Mapping the archival horizon: A comprehensive survey of COVID-19 web collections in European GLAM institutions

Nicola Bingham

Abstract: This chapter analyzes the COVID-19 web collections of Galleries, Libraries, Archives, and Museums (GLAM) across Europe. As the pandemic reshaped the digital landscape, these institutions have worked to archive web content from this unprecedented event. The study explores European GLAM web archiving efforts, focusing on the scope, methods, and challenges of curating COVID-19 content. It features a comparative analysis that highlights similarities and differences in collection strategies, emphasizing the implications for future digital preservation and access to pandemic-related content. This research offers insights valuable to information professionals and researchers studying the role of GLAM institutions in preserving key contemporary historical moments.

Keywords: COVID-19, web collections, digital preservation, GLAM institutions, web archiving

Introduction

Amid the COVID-19 pandemic, cultural heritage institutions embarked upon substantial endeavors to document this epochal period for posterity. In 2021, within the purview of the WARCNet project, a cohort of researchers delved into the dynamics of COVID-19 web archiving within the GLAM sector in Europe. In the summer of 2022, a comprehensive survey[1] was initiated to solicit participation from European counterparts in the GLAM sector, seeking insights into the breadth of their web collections and collection methodologies. The results of this study will be published in 2024. The focal point of the 2022 study encompassed three distinct respondent cohorts delineated by their respective countries of origin: the United Kingdom, Belgium, and other European nations. This segmentation was necessitated by the notably substantial response rates from the UK and Belgium, attributed to the distribution strategy employed for the survey. The current chapter integrates the survey's findings and extends its scope by assimilating an additional comprehensive literature review, thereby enriching the analysis of the gathered data.

The Survey

---

[1] A copy of the survey questions is held in the British Library Research Repository https://doi.org/10.23636/g5n0-pp30

Nicola Bingham, British Library, United Kingdom, nicola.bingham@bl.uk, 0000-0002-5510-9869

The questions included in the survey were informed by a hypothesis about the decisive factors that form web archives. In contrast to physical archives, archiving the web is a relatively novel endeavor characterized by technological advances and little international coherence. Local strategies are governed by specific national legislations and new tools for harvesting are continuously being developed as web archiving technology evolves. The survey aimed to probe the extent to which these differences have affected both rapid response and regular collection strategies regarding the COVID-19 crisis. By addressing questions on juridical and technological factors, we guide respondents' attention towards how these have affected collecting strategies.

The survey, conducted from June 13 to September 5, 2022, spanned approximately three months. It reached around 600 recipients via email addresses and distribution lists such as the Archives-NRA JISC mail and IIPC lists, compiled through the project team's existing contacts, network, and desk-based research. The team targeted individuals knowledgeable about collecting policy or involved in research within Museums, Libraries, Archives, and Galleries across Europe. Distribution favored the UK and Belgium due to team members' affiliations with National Libraries in those countries, resulting in contact with 99 organizations/individuals in the UK and 354 in Belgium. Additionally, efforts were made to engage European National Libraries and Archives affiliated with the International Internet Preservation Consortium (IIPC).

The survey was structured into sections focusing on respondent profiles, motivations behind collecting pandemic-related materials, collection scope, the contextualization of collection activities, access challenges, and creation hurdles. SNAP survey software was chosen to carry out the survey primarily because it was utilized by the British Library (BL) and offered secure data storage.

Respondents

The survey garnered 61 responses from 12 countries, resulting in a completion rate of roughly 10%. Factors contributing to the low response include the timing coinciding with university staff leave over the summer, ambiguity in identifying recipients in large organizations, and resource constraints in smaller ones. Additionally, the survey's title and focus may have deterred non-collecting organizations, leading some to perceive their input as redundant. Acknowledging the low response rate as a limitation underscores the need for a cautious interpretation of the results. To maintain respondent anonymity, countries with fewer than five responses were grouped under 'other European countries', with two unidentified

nationalities included in this category. Therefore, the analysis compares COVID-19 web archiving practices across three regions: the UK, Belgium, and other European countries, with 33% from the UK, 25% from Belgium, and 41% from other European countries among the respondents.

Motivation for collections

The motivation for initiating a special collection on COVID-19 was surveyed through the first two questions of the survey. Participants were first asked to report whether their institution had indeed curated a special collection of web materials about the COVID-19 crisis. The average percentage of respondents who initiated a COVID-19 collection across Europe was approximately 70.67%.

Respondents were asked to describe where the initiative to collect web materials related to the COVID-19 crisis had originated. Motivation was notably consistent across the UK, Belgium, and other European countries. In both the UK and Belgium, roughly half of the respondents indicated that the initiative came from the institution's staff (47% in the UK and 54% in Belgium). Approximately 30% of respondents from these countries mentioned that the impetus came from institutional management (32% in the UK and 31% in Belgium) and in other European countries, a majority (54%) also cited that the impetus predominantly originated from the staff. This similarity in responses across regions suggests a common trend where staff within institutions played a significant role in driving the initiative to collect web materials related to COVID-19.

Collaboration and partnerships

From the results of the survey and a review of additional literature, several insights emerge regarding collaborative partnerships between archiving institutions and external organizations in building COVID-19 web collections.

Collaboration for technical support was observed to varying degrees among institutions in Belgium, the UK, and other European countries in the survey. Belgian institutions showed relatively higher levels of technical collaboration. Chiara Zuanni, (2022) writing in the context of museums, identifies that the urgent need to collect digital items led to rapid developments in this area together with novel collaboration between memory organizations and digital specialists.

Several instances of participatory approaches have been observed, such as the National Library of Luxembourg's call for participation to integrate websites overlooked in the initial collecting phase (Schafer and Els, 2020). Furthermore, the Danish Web collection on coronavirus was part of a general project documenting coronavirus lockdowns in Denmark in 2020.

This effort was a cooperation between several cultural institutions, the National Archives (Rigsarkivet), the National Museum (Nationalmuseet), the Workers Museum (Arbejdermuseet), local archives, and the Royal Danish Library. Also in Denmark, the 'Days with Corona' initiative from the Danish Folklore Archives run by the National Museum and the Royal Library called on citizens to contribute photos, narratives, and websites for inclusion in the web archive, showcasing a broader engagement strategy involving cultural institutions, local archives, and libraries. Netarchive also asked the public for help by nominating URLs of web pages related to coronavirus, social media profiles, hashtags, memes, and any other relevant material (Schostag 2020).

The experience gained from COVID-19 collections has led to plans for future collaborations and thematic collections. For example, the National Library of Luxembourg aims to expand partnerships with the Government IT Centre, the CNA (National Audiovisual Centre) and the CNL on Luxembourg authors and publishers and other cultural institutions to deepen thematic collections. (Schafer and Els 2020)

Overall, these insights demonstrate the varying degrees of collaboration, outreach strategies, and participatory efforts employed across different countries in the context of COVID-19 collection initiatives. These approaches not only enrich the collections but also pave the way for future collaborations and more inclusive, engaging archival practices.

Platforms like the Internet Archive (IA) and the UK Web Archive offer features enabling users to nominate websites, acknowledging that many sites are missed simply because archivists are unaware of them. While participation enriches archives, it remains somewhat sporadic and mainly reaches those already digitally engaged. Efforts for broader engagement are expected to bring different participation forms and contribute to more inclusive collections.

Cui Cui et al. (2023) noted in their conference presentation that participatory web archiving redistributes power among stakeholders e.g., community partners, creators, and users, aiming to overcome limitations in traditional archiving. However, the participatory approach raises questions about effectiveness, mechanisms, and impacts.

Underrepresented groups

As Schafer and Winters (2021) point out, in terms of minorities, and marginalized populations affected by the pandemic, "Web archives include information for and about diverse groups in society and hold out the promise of preserving the voices of individuals who in previous centuries would only have featured in an archive if they interacted with the church, central government, or the law."

The survey highlighted several examples of the various archiving institutions' attempts to contact underrepresented groups, such as engaging mental health charities, using social media, contacting partner institutions, and utilizing internal networks (e.g., LGBTQ staff networks) to connect with underrepresented communities during the creation of COVID-19-related web archives (Geeraert and Bingham 2020).

Religious communities have been the focus of some collecting efforts, for example, the Marian Library, University of Dayton, Ohio built a collection which reflected how the pandemic affected underrepresented and underserved communities of Catholics. The collection gives a glimpse of how the pandemic has affected the faith and religious practice of 'ordinary' Catholics (Harris et al. 2023).

Another example was provided by Ben Els at the BnL, who initially looked at a Muslim community website (shoura.lu) which contained information and recommendations for its community about services in mosques, religious holidays, etc. BnL then expanded its collecting strategy to include other religious communities. Ben Els pointed out "It is important to capture the minority viewpoint as well and I have also included sites for border residents, for example, the site frontaliers.lu, because they face different problems from residents." (Schafer and Els 2020)

There was increased global attention on equality, diversity, and inclusion during 2020 due to the pandemic's spotlight on inequalities and the rise of movements like Black Lives Matter. Initiatives focused on the challenges faced by communities during events like racial injustice protests, struggles within the LGBTQ+ community, and instances of anti-Asian racism. Archivists prioritized capturing these narratives to ensure a more accurate representation of ongoing historical events and discussions (Greenwood 2022).

Meanwhile, Schafer and Winters (2021) noted controversies in web archiving, citing disputes over ethical concerns, such as the collection of sensitive content during events like the George Floyd protests. These debates prompted calls for more inclusive archiving policies, advocating for broader representation, particularly of Black perspectives, underscoring the growing complexity and ethical considerations in archival practices. Schafer and Winters highlight instances of disputes, including lawsuits and debates between entities like the Internet Archive (IA) and editors. One example involves clashes between the IA and Doc Now during the George Floyd protests, where concerns about risks to protesters' safety were raised against the collecting effort, reflecting a divergence of opinions on the ethical aspects of archiving sensitive content.

Scope

Regarding the level of archiving, the results from the GLAM survey indicate a correlation between the level at which institutions operate and the focus of their web archiving efforts. In the UK, a majority of respondents collected at a national scale, which aligns with a significant portion of these institutions operating at a national level. In Belgium, a higher percentage of institutions operated at a regional level, which coincided with more respondents indicating collection efforts at the regional level.

Smaller web archives, like those operated by universities or individual scholars, may lack resources and infrastructure for independent hosting and preservation. Consequently, they often link their content to larger initiatives like the Internet Archive to avoid issues with copyright, data storage, and preservation, enabling their content to be preserved and accessible through such extensive platforms (Priem and Grosvenor 2022).

In many cases specialist collections were built adhering to the institutional scope for example the University of Dayton's collection on Marian Devotion (Harris et al. 2021). However, some web archives are not under the control of libraries, such as arquivo.pt, the Portuguese Web Archive, which is linked to the Portuguese national research and education network.

There is evidence that web collections vary based on specific national or local events. For instance, the Royal Danish Library coronavirus web collection includes web activity related to Queen Margarethe II's 80th birthday celebration during the pandemic (Schostag 2020). This example underscores that local events unique to different nations significantly shape the content and focus of individual web collections. It prompts the question of whether similar examples from other countries, like the UK's 'eat out to help out' initiative, would also impact web collections in distinct ways.

In some cases, Web archives have structured their COVID-19 collections within a broader context of historical and potential future pandemics. However, the selection process has been somewhat subjective, often excluding recent pandemics like swine flu or HIV/AIDS (Priem and Grosvenor 2022). Some initiatives such as the UK Web Archive have organized COVID-19 collections within a larger category covering other pandemics, potentially indicating a placeholder for future pandemic-related collections (Geerart and Bingham 2020).

Similarly, the ISCHE Education & Pandemics Archive acknowledges the broader context of historical pandemics and epidemics, highlighting their significance in local, national, and transnational education histories. These initiatives aim to recognize the 'pandemic century' and the multitude of disruptions and challenges, reshaping historical perspectives that traditionally prioritize human-centric narratives and control (Priem and Grosvenor 2022).

Type of content collected

The survey investigated preferences in collecting social media and website content, revealing distinct tendencies between Belgium and the UK. In Belgium, Facebook tops the list as the most frequently collected social media platform, followed by Twitter, Instagram, and YouTube. Conversely, the UK prioritizes Twitter, blogs, university websites, and non-profit websites. These variations might be attributed to technological limitations, for example, Heritrix, a widely used crawler in UK web archives, faces challenges in capturing Facebook content whereas Belgium's preference for Facebook archiving could stem from the Flemish Institute for Archives research project, meemoo, on social media archiving practices: "Best practices for archiving social media in Flanders and Brussels" (meemoo 2023).

These differences in collection preferences highlight varying priorities and technological constraints between different web archiving institutions. Schostag explains that at the Royal Danish Library "More or less successfully, we tried to capture content from Facebook, Tik-Tok, twitter, youtube, instagram, reddit, imgur, soundcloud, and pinterest. Twitter is the platform we are able to crawl with Heritrix with rather good results. We collect Facebook profiles with an account at Archive-It, as they have a better set of tools for capturing Facebook. With frequent Quality Assurance and follow-ups, we also get rather good results from Instagram, TikTok, and Reddit. […]As Heritrix has problems with dynamic web content and streaming, we also used Webrecorder.io […]. However, captures with Webrecorder.io are only drops in the ocean." (Schostag 2020).

The National Library of Luxembourg (BnL) focused primarily on archiving websites due to the significant technical challenges and higher costs associated with archiving social media platforms. While they included some Facebook pages in their collection, they prioritized news media, websites, and Twitter due to their higher effectiveness in archiving compared to other social media platforms (Schafer and Els 2020).

Sub-categories

The survey asked respondents to comment on the criteria for selection in Covid-19 web collections from the following options; 'based on language(s)', 'top-level domain', 'subject/theme of web content', 'nationality of the creators/owners', 'postal addresses mentioned in the web content', 'specific hashtags', and 'specific social media profiles'.

Belgium and the UK primarily relied on the subject or theme of the web content, nationality of creators/owners, and specific social media profiles for collection scoping. In Belgium, no respondents indicated that the top-

level domain or the postal addresses mentioned in the web content were determining criteria and in the UK none of the respondents indicated that language was used as a selection criterion. This can be partly explained by the fact that there is no top-level domain crawl being done in Belgium as KBR, the Royal Library of Belgium, is waiting for a change in the legal deposit legislation that would make it possible. Overall, the survey indicated that the most common way to scope COVID-19 collections among all the survey respondents was based on the subject/theme of the web content.

News content was a priority, however, it was handled differently according to the collecting institution. At the National Library of Luxembourg between mid-March and mid-July, for example, curators selected individual articles from Luxembourg news websites constituting a collection of just over 26,000 articles. From the beginning of June, they requested a daily capture of each news website from the Internet Archive, which had previously been undertaken only twice a year (Schafer and Els 2020).

In contrast, Schostag (2020) explains that the Danish Web Archive crawls all Danish news media from several times daily to once weekly, so felt there was no need to curate individual news articles in the pandemic event crawl. Rather, they focused on augmented activity on social media, blog articles, new sites emerging in connection to the event, and reactions in news media outside Denmark.

Curators building web archives about the COVID-19 pandemic responded dynamically to unfolding events, employing inclusive approaches in content selection. They focused on various pandemic-related topics such as medical care, government communications, lockdown effects, remote work, and cultural activities. Specific themes were identified in Luxembourg, Belgium, and the UK, encompassing liberalism, solidarity, inequality, and more (Schafer and Els 2020).

The collection scopes varied, including creator-based and local collections, and selection methodologies ranged from keyword searches to open calls for participation. The curated subtopics extended beyond health effects to societal concerns like changing political and economic landscapes, threats to food security, misinformation spread, and governmental exploitation of the crisis. However, some collections, like Cornell University's focus on international labor challenges, leaned more towards an economic perspective and lacked a direct human narrative (Greenwood 2022).

Temporal span

The survey revealed that the temporal span of COVID-19 web collections varied across institutions and countries. Most collections began

their gathering process in March 2020, aligning with the imposition of lockdown measures in Belgium and the UK. Other European countries started slightly earlier, primarily in the first half of March 2020. However, the Danish web archive commenced its efforts about six weeks before the virus reached Denmark, prompted by global reactions to a controversial cartoon. As stated by Schostag (2020), "In a sense, the story of Corona and the national Danish Web Archive (Netarchive) starts at the end of January 2020 – about 6 weeks before Corona came to Denmark. A cartoon by Niels Bo Bojesens in the Danish newspaper 'Jyllandsposten' (26 Jan 2020) showing the Chinese flag with a circle of yellow corona-viruses instead of the stars caused indignation in China and captured attention worldwide. We focused on collecting reactions on different social media and in the international news media. Particularly on Twitter, a seething discussion arose with vehement comments and memes about Denmark."

Regarding the end dates for collection efforts, by the time of the survey conducted between June and September 2022, a quarter of UK respondents and almost half of the Belgian respondents had ceased collecting. This cessation was due to various factors, such as diminished interest, staff reallocation, completion of projects, the end of pandemic-related measures, lack of relevant online content, and technical difficulties in capturing specific social media.

The decision to halt collections was not universally aligned and was often influenced by practical considerations, like budgetary constraints. At the National Library of Luxembourg, for example, Ben Els mentions having ample storage capacity (terabytes), but expresses reluctance to allocate an additional 5 terabytes due to limitations. He had hoped for a slowdown in June but observed continued activity, maintaining the pace of collection similar to that of March. He notes comparing the number of articles collected from media websites to the recorded number of COVID-19 cases as part of their collection analysis (Schafer and Els 2020). The unpredictability of the pandemic's trajectory made it challenging for institutions to definitively determine when to conclude their collection activities, leading to varied responses among institutions.

Some institutions expressed intentions to cease collection efforts once the World Health Organization declared an end to the pandemic. The relationship between these intentions and actual collection cessation could provide valuable insights into the dynamics between collecting practices and the perceived trajectory of the pandemic.

The Danish Royal Library emphasizes the importance of broad, domain-scale archiving to capture a comprehensive view. Their timing for domain crawling during the Danish Corona lockdown, starting on March 14, proved fortuitous. These broad crawls, conducted up to four times a year, capture all Danish-related content across domains. This approach allows for the

inclusion of Corona-related content that curators might not identify using their usual methods, utilizing keyword searches and link scraping tools (Schostag 2020).

Technology and quality assurance

The survey focused on the technologies and quality assurance methods used in web content harvesting. Different tools were employed across countries, with Belgium favoring Webrecorder and Chrome SingleFile, while the UK used Webrecorder and Heritrix. Larger institutions preferred Heritrix for scalability, while smaller entities chose user-friendly tools like Webrecorder, aligning with their specific archiving needs. Quality control primarily involved visual inspection, with some using patch crawling or specific methods.

Webrecorder emerged as the most popular tool across surveyed countries. Organizational tool choices vary due to complex factors, leading some to utilize combinations of tools. Large-scale initiatives opt for robust, scalable tools like Heritrix, capable of capturing extensive content, while smaller projects prefer simpler tools like Webrecorder, requiring less technical expertise.

The survey identified a long tail of less frequently mentioned tools, often open-source and straightforward. For instance, Chrome SingleFile, a browser extension for archiving complete web pages in a single HTML file, appeals to smaller organizations for its offline viewing capabilities. Some smaller entities archive web content without specific legal mandates, utilizing tools like Instaloader for Instagram, operating within platform terms of service and legal compliance.

Access

The landscape of access to web archive collections varies according to country and institution. For example, the access restrictions imposed due to personal data protection laws, as seen in the Danish case, limit public access to COVID-19 web archives. In Denmark, only researchers affiliated with specific institutions can apply for access related to particular research projects. Although one project on values in COVID-19 communication has already been initiated (Schostag 2020).

A common challenge across Belgium, the UK, and other European countries is the underutilization of these special collections for research and teaching. Responses from institutions in these countries indicated that the collections were rarely used or that there was uncertainty about their usage. This lack of insight into collection usage suggests that either these

collections are underused or that there is a gap in understanding their potential.

Few institutions in Belgium or the UK provided data-level access or offsite access, whereas some other European countries did so. Additionally, sharing derived datasets or seed lists with the public was rare in the UK and Belgium, but more common in other European countries. Restrictions on access were often linked to access being limited to the institution's reading room, with additional limitations surrounding intellectual property rights, personal data protection, and legal deposit legislation.

Despite accessibility challenges, efforts were made to promote access. For instance, the Luxembourg National Library (BnL) provided information and starting points on its website and intended to offer interactive statistics and metadata, such as title and keywords, to aid users in navigating the collection (Schafer and Els 2020).

Ben Els (2020) stated that the archive is accessible at the BnL, but on the webarchive.lu website they also try to give information and starting points, to outline the background of the collection and explain the processes to interested people. This implies that interest from academic partnerships in accessing and analyzing COVID-19 collections exists, with the mention of WARCnet and an academic team being the first on the horizon.

Schafer and Winters (2021) highlight the complex interplay between ethical practices, legal frameworks, and protection in web archiving. In many countries, web archiving is regulated by legal deposit and copyright laws. Some countries have explicit laws addressing the archiving of digital content, enabling collection and preservation but potentially restricting access. However, legal frameworks often lag behind technological advancements, posing challenges to good governance criteria. Restrictions on access might lead researchers and the public to rely solely on platforms like the Internet Archive (IA), missing out on more comprehensive collections held by other archival institutions. The IA's copyright policy, allowing unrestricted access unless there are take-down requests, contrasts with legal limitations in other archives. Balancing citizen safeguards with increased access could foster innovation and experimentation in web archiving. This underscores the need for a balance between protecting individual rights and enhancing accessibility and innovation in web archiving practices.

In her conference presentation, Alice Austin (2023) delineated the strategic approach adopted by the Archive of Tomorrow project in addressing the complexities associated with the collection, organization, and accessibility of disputed or obsolete medical information for researchers. The deliberate inclusion of data sourced from dissenting or contentious entities, juxtaposed with content from established authoritative sources, is poised to facilitate forthcoming investigations into the dynamic interplay

between health discourse and the online sphere. Moreover, there is an anticipation that this curated collection will serve as a pivotal experimental ground for scrutinizing the methodologies involved in crafting, administering, and leveraging archived web resources for scholarly inquiry.

Research

According to Priem and Grosvenor (2022), the COVID-19 pandemic has not only sparked a renewed interest in history, it has also focused attention on how the present can be historically preserved. Therefore, it is safe to predict that the COVID-19 crisis and its documentation will be analyzed by future historians, and it will bring about methodological and technological changes that affect our ways of working as historians of education.

Piguet and Montebello (2020) point out that in response to the global Covid-19 pandemic and its impact on societal norms and individual freedoms due to widespread lockdowns, historians, sociologists, and archivists are urging individuals, institutions, and public authorities to document this unprecedented event. They emphasize the importance of archiving this exceptional period, aiming to uncover the overlooked aspects of daily life, often invisible but integral to human societies. The analysis of these archives is expected to illuminate lesser-known lives and contribute to various historiographical fields, the history of solidarities and social policies, the history of public policy, or the history of epidemics. The ultimate goal is to create an inclusive and civic memory of the pandemic, fostering collaboration between potential donors and archives globally.

There was a surge in content creation initiatives during the pandemic, as illustrated by the Made By Us project that mapped over 450 COVID-19 crowdsourcing projects globally (Priem and Grosvenor 2022). These initiatives encouraged the public to self-document and share their pandemic experiences, by posting images, audio files, texts, videos, and other content online. Marta Severo et al. (2023) describe how web archives and cultural institutions actively engaged in memorialization and historicization during the pandemic. Severo identifies two primary types of collections: one capturing born-digital data like tweets and online videos, and another seeking contributions and observations from the public, like photos and stories, to build a memory of this unprecedented time. Initiatives like 'Vitrines En Confinement' ('Windows on Confinement') aim to systematically collect and document lockdown experiences from public spaces.

In some respects, content created during the pandemic serves to project an awareness of history in the making. This content portrays individuals as protagonists in future narratives, akin to the way World War narratives are constructed, mythologized, and remembered. The pandemic is framed as a

critical historical moment, prompting GLAM organizations and various institutions worldwide to create and collect histories related to individual and community responses to COVID-19 (Priem and Grosvenor 2022).

Acker and Chaiet (2020) investigated how web archives were being "weaponized to propagate and preserve health misinformation circulating on platforms like Facebook and Twitter" as a result of the pandemic. They explain that typically most social media content has been gathered in the aggregate on a macro-level basis allowing researchers to understand the pandemic on a large scale, but that the individual's experience, or the micro-level analysis, should be considered as well.

Challenges

Survey respondents were asked what they deemed to be the most important challenges for the COVID-19 special collection. With regard to working from home, this was noted as a challenge for some web archiving institutions, however there is evidence that for others remote working provided an opportunity to devote more time to web archiving projects. Employees at the University of Dayton, for example, began remote working in March 2020 and were able to capture websites related to their library's collection development policy (Harris et al. 2021).

Technological adaptations facilitated work from home, with platforms like Archive-It supporting remote access. For example, Ben Els (2020) stated: "Video conferencing worked well. Our team is very small and for crawls the Archive-It platform is accessible from home. A lot of things that seemed impossible before have now become possible."

Quick policy changes and adaptations were crucial in response to the pandemic's onset, altering appraisal methods, prioritizing digital records, and emphasizing metadata creation and web archiving tools. Despite such shifts, managing increased workloads, securing permissions, avoiding duplication, and handling the enormity of collected material proved challenging. Respondents highlighted the emotional and time-consuming nature of the project, especially in capturing transient websites, requiring a balance between project demands and existing responsibilities. Additionally, while acknowledging the project's importance, respondents emphasized the need for increased investment in web archives to enhance responsiveness, especially considering limited resources and small team sizes in national memory institutions. The evolving situation led to reflections on the effectiveness of adding COVID-19-related content to existing collections, considering user experience.

Other challenges cited included; ensuring that work is not duplicated, making the right choices under the stress of time, data cleaning, training and coding in the midst of a pandemic, not having a platform to share

methodologies with other researchers, the long duration of the pandemic, the sheer size of the material, and ensuring the needs of the COVID-19 collections did not supersede the institutions' other existing web archiving priorities (Greenwood 2022).

Schafer and Winters highlight that the level of responsiveness demonstrated during the pandemic was impressive, but even more remarkable given the small size of the teams involved and the relatively low level of funding provided to web archives of all kinds. The web archives based in national memory institutions are usually staffed by a handful of people, who are competing internally for limited resources. Increased investment in web archives, as custodians of the world's digital cultural heritage, is essential if greater responsiveness is to be delivered.

Finally, rapid response archiving is time-consuming, and can also be emotionally consuming, especially considering the feelings of urgency that come when trying to capture and preserve very ephemeral websites. Archivists must balance the demands of this project with their other work responsibilities at a time when there is already added stress in dealing with the pandemic, personally and professionally (Harris et al. 2021).

Conclusion

This chapter undertook a comprehensive examination of the COVID-19 web collections curated by Galleries, Libraries, Archives, and Museums (GLAM) institutions in Europe. The research drew upon a comprehensive survey initiated by WARCnet researchers, aiming to engage European GLAM counterparts to gather insights into the breadth and methodologies employed in their COVID-19 web collections. Additionally, a literature review complemented the survey results to offer a more comprehensive understanding of the impact of COVID-19-related collecting among GLAM organizations in Europe and beyond.

The survey delved into the initiatives of European GLAM institutions in web archiving, scrutinizing the extent, methodologies, and challenges inherent in curating COVID-19 web collections. Through comparative analysis, this chapter illuminated both the commonalities and disparities in collection development strategies, thereby underscoring broader implications for the future of preserving and accessing pandemic-related web content. This research may provide insights for information professionals and researchers interested in the evolving role of GLAM institutions in documenting and preserving significant contemporary historical events.

However, it should be borne in mind that the survey data may carry potential biases due to specific factors in its distribution. The survey was predominantly circulated among individuals/organizations in the UK and

Belgium, influenced by the project team's proximity and networks. This skewed distribution, particularly with a higher representation of institutions in Belgium and the UK compared to other European nations, has limited the generalizability of findings.

Moreover, it is crucial to note that the survey was conducted during the pandemic, potentially capturing evolving collections in 2022, requiring organizations to adapt to unfolding events and changing priorities. Therefore, this analysis should be considered as reflecting collections in progress rather than finalized.

Despite these limitations, the survey in conjunction with a wider review of the available literature enables us to make several useful conclusions.

The insights gained underscore the diverse approaches, outreach strategies, and participatory efforts implemented across various countries during COVID-19 collection initiatives. These endeavors not only enriched the collections but also laid the groundwork for future collaborations, fostering more inclusive and engaging archival practices. While participatory web archiving aims to redistribute power among stakeholders and overcome traditional archiving limitations, questions remain about its effectiveness, mechanisms, and impacts. Despite efforts to engage users in nominating websites and contributing to archives, participation tends to reach digitally engaged individuals primarily. Broader engagement strategies are anticipated to diversify participation forms and contribute to more comprehensive and inclusive collections.

The research highlights the significance of web archives in preserving voices historically underrepresented in traditional archives. During the COVID-19 pandemic, archiving institutions made efforts to engage marginalized communities, utilizing diverse methods like social media, partner contacts, and internal networks to ensure representation in COVID-19-related web archives.

The pandemic's focus on inequalities led to increased attention on equality, diversity, and inclusion globally. Archivists prioritized documenting events like racial injustice protests, LGBTQ+ struggles, and instances of discrimination, aiming for a more accurate historical representation. However, these discussions emphasized the need for more inclusive archiving policies, particularly to represent diverse perspectives and navigate ethical dilemmas in archival practices.

The access landscape for web archive collections varies across countries and institutions, evident in cases like Denmark, where personal data protection laws restrict public access, limiting it to researchers affiliated with specific institutions and particular research projects. Similar challenges exist across Belgium, the UK, and other European nations, with collections being underutilized for research and teaching purposes, indicating a potential gap in understanding their value. Despite challenges, interest from

academic partnerships in accessing and analyzing COVID-19 collections exists, exemplified by initiatives like WARCnet. However, legal frameworks governing web archiving, often tied to legal deposit and copyright laws, present complexities.

This survey's results might inspire researchers to explore the extensive web archive collections available throughout Europe, particularly those related to COVID-19. By acknowledging biases, understanding diverse approaches, and recognizing limitations, archiving institutions and researchers can leverage and contribute to these valuable collections, enriching the knowledge surrounding COVID-19 and its societal impacts.

References

Acker, Amelia, and Chaiet, Mitch. 2020. "The Weaponization of Web Archives: Data Craft and COVID-19 Publics." *Harvard Kennedy School (HKS) Misinformation Review* 1, no. 3: 2. Accessed December 31, 2020. https://doi.org/10.37016/mr-2020-41.

Austin, Alice. 2023. "Bridging the Gap: Capturing UK trans health discourse in the Archive of Tomorrow." Abstract for the RESAW Conference 2023. https://web.archive.org/web/20240503135419/https://resaw2023.sciencesconf.org/4347 51/

Bingham, Nicola, Karin de Wild, Friedel Geeraert, and Caroline Nyvang. "Surveying the Landscape of COVID-19 Web Collections in European GLAM Institutions." In *The Routledge Companion to Transnational Web Archive Studies*, edited by Niels Brügger. Publisher: Taylor & Francis (Forthcoming).

Cui, Cui, Stephen Pienfield, Andrew Cox, Frank Hopfgartner. 2023. "Participatory Web Archiving: The path towards more inclusive web archives?" Abstract for the RESAW Conference 2023.
https://web.archive.org/web/20240503135603/https://resaw2023.sciencesconf.org/4335 45/

Geeraert, Friedel, and Nicola Bingham. 2020. "Exploring special web archives collections related to COVID-19: The case of the IIPC Collaborative collection, An interview with Nicola Bingham (British Library) conducted by Friedel Geeraert (KBR)." *WARCnet Papers*. Aarhus: WARCnet.
https://web.archive.org/web/20240503135745/https://cc.au.dk/fileadmin/user_upload/W ARCnet/Geeraert_et_al_COVID-19_IIPC__1_.pdf

Geeraert, Friedel, and Nicola Bingham. 2020. "Exploring special web archives collections related to COVID-19: The case of the UK Web Archive, An interview with Nicola Bingham (British Library) conducted by Friedel Geeraert (KBR)." *WARCnet Papers*. Aarhus: WARCnet.
https://web.archive.org/web/20240503140007/https://cc.au.dk/fileadmin/user_upload/W ARCnet/Geeraert_et_al_COVID-19_UKWA__1_.pdf

Greenwood, Amanda. 2022. "Archiving COVID-19: A Historical Literature Review." *The American Archivist* 85, no. 1: 288–311. Accessed October 12, 2023. URL: https://web.archive.org/web/20240503140409/https://meridian.allenpress.com/american -archivist/article/85/1/288/486596/Archiving-COVID-19-A-Historical-Literature-Review.

Harris, Kayla, and Stephanie Shreffler. 2021. "Archiving Catholic Faith on the Web During

the COVID-19 Pandemic." University of Dayton eCommons Marian Library Faculty Publications, March 2021. Accessed November 5, 2023. https://web.archive.org/web/20240503140756/https://ecommons.udayton.edu/imri_fac_pub/206.

KBR. "PROMISE Project." Accessed September 12, 2023. https://web.archive.org/web/20240503141227/https://www.kbr.be/en/projects/promise-project/.

KBR. "BESOCIAL." Accessed September 12, 2023. URL: https://web.archive.org/web/20240503141538/https://www.kbr.be/en/projects/besocial/.

meemoo. "Best Practices for Social Media Archiving in Flanders and Brussels." 2023. Accessed October 14, 2023. URL: https://web.archive.org/web/20240503141733/https://meemoo.be/nl/projecten/best-practices-voor-de-archivering-van-sociale-media-in-vlaanderen-en-brussel.

Piguet, Myriam, and Caroline Montebello. 2020. "Covid-19 : pour une mémoire ordinaire de l'extraordinaire." *Libération*. Accessed October 01, 2023. URL: https://web.archive.org/web/20240503141951/https://www.liberation.fr/debats/2020/04/25/covid-19-pour-une-memoire-ordinaire-de-l-extraordinaire_1786299/

Priem, Karin, and Ian Grosvenor. 2022. "Future Pasts: Web Archives and Public History as Challenges for Historians of Education in Times of COVID-19." In *Exhibiting the Past. Public Histories of Education*, edited by Frederik Herman, Sjaak Braster, and Maria del Mar del Pozo Andrés, 177–96. Berlin, Boston: De Gruyter Oldenbourg. https://doi.org/10.1515/9783110719871-009

Schafer, Valérie, and Ben Els. 2020. "Exploring special web archive collections related to COVID-19: The case of the BnL An interview with Ben Els (BnL) conducted by Valérie Schafer (C2DH, University of Luxembourg)." *WARCnet Papers*. Aarhus: WARCnet https://web.archive.org/web/20240503142044/https://cc.au.dk/fileadmin/user_upload/WARCnet/Schafer_et_al_COVID-19_BnL.pdf

Schafer, Valérie, and Jane Winters. 2021. "The values of web archives." *International Journal of Digital Humanities*, 2, 129–44.

Schostag, Sabine. 2020. "The Danish Coronavirus web collection – Coronavirus on the curators' minds." *International Internet Preservation Consortium Blog*, July 29, 2020. https://web.archive.org/web/20240503142151/https://netpreserveblog.wordpress.com/2020/07/29/the-danish-coronavirus-web-collection/

Severo, Marta, Sarah Gensburger, and Louis Gabrysiak. 2023. "Lockdown collections and web archives: cross-exploration of BNF and INA archives." Abstract for the RESAW Conference 2023. https://web.archive.org/web/20240503142430/https://resaw2023.sciencesconf.org/435367

Zuanni, Chiara. 2022. "Contemporary Collecting in a Pandemic: Challenges and Solutions for Documenting the COVID-19 Pandemic in Memory Organizations." *Heritage* 5, no. 4: 3616–3627.