Labservatory: a synergy between journalism studies and computer science for online news observation

Online Information Review

Received 10 October 2023 Revised 21 February 2024 15 April 2024 Accepted 7 May 2024

Dimitris Trimithiotis

Department of Social and Political Sciences, University of Cyprus, Nicosia, Cyprus, and

Iacovos Ioannou, Vasos Vassiliou, Panicos Christou, Stelios Chrysostomou, Erotokritos Erotokritou and Demetris Kaizer Department of Computer Science, University of Cyprus, Nicosia, Cyprus

Abstract

Purpose – This article explores the synergy between journalism studies and computer science in the context of observing online news. By establishing web applications of online media observatories as research tools, researchers can employ various analytical approaches to gain valuable insights into online news discourse and production.

Design/methodology/approach – Drawing eight months of data (01.08.2022–30.04.2023) from the Labservatory's web application, i.e. over 250,000 news items, the article demonstrates how some of this web application's main functionalities may be useful in implementing (1) news flow analysis, (2) news topic distribution analysis and (3) media discourse analysis.

Findings – The capabilities provided by this web application, (1) to simultaneously analyse the daily news production of ten media outlets with varying features, (2) to rapidly collect a large volume of news items, (3) to identify the news categories as classified by the media themselves, (4) to present the results of the search in relevance order and (5) to automatically generate a search report, highlight the significance of this interdisciplinary collaboration for implementing comprehensive analyses of online news.

Originality/value — The article concludes by emphasising the importance of continuing this joint effort, as it opens new avenues for further research and provides a deeper grasp of the intricate relationship between journalism, technology and society in the digital era. The Labservatory also contributes to society since it may be used by the broader public for immediate access to more pluralistic information and thus for promoting both news media literacy and news media accountability.

Keywords Media discourse, Online journalism, Artificial intelligence (Al), Media observatories **Paper type** Research paper

Introduction

Journalism and media observatories, are initiatives that attempt to monitor and analyse the state of the media industry, including trends in journalism, changes in media ownership and the impact of new technologies in the field. In the last decades journalism and media

© Dimitris Trimithiotis, Iacovos Ioannou, Vasos Vassiliou, Panicos Christou, Stelios Chrysostomou, Erotokritos Erotokritou and Demetris Kaizer. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

In the interest of transparency, data sharing and reproducibility, the author(s) of this article have made the data underlying their research openly available. It can be accessed by following the link here: https://labservatory.eu/

Funding: This work was supported by University of Cyprus Start-Up Scheme Grant.



Online Information Review Emerald Publishing Limited 1468-4527 DOI 10.1108/OIR-10-2023-0531 observatories have become increasingly important since the media landscape has undergone significant changes, such as the proliferation of digital media and online journalism alongside the decline of print journalism, as well as the exponential increase of disinformation and other discriminatory discourses in journalistic and media production processes (i.e. Omar *et al.*, 2021; Villanueva-Ledezma *et al.*, 2020; Xiao and Yang, 2024). The aim of this article is to present the research impact of an online journalism observatory, namely Labservatory. The latter is product of an interdisciplinary synergy between researchers in journalism studies and computer science. Drawing from the data collection and analysis of the first eight months of its "life" (01.08.2022–30.04.2023), the article discusses the potentialities for new research brought by this synergy, as they are reflected through the creation of the Labservatory web application [1]. Specifically, the article examines how the web application of Labservatory, may contribute in news flow analysis (RQ1); news topic distribution analysis (RQ2) and news discourse analysis (RQ3).

The article contributes to expose the significance of the interdisciplinary collaboration between journalism studies and computer science in understanding the evolving landscape of online news. It emphasises the importance of continuing this joint effort, as it opens new perspectives for further research and provides a deeper grasp of the complex relationship between journalism, technology and society in the digital era. The article is structured into three main sections. It begins with a literature review that provides an overview of the main features of journalism and media observatories and their potential in contributing to the journalistic field – both as professional practice and as a scientific research area. The second section presents the methodology on which the Labservatory web application is based in collecting, categorizing and analysing online news. The third section presents the preliminary results based on the data collected during the first eight months of the functioning of the observatory and opens up new potentials for further research in media and journalism studies.

Journalism and media observatories: a research perspective

Evolution, role and challenges

Journalism and media observatories are organisations that monitor and analyse media coverage, with the goal of improving media transparency and accountability. Although existing scientific literature on media observatories is rather scarce, it does nevertheless cover various aspects of their operations, including elements that pertain to their structure, mandate, impact and challenges. This section offers a brief overview of the origins and evolution of journalism observatories and explores their roles and functions, including their input in monitoring and analysing the media industry, advocating for press freedom and media pluralism and providing support and training for journalists. It also examines the impact of journalism observatories on the field of journalism, including their role in shaping media policy and public discourse about the media industry.

Origins and evolution of journalism observatories. Journalism observatories have their roots in the early 20th century when media scholars and activists began to study the media industry and its impact on society. The French Press Institute, founded in Paris in 1925 could be apprehended as one of the earliest examples of media and journalism observatories. The institute's mission was studying the press and its societal role. In the post-World War II era, the focus of journalism observatories shifted from studying the media industry to advocating for press freedom and media pluralism. The International Press Institute (IPI), founded in 1950, was one of the first organisations to focus on press freedom and media pluralism. In the 1970s and 1980s, journalism observatories began to expand their focus to include issues such as media ownership and concentration. In the United States, the Media Access Project (MAP), founded in 1973, was one of the first organisations to focus on media ownership and

concentration. The MAP advocated for increased diversity in media ownership and fought against the consolidation of media companies. In the 21st century, journalism observatories have been increasingly turning their attention to the impact of new technologies on the media industry and journalistic production. The Centre for International Media Assistance (CIMA), founded in 2006, is one example of a journalism observatory that focuses on the effect of new technologies on the media industry and simultaneously provides research and training on digital media and its impact on journalism. Current preoccupations of media and journalism observatories mainly include the rise of digital media and the proliferation of disinformation, hate speech, xenophobia and other types of discriminatory discourses in news media (Villanueva-Ledezma *et al.*, 2020).

Structure and roles of journalism observatories. According to Smith *et al.* (2013), there are three main types of media observatories: (1) independent media observatories, run by civil society organisations or NGOs; (2) academic media observatories, affiliated with universities or research institutes; and (3) state-run media observatories, established by governments or regulatory bodies. Similarly, Cedillo and Carretero (2016) also confirm the existence of a heterogeneous landscape of media and journalism observatories, while also arguing that his said landscape is characterised by a strong focus on auditing activities and observatories generated by universities. Each type has its advantages and limitations, and the choice of structure depends upon the goals and resources of the observatory.

Literature shows that journalism observatories serve a variety of roles and functions in the contemporary media landscape. One of their overarching roles is to monitor and analyse the media industry, including trends in journalism, changes in media ownership and the impact of new technologies in the field. Journalism observatories provide critical data and analyses on the state of the media industry, which can be utilised by journalists, policymakers and the public to better understand the challenges and opportunities facing the field. According to Cammaerts and Carpentier (2007), the main functions of media observatories include monitoring media content, analysing media performance and promoting media accountability. Other studies have highlighted additional roles, such as providing media literacy training, promoting media pluralism, serving as a watchdog for media regulation and overall developing ethical and socially responsible journalism (Villanueva-Ledezma et al., 2020).

Impact and challenges of journalism observatories. Since the late 20th century, media observatories have played an important role among institutions devoted to the promotion of media reform, since they constitute an essential tool for the monitoring and analysis of the media by citizens (Cedillo and Carretero, 2016). Several studies have shown that media observatories can have a positive impact on media performance and transparency and can increase media accountability and reduce bias in news coverage (i.e. Golan and Yang, 2009). Furthermore, according to Villanueva-Ledezma et al. (2020), journalism observatories can even lead to the development of an ethical and social responsibility accreditation that can be granted to news outlets with the best practices for journalism. In terms of scientific research, journalism and media observatories may serve as important tools in raising new research questions related to media discourses and news production values and journalistic deontology (Villanueva-Ledezma et al., 2020).

However, other studies have highlighted the challenges of measuring the impact of media observatories, given the complexity of the media landscape and the difficulty of attributing changes in media behaviour to the observatories themselves. It is also important noting the organisational and ethical challenges faced by media and journalism observatories, such as funding constraints, political interference and the challenging task of maintaining credibility and safeguarding independence. Relevant literature highlights the categorical need for media and journalism observatories to be transparent and accountable in their own operations, in order to maintain their legitimacy and effectiveness. In terms of scientific research, the work of Cedillo and Carretero (2016) – which is based on qualitative interviews with observatories'

directors in Spain – observes that most observatories are characterised by unsustainability and by an irregular research activity. Overall, while there are challenges and limitations to their operations, media and journalism observatories can play a key role in promoting media transparency and accountability and in contributing to a more informed and engaged citizenry through both traditional and innovative scientific research perspectives.

Labservatory as tool for researching online journalism in Cyprus

In our case, the Labservatory's primary goal is academic (Smith *et al.*, 2013). Affiliated with a higher education institution, the University of Cyprus, it aims at generating analytical data related to online journalism in Cyprus and based on these empirical data, to advance theoretical reflections and discussions around the broader journalistic process in online news media.

Over the last years, online news media have become essential in the Cypriot media landscape. Dozens of online-only news sites have emerged during the last decade and all traditional media organisations now have their own news portals. As online journalism process requires different competences from journalists, such as multiskilled staff, social media management and continuous news production have become important characteristics of digital journalism in Cyprus (Sahin, 2022b; Trimithiotis and Stavrou, 2022). The online development of Cypriot journalism has nonetheless emerged within a context of economic recession, contributing to normalising online journalists' poor working conditions and weak professional status. This financial condition of the sector makes online journalism in Cyprus even more vulnerable towards new economic turbulences, such as the one provoked by the covid-19 pandemic (Vatikiotis et al., 2023; Price et al., 2023). Online news media in Cyprus adopt a rather market-oriented mode of production, which lead to promptness to gain more importance in news production and dissemination. This pressure often makes it challenging to balance immediacy with accuracy (Sahin, 2022b). This also affects the overall reporting process and content of online journalism in Cyprus, favouring churnalism as the main mode of news production (Trimithiotis, 2020), which paradoxically contributes to maintain the historically formed close ties of the journalistic field with political and other power elites through the uncriticised reproduction of their press releases and other strategic communication (Christophorou, 2010; Maniou and Photiou, 2017). This negatively impacts the quality of the reporting on both traditional and novel major issues of the Cypriot society such as the Cyprus conflict (Sahin, 2022a, c) and the refugee crisis (Trimithiotis and Voniati, 2023).

This article, examines some of the main types of analysis that the web application of Labservatory, may be used in elaborating them, namely, (1) news flow analysis, (2) news topic distribution analysis and (3) news discourse analysis.

News flow analysis. In media and journalism studies, news flow analysis refers to the study of the rhythm of the news production and distribution within and across media organisations (Lim, 2012). This type of analysis seeks to understand the ways in which news are selected and presented to audiences, as well as to discuss the factors that influence these processes. While it is often associated with the process of globalisation, news flow analysis in media and journalism studies essentially involves shedding light on newsroom routines and editorial practices, as well as on the influence of external macro-factors, such as political and economic aspects, to understand how news are produced and distributed. There are three main research perspectives in understanding news flow (Schudson, 2002). The first perspective involves a political economy approach, linking the news processes to the structure and the features of the macro-economic and political power in different states. The second perspective draws on a sociology of organisations and focuses on the examination of the structure of the news organisations themselves along with the occupational routines of their actors to

understand media content. The third perspective approaches news as a form of culture that incorporates general belief systems, norms and values into news writing. In summary, these three approaches suggest that news is a product resulting from political, economic, social and cultural forces (Schudson, 2002). Furthermore, news flow analysis may also involve analysing the flow of specific news topics across different media platforms, such as traditional and newer news outlets and social media platforms (Karayianni and Psaltis, 2023), to produce insights on the current hybrid media systems (Chadwick, 2017).

Overall, web applications of media observatories that collect and reorganise and present news items from various online news portals are valuable tools in media and journalism studies for proceeding news flow analysis. They are useful methodological instruments for observing news distribution since they can provide insights into the factors that shape news production, distribution and consumption in contemporary media environments.

News topic distribution. News topic distribution is the second type of analysis that web applications of online media observatories are particularly useful in elaborating. This is a method of scrutinising news content that involves identifying and categorising the specific or recurring topics and issues across a set of news stories (Carvalho, 2008). The goal of news topic analysis is to provide an overview of the types of issues and events that are being covered by a particular news outlet or set of news outlets and to identify patterns and trends in news coverage, often in terms of *Meaningfulness* (Van Dijk, 1983, p. 25). To conduct news topic analysis, researchers typically select a sample of news stories from a particular news outlet or set of news outlets over a given period. They then examine these stories to identify the specific topics or thematic areas that are being covered. Conventionally, this involved manually coding the news stories. Web applications of media observatories, such as the Labservatory, use automated text analysis techniques to identify relevant keywords or categories. News topic analysis can be used to provide insights into the news agenda of a particular news outlet or a set of news outlets, as well as to track changes in the types of issues and events that are being covered over time. This type of analysis in media and journalism studies can be particularly useful for understanding how media shape public discourse and perceptions around specific issues and for identifying gaps or biases in news coverage (Bednarek and Caple, 2014).

Media discourse analysis. The third type of analysis that web applications of online media observatories may be useful in realising is the Discourse analysis. News media discourse analysis is a method of analysing the language and symbols used in news media to construct meaning and influence public opinion. This type of analysis is used in media and journalism studies to understand how news media construct and disseminate messages, as well as the ways in which audiences interpret and respond to these messages (Van Dijk, 1983). News media discourse analysis typically involves close observation of news articles or broadcasts to identify patterns of language use, such as the use of specific words, phrases, or metaphors (Carpentier and De Cleen, 2007). It also involves analysing the structure and organisation of news stories (Trimithiotis, 2022), such as the use of headlines, lead paragraphs and quotes from sources (Carvalho, 2008). Thus, through discourse analysis, researchers can identify the underlying messages that are being conveyed in news media. They can also identify the performative character of media discourses. That is, how media outlets and journalists might use language and symbols in distinct ways to construct varying narratives and audiences related to the same or different issues (Chouliaraki, 2008).

Media discourse analysis can be used to explore a wide range of topics, locally and globally, from important social and political issues such as the refugee crisis, political campaigns and social movements to the coverage of specific events such as natural disasters or terrorist attacks. By examining the language and symbols used in news media, researchers can gain insights into the ways in which news media shape public perceptions and attitudes on a variety of events and issues.

In the remaining sections of this article, we will first present the methodology on which the web application of the Labservatory is based on for the collection and analysis of online news discourses. Then, we will present some of the preliminary findings based on the analysis of data collected during the first eight months of the functioning of the observatory and open up potential possibilities for further research in media and journalism studies.

Methodology

Labservatory uses state-of-the-art techniques to collect articles from ten different news portals. Specifically, ten background jobs are executed frequently and in parallel for indexing all the latest news items using web crawlers and Really Simple Syndication (RSS) feeds. Then the content of the news items is extracted, using artificial intelligence (AI) methods and stored in a database. This section presents the sources of the data and details the methodology on which the digital platform of the Labservatory is based for the collection and classification of the online news outlets. The following functionalities are included to describe the approach so that users, both the researchers and the broader public can accurately comprehend the results.

News media

The Labservatory digital platform collects data from ten Greek language Cypriot news portals:

- (1) Alpha News, (2) Cyprus Times, (3) Dialogos, (4) Kathimerini, (5) Offside, (6) Philenews,
- (7) Politis, (8) Reporter, (9) Sigma Live and (10) Thema Online.

The selection of these ten portals provides researchers and users of the Labservatory web application with the opportunity to conduct various types of comparisons among news portals based on their research questions and hypotheses. Specifically, these media differ in relation to their journalism culture, organisational structure and volume, financial strength, degree of establishment and political identity. Thus, some are part of bigger and more established media organisations, which include at least a newspaper, or/and a radio station. or/and a television channel (i.e. Alpha news, Dialogos, Kathimerini, Philenews and Sigmalive). Others, in contrast, are part of a much smaller and internet-based media organisations (i.e. Cyprus Times, Offside and Thema online). From these ten news portals, some mainly delivers advocate journalism and having either a rather right-wing (i.e. Sigma Live) or a left-wing political identity (i.e. Dialogos). Some others while they have a certain political identity, try to preserve the label of "balanced" and "mainstream" media (i.e. Philenews, Kathimerini, Politis and Reporter). Finally, some of these news portals follow a rather tabloid and inflammatory style of reporting which clearly privilege a market-oriented journalism (i.e. Cyprus Times, Offside and Thema Online). These variations provide users-researchers of Labservatory the possibility to observe the extent to which these different characteristics influence their media discourse and news production process. Collecting data from online media with variant features within the same national context allows to scrutinise variations of the journalistic discourse and production beyond their national appurtenance (Dunaway, 2013).

News collection

There are many ways to extract articles depending on what a particular news portal provides. The simplest form of collection is with an RSS Feed. RSS (RDF Site Summary) is a web feed that allows users and applications to access website updates in a standardised, computer-readable format. Some news portals do not give access to their articles with an RSS Feed. For those, it was necessary to develop more sophisticated solutions.

First, we discuss how articles are extracted from a website that provides an RSS Feed. Periodically a computer process that runs in the background performs an HTTP GET request to the RSS server. All the new articles from the request's response are parsed and inserted into our database. The collected fields include titles, text bodies, URLs and categories. Thumbnails are also extracted and hosted on our web server. Although the RSS approach is seemingly simple, we faced issues with many news portals. The problem was that their response included only a portion of the article's body. We solved this problem with a Natural Language Processing (NLP) library that can distinguish and extract text from the core text of a response Document Object Model (DOM). Given any article uniform resource locator (URL), this library can extract the text for us. Remember that the article's URL is known from the RSS feed.

The next category includes websites that do not provide an RSS Feed. For those, we initiated web crawling, a technique widely used on the web by search engines. We have written a program that can interact with the website like humans. During the interaction, this program uses predefined Cascading Style Sheets (CSS) selectors to extract articles from the websites it interacts with. Those CSS selectors are specific to the website, and the limitations are that if any given website decides to change its appearance, we may need to redefine them.

The techniques above are not always reliable since there are rare cases in which a news item cannot be collected. This happens when the server is down, an RSS feed has missing news items, or even the crawler may fail to find some news items. For these reasons, the administrators can add the lost news item manually using a graphical user interface on the Dashboard. They can also update an existing news item if any information is missing.

Keyword generation

During the parsing and processing of each new article, our parser generates the keywords of the articles based on the article title. More precisely, it breaks the news item title into small words, selects words with lengths of more than three letters, then inserts the keywords in the database and assigns them to the article. No new record is created if this keyword is already present in our database.

Category generation

In our system, we do not generate categories. Instead, categories are defined before the parsing procedure of the article collection. The predefined categories are a common set of words taken from the websites, and they are the following:

Κύπρος [Cyprus]; Ελλάδα [Greece]; Διεθνή [International]; Πολιτική [Politics]; Υγεία [Health]; Κυπριακό [Cyprus Problem]; Οικονομία [Finance]; Αθλητικά [Sports]; Κοινωνία [Society]; Επιστήμη [Science]; Επιχειρήσεις [Enterprises]; Ενέργεια [Energy]; Ψυχαγωγία [Entertainment]; Κορονοϊός [Coronavirus]; Εκλογές [Elections]; Αλλα θέματα [Other topics].

It should be noted that not all the categories are common for the ten news portals. For example, some portals don't have separate categories for "Enterprises" or "Greece".

New categories can be created through our Dashboard (please see the subsection related to the "Dashboard" below). A manual restart must happen for the RSS parsers/crawlers to be updated. An important note for the categories is that they must be kept because there may be related articles to the existing categories; therefore, deletion is unavailable.

We use different approaches to extract categories from articles depending on the news portal and how articles are collected. For websites that provide an RSS feed, each RSS feed usually consists of articles from a specific category. Other websites may only have one RSS feed and provide the article category inside the response body. Lastly, for sites without RSS, we use the categories from the actual website that any user can view while browsing.

Analytical category generation

The admin users can manage internal categories. The user can assign a specific "internal category" word to an article and then search by it. The Analytical categories are assigned on the "Edit" page of the articles. Also, the addition/removal/update of the internal categories can be done on the Dashboard.

Search and advanced search

To provide the best search experience possible, we used a tool called Meilisearch. This tool offers many features, but the one we used heavily in this web app is its search capabilities. Eventually, all news items inserted into our database will be indexed by Meilisearch. Our index includes fields like the article's title, body and date. All of those fields are considered when a search query executes. To index all our platforms' articles to Meilisearch, we use Laravel Scout. This tool is responsible for listening for any database changes in our tables and performing the relevant operations to keep the Meilisearch index up in sync with our database.

The search engine behind the scenes tokenises search queries and performs sophisticated algorithms to generate the most relevant result. For example, when a search term includes the word "Cyprus", Meilisearch can identify similar words using its tokenisation technique and may provide results with other word directives such as "Cypriot". Articles with exact much and more repetition of this word will come up higher in the result (relevance order). Search terms can also include phrases, and the same logic will apply. For instance, a search result for "European Union" may consist of articles that specify any combinations of the words "European", "Union" or derivatives. Lastly, exact word matching is available when text queries are wrapped with quotes.

The advanced search feature allows users to enrich their query with specific criteria such as Publication Date, Categories, Newsportal, Internal Categories and Keywords. All required fields are computed during the news items collection phase except Internal categories (Analytical Categories), which can be assigned manually by the researchers.

Dashboard

The Labservatory web application provides a dashboard for managing users and resources (such as Categories, Keywords and Internal Categories) and viewing the system and data collection reports. This section provides detailed information about the resource management components under the Dashboard.

Category

Authorised Users can expand the "Resources" and click "Categories" on the dashboard's sidebar, and a list with categories will appear. Users can create a new category by clicking "New", entering a name and clicking "Create". They can edit a category by clicking the pencil near the category they want to edit, entering a new name and clicking "Edit". They can also export a CSV file that contains all the related keywords by clicking the download icon. Finally, users can delete a category by clicking the trash icon or selecting multiple categories and clicking "Delete".

Keywords

Administrators can create a keyword on the dashboard. Firstly, they can expand "Resources" and click "Keywords", and a table with all keywords will appear. By clicking "New", a modal will appear to enter the name and select the related Category. Then, the users should click "Create" to save the new keyword. They can edit a keyword by clicking the pencil near the

keyboard they want to edit. Finally, they can delete a category by clicking the trash icon near the keyword they want to delete or selecting multiple keywords and clicking "Delete". Online Information Review

Analytical category (internal category)

Registered users can expand the "Resources" and click "Analytical Categories" on the dashboard's sidebar, and a list with analytical categories will appear. Users can create a new analytical category by clicking "New", entering a name and clicking "Create". They can edit an analytical category by clicking the pencil near the category they want to edit, entering a new name and clicking "Edit". Finally, users can delete an analytical category by clicking the trash icon or selecting multiple categories and clicking "Delete".

System and data collection reports

One of the most important functions of the Labservatory is its ability to treat the big data collected in real-time, providing analytical statistics related to the category of each news portal and the number of news published by a news portal. This section describes the procedures which lead to the non-stop generation of the system and data collection reports that show the health status of the system along with the reports that show the statistics related to the news of the news portals.

Lists of the charts that are included in this subsection are the following:

- (1) Today's News portal's counts distribution
- (2) Today's Categories counts distribution
- (3) News portal's counts distribution
- (4) Category's counts distribution
- (5) News portal's views distribution
- (6) Each newspaper's category distribution
- (7) Each category's newspaper distribution
- (8) Keywords per newspaper distribution
- (9) Keywords per category distribution

System reports

This section of the web application is executing the system analysis and provides a system status report of Labservatory. Thus, is responsible for assisting with the visualisation of news crawler performance. The first pie chart determines from which news source the articles we collected today originated from and how many in order to show, in a case of issue with crawler, zero articles for the erogenous crawling module related to the specific news portal. As shown in the pie chart below, the distribution of news sources is fairly even.

Figure 1 shows that all 10 news portals added their articles to our site. This helps us find out when there is an issue with the news crawlers. If one day a user identifies that for example the sigma live did not contribute in the numbers of articles and have zero articles, then something erogenous or bug instantiated in the process and the IT must investigate it.

Moreover, an additional Figure 2 is created of type pie chart, indicating the approximate category distributions of the articles collected today. This figure shows a quick overview of the news categories that are essential/hot in the Cyprus news scene for the current day.

Additionally the system provides three more pie charts with which a user can select an article report with choices to be based on the time frame of reference, either weekly, monthly, yearly or

Today's News portal's counts distribution:

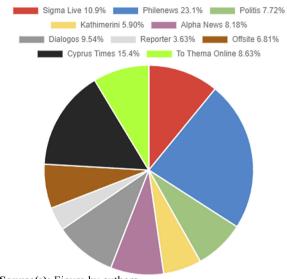


Figure 1. Pie chart of each News Portal's today's articles percentage

Source(s): Figure by authors

Today's Categories counts distribution:

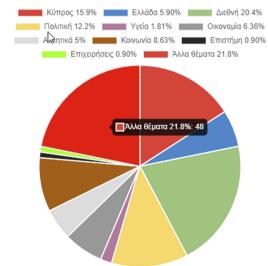


Figure 2. Pie chart of each Category's today's articles percentage

Source(s): Figure by authors

all the articles. These pie charts help our researchers comprehend the percentage of news portals and article categories that have been indexed during this time period. In Figure 3, a user will be able to determine, for instance, if one news portal has added significantly more articles this week than its competitors and/or if a category has been more popular for a given time period. In the final pie chart, we can also see which news portal's articles are accessed by researchers.

Lastly, the Figure 4, is a bar plot in which a user can select a number of news portals and the report will generate in a side-by-side style the articles count of each one.

Data collection reports

In this section, the data collection reports focus on the relationship between a category and news portals and news portals and categories with the utilisation of two pie charts. In the first, a user can select a news portal and a time period to view the categories from which this portal has been publishing articles. A user can also do the opposite with the subsequent pie chart.

In Figure 5, a user can see for the last month statistical data related to the Cyprus news portals on how many times has written the most articles about specific categories such as Greece.

Figure 6, with keywords pie chart reports for each category a user selects, the number of articles per news portal or category that contain it for a specific time range.

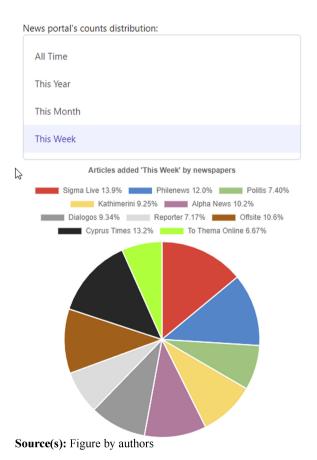


Figure 3.
Pie chart of each News
Portal's weekly article
percentage

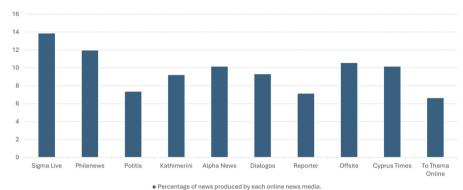


Figure 4. News Portal's weekly contribution

Source(s): Figure by authors

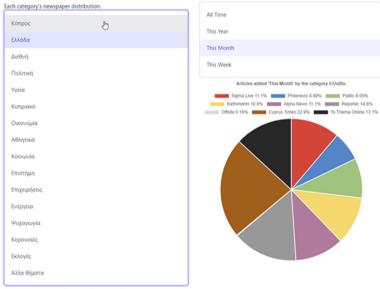


Figure 5. Pie chart of this week's News Portal Greece Article's distribution

Source(s): Figure by authors

As an example shown in Figure 7, an example of the barplot figure display, Sigma Live appears to produce more articles with the keyword "Ukraine" [Oukpavia], and the number of articles published by news sites has decreased since the beginning of the year.

Preliminary results and further research

This section presents three preliminary analyses based on data collected during the first eight months of the Labservatory web application's operation. These preliminary analyses aim to



Figure 6.
Pie chart of this
month's News Portal's
International News
distribution

explore potentialities for further research in media and journalism studies. Specifically, the section consists of three subsections: the first examines the Labservatory's potential contribution to news flow analysis (RQ1); the second scrutinises its role in elaborating news topic distribution analysis (RQ2); and the third explores its contribution to news discourse analysis (RQ3).

News flow analysis: questioning production strategies

From the beginning of the collection process (08/2022) until now (04/2023) they have been collected over 250,000 news items. 45,689 news items were collected from Cyprus Times, 37,942 from Offside, 34,307 from Sigma Live, 29,658 from Dialogos, 21,381 from Philenews, 20,471 from Kathimerini, 19,599 from Reporter, 17,719 from Politis, 17,557 from Alpha News and 14,046 were collected from Thema Online (see also Figure 8). This involves a daily news production for each online media, approximately, between 50 and 150 news items. If one takes into consideration the relatively small number employed in the online news media in Cyprus (Sahin, 2022b) the number of the news items which are daily produced by the most "productive" online media seems huge. This finding calls for a closer examination of the immediacy as the defining characteristic of online news (Omar *et al.*, 2021), and its connection with churnalistic practices (Van Leuven, 2019) within online media newsrooms in Cyprus.

The observation of the news production rhythm indicates important variations between the ten media outlets under scrutiny. The most "productive" media produce approximately three times more than the less "productive". A possible explanation of this differentiation regarding the production strategy between the online news media could be based on the features of their structural appurtenance and ownership. However, a closer examination of this finding shows that these variations cannot be fully explained based on whether the online media are part of an established media organisation or part of an internet-based media organisation. In fact, between the three most "productive" online media, the first two come from internet-based organisations

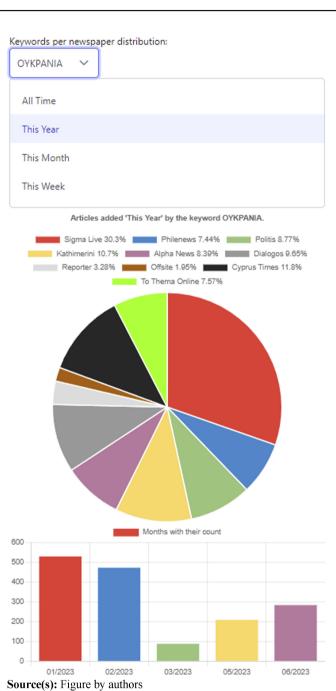


Figure 7.
Pie and bar chart of
News Portals articles
distribution for
Ukraine keyword



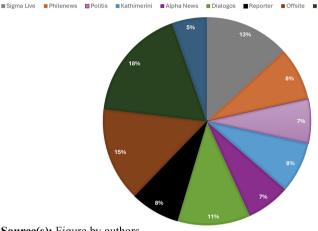


Figure 8. Percentage of News items produced by each of the ten media under examination for the period 08/2022-04/2023

Source(s): Figure by authors

and the third from an established organisation. The same applies to the less two productive online news media. This indicates that online news media in Cyprus implement their production strategy and in particular their investment in a more or less important churnalistic production independently of their structural features. This observation could be developed in relation to previous findings suggesting that media organisations that combine print and digital activities or operate other mass media businesses have a higher probability of surviving in the market than internet-based media (Mangani and Tarrini, 2017).

Of course, this assumption involves a closer observation on the newsroom routines, the editorial practices, and the business plans of these media to fully understand the decision processes behind the adoption of variant production and dissemination strategies. This could also involve studying the influence of external factors, such as economic pressures, on the media production strategies (Schudson, 2002). How is this phenomenon related to the Click bait and metrics culture of the online journalists?

Overall, the news flow analysis provided by the web application of the Labservatory appears as a valuable tool for understanding how news are produced, distributed and consumed in contemporary media environments. It can provide comprehensive insights into the news production rhythm of the online news media based on a big amount of data and indicate possible connections of these insights with other factors resulting from different methods of data collection, such as ethnographic observation, interview survey and questionnaire.

News topic distribution analysis: understanding the dominance of international news In general, news topic analysis aims to provide an overview of the main categories of topics that are being covered by a set of media outlets during a given timeframe. This section presents the results of the news topic analysis operated with the use of the web application of the Labservatory using its functions: Category Distribution and Advanced Search [2]. This consists of identifying the main categories of topics as they have been classified by the media themselves. In order to delimitate our simple, we collected the first 1,000 news items published by each online media under examination for the month of April 2023. Thus, 10,000 news items have been collected and scrutinised for the purposes of the news topic analysis. Due to their different production rhythm, as it has been explained earlier, some media needed less than a week to reach the number of 1,000 news items while other media needed more than three weeks. The results presented here are not exhaustive, in any case. Only some of the main features are presented in this section aiming at exemplifying the potentials of the web application of the Labsorvatory in operating such type of analysis.

As the Table 1 below shows, they have been identified 14 categories in which the ten online media opted to classify their news items produced during the first weeks of April 2023. Only two of them were used by all the online media though, namely, "international" and "politics". In contrast, the categories of "corporate" and "coronavirus" were used only by one media each. The most used categories by media to classify their news items were "international" (2,167 times), "Cyprus" (2,074), "sports' (1,214) and "politics" (1,262). In contrast, the categories of "energy" and "coronavirus" are attributed only to 4 news items each.

The supremacy of the category "international" is certainly a finding that merits further consideration and, also, in an ethnographic manner. The dominance of the international news is even clearer if we consider that there is another category which should be associated with, namely "Greece". These two categories, together, comprise nearly the 30% of the 10,000 news items produced during the given period.

Is the dominance of international news related to the process of classification, association and prioritisation of news used by algorithms (Choi, 2019; Spyridou *et al.*, 2022)? Or it is rather related to the increasing churnalistic practices of online news production? Does it respond to the "need" of the online media for a 24/7 news flow and thus to the demand for prepackaged news published by international news agencies? Also, in this perspective, the important number of daily news related to Greece has rather linguistique explanations? Is it involving the attractiveness of the Cypriot online media for ready news published by Greek media? Could this phenomenon be further investigated in terms of media glocalization (Kraidy, 2003) and information neo-colonialism (Guo and Vargo, 2017)?

Towards a discourse analysis: media discourses on refugees

Media discourse analysis is a method for analysing the language and symbols used in media to construct meaning in relation to specific issues. As such, media discourse analysis can be used to explore a wide range of topics. In this section we present the preliminary results of the media discourse analysis on "refugees". In doing so, this section focuses rather to illustrate the usefulness of the *Advance Search* functionality of the web application of Labservatory for the media discourse analysis than to present the findings of the analysis in an exhaustive manner.

	Categories	Number of news items per category	Number of media using each category					
1	International	2,167	10					
2	Cyprus	2,074	8					
3	Politics	1,262	10					
4	Sports	1,214	4					
5	Society	1,005	5					
6	Other topics	844	3					
7	Greece	630	8					
8	Economy	564	8					
9	Health	156	6					
10	Science	38	2					
11	Corporate	21	1					
12	Cyprus problem	17	2					
13	Energy	4	2					
14	Coronavirus	4	1					
Source(s): Figure by authors								

Table 1. Category distribution in relation to the news portals

By using the functionality of "keyword search" in "advance search" mode, we looked for news items containing the term "refugees" $[\pi\rho \acute{o}\sigma \phi \nu\gamma \epsilon\varsigma]$ during the period 01/08/2022–30/04/2023. The options provided by the web application of the Labservatory accelerate considerably the data collection and delimitation procedures. To be precise, we opt to focus only on news items published by four online media, (1) Dialogos, (2) Philenews, (3) Sigmalive and (4) To Thema. We make this choice for comparative purposes of the study with previous work which has been focused only on these four media outlets. The option to present the results in "relevance order" gave us the possibility to further delimitate our corpus and focus only on the most relevant news items for a more profound discourse analysis.

This process led to the collection of 362 news items of which 109 texts were collected from Dialogos, 84 from Philenews, 131 from Signalive and 38 from To Thema.

The report of the advance search which is generated automatically by the web application of the Labservatory includes two more categories, allowing for the mapping of the results, namely, the "categories" and "keywords" (see Table 2). The first, indicates that the news on refugees were categorised by the media mostly in four categories, International (163), Society (72), Politics (50) and Cyprus (38). The second, reveals somehow the main elements of the co-text of the word "refugees". It shows that the news stories related to refugees are mainly associated with the words "Ukraine/Ukrainians", "migrants", "dead", "borders" etc. This indicates a shift in the media construction of refugees, since they are identified mostly as "Ukrainians" in the international news while in the recent past, they were mostly identified as Syrians in local news. Thus, the advanced search functionality revealed an important element of the current discursive construction of refugees that merits further consideration.

Labservatory enables to export the search results in a word document, and this limits consequently the time spent to prepare the collected texts for further systematic qualitative discourse analysis. For example, it could be scrutinised whether this shift related to the identity of the "refugees", goes along with more positive and humanitarian discourses on the Ukrainian refugees in comparison with "other" refugees that have been in the epicentre of the media discourses during the last years, prior the war in Ukraine.

Discussion and conclusion

This article shows that the synergy between journalism studies and computer science offers a powerful and insightful approach for observing online journalism. In presenting some of the main features and functionalities of the web application of the Online Journalism Observatory: Labservatory, the article examines how researchers can scrutinise various aspects of online

Total articles: Total news portals:	362 4	Total categories:	10	Total keywords:	472
Sigma Live	131	International [Διεθνή]	163	Ukraine [Ουκρανία]	14
Dialogos	109	Society [Κοινωνία]	72	Migrants [Μετανάστες]	13
Philenews	84	Politics [Πολιτική]	50	Refugees [Πρόσφυγες]	9
To Thema Online	38	Cyprus [Κύπρος]	38	Ukrainians [Ουκρανοί]	9
		Other issues [Άλλα θέματα]	23	Dead [Νεκροί]	9
		Greece [Ελλάδα]	8	Borders [Σύνορα]	7
		The Cyprus Issue [Κυπριακό]	3	Pournara [Πουρνάρα]	6
		Economy [Οικονομία]	3	Syria [Συρία]	4
		Energy [Ενέργεια]	1	Merkel [Μερκελ]	4
		Health [Υγεία]	1	English Chanel [Μαγχη]	4
Source(s): Figure by authors					

Table 2.
Media discourse on
"refugees" according to
news portals,
categories and
keywords

news discourse, through three different preliminary studies. Precisely, the first study, demonstrates the contribution of the web application in the news flow analysis (RQ1). The preliminary analysis revealed that immediacy is a crucial characteristic of online news media in Cyprus, similar to other outlets in various media environments (Omar *et al.*, 2021). However, within a media sector experiencing economic recession, immediacy negatively impacts news values by reinforcing a churnalistic news production process (Saridou *et al.*, 2017). The possibility offered by the Labservatory to simultaneously analyse the daily news production of ten media outlets with varying features demonstrated that this effect persists regardless of the size of the media outlet or whether they are part of an internet-based media group. The adoption of similar production strategies by media outlets with differing characteristics within the same media ecosystem raises questions about the assumption that hybrid media groups are more likely to survive in the market than internet-based media (Mangani and Tarrini, 2017).

The second study examined the contribution of the Labservatory web application to news topic distribution analysis (RQ2). The capabilities provided by the Labservatory's web application, including the rapid collection of a large volume of news items and the identification of news categories as classified by the media themselves, revealed interesting insights not typically discernible through other methods of news topic distribution analysis. Drawing from a corpus of 10,000 news items, the study unveiled the unexpected dominance of the "international" category within the content of Greek Cypriot online news media. This finding raises significant research questions, such as those concerning the processes of classification, association and prioritisation of news utilised by algorithms (Choi, 2019), as well as the impact of media glocalization (Kraidy, 2003) and information neo-colonialism (Guo and Vargo, 2017) on small-scale media sectors in peripheral regions.

Finally, the third preliminary study investigates the contribution of the Labservatory to news discourse analysis (RQ3). It demonstrates how the Labservatory's web application's capability to present the results of the "advanced search" in "relevance order" allows for the delimitation of the sample, facilitating a more in-depth qualitative discourse analysis. Furthermore, it illustrates how the automatically generated "advanced search report" aids in mapping the results by providing a co-textual analysis associating search terms with "categories" and "keywords". Specifically, based on a delimited corpus of 362 news items containing the term "refugees", the study reveals a shift in the media's portrayal of refugees in recent years. They were predominantly identified by the media as "Ukrainians" in "international news" with positive connotations in 2022–2023, whereas in the recent past, they were mostly depicted as "Syrians" in "local news" with predominantly negative connotations (Trimithiotis and Voniati, 2023).

Of course, Labservatory's web application has also limitations. Some of the most important are listed below. First, due to system limitations, the web application cannot display or export more than 1,000 search results at a time. Thus, in some cases the researcher needs to repeat the same search various times including different media outlets or timeframes to collect the totality of the results. Second, it only collects news from Greek Cypriot media. It will be very fruitful for the research in media and journalism studies to collect news also from online media based in other countries or regions. This would be extremely beneficial for comparative perspectives in research. Third, it collects news from ten online media and not from all the online media outlets of the Republic of Cyprus. Even if the included ten media outlets have different characteristics, this limitation could raise questions on the representativity of the database. As any technological tool, the web application of the Labservatory is dynamic. Its progress is related to the available resources and funds. The objective is to erase these limitations by developing the next version of the Labservatory in the next few years.

Overall, this article showed that the web application of the Labservatory enables researchers to combine the strengths of journalism studies and computer science to continue uncovering deeper

insights into the complexities of the evolving landscape of online news and its societal impact. The preliminary findings presented in this article highlight the significance of this interdisciplinary collaboration and indicate new perspectives for further research, including ethnographic approaches. It is also important to mention the contribution of the Labservatory to the society, since it may also be used by the broader public for an immediate access to a more pluralistic information and for promoting both news media literacy and news media accountability and thus maybe contributing strengthening healthy news media scepticism (Xiao and Yang, 2024).

Notes

- 1. The web application can be found in the following address: https://labservatory.eu/
- We should note that due to space limitation, in this article, we do not use the functionality keywords distribution, provided by the web application of the Labservatory which is also very fruitful for the news topic analysis.

References

- Bednarek, M. and Caple, H. (2014), "Why do news values matter? Towards a new methodological framework for analysing news discourse in critical discourse analysis and beyond", *Discourse & Society*, Vol. 25 No. 2, pp. 135-158, doi: 10.1177/0957926513516041.
- Cammaerts, B. and Carpentier, N. (2007), Reclaiming the Media: Communication Rights and Democratic Media Roles, Intellect Books, Bristol.
- Carpentier, N. and De Cleen, B. (2007), "Bringing discourse theory into media studies: the applicability of discourse theoretical analysis (DTA) for the study of media practises and discourses", *Journal of Language and Politics*, Vol. 6 No. 2, pp. 265-293, doi: 10.1075/jlp.6.2.08car.
- Carvalho, A. (2008), "Media (ted) discourse and society: rethinking the framework of critical discourse analysis", *Journalism Studies*, Vol. 9 No. 2, pp. 161-177, doi: 10.1080/ 14616700701848162.
- Cedillo, G.R.and Carretero, A.B. (2016), "Analysis of media observatories in Spain. A tool for civil society in media reform processes", *Revista Latina De Comunicación Social*, Vol. 71, pp. 443-469, doi: 10.4185/RLCS-2016-1104en.
- Chadwick, A. (2017), The Hybrid Media System: Politics and Power, Oxford University Press, New York, NY.
- Choi, S. (2019), "An exploratory approach to the computational quantification of journalistic values", Online Information Review, Vol. 43 No. 1, pp. 133-148, doi: 10.1108/OIR-03-2018-0090.
- Chouliaraki (2008), "The media as moral education: mediation and action", *Media, Culture and Society*, Vol. 30 No. 6, pp. 831-852, doi: 10.1177/0163443708096096.
- Christophorou, C. (2010), "Greek Cypriot media development and politics", The Cyprus Review, Vol. 22 No. 2, pp. 235-245.
- Dunaway, J. (2013), "Media ownership and story tone in campaign news", American Politics Research, Vol. 41 No. 1, pp. 24-53, doi: 10.1177/1532673x12454564.
- Golan, G.J. and Yang, S.U. (2009), "Media accountability systems: evaluating the impact of media councils in Peru and Korea", *Journalism & Mass Communication Quarterly*, Vol. 86 No. 3, pp. 627-645.
- Guo, L. and Vargo, C.J. (2017), "Global intermedia agenda setting: a big data analysis of international news flow", *Journal of Communication*, Vol. 67 No. 4, pp. 499-520, doi: 10.1111/jcom.12311.
- Karayianni, C. and Psaltis, C. (2023), "Tweet for peace: twitter as a medium for developing a peace discourse in the hands of the Greek-Cypriot and Turkish-Cypriot leaders", *Online Information Review*, Vol. 47 No. 5, pp. 925-943, doi: 10.1108/oir-03-2022-0161.

- Kraidy, M.M. (2003), "Glocalisation: an international communication framework?", Journal of International Communication, Vol. 9 No. 2, pp. 29-49, doi: 10.1080/13216597.2003.9751953.
- Lim, J. (2012), "The mythological status of the immediacy of the most important online news: an analysis of top news flows in diverse online media", *Journalism Studies*, Vol. 13 No. 1, pp. 71-89, doi: 10.1080/1461670X.2011.605596.
- Mangani, A. and Tarrini, E. (2017), "Who survives a recession? Specialization against diversification in the digital publishing industry", Online Information Review, Vol. 41 No. 1, pp. 19-34, doi: 10. 1108/OIR-09-2015-0310.
- Maniou and Photiou (2017), "Watchdog journalism or hush-puppy silencing? Framing the banking crisis of 2013 in Cyprus through the press", Catalan Journal of Communication and Cultural Studies, Vol. 9 No. 1, pp. 43-66.
- Omar, B., Al-Samarraie, H. and Wright, B. (2021), "Immediacy as news experience: exploring its multiple dimensions in print and online contexts", Online Information Review, Vol. 45 No. 2, pp. 461-480, doi: 10.1108/OIR-12-2019-0388.
- Price, L.T., Clark, M., Papadopoulou, L. and Maniou, T.A. (2023), "Southern European press challenges in a time of crisis: a cross-national study of Bulgaria, Cyprus, Greece and Malta", *Journalism*, Vol. 0 No. 0, doi: 10.1177/14648849231213503.
- Şahin, S. (2022a), "News media and the conflict in Cyprus", in Reporting Conflict and Peace in Cyprus, Springer International Publishing, Cham, pp. 53-86.
- Şahin, S. (2022b), "Digital journalism in Cyprus", in Reporting Conflict and Peace in Cyprus, Palgrave Macmillan, Cham, doi: 10.1007/978-3-030-95010-1_6.
- Sahin, S. (2022c), Reporting Conflict and Peace in Cyprus: Journalism Matters, Springer Nature.
- Saridou, T., Spyridou, P. and Veglis, A. (2017), "Churnalism on the rise?", Digital Journalism, Vol. 5 No. 8, pp. 1006-1024, doi: 10.1080/21670811.2017.1342209.
- Schudson, M. (2002), "The news media as political institutions", Annual Review of Political Science, Vol. 5 No. 1, pp. 249-269, doi: 10.1146/annurev.polisci.5.111201.115816.
- Smith, A.J., Wright, K. and Martin, F. (2013), "Global media monitoring and observatories", International Journal of Communication, Vol. 7, pp. 1346-1364.
- Spyridou, P., Djouvas, C. and Milioni, D. (2022), "Modeling and validating a news recommender algorithm in a mainstream medium-sized news organization: an experimental approach", Future Internet, Vol. 14 No. 10, p. 284, doi: 10.3390/fi14100284.
- Trimithiotis, D. (2020), "The persistence of ethnocentric framing in online news coverage of European politics", *Digital Journalism*, Vol. 8 No. 3, pp. 404-421, doi: 10.1080/21670811.2019.1598882.
- Trimithiotis, D. (2022), "A multilevel contextual discourse analysis of online news: news on Europe in Cypriot online media", Sage Research Methods: Doing Research Online, doi: 10.4135/9781529601459.
- Trimithiotis, D. and Stavrou, S. (2022), "Digitalisation as discursive construction: entrepreneurial labour and the fading of horizons of expectations for newcomer journalists", *Journalism Studies*, Vol. 24 No. 1, pp. 88-107, doi: 10.1080/1461670X.2022.2143866.
- Trimithiotis, D. and Voniati, C. (2023), "(Un)Reporting Xenophobia: normalising and resisting officials' discriminatory discourse on migration in online journalism in Cyprus", *Journalism Practice*, pp. 1-21, doi: 10.1080/17512786.2023.2279336.
- Van Dijk, T.A. (1983), "Discourse analysis: its development and application to the structure of news", Journal of Communication, Vol. 33 No. 2, pp. 20-43, doi: 10.1111/j.1460-2466.1983.tb02386.x.
- Van Leuven, S. (2019). "Churnalism", in Vos, T.P. and Hanusch, F. (Eds), The International Encyclopedia of Journalism Studies, Wiley-Blackwell, Hoboken, pp. 1-5.
- Vatikiotis, P., Maniou, T.A. and Spyridou, P. (2023), "Towards the individuated journalistic worker in pandemic times: reflections from Greece and Cyprus", *Journalism*, doi: 10.1177/14648849231207670, available at: https://journals.sagepub.com/doi/full/10.1177/14648849231207670

Villanueva-Ledezma, A., Machin-Mastromatteo, J.D., González-Quiñones, F., Cordero-Hidalgo, A. and Flores-Flores, J. (2020), "Ethics, human rights and violence in Chihuahua's digital journalism: evidence from a media observatory", *Digital Library Perspectives*, Vol. 36 No. 1, pp. 55-68, doi: 10.1108/dlp-09-2019-0035.

Online Information Review

Xiao, X. and Yang, W. (2024), "There's more to news media skepticism: a path analysis examining news media literacy, news media skepticism and misinformation behaviors", Online Information Review, Vol. 48 No. 3, pp. 441-456, doi: 10.1108/OIR-04-2023-0172.

Corresponding author

Dimitris Trimithiotis can be contacted at: trimithiotis.dimitris@ucv.ac.cv