Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive
by Andreas Lenander Ægidius

Abstract
This article uses Internet archives to explore the emergence and spread of terms with the stem ‘stream’ in the Danish Web domain from 2006 to 2021, focusing on the actors that contributed to its evolution. I test three proposed methods for investigating the Web pages and Web sites that employed a given term. My findings highlight temporal developments in the use of ‘stream*’, ‘streamingtjeneste’, and ‘streaming service’ with diverse actors using it, though news Web sites clearly dominated. This research attends to challenges in working with Web archive data, and evaluates methods with regard to large datasets, that others can use to engage empirically with Internet archives, which remain vast, but largely under-exploited resources.

Contents
Introduction
Methodology
Findings
Discussion
Conclusion

Introduction
This paper presents findings that are part of an investigation of the collection of streaming at a legal deposit institution. New distribution methods challenge the collection of online digital cultural heritage. Streaming as an everyday term covers a paradigm shift in media culture, its infrastructures, and business models. Statistics indicate that Denmark has high Internet penetration, with 95 percent of homes having access (Statistics Denmark, 2022a). In 2020, 64 percent of Danes streamed films and television programs, while 70 percent of Danes streamed music (Statistics Denmark, 2022b). The Danish Broadcasting Corporation’s (DR) analysis of media development noted that in 2020, 52 percent, 48 percent, and 42 percent of Danes (aged 4+) used the streaming services YouTube, Netflix, and DRTV at least once a week (DR, 2020). Streaming is also a concept that connects media use and constitutes cultural heritage across industries, practices, and content. We ‘stream’ over a wide range where we used to ‘listen’ to music, ‘read’ books, and ‘watch’ TV. The spread of the streaming concept across media sectors and everyday activities tells us about the media development and can guide the documentation and collection for research purposes and preservation for posterity. However, streaming is a complex and composite concept that dissolves and reinvents distribution and media use in continuous exchange between Internet media and traditional flow-
based broadcast media (Spilker and Colbjørnsen, 2020). Web archives constitute a place to find these entities in an attempt to map and explore their evolution online. This endeavor is not helped by the fact that the Web archives exhibit a similar degree of complexity as archivists encompass the many facets of the Web and our online activities. The archived Web is a reborn digital medium. This means that three interdependent constructions establish the Web as a research object in a Web archive: collection, preservation, and making available [1].

A number of collaborative efforts have addressed Web archives for research purposes in recent years opening the field of Web history and archives studies to researchers with an interest in digital humanities (Brügger and Laursen, 2019; Brügger and Milligan, 2019; Gomes, et al., 2021). Among them, Morris (2019) addresses the sonic properties of the past Web concerned with the sound files, formats, and soundscapes that are available to digital historians. The introduction of streaming is emblematic of the Web-based technologies that are prizing the mobile and the ubiquitous, the dynamic over the static, causing difficulties for those trying to write histories of sound on the Web or anyone interested in listening to the past [2]. I have previously investigated how video-streaming services challenges the collection of digital cultural heritage (Ægidius, 2021). I discuss a number of methods for collecting content and its context that is subject to this highly dynamic concept. What we also need is a better map of streaming history to inform our understanding of the streaming concept. Here, I trace the evolution of the use of stream-related terms online as a way of documenting the general media history of the streaming era in Denmark, not exclusively auditory or visual or various combinations.

Streaming exhibits adaptive powers in the way it settles into, or occupies, the existing digital media ecology. Spilker and Colbjørnsen (2020) theorize five dimensions of streaming and analyze their dynamics as a way to outline the evolution of streaming. The borders and distinctions of streaming are under negotiation and in flux: (1) between legal and illegal; (2) between live and on-demand; (3) between user-generated and professional content; (4) between single-purpose and multi-purpose streaming services; and, (5) between appeals to niche and general audiences. Niche appeal, for example, can be determined by language or geography, as well as an audience’s preferences for content genres [3]. I look for evidence of these dimensions and map my findings according to these dimensions of streaming in this empirical study of the mention of streaming services on archived Web pages in the Danish Web archive (Netarkivet).

The proposed methods are explorative as they include all data on the historical Danish Web that mentions a given term in an analysis. This means that the phenomenon reveals itself through its online presence rather than in relation to pre-selected criteria. In this sense, the methods capture messiness and do not assume any grand purpose or teleology, as highlighted by Brügger (2016) who describes the value of inductive and inclusive approaches in Web archive research.

The overall aim of my study is two-fold. I test the proposed methods during which I explore the emergence and evolution of the terms ‘streaming service’ and ‘streamingtjeneste’ in the Danish Web domain identifying the actors that were part of its evolution. In so doing, I explore the communication about one digital media concept (streaming) in another (Web archives). The following three research questions frame

I test the methods proposed by Fage-Butler, et al. (2022). They explore the emergence and evolution of the term ‘mHealth’ in the Danish Web domain, identifying the actors that were part of its evolution. The purpose of doing so is threefold. First, specific to my topic of interest, I extract similar findings by re-using their aforementioned research questions that are closely tied to the tools and type of dataset used. Second, I aim to contribute to the validation of their proposed methods. I follow the steps of the proposed methods with datasets based on a term with more semantic variants, wider usage, and a resulting larger number of extracted Web sites (see the section on methodology). Third, the Web curators of the Danish Web archive framed their assistance to my investigation as an informal test of in-house skills. Specifically the skills needed for the steps following data extraction, which are not part of their ‘normal’ skill set, e.g., data cleaning and data treatment with notebooks in the software Rstudio. In this way, the curators analyzed and negotiated the skills that recipient researchers typically need and often solicit assistance for, when working with data extracted from the Web archive. This in-house skills test will not be the subject of this article [4].
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

the analysis. First, when and how often did the terms ‘streaming service’ and ‘streamingtjeneste’ occur on the Danish Web? Second, which actors (Web sites) on the Danish Web used ‘streaming service’ and ‘streamingtjeneste’ 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 (2009), I track the life of individual Web elements — the terms ‘streaming service’ and ‘streamingtjeneste’ — across the Web sphere constituted by the discursive Web space related to streaming services on the Danish Web from 2006 to 2021 [5]. I have opted to mirror the article progression from Fage-Butler, et al. (2022) and its sections for easier comparison. I note digressions from their framework in the relevant sections.

Methodology

Data from the national Web archive

I extracted data from the Danish Web archive (Royal Danish Library, n.d.a). The Danish Web archive is responsible for collecting and preserving the Danish part of the Internet under the provisions of the Danish Legal Deposit Act (for more information on the Danish Web archive see Brügger, et al. [2019]; Royal Danish Library [n.d.a]). The method that I test tracks temporal (diachronic) changes to capture the evolving occurrence of one term over time on the Web and any changes or stability in the actors who referred to it. I show how the proposed methods can track two or more terms with few modifications to the method. Furthermore, I extract a larger dataset because the term ‘stream’ is more widely mentioned on Web sites over time in the archive.

I produced and prepared the dataset for analysis using the following steps. Initial explorative free text searches via the Web archive’s SolrWayback interface provided indications of a widespread use of the terms in the Danish Web. During the first step, I learned that Web pages that mention terms beginning with ‘stream’ would produce a large dataset that would not be suited for a test of the proposed methods. I included ‘streamingtjeneste’ and ‘streaming service’ to produce a subsection of the dataset with a scope that addressed my main interest in streaming services. Then, in collaboration with the Web curators, I optimized and translated the terms into search queries, taking into account variations with and without spaces [6]. A data specification document is a mandatory addendum to the data delivery agreement between researchers and the Danish Web archive. The document specified search queries to be performed that would identify Web pages in Danish and English that included terms beginning with ‘stream’ as well as Web pages that mention ‘streamingtjeneste’ and Web pages that mention ‘streaming service’ from 2003–2021; see Ægidius (2023) for data specifications.

This resulted in a dataset containing 130,671,109 Web pages in Danish and 76,227,896 Web pages in English that mention at least one term beginning with ‘stream’. The result for ‘streamingtjeneste’ was 9,434,009 Web pages and the result for ‘streaming service’ was 62,911 Web pages in a separate dataset.

The data specification stated the formats of the extracted data with reference to the data fields of the archive. As the name implies the format metadata is data about the Web pages while the formats text content and links contain textual content and all links from the Web pages. It is important to note that the ‘mHealth’ study only utilizes the two formats metadata and links. I have not modified the method to include the text content from the Web pages [7].

With reference to Gorsky (2015), Fage-Butler, et al. (2022) commented on their choice of de-duplication that distinguished between duplicates and versions. Duplicates are identical files while files that are slightly different from one another are versions. Following the method, I removed duplicates and selected the first archived version. This means that if a given Web page a exists in three versions in the Web archive — a1, a2, and a3 — but a2 was archived before a1, then a2 was selected, even if it was not part of the archiving when the entire Web site was archived. It is worth underlining that versions can be dealt with in different
ways, each with their own advantages and disadvantages [8]. Removing duplicates and choosing what to do with archived versions of a document is part of ‘cleaning’ the dataset.

Fage-Butler, et al. (2022) published a script in the format R markdown (.rmd). This type of open and editable tool for data treatment is dubbed a ‘notebook’. As part of testing their method, I initially had to define the file path of my data to load into the notebook using the software Rstudio. The script identifies duplicates, counts them, and removes them. The script handles versions by grouping the data into months and then remove any alternate versions within that month besides the first version. I did have to further adjust the script to encode the data properly to account for special characters and separate the rows of target links meaningfully, more in detail below. The modified script is included in an online open access data repository along with other supplementary files for this article (Ægidius, 2023).

I extracted the large ‘stream’ dataset but its size demanded more processing resources than a typical work PC with 16GB RAM could provide to run Rstudio satisfactorily. Two factors led me to abandon the processing of the large dataset. Firstly, the library could not provide the necessary computing resources within the timeframe of the project and I had not budgeted for data processing by an external provider like Azure or AWS. Secondly, the Danish Web archive maintains a growing toolbox in its SolrWayback interface for researchers. It includes a software tool, Ngram Netarchive, which can visualize search queries by year (Royal Danish Library, n.d.a). This tool provided approximately the same output of the first step in the method, which is a visualization of the frequency of a term’s appearance over time.

Using the script, I removed 888,871 duplicates from the ‘streamingtjeneste’ subsection of the dataset. Then, I removed 23,880 duplicates from the ‘streaming service’ subsection of the dataset. The combined ‘streamingtjeneste’ and ‘streaming service’ dataset contained 8,584,169 unique records. Approximately 37 percent of the ‘streamingtjeneste–streaming service’ dataset were identified as versions and removed from the dataset. These choices made during the data treatment and the proposed steps for tidying up the data led to a ‘clean’ dataset ready for analysis.

Since I am testing newly proposed methods, it is worth comparing the numbers. The ‘mHealth’ dataset contained 159,000 documents (Fage-Butler, et al., 2022). The ’streamingtjeneste–streaming service’ dataset is 60 times larger than the ‘mHealth’ dataset. At the beginning of 2023, the Danish Web archive contained approximately 900 TB and over 38 billion objects (Royal Danish Library, n.d.a). Relative to the size of the Danish Web archive my datasets account for less than one percent of the Web archive and I only test all the steps of the method with a smaller part of the dataset containing Web pages that mention ‘streamingtjeneste’ or ‘streaming service’. As I mentioned earlier, the method that I test leaves aside the textual content on the Web pages. This has downsized the size of the total size of my datasets from approximately five terabytes to approximately 100 gigabytes and six gigabytes. In terms of the size of datasets, I have chosen smaller rather than larger. Helles (2013) argues that downsizing of large–scale datasets in combination with qualitative and/or small–scale quantitative procedures may provide qualitatively better understandings of macro phenomena than purely automated, quantitative approaches.

**Testing a step-wise methodology**

The method that I test proposes three steps in the analysis of data extracted from a Web archive. The first step is a calculation of the total occurrences of the term per year and per month. The second step distributes the occurrences per month against the actors (Web sites). The third step is a network analysis of relations between actors (Web sites). The steps are interdependent in the sense that the script prepares the data stepwise. The outputs are lists (.csv) of domain statistics per year and lists of links per year. The authors point out the three steps of their approach could be used separately in other studies of different topics and their historical Web presence (Fage-Butler, et al., 2022).

The first step of the method involves a simple calculation of the total number (sum) of Web pages per month where one of the terms appears at least once. As such, the figures do not reflect how many times the terms appeared on each Web page: a Web page with only one mention has the same weight as a Web page
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

where a term is present 100 times. This means the method prioritizes identifying Web pages that referred to a term, not how much each Web page referred to a term. The latter would entail analysis of the extracted format textual_content, which the proposed methods do not support nor do I include it in my test of the methods.

The second step of the 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. The calculation performed here involved determining the quantity of Web pages within each Web site where the specified terms were present, on a monthly basis. Following their detailed steps for data preparations, I first replaced www. With a blank in the CSV files that included the actors with the greatest number of occurrences of the term in all subsections of the dataset (≥5). This allows for the aggregation of identical versions of the Web address in a month that both included and excluded ‘www.’ in their URL. After this, I used the digital tool called Triangulation, available from the Digital Methods Initiative (2008) to identify the recurrent actors across time. Finally, this method entails an interpretive step of categorizing the actors (Web sites) to help characterize who the main actors were news Web sites, governmental Web sites, corporate Web sites, private bloggers, etc. The categories will be a function of the focus of the given study and vary accordingly; see Brügger (2022) for a discussion of actor categorization.

The third step of the method is network analysis of relations between actors (Web sites). Few historical network analyses of the archived Web exist, for instance, Brügger (2022); Webster (2021); and Weltevrede and Helmond (2012). Consequently, methods for making this type of study are not widespread. For an introduction to network analysis for Web history, see Stevenson and Ben-David (2019). What the network analysis of the archived Web have in common is that they describe the relations between Web sites via analysis of the hyperlinks between them. It is possible to determine the centrality of a Web site in a network of Web sites based on the number of other Web sites linking to it (indegree centrality), or based on the number of Web sites it links to (outdegree centrality). An actor might be notable because it interconnects (bridges) two or more clusters in a network that might signify relations between different categories of actors or otherwise significant clusters in the network. To identify the extent to which Web sites that mentioned a searched for term were closely or not closely related through hyperlinks, I used several layouts in the visualization software Gephi 0.9.7 (2022). Using visualization software for network analysis helps identify clusters of actors based on different parameters, e.g., repulsion and attraction (gravity). The difference in the sizes of my networks over time compared to the relatively smaller scale of the networks analyzed in the proposed methods led me to choose different layouts depending on the number of actors (nodes) and links between them (edges). The analysis focused on ingoing links to identify the actors to whom most other actors linked. The analytical process then centered on outgoing links to identify the actors that linked the most to other actors.

Ethical considerations

The material analyzed has been or is currently publicly available online. It may be or have been accessible behind paywalls. The Royal Danish Library is responsible for the material and ascertain ethical and security risks involved when extracting and delivering material from the Danish Web archive. A data delivery agreement between the researcher and the library formalized this procedure. As an employee at the Library during the project, I applied for access and subsequently signed a data delivery agreement stating the project details and what data I wished to search for and extract. The data are stored on a closed server at the library to which only the Web archive curators and I have access.

As for the use of visualizations of data, I refer to Ng (2022) for an excellent discussion of how all visualizations are “sensitive to bias and underlying assumptions during data collection and processing; presentation and design are susceptible to distortion and misinterpretation”. I mitigate this by stating clearly, when the use of metrics or layouts vary from what is usually expected of a figure or layout. I do this with the aim of providing transparency. This is essential for scientific rigor and replicability as well as the participatory potential of data visualizations (Ng, 2022). At the very least, my test of the proposed methods should maintain the same level of transparency to facilitate future use, further testing, or modification of the
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

proposed methods.

Findings

Occurrences of stream-related terms on Web pages

First, I identified all of the Web pages where terms beginning with stream appeared at least once in the datasets based on the content language identifier on the Web pages. Figure 1 shows the development of the use of ‘stream’ and words beginning with ‘stream’ in Danish and English Web sites. The ‘Ngram Netarchive’ tool shows all data in the Danish Web archive from the earliest collected Web page to the newest as percent of the total archive. It does not remove duplicates or versions. This made it ‘noisy data’ and served as a reminder to consider the historical usage of specific terms in different settings [9]. I include this preliminary look at the ‘stream’ dataset to show the development I most likely would have produced had I followed the first step of the method. The 2004 peak in Danish language Web pages shows a large single-shot Web collection of a popular music community site named mymusic.dk at the time. The 2014–2015 peak in the Danish language Web pages is mostly noise. This is partly due to a HTML reference to a stream object in facebook.com URLs (ref=stream) and partly due to the term ‘streaming’ recurring as a topic in a top menu on the popular news Web site bt.dk. However, Figure 1 shows that the frequency of occurrence ran parallel in both languages and rose across the years with a higher number of occurrences on the Danish language Web pages.

Second, I identified all the Web pages where the terms ‘streaming service’ and ‘streamingtjeneste’ appeared at least once. Figure 2 builds on the dataset processed with the script and shows the development of the terms ‘streaming service’ and ‘streamingtjeneste’ from 2003–2021, using log plots. Log plots facilitate an overview of data despite large differences between the lowest and the highest number of occurrences. Figure 2 shows an increase in the number of pages that mention the terms, which peaked in 2013 and 2015. Web pages in the archive have not used ‘streaming service’ nearly as much as they have used the Danish term ‘streamingtjeneste’. This is expected for data extracted from a Danish Web archive with its focus on

Figure 1: Occurrence of words beginning with stream on Danish and English language Web pages.
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

the Danish top-level domain (.dk). Figure 2 does show that the frequency with which the terms are mentioned year over year rose in parallel.

**Figure 2:** Occurrences of ‘streaming service’ and ‘streamingtjeneste’ in the Danish Web archive.

The script produced domain statistics that included a count of Web pages per Web domain per month which I summed and visualized using excel to enable the more granular approach shown in Figure 3, again using log plots. The occurrence of Web pages that included ‘streamingtjeneste’ fell into five phases:

1. few occurrences (2003 — 2009 March),
2. a rise in occurrences (2009 April — 2012 December),
3. steady peak occurrences (2013 — February 2016),
4. lower occurrences relative to adjacent phases and more variability (2016 March — 2019 August),

I made a similar analytical partition of the Web pages that mention ‘streaming service’. For the sake of brevity, these are not included in this paper.
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

However, it is worth noting how two trends characterize both developments in occurrences shown in Figure 3. First, the early years 2003–2009 show few occurrences, which matches the general history of streaming services in Denmark (first music streaming service by Danish telecom TDC in 2008, Spotify launch in 2011, and Netflix launch in 2012). Second, there is a match in lower occurrences of both terms in the years 2016–2017, in April 2018, as well as in July–August 2019. This could indicate a periodical trend of lower quantities of collected Web pages across the entire Danish Web archive. Interestingly, perhaps most clearly in Figure 2, there are trends shared with the parallel graphs produced by Fage-Butler, et al. [10]. The graphs show the highest counts for ‘streamingtjeneste’ and ‘mHealth’ around 2013–2015 and an identical dip in the year 2017. I discussed the possible global trends with the Web curators, here acting as knowledge experts on the Danish Web archive. They summarized three main reasons why 2013–2015 would appear as peaks and 2017 would appear as a dip. Firstly, during 2013–2015 they successfully ran four cross-sectional collections with a high data limit per domain. Secondly, the Web curators explained how time-consuming data compression activities within the existing archive limited the scope of the cross-sectional collections resulting in the dip of 2017. Thirdly, later years had fewer cross-sectional collections on average with optimized collecting tools and techniques that has lowered the quantity of data per year and raised the quality of data.

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

This part of the analysis investigated which Web sites (actors) used ‘streamingtjeneste’ or ‘streaming service’ the most, and how this evolved across the phases. The purpose of this is to elaborate on how the terms emerged and spread online in Denmark. The premise of the method is that a higher occurrence of the term ‘streamingtjeneste’ or ‘streaming service’ per Web site is an indication that the actors were important and had influence in the area. I identify actors that included the term ‘streamingtjeneste’ or ‘streaming service’ five or more times across each temporal phase (month).

I then classified the actors in categories corresponding to the actors’ identity and the type of content published on the Web site. I sometimes had to look up the actor in the Web archive if I was in doubt about the appropriate category for a specific actor. Tables 1 and 2 presents the results of this second method applied to my dataset.

In Table 1, I include the top 20 main actors for each phase. This mapping of actors involved in the evolution of the use of the term ‘streamingtjeneste’ seems to record the discourses about streaming services rather than discourse from the streaming services themselves. A heterogeneous collection of actors characterizes the five phases. However, news Web sites dominate the phases, especially those in the
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

category general news and business news, both of which are common to four of the five phases. This includes the public service media the Danish Broadcasting Corporation (DR), which is also a prominent general news Web site [11]. In the second phase, Web sites of music-related organizations and music news appear. Nevertheless, in the third phase general news and tech news from Web sites that specialize in video-based streaming services (recordere.dk, flatpanels.dk) supersede music-related Web sites. This suggests that streaming becomes a general topic of interest covered by common top news sites and technology-oriented news sites. This correlates with the general development in streaming services since their inception where music spearheaded the expansion while subscription-based video streaming services (SVODs) took over in terms of number of services and number of subscribers.

Another noteworthy development gleaned from the actors that mention ‘streamingtjeneste’ the most is the persistence of business news among the top actors, e.g., the overlapping presence of the nationwide (omnibus press) business newspaper, borsen.dk, and specialized media news service, mediawatch.dk, in four of the five phases. What I cannot say with this method is whether they are mentioning international or national streaming services.

A number of factors are notable, when comparing to the more specialized term ‘mHealth’ whose evolution was explored in the proposed methods (Fage-Butler, et al., 2022). First, the number of actors, especially news Web sites suggests that ‘streamingtjeneste’ is a more common and widely used term on the historical Danish Web. Second, the popularity of the term and the widespread use seems to reflect how news Web sites is one of the categories of Web sites that the Danish Web archive has consistently archived in depth (selective collection) since it began collecting Danish digital cultural heritage online in 2005 (Royal Danish Library, n.d.b).

YouTube is the only streaming service to appear among the actors and it did so in the fourth phase. YouTube is also the most widely used general purpose streaming service, the use of which is able to encompass all the dimensions of streaming (Spilker and Colbjørnsen, 2020). YouTube hosts the legal and the illegal, the professional and the user-generated content both live and on-demand. All the while, YouTube caters to the general audience and to niche audiences, who use it for focused stream viewing or mixed interactions that include commenting and publishing. The Danish Web archive uses a specialized collection method to collect Danish digital cultural heritage in the shape of selected YouTube profiles, YouTube channels, and their content. It is a manual method alongside its general automated Web collecting tools. Besides YouTube, the absence of the most popular video streaming services among the actors seems to relate to them not having Web sites on the Danish top-level domain (.dk). With 85 percent of the 24.2 billion Web pages in the Web archive, the Web pages on the .dk domain dominate the collection. This skews the analysis. Even though the streaming services Netflix and Prime Video have Web sites in Danish, these sites are not collected to the same extent as general news Web sites or they simply did not use the term ‘streamingtjeneste’.

Table 1: Five phases and their main actors.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Actors</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1: 2003 — 2009 March (total actors: 6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Governmental Web site (children’s culture)</td>
<td>boernekultur.dk</td>
<td>2</td>
</tr>
<tr>
<td>Corporate Web site</td>
<td>sony.dk</td>
<td>2</td>
</tr>
<tr>
<td>Danish film</td>
<td>dfi.dk</td>
<td>2</td>
</tr>
</tbody>
</table>
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

<table>
<thead>
<tr>
<th>Institute</th>
<th>Web Site</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile phone vendor Web site</td>
<td>nds1.nokia.com</td>
<td>2</td>
</tr>
<tr>
<td>Corporate Web site</td>
<td>dk.playstation.com</td>
<td>1</td>
</tr>
<tr>
<td>Online Web shop mobile phones</td>
<td>goblue.dk</td>
<td>1</td>
</tr>
</tbody>
</table>

**Phase 2: 2009 April — 2012 December (total actors: 327)**

<table>
<thead>
<tr>
<th>Phase 2</th>
<th>Web Site</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>News, tech</td>
<td>comon.dk</td>
<td>28</td>
</tr>
<tr>
<td>Public service</td>
<td>media.dr.dk</td>
<td>28</td>
</tr>
<tr>
<td>News, general</td>
<td>politiken.dk</td>
<td>26</td>
</tr>
<tr>
<td>Public service media, sub site about tech</td>
<td>beep.tv2.dk</td>
<td>21</td>
</tr>
<tr>
<td>Private blog, music</td>
<td>capac.dk</td>
<td>18</td>
</tr>
<tr>
<td>News, general</td>
<td>information.dk</td>
<td>14</td>
</tr>
<tr>
<td>News, general</td>
<td>b.dk</td>
<td>14</td>
</tr>
<tr>
<td>News, business</td>
<td>business.dk</td>
<td>13</td>
</tr>
<tr>
<td>Music trade organization</td>
<td>ifpi.dk</td>
<td>13</td>
</tr>
<tr>
<td>News, business</td>
<td>epn.dk</td>
<td>11</td>
</tr>
<tr>
<td>News, tech</td>
<td>bitzonen.dk</td>
<td>10</td>
</tr>
<tr>
<td>News, business</td>
<td>borsen.dk</td>
<td>10</td>
</tr>
<tr>
<td>News, general</td>
<td>news.dk</td>
<td>10</td>
</tr>
<tr>
<td>News, tech</td>
<td>newz.dk</td>
<td>10</td>
</tr>
<tr>
<td>News, general</td>
<td>overskrift.dk</td>
<td>10</td>
</tr>
<tr>
<td>News, tech</td>
<td>gadgetzonen.dk</td>
<td>10</td>
</tr>
<tr>
<td>News, general</td>
<td>ekstrabladet.dk</td>
<td>9</td>
</tr>
<tr>
<td>News, tech</td>
<td>mobil.nu</td>
<td>9</td>
</tr>
<tr>
<td>Corporate Web site</td>
<td>dk.playstation.com</td>
<td>9</td>
</tr>
<tr>
<td>Private company and blog, music</td>
<td>promote-it.dk</td>
<td>9</td>
</tr>
</tbody>
</table>
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

**Phase 3: 2013 — 2016 February (total actors: 1,940)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Source</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>News, general</td>
<td>politiken.dk</td>
<td>38</td>
</tr>
<tr>
<td>Social media</td>
<td>twitter.com</td>
<td>38</td>
</tr>
<tr>
<td>News, general</td>
<td>jyllands-posten.dk</td>
<td>38</td>
</tr>
<tr>
<td>News, general</td>
<td>b.dk</td>
<td>38</td>
</tr>
<tr>
<td>News, general</td>
<td>bt.dk</td>
<td>38</td>
</tr>
<tr>
<td>Public service</td>
<td>media dr.dk</td>
<td>38</td>
</tr>
<tr>
<td>News, business</td>
<td>mediawatch.dk</td>
<td>37</td>
</tr>
<tr>
<td>News, general</td>
<td>ekstrabladet.dk</td>
<td>36</td>
</tr>
<tr>
<td>News, general</td>
<td>information.dk</td>
<td>35</td>
</tr>
<tr>
<td>Social media</td>
<td>facebook.com</td>
<td>33</td>
</tr>
<tr>
<td>Radio station archive</td>
<td>arkiv.radio24syv.dk</td>
<td>33</td>
</tr>
<tr>
<td>News, general</td>
<td>dagens.dk</td>
<td>31</td>
</tr>
<tr>
<td>News, business</td>
<td>borsen.dk</td>
<td>30</td>
</tr>
<tr>
<td>News, tech</td>
<td>flatpanels.dk</td>
<td>30</td>
</tr>
<tr>
<td>Public service media, sub site about tech</td>
<td>beep.tv2.dk</td>
<td>30</td>
</tr>
<tr>
<td>News, general</td>
<td>nordjyske.dk</td>
<td>29</td>
</tr>
<tr>
<td>Regional news, general</td>
<td>fyens.dk</td>
<td>29</td>
</tr>
<tr>
<td>News, general</td>
<td>news.dk</td>
<td>27</td>
</tr>
<tr>
<td>News, tech</td>
<td>recordere.dk</td>
<td>27</td>
</tr>
<tr>
<td>News, tech</td>
<td>computerworld.dk</td>
<td>26</td>
</tr>
</tbody>
</table>

**Phase 4: 2016 March — 2019 August (total actors: 2,146)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Source</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>News, general</td>
<td>bt.dk</td>
<td>41</td>
</tr>
<tr>
<td>News, general</td>
<td>ekstrabladet.dk</td>
<td>41</td>
</tr>
<tr>
<td>News, media business</td>
<td>mediawatch.dk</td>
<td>41</td>
</tr>
<tr>
<td>News, culture</td>
<td>soundvenue.com</td>
<td>41</td>
</tr>
<tr>
<td>News, general</td>
<td>sondagsavisen.dk</td>
<td>41</td>
</tr>
<tr>
<td>Social media</td>
<td>twitter.com</td>
<td>41</td>
</tr>
<tr>
<td>News</td>
<td>finans.dk</td>
<td>40</td>
</tr>
</tbody>
</table>
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

<table>
<thead>
<tr>
<th>Phase 5: 2019 September — 2021 December (total actors: 3,461)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public service media</td>
</tr>
<tr>
<td>News, general</td>
</tr>
<tr>
<td>News, general</td>
</tr>
<tr>
<td>News, general</td>
</tr>
<tr>
<td>News, business</td>
</tr>
<tr>
<td>News, politics</td>
</tr>
<tr>
<td>News, gossip</td>
</tr>
<tr>
<td>Regional news, general</td>
</tr>
<tr>
<td>News, general</td>
</tr>
<tr>
<td>News, general</td>
</tr>
<tr>
<td>News, business</td>
</tr>
<tr>
<td>News, gossip</td>
</tr>
</tbody>
</table>
I provide a simple comparison of actors and categories involved in the use of the two terms. Given the small number of actors that have mentioned ‘streaming service’, I chart these in Table 2 without dividing into phases. Both terms first occurred in 2003 but the term ‘streaming service’ occurred three months before ‘streamingtjeneste’. The earliest occurrences were on Web sites in the Danish top-level domain (.dk).

<table>
<thead>
<tr>
<th>Categories</th>
<th>Actors</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blog by independent trade union</td>
<td>blog.it-faggruppen.dk</td>
<td>38</td>
</tr>
<tr>
<td>Corporate blog about children’s TV</td>
<td>bornetv.dk</td>
<td>26</td>
</tr>
<tr>
<td>Private blog, consumer’s saving tips</td>
<td>mest-for-pengene.dk</td>
<td>26</td>
</tr>
<tr>
<td>The Royal Danish Library’s old Web site</td>
<td>statsbiblioteket.dk</td>
<td>24</td>
</tr>
<tr>
<td>Social media</td>
<td>twitter.com</td>
<td>19</td>
</tr>
<tr>
<td>Private blog, tech</td>
<td>blog.digital-</td>
<td>16</td>
</tr>
</tbody>
</table>
A heterogeneous collection of actors also characterizes the smaller ‘streaming service’ dataset. The majority of these Web sites are on the Danish top-level domain (.dk). I assume these Danish actors either write in English or chose to use the English word. The two datasets have the following actors in common: YouTube, Twitter, and Danish public service media TV2. These are large Web sites with high numbers of contributors and it seems a fraction of them chose to write ‘streaming service’ instead of

<table>
<thead>
<tr>
<th>and fashion</th>
<th>kingdom.dk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate support Web site</td>
<td>support.micrsoft.com 15</td>
</tr>
<tr>
<td>Private blog, musician</td>
<td>sorenraaschou.dk 15</td>
</tr>
<tr>
<td>Online forum, music and tech</td>
<td>hoved-fi.dk 15</td>
</tr>
<tr>
<td>News, tech</td>
<td>techstart.dk 14</td>
</tr>
<tr>
<td>Private blog, TV-series</td>
<td>gode-serier.dk 12</td>
</tr>
<tr>
<td>Music festival Web site</td>
<td>roskilde-festival.dk 12</td>
</tr>
<tr>
<td>News, films</td>
<td>moovy.dk 12</td>
</tr>
<tr>
<td>Public service media, sub site about tech</td>
<td>beep.tv2.dk 11</td>
</tr>
<tr>
<td>News, business</td>
<td>trendsonline.dk 10</td>
</tr>
<tr>
<td>Corporate consultancy, radio</td>
<td>radioassistant.com 10</td>
</tr>
<tr>
<td>News, tech and culture</td>
<td>railgun.dk 9</td>
</tr>
<tr>
<td>Streaming service</td>
<td>youtube.com 9</td>
</tr>
<tr>
<td>Online forum, music</td>
<td>svingninger.dk 8</td>
</tr>
<tr>
<td>Private blog, writer</td>
<td>gittemieeriksen.dk 8</td>
</tr>
</tbody>
</table>
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

‘streamingtjeneste’.

Interestingly, one of the Royal Danish Library’s own Web sites (statsbiblioteket.dk) is among the main actors that use the word ‘streaming service’ and this merited closer investigation. From 2012 to 2015 Bibzoom.dk was a music streaming service based on a collaboration between the Danish music industry and the public libraries. The Royal Danish Library is part of the consortium of libraries that run Bibzoom.dk, which is no longer a music streaming service. Bibzoom.dk was by no means the largest streaming service in its time. It seems to appear here because it used the term streaming service while most other Danish Web sites used the Danish term. Moreover, Bibzoom.dk is an internally developed service, which could provide the basis for an optimal collection of the Web site and its Web pages. This could lead to its place among the main actors that mention ‘streaming service’.

‘Streamingtjeneste’ actor connections in networks

The third part of the proposed methods involves network analysis based on hyperlinks. This offers a mapping of which actors were the most important and how this has changed over time. A network analysis provides insights into the discursive Web environment in which the actors, identified in the second part, were embedded and whether they were important. The proposed methods provides a set of data that allows various scales of analysis across years and months. To prove the method works for a larger dataset I provide three network graphs for 2010, 2014, and 2021 of the phases two, three, and five, see Table 1. I focus on selected months of the years for ‘streamingtjeneste’ that showed the greatest variety of actors. Here, I do not analyze the ‘streaming service’ data using network graphs.

With a degree of edge pruning, we can better understand large network graphs. The Gephi network visualization tool allows me to prune the network graphs by not including the Web sites with a small number of links. This makes better sense of the most linked Web sites and ‘clears up’ the network visualization to allow better analysis of patterns such as clusters or recurring actors. The resulting graphs show between four percent and 0.8 percent of the Web sites that I extracted from the dataset. This is because I exclude Web sites with few links relative to the number of nodes and single out one month to analyze. The resulting network graphs and the data behind their creation are included in the open access data repository (Ægidius, 2023).

Following the suggestion from the proposed methods, I compare the graphs for outgoing and ingoing links. Figures 4 and 5 show network graphs for December 2014. The size of a node is set to indicate the number of links to or from it (edges). The bigger nodes have a greater number of edges.
Figure 4: Network of hyperlinks, ‘Streamingtjeneste’, December 2014, outgoing links.
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

The network graphs confirm the large number of collected news Web pages in the Danish Web archive that use the term ‘streamingtjeneste’ as summarized by the second method.

Social media, large public service media Web sites, and streaming services have many links pointing to them. The network graphs for 2010 and 2021 mirrors the centrality of social media evident in the 2014 network graphs. The primary reason social media feature so prominent in terms of ingoing links is most likely because the majority of the news Web sites in Denmark that dominate the dataset have links to Facebook, Twitter, etc. It is worth noting that social media Web sites are hubs for user-generated content related to and perhaps coming directly from streaming services of various kinds. However, the method does not include analysis of textual content that would let us know which streaming services are being talked about and why this may be.

Web sites of streaming services have ingoing links from other Web sites shown as medium sized nodes in Figure 5 and few outgoing links shown as small nodes in Figure 4. The hybrid video streaming service YouTube shows the largest difference between ingoing and outgoing links. The pattern is similar for the video streaming services Netflix, HBO Nordic, and Viaplay as well as the music streaming services Spotify and Wimp.

Gephi calculates the relationship between the Web sites and suggests clusters. Therefore, we cannot relate
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

the color-coded clusters in Figure 4 and Figure 5 to my categorization of actors in Table 1 and Table 2. Unfortunately, Gephi assigns colors at random. This means that clusters across the years have different colors and placements. However, there are similarities in broad terms since typical general news Web sites form a cluster and tech news Web sites form a cluster, the blue and green clusters in Figure 4. The subscription-based video streaming services are part of the green cluster on the left side of the 2014 network while music streaming services and music-related Web sites are part of the teal and dark grey clusters on the right side of the 2014 network. The Web pages related to music streaming that were prominent in the early phases linger in 2014. Figure 4 shows that the music magazine and Web site Gaffa.dk and the private music blog Capac.dk have high numbers of outgoing links.

I look across the three sampled years, when analyzing the presence of the dimensions of streaming outlined by Spilker and Colbjørnsen (2020). In 2010, 2014, and 2021 the streaming services that have been collected lean into the following dimensions of streaming: the legal and on-demand with professional content and I identify mostly single-purpose services that appeal to general audiences. The biggest deviations from those dimensions are YouTube, as already mentioned, along with DR and TV2. The latter are both public service media. As multi-purpose services, a part of their Web sites provides a broad scope of news coverage while other parts of their sites host their respective live and on-demand video streaming services with an equally broad selection of content including news and entertainment. From crosschecking manual searches in the SolrWayback interface with the outputted domain statistics, I know that the method counted URLs that used the slash parameter, e.g., dr.dk/drtv [12]. The TV2 SVOD operates on a subdomain, play.tv2.dk. However, it has been pruned because a low number of its Web pages mentions ‘streamingtjeneste’. This suggests that differences in naming logics for Web sites and the high variance of Web site architecture and structures challenge the identification of streaming services’ Web pages that exist as part of large Web sites.

Size of the node of Netflix, the single-purpose legal on-demand service with professional content, suggests its dominance (measured ingoing links) among SVODs of the same type and supports its historic and current top market position in Denmark.

I note the absence of sports Web sites and pornography Web sites among the main actors, both of which have played a massive role in transforming streaming technologies (Hutchins, et al., 2019; Paasonen, 2019). I assume they simply do not use the particular terms and favor terms such as ‘watch’, ‘stream’, and ‘live’. It is also possible that this is a bias of the collection itself since the Web curators optimize the collecting of the Danish Web in adherence to selection parameters that do not include special attention to live streaming nor sports or online pornography (Royal Danish Library, n.d.b).

---

Discussion

I refer to the discussion of the proposed methods by Fage-Butler, et al. (2022) given the immediate applicability of most of their points for this test of their method [13]. I add a number of discussion points related to the test, the size of the datasets, my search terms with differences in semantic properties, and finally my findings about the emergence and evolution of the terms ‘streamingtjeneste’ and ‘streaming service’ in the Danish Web domain.

The method, specifically the notebook (Rmarkdown script) can be modified to process several search terms at the time. This would entail modifications to several chunks of script and adjustments in their order. My test suggests that the script, which provides output for all three of the tested methods, is a suitable base for further investigation of the archived Web given similar or adjustable data fields. Based on my experience with adapting the script chunks, I suggest the notebook could be adapted to sample the data or split datasets into manageable batches if you want to process larger datasets, like the ‘stream’-dataset described above in the section on methodology. For strategies for working with big data in Rstudio, see Gold (2019). I found larger network graphs require more experiments with layouts and especially filtering for analytical
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

I made two crucial adjustments that relate to working with larger datasets of metadata about the archived Web. First, the adjustments have made the script accept (rather than transform) datasets that contain special characters. Second, the transformation of the links dataset is now properly set up to output a link count importable to Gephi without further external processing in Rstudio or Excel. I have annotated the adjustments to the script in the .rmd file that is available in the open access data repository (Ægidius, 2023).

It was my aim to test the proposed methods on larger datasets. There are limits to the size of dataset that a standard work PC can process at any given time. This test can serve as a simple benchmark where a 64-bit system with 16 GB RAM managed to process .csv files with approximately eight million rows in Rstudio.

The terms ‘streamingtjeneste’ and ‘streaming service’ are so widely mentioned that they can produce huge complex network graphs. Many analytical experiments with layouts and edge pruning were necessary to achieve meaningful insights in the network analysis. I generated insights about actors, their position and relation to other actors. In this regard, analysis of network clusters is a powerful tool to determine relations between actors. The dominance of various types of news Web sites meant that this category of actors was prominent. I assume that we could achieve a more granular view of the evolution of the Web sites of streaming services providers by way of a more closed approach with an outset in the specific URLs of streaming services.

Likewise, adjacent approaches include analysis of text content from the Web sites alongside contextual and extended searches. Of particular interest are similar Internet historical terms like ‘Webcasting’, live, or ‘on-demand’ by themselves or in collocation with stream-based terms. As already mentioned, it is possible to adapt the method to process several search terms at the same time.

Commercial news Web sites is one of the categories that the Danish Web archive collects more frequently in the selective collections while they run three–four cross-sectional collections yearly (Royal Danish Library, n.d.b). The removal of duplicates and choice of versions should mitigate this. However, dynamic Web sites have minute changes that makes otherwise identical Web pages distinct during collection. When discussing this with the Web curators they pointed out that dynamic Web sites present a number of challenges. The non-commercial music blog capac.dk and the local antenna association aovnet.dk in Figure 4 are two examples of relatively small Web sites, based on the WordPress platform, which the Web curators know cause problems. Web archiving professionals refer to this phenomenon as crawl traps that can occur when dynamic functionalities of the Web site create feedback loops that trap the Web crawler software that collects Web pages (Maemura, et al., 2018). Hence, crawl traps are a known noise source in large datasets of archived Web sites. Deduplication and filtering of versions built into the method addresses this issue. Another development that introduces noise and bias concerns trends in management of outgoing links. Non-commercial actors see no harm in sharing information in the form of links other Web sites while commercial entities will try to minimize the use of outgoing links. The latter maximizes customer retention on their own Web site(s) and minimizes the risk of having ‘dead links’ that score negatively during search engine optimization. I recommend discussions with Web curators since analytical insights garnered from collection issues and technological developments evolve as fast as the Web itself.

---

Conclusion

The overall aim of my study was to test and help validate a method to explore the emergence and evolution of the terms in the Danish Web domain. Simultaneously, I used the same method to explore selected terms with semantic variants that resulted in larger datasets. I mapped the terms ‘stream’, ‘streaming service’, and ‘streamingtjeneste’. Terms beginning with ‘stream’ are widely distributed in the Web Archive from its earliest collected material in the late 1990s to the present. The frequency of ‘stream’ rises from 2003. It has
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

been used in the invisible Web as a function in HTML code and in the visible Web for a range of terms unrelated to streaming services. The mention of the terms ‘streamingtjeneste’ and ‘streaming service’ first appeared and rose in frequency from 2003. I identified the actors that were part of their evolution. Because of the size of dataset, I did not process the ‘stream’ dataset using the script. The more specific terms ‘streaming service’ and ‘streamingtjeneste’ were predominately mentioned on news Web sites though a heterogeneous collection of actors characterized the five phases.

Web sites of streaming services were among the actors that came to the fore in the network graphs that showed ingoing links. Although news Web sites dominated the network graphs, a number of legal on-demand single purpose streaming services with professional content seemed to remain in the network from 2014 onwards. Furthermore, they lean towards a general audience appeal. I associate the representation of these dimensions of streaming with the large number of general, tech and business news Web sites that link to them. YouTube and the public service media stood out from this pattern and formed an alternative fraction of the dimensions of streaming. They were fewer in number but measured in amount of ingoing link they were the more dominant dimensions of streaming represented in the network analyses. Taken together, YouTube, DR and TV2 represented the user-generated dimension and the multi-purpose service dimensions of streaming. I have presented an open explorative approach following the approach of the proposed methods. An urgent follow-up should be a more closed approach to look for specific Web sites that represent the various dimensions of streaming to track their history and evolution in the archived Web.

It is not possible to elaborate fully on whether the method is applicable at a global scale. For instance, the Internet Archive has a different setup, e.g., their commercial branch (Archive-IT, 2014). Fage-Butler, et al. (2022) anticipate that Web archives with similar data structures and practices could implement and further develop their proposed methods that I have tested. The data fields of your local Web archive index may differ. Therefore, it is best to consult your local Web archivist.

About the author

Andreas Lenander Ægidius, Ph.D., is a research librarian at the Department of Digital Cultural Heritage at the Royal Danish Library, Denmark. His research interests include streaming media, digital cultural heritage, Web and media history, and digital methods.
ORCID: https://orcid.org/0000-0002-7640-1236
E-mail: anae [at] kb [dot] dk | andreas [dot] aegidius [at] gmail [dot] com

Funding

A grant from the Ministry of Culture Denmark funded the research project that provided insights for this paper: FPK-2021-0004.

Acknowledgements

I would like to thank the following staff at the Royal Danish Library for their invaluable support and curiosity: program manager for The Danish Web archive (Netarkivet) Anders Klindt Myrvoll, Web curators and technicians Thomas Martin Elkjær Smedebøl, Stephen Hunt, and Thomas Eegense. I am also thankful for the assistance and advice I received from IT developer Ulrich Karstoft Have at the School of Communication and Culture at Aarhus University, and special consultant Christian B. Knudsen at KUB Datalab at the Royal Danish Library.
Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

Notes

4. The in-house skills test was part of different work package of the research project based on action research (M.C. Sørensen, and M. Snorre Wilms Boysen, 2021).
6. This is the search query for ‘streamingtjeneste’: (text:streamingtjeneste* OR text:streaming?tjeneste* OR url_search:streamingtjeneste* OR url_search:streaming?tjeneste*).
7. Personal correspondence with the authors of the proposed methods confirmed that they did not use the text content even though the article states that “… only the text content and link data were used” (Fage-Butler, et al., 2022). This I confirmed by the fact that the data treatment performed with the Notebook (Rmarkdown script) does not read the textcontent.csv.
9. E.g., English literature published online that reference water streams, a commercial ad print to put on the side of vehicles (streamers), the carbonation home consumer product SodaStream, the car Honda Stream, a type of fly-fishing hook (streamer), a person publishing live recordings online of themselves and their activities (streamer).
10. Fage-Butler, et al., 2022, Figure 2.
11. In 2021–2022, the domain DR.dk was among the top five most viewed and visited Web sites in Denmark according to the “Danish online index” published online by the Web analytics company Gemius (n.d.).
12. Searching for specific URLs in the raw dataset is the alternative method of checking to identify the inclusion of the DR URLs, from 2012–2014 “/nu”, 2014–2016 “/drtv”.

References


Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive

4, pp. 962–982.


Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive


Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive


Editorial history

Received 10 May 2023; accepted 26 November 2023.

This paper is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Testing methods to explore the evolution of ‘stream’-related terms in the Danish Web archive by Andreas Lenander Ægidius.

First Monday, volume 28, number 12 (December 2023).

doi: https://dx.doi.org/10.5210/fm.v28i12.13189