every year, such as higher education or trade unions. And they have a lot of COVID content in there as well. But they weren't included in the specific COVID-19 collection, because they are done annually.

Maria Ryan: There are regular collections that we do on an annual basis, things like the local government's response to the pandemic that will be in our local authorities collection. So, it really feeds into the larger collection of COVID-related material because the actual nature of the pandemic is so all encompassing. There isn't one aspect of Irish Life that hasn't been touched by the pandemic, and that's really reflected in our collections as a whole.

ACCESSIBILITY AND SEARCHABILITY

Can we talk about the accessibility and searchability of these data?

Joanna Finegan: Our technical partner is Internet Archive and our collection is presented through Archive-it. That means the collection is fully available with a full text search function. We are also looking into new ways of exploring the collection. Our goal was to make the collection publicly available as quickly as possible.

PARTNERSHIPS AND USES

Are researchers already asking you about the COVID-19 collection and wanting to analyse it?

Joanna Finegan: I think you guys [Warcnet] are probably the first that have approached us in that regard.

Maria Ryan: It's early days, and I'm not surprised that nobody has made official requests yet. But they will and I can see that happening. We're still in the middle of it, that makes it hard for researchers. The collection is of course incomplete as well, because it is still ongoing. So I can see people wanting to chart the pandemic online, eventually. I can imagine that those requests will come. But you know, people use our collections every day and they do not need to formally request that data from us.

How did you communicate about this special collection ? Did you have some partnerships? (with local stakeholders, with Archive-It? With the IIPC?)

Maria Ryan: The NLI's technical partner is Internet Archive and we're also members of the IIPC and the DPC. It is great to have that community, particularly with the IIPC. So it was great to see what other people were doing and to be able to talk about that like at the conference this year and the year before. We didn't feel like we were alone. We also consulted with Dr. Ida Milne, She's a professor here in Ireland. And she advised us on what to collect, looking at our collection as a whole. That's something that we always try to do in the National Library, develop partnerships with subject specialists, and experts. And so that's what we wanted to do with this one as well.

We also communicated about our Covid19 Collection through NLI social media channels, including seeking nominations for the collection from the public. The library's event to mark World Digital Preservation Day 2020 was 'Panic and potential: challenges and opportunities in Pandemic collecting' and we joined with other Irish organisations in discussing collecting during Covid19 times.

What were issues, challenges or limites of this collection activity?

Joanna Finegan: There's never an interview or an opportunity to engage about web archives in Ireland that we don't mention this: The single biggest challenge is the legislation in relation to archiving the Irish web, that is still an enormous challenge and a constant source of surprise that we're still at that point where we don't have a domain crawl. You know, particularly with the public remit as, as copyright library, were the only Irish copyright library founded with a specific public benefit remit. Lots of material is lost to the public, ultimately. Other than that, we had issues of scoping and providing a fair representation of the impact across society.

Maria Ryan: The legal situation has a massive impact, because of just the pressure then on the selection and trying to get as much as we can. And that's very difficult. Having a domain call would help us focus our collecting. It will also give us the security of knowing that we have captured more of the Irish web. And what, apart from that, one of the main challenges of COVID-19 is that it's different from any other collection that we have done. I think Brexit is probably our other really complicated collection, and all encompassing, but COVID-19 was on a completely different scale. And it just impacted on everything. And that was one of the massive challenges was trying to keep on top of everything, while getting a representative picture of what life in Ireland is with COVID.

How easy was it to manage collecting activity from home?

Maria Ryan: We adapted quite easily because all our systems are online anyway. And we manage our collecting online. I think we did quite well, we're probably one of the easiest departments in the National Library to put online. We were so focused we just had to hit the ground running. So it really just gave us that push.

Joanna Finegan: Maria, you're the person managing in real time the pressure of recording an evolving situation and also collecting an evolving situation. That was really, really considerable, especially as the only member of full time staff on web archiving.

How do you archive nationally something which is fundamentally global?

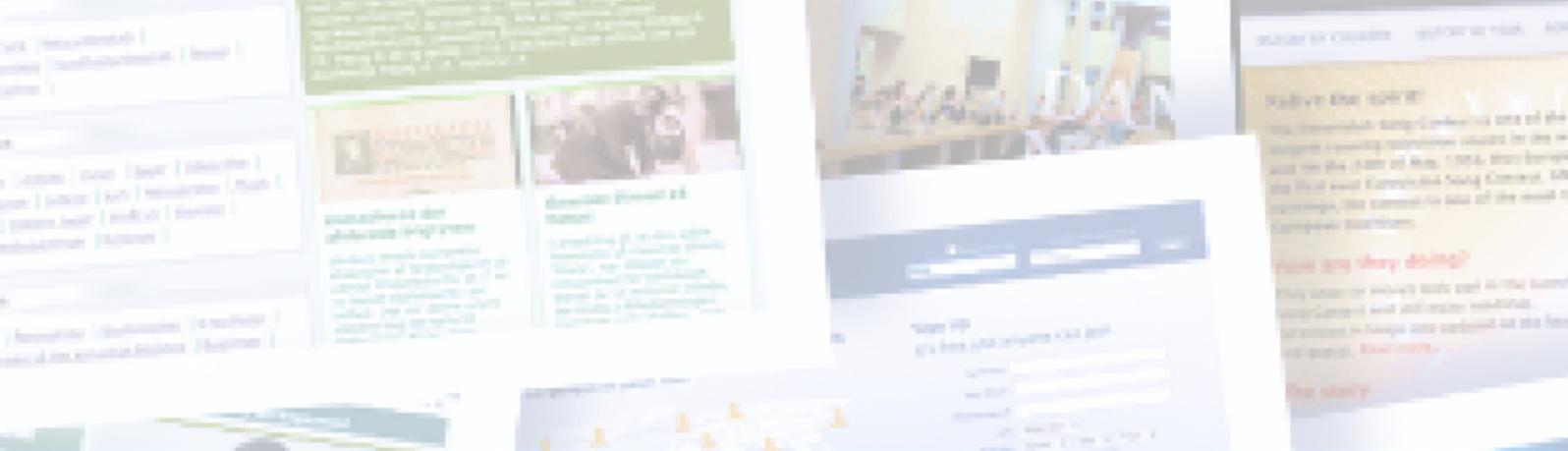
Joanna Finegan: We all have to play our part, I suppose. There are some global web archiving initiatives. But ultimately, digital preservation is a network.

Maria Ryan: Nearly every national library in the world was doing what we were doing. So we just needed to focus on what we could do and what we could archive. I suppose that's all we don't have a global remit. We're not the Internet Archive. And we've always just looked at our Ireland and the experience of the Irish people. Both at home and abroad, we did look at the Irish diaspora, and their experience. We just focused on collecting the Irish experience. And we knew from our work with the IPC and our membership, that other people were doing the same so that they could fill in those gaps. And we also contributed to the IPCs own collection about COVID. So Ireland is represented in that kind of global collection. We just had to focus on Ireland and collect what we could collect, and look at that Irish experience, because ultimately, that's the remit of the National Library.

Would you like to make some additions?

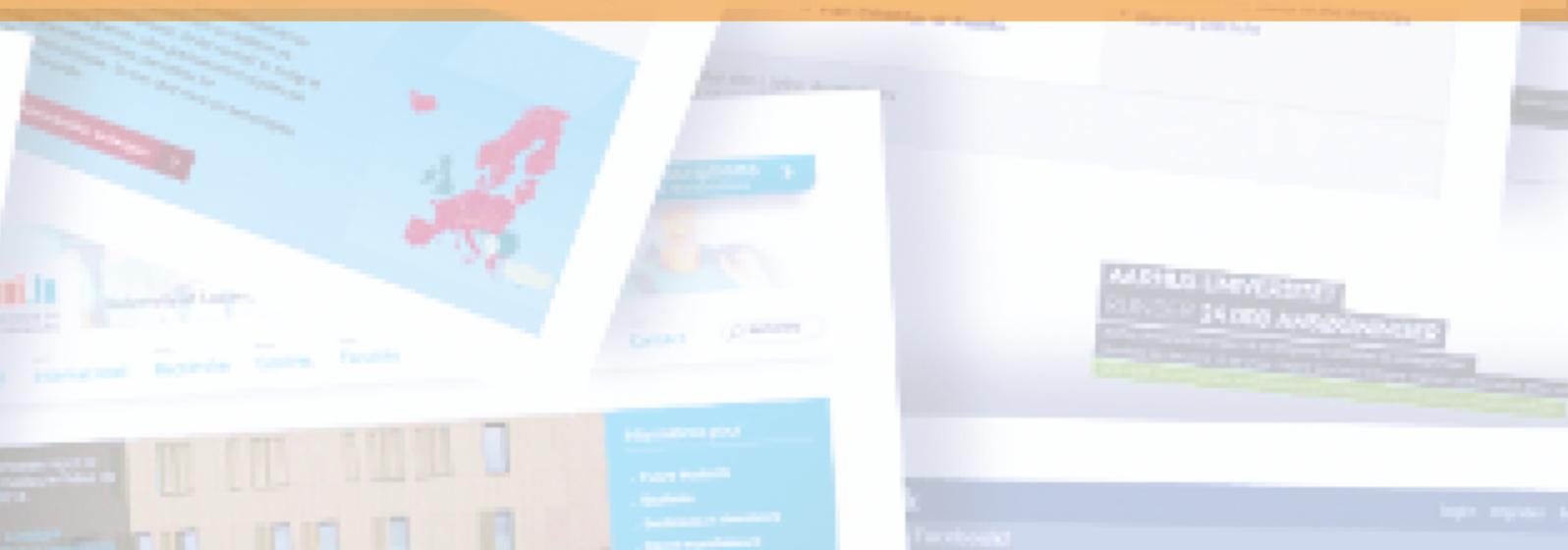
Joanna Finegan: There is a real drive to collect the Irish language. And the library has outstanding collections of published works in the Irish language, as evidenced by the extent of the library's holdings represented in *Clóliosta* (*Clóliosta: Printing in the Irish language, 1571–1871. An attempt at narrative bibliography by Richard Sharpe and Mícheál Hoyne*). And so that is something that we are really conscious of in relation to the unique Irish language content in the web archive, and we are looking to develop that further. That's obviously an incredibly distinctive part of our national culture as well. So we are very aware of that.

The other thing to say is in terms of accessibility, our online portal is provided by archive-it, and you can also search the NLI's online catalog, and it will return searches from archive-it as well. There's a unified search there through that general catalog, and that's returning a search across all the Library's collections. The diversity and variety of our collections, both published, and archival, physical and digital is represented in that. A really distinctive part of NLI collections is their breadth. To have the web archive and online record preserved and accessible via our catalogue along with the record of the pandemic through other formats such as books and photographs gives a more complete, rounded picture of the impact of COVID19 on our country.



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