Proposing methods to explore the evolution of the term ‘mHealth’ on the Danish Web archive
by Antoinette Fage-Butler, Loni Ledderer, and Niels Brügger

Abstract
This article uses Internet archives to explore the emergence and spread of the term ‘mHealth’ (mobile health technologies) in the Danish Web domain from 2006 to 2018, focusing on the actors that contributed to its evolution. We propose three methods for investigating the Web pages and Web sites that employed the term ‘mHealth’. Our findings highlight temporal developments in the use of ‘mHealth’, with diverse actors using it, though none clearly dominated. The article attends to challenges in working with Web archive data, and presents methods that can be used by others wishing to engage empirically with Internet archives, which remain vast, but largely under-exploited resources.

Contents
Introduction
Methodology
Findings
Discussion
Conclusion

Introduction
Web archives preserve versions of Internet material. In recent years, Web archives have started to be used as source material for Internet studies, particularly within the larger field of Web history (Brügger and Milligan, 2019; Brügger and Schroeder, 2017; Brügger and Laursen, 2019). Web archives have been used to study historical developments within a range of areas, such as advertising, news media, immigration, religion, politics, business, and social media [1]. However, as the field of Web archive research is still relatively young, and due to the complexity of Web archive data and systems, there remains the methodological challenge of developing adequate methods to unlock the archived Web as a source with its own particular features and as empirical data (Brügger, et al., 2020).

Despite the challenges involved, it is important to develop methods to explore Web archive data as our knowledge of the past is increasingly and nearly exclusively stored online. Non-digital media like print newspapers, documents, film, radio and television are being digitised, born-digital media like the Web and social media continue to grow, and collections of born-digital media such as Web archives continue to expand as they compile all previous instances of born-digital media (Brügger, 2016a).
Proposing methods to explore the evolution of the term 'mHealth' on the Danish Web archive

A media-historical approach that draws on Internet archives has been recognised as highly valuable for the field of public health, particularly with respect to its recent history (Evans, 2018), as much public health communication appears in socioculturally significant documents that are uploaded on the Internet (Baguia, 2020; Lupton, 2020). This article focuses on the term ‘mHealth’ (mobile health technologies). mHealth characterizes the use of digital apps in health care. It has been defined as ‘medical and public health practice supported by mobile devices’ by the World Health Organization (WHO) [2], and incorporates technologies (Lucivero and Jongsm, 2018) that are financially and politically supported by industry, health professionals and policy-makers (Petersen, 2019). mHealth can be considered a subset of eHealth which in turn is a subset of the larger framework of digital health (WHO, 2019, 2011). It has been increasingly associated with health apps (Estrin and Sim, 2010), but it is not used exclusively when referring to apps, as the term ‘eHealth apps’ also exists (Danish Ministry of Health, 2018). mHealth has been seen as complementary to usual healthcare provision (WHO, 2011), as a new paradigm for evidence generation in healthcare (Kumar, et al., 2013), and as potentially revolutionising healthcare (Lucivero and Jongsm, 2018), with the digitisation of healthcare systems often characterised as an inevitability (Danish Ministry of Health, 2018).

The apparent flux and inconsistent terminology just described reflect the field’s relative immaturity, and the fact that meanings of ‘mHealth’ evolve to keep up with input from various fields, including technology-driven developments. The evolution of mHealth as a new communication technology has been shaped by industry, policy-makers and a host of other players that publish online (Lucivero and Jongsm, 2018). To date, research has focused mainly on the benefits and the drawbacks of health-related smartphone technology, particularly with respect to the efficacy of mHealth technologies for managing chronic conditions [3]. This prioritization has resulted in large research lacunae on social, cultural and policy-related aspects of mHealth media (Ledderer, et al., 2019).

One way of investigating mHealth as a mediatized phenomenon is to track the evolution of its online presence. This is highly relevant for two main reasons. First, a Web-based approach makes it possible to capture ‘mHealth’ in its embeddedness on the Web. Second, like other branches of smartphone technology, mHealth is a dynamic object that has evolved over time. It has been highlighted and promoted by players that have used the term on the Web — and Web archives have the great benefit of documenting a terms salience through online mentions over time. As the term is of recent origin, it is possible to track the complete evolution of the term ‘mHealth’ over time in Web archive data. Our explorative approach of including all data on the historical Danish Web that mentioned ‘mHealth’ in our analysis means that the phenomenon revealed itself through its online presence rather than in relation to pre-selected criteria. In that sense, we apply a genealogical approach to historical analysis (Foucault, 1984), capturing messiness and not assuming any grand purpose or teleology, as highlighted by Brügger (2016b) who described the value of inductive and inclusive approaches in Web archive research.

Denmark is an interesting case to explore the term ‘mHealth’, as it has one of the most digitised healthcare services in the world (Danish Ministry of Health, 2018). Recent statistics indicate that Denmark was in the top three countries for individuals using the Internet in the years that were studied (World Bank, 2021). In 2015, Denmark was deemed the country with the greatest market readiness for mHealth business (Research2Guidance, 2021), and Denmark has promoted itself as a digital health hub (Danish Ministry of Health, 2018).

The overall aim of our study is to explore the emergence and evolution of the term ‘mHealth’ in the Danish Web domain, identifying the actors that were part of its evolution. In so doing, we explore the communication about one digital medium (mHealth) in another (Web archives). Our aim is operationalised through three research questions. First, when and how often did the term ‘mHealth’ occur on the Danish Web? Second, which actors (Web sites) on the Danish Web used ‘mHealth’ the most, and how did this evolve? Third, who among these actors (Web sites) were the most important, and how did that change over time? In line with Brügger [4], we track the life of one individual Web element — the term ‘mHealth’ — across the Web sphere constituted by the discursive Web space related to mHealth on the Danish Web from 2006 to 2018.
Methodology

Data from the national Danish Web archive

Our data come from the national Danish Web archive called Netarkivet (for further information on Netarkivet, see Laursen and Møldrup-Dalum, 2017). Netarkivet was established in 2005. It collects and preserves all Web sites on the national Danish top-level domain .dk, from approximately 600,000 to 1.2 million Web sites between 2005 and 2015 [5]. We were interested in tracking diachronic (temporal) changes to capture the evolving occurrence of the term ‘mHealth’ over time on the Web and any changes or stability in the actors who referred to it. As pointed out by Brügger [6], ‘a temporal dimension with focus on developments and trends has to be added to studies of a nation’s Web, just as an archived Web has to be at the core of the study’.

The data set was prepared for analysis using the following steps. First, an explorative free text search in Netarkivet’s Wayback Machine confirmed that the term ‘mHealth’ had been used on the Danish Web. After this, an ETL specification was drafted that specified exactly what should be extracted from Netarkivet. ETL stands for Extract, Transform, Load; the terms relate to how a subset is extracted from a collection, transformed to fit an analytical purpose, and loaded into hardware/software where the analysis is performed — see Have (2018). The ETL specification stated that a search query should be performed that identified Web pages in Danish and in English (excluding content from Twitter or YouTube) that included the term ‘mHealth’ and were archived from 2006–2018. This resulted in a total of 54,131 Danish-language documents, and 104,856 English-language documents. These Web pages were then extracted in three formats: metadata, text content and links. The metadata extraction included information about the archiving process, the text content data set contained the text of all the Web pages, and the links data set listed all of the hyperlinks on all of the found Web pages. In the analysis conducted in this article, only the text content and link data were used.

The final stage, where the extracted data were cleaned to prepare them for analysis, deserves further comment. Archiving processes mean that the same Web entity — a Web page, an image, etc. — may be archived more than once. This is because Web archiving is performed by following hyperlinks, and if numerous hyperlinks point to the same entity, it is archived several times. Gorsky [7] raised this as an inherent methodological challenge when he asked: On what basis should de-duplication proceed?. It is important to bear in mind that the same can be understood in two ways: either a Web element such as an image file can be identical to an image file that is already in the archive; alternatively, a Web element such as a Web page can be almost identical to a previously archived Web page. We characterise the first as a duplicate and the latter as a version. In order to avoid skewing the findings that would occur if too many copies of some Web pages were included, duplicates and versions had to be removed. Removing duplicates was easily done because every file in the Danish Web archive has a unique identifier (a so-called hash value), so all of the files with a hash value that were there more than once were removed. Removing versions can be done in different ways, each with their own advantages and disadvantages [8]. The approach we adopted here involved selecting the first version that was archived. In other words, if a given Web page a exists in three versions in the Web archive — $a_1$, $a_2$, and $a_3$ — but $a_2$ was archived before $a_1$, then $a_2$ was selected, even if it was not part of the archiving when the entire Web site was archived.

The result of these tidying processes was that 18,456 duplicates were removed from the Danish data set, and 19,163 from the English data set, approximately 25 percent of the initial data set of 158,480 Web pages. As to versions, approximately 45 percent of the Web pages were identified as versions and were thus removed. As a result of these processes, the data sets were made as ‘clean’ as possible, and ready for analysis.

Proposing a step-wise methodology

Three methods were used in the overall analytical design that is illustrated in Figure 1.
Proposing methods to explore the evolution of the term 'mHealth' on the Danish Web archive

Figure 1: The overall analytical design. A is the calculation of total term occurrences per month, B is the term occurrences per month distributed against Web sites (actors), C is a network analysis of relations between the Web sites (actors), and the arrow illustrates the development over time.

The first method involves a simple calculation of the total number (sum) of Web pages per month where the term ‘mHealth’ is found at least once. As such, the figures do not reflect how many times the term appeared on each Web page: a Web page with only one mention has the same weight as a Web page where the term is present 100 times. We decided on this method as our priority was identifying Web pages that referred to ‘mHealth’, not how much each page referred to ‘mHealth’.

The second method used this basic quantitative information as its point of departure and took the next step of identifying which actors used the term the most within each month. What was calculated here was the number of Web pages on each Web site where the term ‘mHealth’ appeared, month by month. First, we removed www. from CSV files that included the actors with the greatest number of occurrences of the term ‘mHealth’ in both languages (≥5) by replacing www. with a blank. We did this in case there were versions of the Web address in each month that both included and excluded www.; these could then be aggregated. After this, we used the digital tool called Triangulation (Digital Methods Initiative, 2008) to identify the most recurrent actors across the phases. The output took two forms: a summary that identified the number of occurrences of the term ‘mHealth’ from the greatest — across all the months, to the least — one occurrence only, as well as a colour-coded chart that indicated the trends across all the months in the phase visually. The files were saved as htm files and pdf files. Finally, we drew on Stroobant’s (2019) typology of senders of Web pages (actors) in the biomedical field to help us characterise the main actors that were the senders of the Web pages that referred to ‘mHealth’.

The third method involved performing a network analysis of relations between Web sites (actors). Network analysis, which has a long history within the social sciences, can involve a variety of source types (Wasserman and Faust, 1994; Prell, 2012; Park and Thelwall, 2005; Stevenson and Ben-David, 2019). A network analysis can be used to identify the importance of actors based on the links between them. Using the terminology of network analysis, actors (‘nodes’) are related to other actors by connections (‘edges’), and the constellation of nodes and edges form a network. If the connection is directed — for instance, going from node A to node B — one can determine the centrality, or importance, of a node in the network in two ways: either it is central
because many other nodes point to it (indegree centrality), or because it points to a lot of nodes itself (outdegree centrality). In both cases, a central node can also be considered a ‘bridge’ — a node that connects two or more clusters of nodes, thus facilitating connection between nodes that are not directly connected.

In performing the network analysis, we used the link data set from the ETL process, with duplicates and versions removed, as described earlier. Based on this data, all hyperlinks on all Web pages were listed year by year and month by month in the following format ‘Source, Target, Count’, where Source was the Web site where the link started, Target was the Web site to which the link pointed, and Count was the number of links from Source to Target. To identify the extent to which Web sites that mentioned ‘mHealth’ were closely/not closely related through hyperlinks, we used the ForceAtlas layout from Gephi (2017), a network analysis and visualization software, as it identifies clusters of nodes based on their repulsion and attraction/gravity. This involved analysis focusing on ingoing links with a view to identifying the actors to whom most other actors linked, and analysis centring on outgoing links where the actors that linked the most to other actors were identified.

Ethical considerations

We took steps to ensure we met ethical standards in using material from the Danish Web archive in our study. We applied to the Royal Library in Denmark for permission to access the Web archive (according to the Copyright Act paragraph 16a, subsection 3). The application included a description of the project: ‘mHealth in Denmark: Findings from the Web archive’ and a description of the data we wished to extract from the archive. The project was accepted, and we were granted permission to search and harvest data from the archive to use in our analysis. The data are stored on a closed server at Aarhus University to which only the project group has access.

According to Danish legislation and as stated by the National Board of Health and the Scientific Ethics Committee, the study did not need approval from the Danish Biomedical Research Ethics Committee System. However, the study was approved by the Danish Data Protection Agency, Aarhus University (journal number 2016-051-000001, number 1238).

Findings

Occurrences of the term ‘mHealth’ on Web pages

First, we identified all of the Web pages where the term ‘mHealth’ appeared at least once in the English and Danish data sets. Figure 2 shows the development of ‘mHealth’ in Danish and English Web sites from 2006 to 2018, using log plots. Log plots facilitate an overview of data despite large differences between the lowest and the highest number of occurrences. As can be seen in Figure 2, from 2008, there was an increase in the number of Web pages that used the term ‘mHealth’ which peaked in 2015. Figure 2 demonstrates that the frequency ran parallel in both languages, though with a continuously higher number of occurrences in the English than in the Danish data set.
Proposing methods to explore the evolution of the term 'mHealth' on the Danish Web archive

Figure 2: Occurrence of ‘mHealth’ in Danish and English language Web pages.

We then looked more closely at the details of each month. Figure 3 shows this more granular approach.
Proposing methods to explore the evolution of the term 'mHealth' on the Danish Web archive

**Figure 3:** Occurrences of ‘mHealth’ in Danish and English language Web archive data over time. Stippled lines indicate the three phases.

In the Danish data (the first chart in Figure 3), the occurrence of Web pages that included the term ‘mHealth’ fell into three phases: an early phase (2006 — March 2012) with few occurrences of the term ‘mHealth’, a second phase of consolidation (April 2012 — October 2015), and a third phase (November 2015–2018) that showed more variability than the previous phase (notably, in March and July of 2018, there was an upsurge in the number of occurrences).

In the English Web pages (the second chart in Figure 3), the term ‘mHealth’ occurred one year before it did in the Danish Web pages. We characterised the trending again according to a three-way division. The first phase ranged from October 2006 — April 2011; there were lower levels of activity throughout the first phase, as in the Danish first phase. The second phase extended from May 2011 — February 2016 and was characterised by continued and increasing activity and a higher level of activity than in the Danish data set. The third and final phase started in March 2016; it was characterised by sporadic occurrences that were most evident around November 2017.

The phases, indicated by the stippled lines in Figure 3, are therefore not identical, as the numbers of occurrences spiked earlier in the second phase in the English language data, possibly reflecting that the term originated outside Denmark; it may also suggest the primacy and influence of English-language Web sites in the field of mHealth.

**Actors (Web sites) using the term ‘mHealth’ on Web pages**

The second part of the analysis investigated which actors (Web sites) on the Danish Web used ‘mHealth’ the most, and how this evolved over the phases. We undertook this analysis as we wished to identify who the actors were, as this could promote understanding of how the term ‘mHealth’ emerged and was propagated in the Danish online setting. We interpreted a higher occurrence of the term ‘mHealth’ per Web site as an indication that the actors (Web sites) were important and had influence in the area. Specifically, for this
analysis, we identified the actors associated with Danish and English Web sites that included the term ‘mHealth’ five or more times across each temporal phase (month).

We then classified the Web sites into actor categories so that we could identify trends in the kind of actors that discussed ‘mHealth’ online. In this, we were inspired by Stroobant’s (2019) characterisation of actor categories, as mentioned earlier. In cases of doubt when categorising, we checked in the Danish Web archive or consulted the current form of the Web site using a Web browser. Three of the actors could still not be categorised either via Web archive data or Internet searches. Our findings are presented in Tables 1 and 2.

In Table 1, which shows the Danish Web sites, a heterogeneous set of actor categories is evident. There are, in particular, many occurrences of news media Web sites from early on (e.g., dr.dk is prevalent in the first two phases), as well as a mixture of government and private institutions. It is noteworthy that networks of organisations and personal Web sites started to be prominent in the second phase, followed by academic institutions in the third phase, indicating that the term ‘mHealth’ was gaining broader discursive traction in the Danish Web environment.

<table>
<thead>
<tr>
<th>Table 1: Danish data — phases and main actors.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories</td>
</tr>
<tr>
<td>Phase 1</td>
</tr>
<tr>
<td>News Web site</td>
</tr>
<tr>
<td>Phase 2</td>
</tr>
<tr>
<td>Industry</td>
</tr>
<tr>
<td>News Web site</td>
</tr>
<tr>
<td>Government institution</td>
</tr>
<tr>
<td>News Web site</td>
</tr>
<tr>
<td>Government institution</td>
</tr>
<tr>
<td>Network of public organisations, private companies and research organisations</td>
</tr>
<tr>
<td>Personal Web page facilitating constructor</td>
</tr>
<tr>
<td>Personal Web page facilitating constructor</td>
</tr>
<tr>
<td>Private blogger</td>
</tr>
<tr>
<td>Private company</td>
</tr>
</tbody>
</table>
Table 2 shows the English-language Web sites where five or more Web pages per month included the term ‘mHealth’. Comparison of Tables 1 and 2 reveals that there are more English-language actors than Danish-language actors whose Web pages (combined) included ‘mHealth’ five times or more in a month. Interestingly, no single actor in the English data included the term ‘mHealth’ at least five times or more in a month in Phase 1, which contrasts with the Danish data, where dr.dk was the only one. In Phase 2 of the English language data, there are many more news Web sites, social media actors and private companies than in the Danish language data (Table 1).
<table>
<thead>
<tr>
<th>Association of health professionals and hospitals</th>
<th>memorialhermann.org</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private company</td>
<td>pa-consulting.dk</td>
<td>10</td>
</tr>
<tr>
<td>Private company</td>
<td>healthbridge2013.dk</td>
<td>9</td>
</tr>
<tr>
<td>Association of private, public and research institutions</td>
<td>appliedhealth.eu</td>
<td>7</td>
</tr>
<tr>
<td>Network organisation between Sweden and Denmark</td>
<td>mva.org</td>
<td>7</td>
</tr>
<tr>
<td>N/A</td>
<td>plancast.com</td>
<td>7</td>
</tr>
<tr>
<td>Academic</td>
<td>findresearcher.sdu.dk</td>
<td>7</td>
</tr>
<tr>
<td>Private company</td>
<td>mobilemarketingwatch.com</td>
<td>6</td>
</tr>
<tr>
<td>Private company</td>
<td>goldengekko.com</td>
<td>6</td>
</tr>
<tr>
<td>Private company</td>
<td>scoop.it</td>
<td>6</td>
</tr>
<tr>
<td>Social network service</td>
<td>storify.com</td>
<td>6</td>
</tr>
<tr>
<td>EU Web site</td>
<td>ec.europa.eu</td>
<td>6</td>
</tr>
<tr>
<td>Academic</td>
<td>findit.dtu.dk</td>
<td>6</td>
</tr>
<tr>
<td>Network of public and research organisations</td>
<td>cimt.dk</td>
<td>6</td>
</tr>
<tr>
<td>Private company (DK)</td>
<td>monsenso.com</td>
<td>6</td>
</tr>
<tr>
<td>N/A</td>
<td>mhealtheconomics.com</td>
<td>6</td>
</tr>
<tr>
<td>Social media</td>
<td>uk.linkedin.com</td>
<td>5</td>
</tr>
<tr>
<td>Academic</td>
<td>iotinternetofthingsconference.com</td>
<td>5</td>
</tr>
<tr>
<td>Private company</td>
<td>apurebase.com</td>
<td>5</td>
</tr>
<tr>
<td>Research institution</td>
<td>globalhealthstudentsblogs.ku.dk</td>
<td>5</td>
</tr>
</tbody>
</table>
A comparison across the phases of the English data set reveals little overlap between the actors in the different phases. The sporadic emergence of the actors, their appearance and subsequent disappearance, suggests lack of maturity in the field of mHealth. A counter-example to the instability across the phases, however, is findresearcher.sdu.dk (University of Southern Denmark) which is present in phases 2 and 3, indicating a consolidation of research engagement with mHealth in more recent years.

Comparison of the two tables revealed that some academic Web sites (actors), such as vbn.aau.dk and findresearcher.sdu.dk, are present in phases 3 of both the Danish and English data. This may be due to internationalisation in universities and the use of parallel language policies. Organisations that facilitate networks between private and public/research organisations first appear in Phase 2 of both tables; this is significant as it indicates the arrival of a new actor type, as well as the growing business importance of mHealth technologies, and the need for new structures and interfaces between private and public institutions.

Interestingly, the Danish actor cimt.dk, a research and innovation centre, is present in Tables 1 and 2, though in different phases (phase 2 for the English data set, and phase 3 for the Danish). The transition that cimt.dk takes from English to Danish Web pages suggests a Danish actor’s integration of an evidently English-language term ‘mHealth’ in the Danish context; the term ‘mHealth’, it seems, is ‘going native’ and becoming normalised in online Danish discourse.

### ‘mHealth’ actor connections in networks

In the third part of the analysis, we performed a network analysis to identify who among the identified actors were the most important, and how this may have changed over time. Network analysis made it possible to shed new light on the discursive Web environment in which the actors were embedded as it indicates the extent to which the actors linked to each other using Web links; the method also helped to identify to what extent the actors identified in the second part of the analysis were important. For the purpose of this article, we have delimited the investigation to a selected month from the phases identified in the first part of the analysis. The methodological insights we gained in the process mean that a full-scale analysis of the entire data set could be performed, an endeavour that lies beyond the scope of the present article. However, the following illustrates the method.

We demonstrate the network analysis using June 2013 of phase 2 of the Danish data set, as this month showed the greatest variety of actor types (cf., the typology in Table 1). Developments in the Danish networks will be briefly recapitulated on the basis of analyses that are included in an online open access repository (Figshare) [9]. The English language networks are not scrutinised here, but are also included in the open access repository.

In our network graphs, the size of a node reflects its number of edges (links to or from it): the bigger the node, the greater the number of edges. Similarly, the thickness of an arrow between nodes indicates the weight of the

### Phase 3

<table>
<thead>
<tr>
<th>Academic</th>
<th>vbn.aau.dk</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research institution</td>
<td>ssi.dk</td>
<td>8</td>
</tr>
<tr>
<td>Academic</td>
<td>pure.au.dk</td>
<td>7</td>
</tr>
<tr>
<td>Academic</td>
<td>orbit.dtu.dk</td>
<td>7</td>
</tr>
<tr>
<td>Academic</td>
<td>pure.itu.dk</td>
<td>6</td>
</tr>
<tr>
<td>Academic</td>
<td>findresearcher.sdu.dk</td>
<td>5</td>
</tr>
<tr>
<td>Private company</td>
<td>advantageaustria.org</td>
<td>5</td>
</tr>
</tbody>
</table>
Proposing methods to explore the evolution of the term 'mHealth' on the Danish Web archive

edge: the more links between two nodes, the thicker the arrow. The colouring of the nodes shows the different clusters. Finally, the nodes for the actors identified in the second part of the analysis and their edges have been marked in red. The network graph and the data behind its creation are included in the open access repository.

The network is best understood by comparing the graphs for outgoing and incoming links (Figures 4 and 5), as the networks are identical apart from the indicated directionality of the edges.

![Network graph](image)

**Figure 4:** Network of hyperlinks, Danish data set, June 2013, outgoing links.
The network is dominated by four clusters with a limited number of nodes, but with a large number of links between the nodes. vis.dk, which is a forum for professionals in the healthcare system, is the central node in one cluster (Figure 4), with many links pointing to the Web sites of ‘regions’, or counties (Figure 5). selvmaal.dk, a Web site related to a book about the quantified self (in Figure 4), is central in a cluster with Web sites such as virkeligheden.dk, quantifiedself.com, pandodaily.com and Twitter (Figure 5). emporiamobile.dk (Figure 4), a telecommunications company, has many links to only three nodes, emporiatelecom.com, YouTube and Facebook (Figure 5). The Web site of a personal blogger, thildevesterby.dk (Figure 4), is the central node in the fourth cluster. Around these clusters, one finds a few smaller clusters with a central node as a bridge, but only one/a few links connect the bridge to the rest of the network (Figure 5). It is worth noting that the clusters — big as well as small — are generally not well connected to each other, creating an impression of disjointed networks.

Regarding the role of the actors in the network identified in the second analysis (Table 1, phase 2), eight of the 10 actors are present. Four of them are very small, whether one looks at incoming or outgoing links, namely welfaretech.dk, flyt-te-fir-ma.dk, dr.dk, and her-hos-dig.dk, and they do not play any significant role in the network, since they have very few links and are not connected to the other clusters. Each of the other four actors are central nodes in each of their clusters: regionh.dk and vis.dk in the same cluster, emporiamobile.dk in a cluster with only three other nodes, and thildevesterby.dk is the central node in a cluster with a large
number of links to different clusters, making this node more important than the other actors in the network. However, there are only two connections between the actors (one in the self-contained cluster with vis.dk and regionh.dk, and one from thildevesterby.dk to regionh.dk). Thus, in June 2013, the actors did not constitute a coherent discursive environment.

As mentioned earlier, our network analysis extended beyond the one we present here (June 2013, phase 2 of the Danish data); we have included additional files and network analysis graphs in the online open access repository for months where most of the identified actors were present showing the network and the actors’ role in it in months where we could expect the strongest presence of the actors. Based on this method, the following four months were selected for further analysis:

- Danish: March 2011 (phase 1), September 2016 (phase 3) — NB: June 2013 (phase 2) is covered above
- English: February 2016 (phase 2), March 2017 (phase 3) — NB: no actors included the term ‘mHealth’ five times or more in phase 1

This allowed us to make more general statements on the development of the role of actors in the networks over time, which we do for the Danish set, which is more complete with three phases.

The following three trends were clear in the Danish set. First, actors (identified in the second phase of the analysis) do not tend to play a big role in the other networks (with the exception of dr.dk in phase 1), despite the fact that some of them have a large number of incoming and/or outgoing links. However, these links relate to a very limited number of nodes or they remain within the boundaries of a given cluster, making the actors less central and rather isolated from the rest of the network. Second, the role of the actors tends not to persist across the three phases; for instance, the central node in the first phase (dr.dk) does not play any role in subsequent networks, and the same goes for a number of actors from the other two phases. Important clusters including central actors such as welfaretech.dk or thildevesterby.dk only play an important role in one network, although they may be present in more than one. Third, in almost none of the phases are the actors linked to each other. As such, they fail to constitute a network with other actors within or between the actor groups identified by Stroobant (2019). Actors remain isolated islands, indicating a striking lack of a shared discursive environment about mHealth.

Discussion

Our aim was to explore the emergence of the term ‘mHealth’ in the Danish Web archive. In the first analysis, we identified occurrences of the term ‘mHealth’ over phases for both Danish and English Web pages. In the second analysis, we saw that the number of actors using ‘mHealth’ increased throughout the phases, that the characteristics of actors changed from typically being a private, public or civic organisation to partnerships, and that the actors tended not to endure between the phases. In the third analysis, several minor networks appeared and disappeared again. Looking closely at one network (Danish data, June 2013), we found that central actors played a role in important clusters but that these were decoupled from other networks during the same time period.

All in all, these findings suggest that ‘mHealth’ became a more popular and mainstream term from 2006–2018, although both an increase and decrease in its prevalence was evident during the timeframe, which indicates a minor to moderate impact. The networks of mHealth actors had a transitory nature, as networks appeared and disappeared again. The isolated quality of the networks (four clear networks in the month we analysed) suggests the lack of overarching institutional support that could have consolidated the networks.

In our article, we highlighted various methodological challenges in relation to using Web archive data with respect to our research aim, and we illustrate in this article how we addressed them. We demonstrated how we used the three methods of our approach together in this article, but they could be used separately in other studies to explore other topics in relation to their historical Web presence.
Since ‘mHealth’ is a rather specialised term that is not part of the vernacular, it had the advantage of not having many semantic variants, allowing us to focus on this term only. Quantification of the term occurrences provided a quick overview of when and how many times a term was used, and enabled the identification of trends and phases over time. However, searching for only one term may have resulted in a ‘noisy’ data set, because one cannot be sure that one mention on a Web page makes the document relevant. Therefore, this ‘raw’ identification of Web sites based on the inclusion of only one term could in a future study be supplemented by a more contextual and extended search, looking at other relevant terms in the proximity of ‘mHealth’ — for example, using a corpus linguistics collocation analysis approach (Brezina, 2018).

The importance of tidying the data to avoid having too much of the same was previously identified by Gorsky [10]. He used the term ‘duplications’ to mean ‘effectively identical pages, but with minute differences’, and thus did not distinguish between duplicates and versions as we have done in this article. In this article, we presented our answer to the challenge: duplicates can be quite readily removed, because they are exact copies, but versions are harder to handle, and researchers have to choose between various approaches. Versions are an inherent challenge in national Web archive research and need to be addressed to avoid skewed results. In other words, the presence of versions reflects the nature of the archived Web, and can be dealt with in different ways, thereby increasing the methodological reflectiveness of a study. For instance, depending on the study, one can opt to choose the biggest version (in number of objects or megabytes), or the version closest to a specific point in time that is important for the analysis.

Collections of archived Web, and in particular national Web archives, include a cornucopia of information that at present remain largely unexplored. However, to use more fully the largely untapped potential of the archived Web, new quantitative research methods must be developed [11]. As explained in this article, Web archive material has distinctive qualities. One of the major differences between a digitised collection (of newspapers, for instance) and a Web archive is that in a Web archive (and on the online Web, for that matter) the hyperlink is an integral and constitutive feature — remove the hyperlink and the Web does not function. However, hyperlinks also provide a valuable object of study in themselves as they can be easily identified in and extracted from HTML source code. It is therefore not surprising that investigations of hyperlinks are found in academic studies of health-related Web archives (e.g., Gorsky, 2015; Lacy-Nichols, et al., 2020) and have been highlighted as a means of adding an alternative dimension to qualitative approaches. For instance, Millward [12] states that instead of ‘conducting large-scale searches of the whole archive, link analysis could be a more accurate way of assessing the relative reach and influence of different organisations.’

Hyperlinks can be studied in various ways, from calculating the number of links between preselected Web sites or groups of Web sites, as in Gorsky (2015) where the focus is on Web sites on the second-level domain names ‘nhs.uk’ and ‘gov.uk’, to constructing a link graph with a view to making a hyperlink network analysis, as we do in this article. In the first instance, one starts with a predefined list of actors (or groups of actors) and limits the analysis to links between these actors; in the latter instance, one takes a more explorative and grounded approach by letting the data ‘talk’ to map the Web environment in which the actors were embedded. The advantage of the first approach is that the analysis is precise, but it misses a strength of the latter approach: actors that may not initially have been deemed relevant for the analysis may, in fact, be important actors in a network.

However, network analyses come with limitations. Some of these limitations are well-known within hyperlink studies of the online Web. A link analysis based on the source code for ‘hyperlinks’ only examines the relation between nodes and not the content on the nodes (Brügger, 2013a), and therefore does not shed light on the actual content of the interlinked Web sites. It can therefore be difficult to determine to what extent a link on a Web page where the term ‘mHealth’ occurs actually has anything to do with mHealth. Also, the equation ‘relation of hyperlinks = importance’ can be questioned since importance can also be investigated by other measures, such as number of users. In addition to these limitations, hyperlink analysis of the archived Web comes with at least one great challenge which is a function of the fact that the archiving process takes time. Specifically, the link source and the link target may not have been archived at the same time which means that our temporal divisions in months may remove the link source from the link target, thus creating a temporal inconsistency in the network as a whole.
Conclusion

The aim of this article was to explore how the single term ‘mHealth’ featured in the historical Danish Web, identifying its emergence, prevalence over time and which actors played a role in its promotion. We assembled methods to explore these questions. Our main contribution is methodological as the methods have general value and applicability. They can be used with Web archive data to identify the cultural salience of historical Web pages, at least with respect to occurrences and linkages. In a similar vein, Gorsky [13] observed that ‘[t]he nature of these online documents means that a project concerned with health policy may end up in part an exercise in cultural studies’. Indeed, we argue that Web archive data create opportunities for new types of research questions (such as those we have presented here), while contributing to fields such as cultural studies (Fage-Butler, 2018) and Internet studies (Brügger, 2013b). We anticipate that countries with similar national Web archives to the Danish Web archive (like the U.K., France or Portugal, cf., the overview in International Internet Preservation Consortium, 2021) could implement — and further develop — the methods we have presented in this article.

About the authors

Antoinette Fage-Butler is an Associate Professor at the Department of English, School of Communication and Culture, Aarhus University, Denmark. Her research interests include cultural aspects of public health and knowledge communication in online data.
ORCID: https://orcid.org/0000-0002-4849-0227
E-mail: age-butler [at] cc [dot] au [dot] dk

Loni Ledderer is an Associate Professor at the Department of Public Health, Aarhus University, Denmark. Her research interests encompass social, organisational and technological aspects of public health and the development of methodologies to study these.
ORCID: https://orcid.org/0000-0002-2685-4579
E-mail: loni [dot] ledderer [at] ph [dot] au [dot] dk

Niels Brügger is a Professor at the Department of Media and Journalism Studies, School of Communication and Culture, Aarhus University, Denmark. His research interests include Web and media history, Web archiving, and digital methods.
ORCID: https://orcid.org/0000-0003-1787-1980
E-mail: nb [at] cc [dot] au [dot] dk

Funding

This work was supported by DIGHUMLAB/NetLab, a Danish research infrastructure with six weeks of IT assistance: https://dighumlab.org/about/.

Acknowledgements

The authors wish to express their gratitude to NetLab’s IT-developer Ulrich Karstoft Have for valuable support, data cleaning, analysis and advice during the different phases of this project. The authors want to thank Kristoffer Laigaard Nielbo, leader at the Center for Humanities Computing Aarhus and student assistant Lasse Hansen for valuable advice and assistance with the data analysis.
Proposing methods to explore the evolution of the term 'mHealth' on the Danish Web archive

Notes


9. Supplementary files are available at the online open access repository Figshare: https://doi.org/10.6084/m9.figshare.c.5184977.v4.


11. Evans, 2020, p. 16.


References


Proposing methods to explore the evolution of the term 'mHealth' on the Danish Web archive


Research2Guidance, 2021. “Denmark is a leading EU country in mHealth market readiness — Preliminary


Editorial history

Received 3 April 2021; accepted 7 January 2022.

Copyright © 2022, Antoinette Fage-Butler, Loni Ledderer, and Niels Brügger. All Rights Reserved.