This is a post-print version of the article published in Communication Methods and Measures and accessible via the following link:

https://doi.org/10.1080/19312458.2021.1907841

To cite this article:

Christner, C., Urman, A., Adam, S., & Maier, M. (2022). Automated Tracking Approaches for Studying Online Media Use: A Critical Review and Recommendations. *Communication Methods and Measures*, *16*(2), 79–95. https://doi.org/10.1080/19312458.2021.1907841

Automated Tracking Approaches for Studying Online Media Use: A Critical Review and Recommendations

Clara Christner, Aleksandra Urman, Silke Adam, Michaela Maier

Abstract

With the increasing importance of online information environments, researchers have started investigating direct measures of online media use, such as online tracking. Most existing studies using tracking data have so far relied on commercial solutions, but these have limitations in terms of their costliness, replicability, and applicability to certain research questions. Hence, different research groups are developing their own tracking solutions for academic purposes. In this paper, we provide a critical review and classification of the existing approaches, apt to guide research decisions on the appropriate tracking approach and tool. First, we develop criteria to distinguish different user-centric desktop and mobile tracking approaches and tools (the type of information, technical complexity, privacy implementation, user experience, and availability). Second, we describe different tools and approaches—separately for desktop and mobile tracking—with concrete examples and evaluate them using the aforementioned criteria. Finally, we discuss how different mobile and desktop tracking solutions can complement each other and provide recommendations for future research.

Keywords: data collection, desktop tracking, mobile tracking, online media use, computational methods

Introduction

New digital media offers opportunities but also poses challenges to researchers in communication science since existing measures of media use, such as self-reports and media diaries, have proven to be not completely reliable (e.g., Araujo et al., 2017; Prior, 2009; Stier et al., 2020). Researchers in the field of communication science started investigating direct measures (de Vreese & Neijens, 2016) and used the tracking of online media use to investigate a broad variety of research questions, such as those regarding online political information searches (e.g., Flaxman et al., 2016), the effects of algorithmically personalized digital media (e.g., Haim & Nienierza, 2019), and online incidental news exposure (e.g., Möller et al., 2020).

The central challenge in this regard concerns the collection of media use data. While browser activity tracking and clickstream data are extensively used by businesses to place targeted advertising and personalize content (e.g., Coffey, 2001), the academic community has limited access to online tracking data. In academic research, most existing studies have relied on the data collected by commercial companies (e.g., Araujo et al., 2017; Mukerjee et al., 2018). These solutions, however, are costly and usually offer only domain- or URL-level data, resulting in limited applicability to research topics, such as algorithmic personalization. Hence, different research groups are developing their own tracking solutions for academic purposes (e.g., Adam et al., 2019; Bodo et al., 2017; Kleppe & Otte, 2017; Menchen-Trevino & Karr, 2012).

So far, only a few studies offer an overview of the field, and they have mostly focused on data structures and the description of particular tools, some of which are now obsolete due to technical changes (e.g., Bruns, 2013; Jünger, 2018; Wieland et al., 2018). Therefore, we lack an up-to-date, comprehensive review of the existing approaches and tools with systematic criteria to compare them.

In this paper, we address this gap in the literature. Our goal is to synthesize literature on tracking solutions in the academic field and create a comprehensive resource for communication scholars interested in using and/or creating online tracking tools. Broadening the general understanding of the opportunities and limitations of tracking solutions (e.g., Jünger, 2018) is essential for the planning and designing of future studies. Therefore, we first define user-centric tracking data and develop a categorization to evaluate different desktop tracking tools and mobile tracking approaches. Next, we apply these criteria to address the question of identifying the different approaches and tools for desktop and mobile tracking that could be suitable for specific applications in communication science. Finally, these insights are translated into concrete recommendations for the selection of the right tracking approach and tool, for a combination of desktop and mobile tracking, and “nice-to-have features” for future tools.

Defining User-Centric Tracking Data

Digital tracking data result from “every procedure intentionally applied to trace the usage of digital media aiming to analyze the collected data for research purposes” (Wieland et al., 2018, p. 134). Such data enable researchers to get the actual content participants saw and to address a variety of research questions, such as those related to personalization, algorithmic selection, and incidental news exposure on social media (e.g., Bodo et al., 2017; Haim & Nienierza, 2019; Stier et al., 2020). However, their collection has to therefore be simultaneous with actual media usage, which poses the risk of behavioral adjustment on the participants’ side as they are aware of being observed. To decrease potential adjustments, a tracking tool must be as unobtrusive as possible (de Vreese & Nejens, 2016; Wieland et al., 2018). Another crucial feature of tracking data is that they are user-centric. User-centric data can be defined as information resulting from procedures applied to tracing the comprehensive media usage of an individual on the client side (Wieland et al., 2018). Comprehensive user-centric data are necessary for testing existing theories in the context of online media use, but also for developing and testing new theories (de Vreese & Neijens, 2016; Stier et al., 2020). For such research questions, it is crucial that the tracking data contain information usage behavior across different platforms and the actual content individuals have seen, and that these data can be combined with other data, such as surveys.

Data Collection and Methodology

In this review, we used a scoping review approach (Munn et al., 2018) as we aimed to establish how online tracking research is conducted in the field. For both desktop and mobile tracking, we included empirical studies using a specific tracking tool as well as technical manuals, articles on the technical development of tracking tools, and existing reviews. Two exclusion criteria were applied: articles not focusing on the tracking devices that are used to study media use in the field of communication sciences and articles for which researchers relied on commercial tracking data. We searched the EBSCO, Web of Science, and JSTOR databases in April 2019 and updated the information based on a database search in December 2019. To make our search as exhaustive as possible, we started with queries such as "tracking tools," "mobile tracking," and "desktop tracking," and augmented them with additional key terms found in the literature (i.e., specific types of mobile tracking tools, such as "smartphone loggers" and "screen recorders"). In addition, we supplemented and updated the literature based on information gathered from researchers working in the field, both during the data collection and evaluation processes, whenever the information retrieved from the literature was not sufficient to systematically compare the tracking tools and approaches.

After collecting the data, we prepared an overview and constructed a set of criteria for the systematic comparison and evaluation of the existing tools and approaches. The criteria were constructed based on the insights we gathered during the initial overview of the existing tools as well as on observations from related literature (e.g., Jünger, 2018; Wieland et al., 2018) and discussions within the scholarly community during a number of dedicated workshops[1]. Then, through a series of discussions between the authors, we decided on a set of criteria based on which tools and approaches are best distinguishable and which of them we consider to be most relevant for the development of a tracking tool. Although not exhaustive, the lists offer crucial criteria for systematically comparing existing approaches and tools and for expanding the planning process of future studies.

1 The precise names of the workshops are omitted here for the blind review but will be included in the final version of the manuscript if it is accepted.

Due to the state of the development of mobile tracking, we focus on approaches but describe and evaluate existing tools where applicable. At the moment, there are very few currently functioning mobile tracking tools (Krieter, 2019a; Reeves et al., 2021; van Damme et al., 2020); however, other approaches, such as browser extensions, are proposed by scholars. We suggest it is worthwhile reviewing them and evaluating their potential for the future development of corresponding tools.

Criteria to Distinguish and Evaluate User-Centric Tracking Tools

To characterize and distinguish different user-centric tracking tools and approaches, we developed five criteria: 1) types of information, 2) technical complexity, 3) privacy features, 4) user experience, and 5) availability.

First, we distinguish between the **types of information** that can be collected by a tracking tool: the usage patterns and the actual content seen by the user. Web usage mining is a technique for discovering usage patterns from web usage logs (Liu, 2008). By investigating users' browsing and navigation, conclusions can be drawn about their interests and the frequency of their visits to particular websites, as well as about their usage patterns (Kleppe & Otte, 2017; Liu, 2008; Wieland et al., 2018). To study the content seen by the user, web content mining needs to be applied. Web content mining extracts specific information from websites in the form of texts and images (Liu, 2008). These data can be further analyzed and linked to usage data and are especially useful for questions concerning media usage behavior in algorithmically curated media environments and for questions where social behaviors are linked to information content (Kleppe & Otte, 2017; Liu, 2008). However, the moment when the information extraction takes place is decisive. If the information is gathered post-factum, for example, based on previously collected URLs, personalization cannot be investigated. This is only possible if the information is extracted simultaneously. For mobile tracking, we additionally consider the potential of a given approach for collecting data on app usage and in-app-browsed content. The possibility of collecting in-app browsing information is becoming increasingly important for communication research as more and more people are using mobile apps for news (Newman et al., 2018).

Our second criterion deals with **technical complexity** with regard to developing or maintaining such a tool. The development and maintenance of tracking tools requires close cooperation between social scientists and developers to balance societal and individual interests (Bodo et al., 2017) and deal with technical challenges or restrictions. Regarding their development, we mainly discuss to what extent existing structures can be used and cross-platform- or cross-browser functionality. Cross-platform functionality is especially relevant in the context of mobile trackers since the development of mobile trackers for Android and iOS require vastly different pipelines, thus complicating the production of cross-platform trackers.

The third criterion comprises the **privacy features** within the tool. As tracking raises substantial privacy concerns, the protection of participants' privacy and the definition of which data should not be collected is crucial. A first relevant feature is that the tool offers login options (i.e., anonymized identification codes, passwords) to ensure that only the data of the targeted participant are tracked and that the participant's identity is anonymized from the beginning (i.e., no personal data, such as names and/or personal email addresses, are necessary to log in). That is important in ensuring data quality but also in terms of following ethical and legal requirements (e.g., Bodo et al., 2017). The second relevant feature is the way in which the tool allows the definition of websites that should or should not be tracked. This can be obtained using either whitelists (tracking only predefined websites) or blacklists (defining websites that will not be tracked). Whitelisting allows researchers to significantly limit the scale and magnitude of the data collection necessary for the research question (Bodo et al., 2017). However, only using data from whitelisted websites is insufficient for certain research questions, such as exploratory approaches where participants are expected to use non-mainstream media that researchers do not necessarily know about when designing the study (e.g., Adam et al., 2019). In such cases, relying on blacklists, which can be client-specific or system-wide, is recommended. Implementing a client-specific blacklist allows participants to individually specify which pages or sites are blacklisted. A system-wide blacklist allows researchers to block irrelevant sites, such as adult sites or e-banking (e.g., Menchen-Trevino & Karr, 2012). Selectors are an additional privacy feature and are especially useful for tracking social networks as they allow researchers to only collect public information from specific domains (i.e., only publicly available posts from Facebook). This is particularly important as it ensures the privacy of the participants without blocking these platforms (e.g., Haim & Nienierza, 2019). We do not discuss privacy implementation in relation to mobile tracking since the corresponding features can be integrated in numerous ways regardless of the general tracking approach. Thus, this criterion is only relevant for the evaluation of specific tracking tools.

Fourth, we turn to the **user experience** and ask how user-friendly the tool is and how strongly the tool impacts users' normal behavior. With tracking users' online media use, the goal must be to influence participants as little as possible; however, two aspects argue against an unobtrusive measure: First, research ethics and data privacy policies require participants' informed consent to their data collection. Second, all tracking approaches and tools depend upon the active installation of the tracking software in browsers or on devices (Wieland et al., 2018). Consequently, the design of the study and the tracking tool involves a trade-off between interrupting participants to get additional information and minimizing users' reactivity and awareness (de Vreese & Neijens, 2016; Wieland et al., 2018). We evaluate this in the context of user experience based on the difficulties of (un-)installing the tool and the extent to which participants are influenced or limited in terms of their usual online behavior by the tool. When assessing the user experience for mobile tracking, in the absence of specific tools, we focus on the features that make a given approach more or less intrusive and, thus, that can affect participants' experience and mobile usage to a higher or lower extent.

Our final criterion deals with the **availability** of the tool. Availability means that the tool is openly available (either for usage and/or the code). Since only a few functioning tools exist for mobile tracking, we refer to prototypes as well as to non-academic tracking tools that can potentially be repurposed for academic use where applicable.

User-Centric Tracking Tools for Desktop Tracking

Two approaches to collecting user-centric tracking data on desktops can be distinguished: the transparent proxy approach (Bodo et al., 2017; Kleppe & Otte, 2017; Menchen-Trevino & Karr, 2012) and the screen-scraping approach (Adam et al., 2019; Haim & Nienierza, 2019).

Transparent Proxy Approach

The first method is the transparent proxy method, with which the data are collected through a virtual proxy in the participants' Internet browsers (Coffey, 2001). Proxy server-based tools allow researchers to intercept participants' requests and save a full record of the content (Menchen-Trevino & Karr, 2012). A transparent proxy "serves as an 'invisible' link in the chain of computers between a user and a website, through which all the traffic of all the participants flow through" (Bodo et al., 2017, p. 147). Unlike with a non-transparent proxy, no proxy settings are required on the users' devices.

To our knowledge, there are three tracking tools in the academic community based on the transparent proxy approach: *Roxy* (Menchen-Trevino & Karr, 2012), *Robin* (Bodo et al., 2017), and *Newstracker* (Kleppe & Otte, 2017). Depending on their programming, the proxies can capture all web requests that contain technical information, such as the IP address of the other computer, the date, the time, the requested URL, and the text content, influencing the **type of information** that can be collected. *Roxy* logs each HTTP request, and the data contain the text content, the URL, the requested page, and the time and date of the request. However, *Roxy* does not log any encrypted (HTTPS) content for data protection reasons (Menchen-Trevino & Karr, 2012). Since most websites switched from HTTP to encrypted HTTPS, the original version is not fully operational anymore and, to our knowledge, is being revised. *Robin* (Bodo et al., 2017) was built to study information behavior in personalized information environments and captures not only HTTP but also HTTPS traffic, including news items, search query results, personalized ads, and user interactions (i.e., comments, likes, shares, etc.). *Newstracker* (Kleppe & Otte, 2017) was set up to analyze online news consumption on news websites on users' desktop computers. The processes of data collection and content extraction do not occur simultaneously but are instead divided into two separate phases. In the data collection phase, all HTTP requests are recorded in a log file containing metadata from the request, such as the time, the username, and the requested URL. The actual content of the visited websites is extracted in the pre-processing phase, in which the content of the visited URLs is extracted using a content extraction application that allows the extraction of specific information. Kleppe and Otte (2017) created extraction templates for predefined news websites to extract information such as the title and main text of the requested news item. Although this reduces the data collected to the data needed to facilitate further data processing, the subsequent extraction of content based on the URLs does not allow conclusions to be drawn about personalization.

Researchers interested in analyzing the content of online media behavior have to construct their own proxy (Menchen-Trevino & Karr, 2012) and filter out and sometimes decrypt relevant traffic, resulting in high levels of **technical complexity**. For both *Roxy* and *Newstracker*, the researchers created a custom-built, transparent proxy that participants have to install on their computer. After the installation and configuration, all Internet traffic generated by the user's Internet browser is rerouted through the proxy server, where each request is logged (Kleppe & Otte, 2017; Menchen-Trevino & Karr, 2012). The above-described content extraction process in *Newstracker*, using extraction templates, facilitates data processing. At the same time, this complicates the maintenance of the tool as extraction templates make the tool vulnerable to changes in the HTML structure. *Robin*, on the other hand, uses a browser plugin that participants have to download and install, which then routes the browser traffic through an enhanced transparent proxy (Bodo et al., 2017).

All three tracking tools offer a number of **privacy features**, such as login credentials to protect participants' privacy and to ensure that data are only collected from participants who have given their informed consent (Bodo et al., 2017; Kleppe & Otte, 2017; Menchen-Trevino & Karr, 2012). In terms of the approach to excluding data that should not be collected, the tools differ significantly. *Roxy* contains a user-specific and system-wide blacklist and, moreover, the option for participants to review their browsing history and remove information that they are not willing to share. Participants can further choose between a regular session, in which all pages except the blacklisted, secure (HTTPS) pages and pages with a lock symbol are recorded, and a private or guest session, in which no information is logged (Menchen-Trevino & Karr, 2012). *Robin* and *Newstracker* both apply a whitelist approach. In the case of *Robin*, the traffic of these whitelisted URLs is intercepted, copied, filtered, and finally stored in secure data storage. Further, best-effort anonymization scripts are used to eliminate sensitive information (Bodo et al., 2017; Möller et al., 2020). In the case of *Newstracker*, a first data cleaning process is performed before the content is extracted. The raw data are first cleaned, and the URLs are matched against a whitelist before the content of the visited URLs is extracted. As *Newstracker* only extracts the specific information that is required for the research interest, this further increases the participants' privacy. Unlike *Roxy*, *Newstracker* does not support a private mode;

however, participants are informed that they can shut the proxy down if they do not want to be tracked (Kleppe & Otte, 2017).

With regard to **user experience**, there are many overlaps between the three tools. *Roxy* and *Newstracker* require participants to install the tool on their device and configure the tool in their Internet browser (Bodo et al., 2017; Menchen-Trevino & Karr, 2012). Besides general information for system configurations, *Roxy* provides a test tool to check whether the setup was successful and gives descriptions of the privacy tools through a web application (Menchen-Trevino & Karr, 2012). A disadvantage of the proxy approach is the dependency of the browser as it is necessary to incorporate proxies into each browser version on the participant's PC. This means that with new browser versions, an update of the tracking tool software might be necessary, which has to be distributed to the sample. Therefore, the regularity with which new browser versions are introduced leads to a big effort to ensure that the sample stays fully installed (Coffey, 2001). In addition to the technical challenges, this also has a negative effect on the user experience as participants have to be contacted again and asked to install a new update.

Our last criterion concerns the **availability** of the presented tools. The code for *Roxy* is available open-source (Karr, 2014), whereas, to our knowledge, so far, neither *Newstracker* nor *Robin* has been made available open-source for the research community.

Screen-Scraping Approach

The second approach is what we call a screen-scraping approach. Scraping is used to automatically collect online data by extracting information from websites that are based on standardized formats (HTML or XML) (Jünger, 2018; Marres & Weltevrede, 2013). Tools that fall into this category directly scrape the content accessed by users via a browser plugin (extension), which allows researchers to directly access and save the content seen by the users (i.e., text, pictures, videos, posts, etc.).

Currently, there is one working tool based on the screen-scraping approach (*Eule*; Haim & Nienierza, 2019) and one tool that is under development (*WebTrack*; Adam et al., 2019). The tools differ in terms of the **type of information** that can be collected: *Eule* was developed to analyze news exposure within algorithmically curated media environments and to collect public information encountered on Facebook. Besides meta information (i.e., timestamps), the data entail information about each viewed post (i.e., content, embedded links, images, usernames, and the number of likes and comments) and captures interactions (i.e., clicking on a link, liking or sharing the post) (Haim & Nienierza, 2019). Although *Eule* is designed to collect Facebook data, the general idea of screen-scraping can be adapted for any other web or social media site. A second tool that is still in its development phase is currently doing this. *WebTrack* (Adam et al., 2019) uses the screen-scraping approach to track participants' entire online information behavior. The tool combines standard screen-scraping, capturing the full content (full HTML) of publicly available websites (e.g., news websites), with website-specific selectors, like *Eule*, to gather only publicly available information from social media sites, such as Facebook, Twitter, and YouTube (Adam et al., 2019).

There are some **technical complexities** inherent to tools using the screen-scraping approach. First, a separate plugin must be developed for each browser. Although most of the common browsers allow the installation of third-party extensions (e.g., Apple Safari, Google Chrome, Microsoft Edge, Microsoft Internet Explorer, Mozilla Firefox, Opera), only Google Chrome, Mozilla Firefox, and Opera follow the so-called WebExtensions standard. The WebExtensions standard allows the development of plugins for multiple browsers with little additional effort. However, adapting and transferring extensions compatible with Apple Safari and Microsoft Internet Explorer, which follow their own technologies, platforms, and guidelines for the development of plugins, is difficult and requires additional development (Haim & Nienierza, 2019). Second, although technically rather flexible, this approach is still dependent on technological changes. To identify specific features relevant to tracking (e.g., Facebook "likes"), screen-scraping relies on CSS or XPath selectors, which represent a hierarchical path along the HTML structure. Every HTML layout change of the website under observation requires an update of the CSS and XPath selectors since they might otherwise miss the mark, which can lead to data loss. Such updates may either be hand-coded (with the requirement to reapply for the upload in the plugin stores) or server-made, allowing faster customization towards minor changes (Haim & Nienierza, 2019).

Tools that apply the screen-scraping method can integrate various **privacy features**. As plugins allow for the assignment of login credentials, researchers using *Eule* and *WebTrack* can ensure that only the targeted person is tracked and that participants have control over their information. Further, *WebTrack* uses a system-wide blacklist to exclude websites that should not be tracked, such as banking websites containing highly sensitive data (Adam et al., 2019). The exclusion of websites is not relevant for the website-specific tool *Eule*. However, both tools have implemented privacy features to define what information from the websites, especially social media websites, is extracted. *Eule* uses website-specific selectors to only capture publicly available Facebook posts (Haim & Nienierza, 2019). *WebTrack* combines standard screen-scraping, by capturing the full content (full HTML) of publicly available websites (e.g., news websites), with website-specific selectors, like *Eule*, to gather only publicly available information from social media sites, such as Facebook, Twitter, and YouTube. An extensive data-cleaning process following the data collection further ensures the privacy of the participants. Further, *WebTrack* offers participants the option of selecting a private mode in which no information is tracked (Adam et al., 2019).

From the users' perspective, participants only have to install a browser plugin on their device, which has a positive effect on the **user experience**. However, as not all browsers allow the installation of custom plugins and as others require the additional adaptation of the developed plugins, this can result in cases where no plugin exists for the browser the participants normally use. This might cause problems regarding the generalizability of this approach as it might be difficult to recruit participants that use a default browser for which no plugin exists (e.g., Apple Safari; Haim & Nienierza, 2019). Participants either have to be persuaded to use a different browser, which means additional effort that can also influence their browsing behavior, or there is a risk of excluding a large part of the population based on their default browser. Finally, we evaluate the **availability**. Currently, only the code for *Eule* is available open-source (Haim, 2019).

User-Centric Tracking Approaches for Mobile Devices

With this increasing use of mobile platforms (Clement, 2020), research into online media consumption requires solutions for mobile tracking as well, especially as research has shown that the tendency to overreport information usage is particularly high among mobile users (Jürgens et al., 2020). However, as mobile devices employ sophisticated encryption (Teufl et al., 2013), mobile tracking is difficult, and when it comes to the browsing performed through apps, almost impossible. Below we discuss potential ways to at least partially overcome this challenge.

Smartphone Log Data

Smartphone log data are accessible directly from users' smartphones. To get such data, it is necessary to install special software—a logger—on a user's mobile device. Depending on the programming of the logger, the **type of information** that can be collected differs: it is possible to monitor a wide range of user activities, including call and SMS histories, GPS data, and

Table 1. Overview of existing tools for tracking of online media use on desktop devices.

Available tools	Approach	Types of information	Technical complexity	Privacy features	User experience	Availability
Roxy (Menchen-Trevino & Karr, 2012)	Proxy	actual content, but not from encrypted websites (HTTPS)	high	user-specific and system-wide blacklist; log-in option	relatively complex installation; relatively high level of intrusiveness	code made available open-source
Newstracker (Kleppe & Otte, 2017)	Proxy	content, but no personalization	high	whitelist; log-in option	relatively complex installation; relatively high level of intrusiveness	not open-source
Robin (Bodo et al., 2017)	Proxy	actual content & usage	high	whitelist	relatively easy installation; relatively high level of intrusiveness	not open-source
Eule (Haim & Nienierza, 2019)	Screen-Scraping	actual content & usage of publicly available Facebook posts	medium/high	whitelist; log-in option	relatively easy installation; relatively low level of intrusiveness	code made available open-source
WebTrack (Adam et al., 2019)	Screen-Scraping	actual content & usage	medium/high	blacklist & private-mode option; log-in option	relatively easy installation; relatively low level of intrusiveness	under development

information about visited URLs and app usage (i.e., the duration of app use, the number of notifications received from apps, whether a user opened the app after clicking on the notification from it). However, a logger cannot give researchers information about the actual content browsed through the apps.

Examples of smartphone loggers previously used by researchers to study users' media use include *DeviceAnalyzer* (Wagner et al., 2014), *The LiveLab Project* (Shepard et al., 2011; Tossell, Kortum, Rahmati, et al., 2012; Tossell, Kortum, Shepard, et al., 2012), an unnamed logger similar to LiveLab (LiKamWa et al., 2013), a parental control app *Kidlogger* (van Damme et al., 2015), and a *MobileDNA* app developed by researchers at the University of Ghent to track app usage on Android devices (Research Group for Media, Innovation, and Communication Technologies, 2019; van Damme et al., 2020).

Smartphone loggers, in part due to their customizability, show high levels of **technical complexity**. None of the loggers mentioned in this section are openly available in the academic community. Importantly, it is necessary to design separate loggers for Android and iOS. To our knowledge, the loggers described above only functioned on one of these platforms: *DeviceAnalyzer*, *Kidlogger*, and *MobileDNA* on Android and *LiveLab* and the logger modeled on Livelab on iOS. Second, smartphone platforms evolve rapidly, and with every major update some functionalities of a logger might be deprecated. Hence, smartphone loggers require continuous support. As of 2020, of the above-mentioned loggers developed by academics, only the recently created *MobileDNA* logger is in use and theoretically **available**. The others are obsolete due to their incompatibility with platform updates.

Depending on their implementation, smartphone loggers can be highly intrusive, affecting the **user experience**. Some functionalities of the loggers might require admin access to a device and privilege escalation. For instance, loggers on rooted smartphones could access private data, such as the internal folders of the installed apps, thus obtaining private information, including the log files of the apps. In essence, rooting is hacking users' smartphones, which is difficult, illegal, and unethical when it comes to the devices the participants already own. Another option is giving the participants already hacked devices with pre-installed logging software, as the creators of *LiveLab* did (Shepard et al., 2011). This, though, is expensive since it requires purchasing new smartphones for all participants. In addition, the creators of *LiveLab* conducted their study in 2011, when smartphones were less widespread and when none of their 24 participants owned a smartphone before the study. With the current penetration rate of smartphones in developed countries (Newzoo, n.d.), it might be difficult to convince people to switch from their devices to those provided by researchers.

Still, loggers that do not require admin access to a user device are rather non-intrusive as they typically come in the form of a regular mobile app (e.g., *Kidlogger*). Such apps normally work in the background and thus do not affect user experiences, with the only potential issue being battery usage. If a tool severely (in terms of participants' perception) drains the battery of a mobile device, participants might decide to uninstall it and thus drop out of the study. Since battery usage depends solely on the configuration of a particular logging tool, that should not be seen as a criticism of the logging approach per se but rather something to keep in mind during logger development.

Transparent Proxy Approach

The transparent proxy approach on mobiles works similarly to the proxy approach on desktop devices described above. To our knowledge, in the academic community, proxy-based tools currently only exist for desktop devices. However, they can also be adapted to provide mobile solutions (Bodo et al., 2017), which is already reflected in the existence of commercial tools.

Charles (von Randow, 2020) is a relevant example of how a (commercial) proxy-based tool might look. *Charles* works for both mobile and desktop devices. Similarly, proxy-approach-based tools designed for desktops and smartphones and for academic use can work within the same infrastructure, making merging desktop and smartphone data easier. Still, similarly to smartphone loggers, versions of the tool for Android and iOS will have to be designed separately. In addition, this approach has limitations in terms of **user experience** since it requires technically challenging pre-configuration by the participants (see Mosaic, 2018 for details on how *Charles* needs to be pre-configured on mobile devices). Further, *Charles* cannot be directly repurposed for academic studies on mobile information consumption as it only allows the tracking of mobile traffic when the smartphone is connected to the same network as the laptop on which the desktop application of *Charles* is running. This means that mobile devices can only be tracked on participants' home networks with this tool, resulting in major data loss and somewhat defeating the basic idea behind mobile tracking. We cannot state with certainty whether different, custom-developed, proxy-based tools would be able to overcome this limitation. In terms of the **types of collected information**, proxy-based tools would be able to collect all the Internet traffic going through the smartphone, thus gathering data on the content accessed through browsers as well as some in-app browsing data. Still, as stated above, right now, no fully-functioning academic tools, or even third-party tools, suitable for academic research exist, and we are not aware of any prototypes being developed in the academic community; thus, proxy-based mobile tracking tools are **not available** at the moment. Thus, we suggest that at the moment the cons of this approach outweigh its advantages.

Custom Browser or App

The custom browser-based approach requires the creation of a standalone mobile application that mimics the look and behavior of a standard browser, but also has a tracking component. So far, the custom browser approach has not been fully implemented in the academic community; hence, there are **no available** tools. Still, this approach has been suggested and discussed by scholars in the context of mobile tracking (e.g., van Atteveldt et al., 2019). However, a standalone app *Habito News*, mimicking not a standard browser but a news app—specifically, a BBC app—was successfully created and used by communication scholars in the past (Constantinides & Dowell, 2016; Constantinides et al., 2015).

A custom browser or app has some pros in terms of the **type of information** collected. Its major advantage is that it allows the content to be accessed by users. A custom news app is also highly customizable to specific research questions and allows the exploration of, for instance, how exactly users navigate a news app and browse news through it or how personalization affects their browsing behavior. Customizability offers an opportunity to conduct controlled experiments, for example, testing whether changing the layout affects how users select news stories to click on and read in more detail. Though such experiments are beyond the scope of tracking research per se, they might be interesting for communication scholars who examine topics such as personalization and user interface effects on cognition and media consumption (e.g., Constantinides & Dowell, 2016).

One major limitation of a custom browser or news app is that it is likely to significantly alter the **user experience**. Participants of a tracking study would have to change their usual habits and start browsing the Internet or reading the news using a new application, which might affect what and how they browse. For this reason, they might be especially aware of being tracked, which will make them susceptible to the Hawthorne effect (McCambridge et al., 2014)—the users might modify certain aspects of their normal behavior when they are aware that they are being observed. Besides, a standalone browser or news app does not allow researchers to get any information on the content browsed through other apps or even statistics about the usage of other apps. Finally, on certain mobile operating systems, a standalone browser with a tracking component cannot be installed at all (van Atteveldt et al., 2019). And, as in the case with mobile loggers, even if it is possible to overcome iOS limitations, a standalone tracking browser or news app has to be created separately for Android and for iOS, posing additional challenges to researchers in terms of **technical complexity**. Right now, for instance, the only tool based on this approach—*Habito News*—only works on Android systems.

Browser Extension

Technically, tools based on this approach work similarly to tracking browser extensions for desktop devices. An extension installed in a user's browser records all the URLs accessed by a user and scrapes their content. Since this is a browser extension, it is not possible to collect any data about in-app browsing and/or the usage of the apps other than for the browser where the extension is installed.

Currently, working tools based on this approach are **not available**. A prototype of such a tool was presented as an option for mobile tracking by a group of researchers studying news personalization (van Atteveldt et al., 2019). However, to the best of our knowledge, the prototype has not been developed into a fully working tool (as of January 2021, at the moment of writing).

One advantage of the browser extension approach is that such tools might be easier to develop than standalone browsers with regard to **technical complexity** but, in theory, can offer similar functionality and provide access to similar **types of information** for researchers. Besides, if installed in a user's default browser, an extension is not going to affect the **user experience** in a major way, unlike a standalone browser or news app; thus, browser extensions are better than standalone apps in terms of the user experience. The caveat is that most popular mobile browsers do not support extensions. Google Chrome for mobile does not support them at all, though some custom-made browsers that run based on Chrome architecture do (e.g., Kiwi browser). Firefox only supports extensions on Android since iOS has a closed app extensions system (Apple Developer, n.d.). Safari supports extensions on mobiles, but due to the technical differences between iOS and other platforms, they would have to be programmed separately and cannot be easily adapted from extensions initially created for other browsers and platforms. This issue is very pressing since Chrome is the most popular mobile browser worldwide with over 60% of market share, followed by Safari with approximately 20%. Firefox, which supports mobile extensions, on the other hand, has under 0.5% of mobile browser market share as of July 2020 (Statcounter GlobalStats, n.d.).

We suggest that the limited support of extensions on mobile browsers trumps the advantages of browser extensions over standalone browser mimicking apps. First, since the most popular mobile browser, Chrome, does not support extensions, the majority of users would likely have to install a mobile version of Firefox, which supports extensions. This would force them to change their standard habits in the same way as the installation of a standalone tracking browser would. Second, due to the differences between platforms, extensions would have to be developed separately for Android and iOS versions, and these cannot be easily adapted from desktop extensions in the case that they are already developed for tracking based on the screen-scraping approach. Another disadvantage is that, at some point, Safari or Firefox might deprecate the extensions' functionality, making an extension-based tracking tool obsolete.

Screen Recording

The final approach to mobile tracking that we would like to discuss is screen recording. Tools based on this approach are **available** for Android phones and either take screenshots of users' smartphone screens with high frequency (e.g., every 5 seconds; Reeves et al., 2021) or record videos of users' screens and then extract usage logs via computer vision and machine learning

(Krieter, 2019a). At the moment, screen recording is the only approach that allows the logging of the full range of **types of information** in terms of users' activities on their smartphones, including everything that is browsed through apps. Tools based on this approach are relatively user-friendly, and the tool that uses screen video recording on Android phones is open-source (Krieter, 2019b). Since this approach does not require any admin permissions, tools based on this approach can be relatively easily implemented. The **technical complexity**, though, varies depending on the functionality of a specific tool; that is, tools that simply take screenshots of the user's screen (Reeves et al., 2021) are less complex in terms of functionality and, consequently, implementation than those that include the integrated processing of the data using machine learning techniques (Krieter, 2019a). Such tools have to be developed separately for Android and iOS.

The major pitfalls of the screen recording approach are privacy issues and difficulties relating to the subsequent data processing. Since the tools record everything users look at on their devices, they can potentially record a lot of sensitive data, such as private messages, emails, or banking data. The video recording tool (Krieter, 2019a) aims to solve this privacy problem by pre-processing the videos directly on the users' phones, so only non-sensitive logs are transmitted to the researchers. The other tool (Reeves et al., 2021) does not pre-process data on users' phones. Instead, it sends encrypted screenshots to secure university storage, and then researchers apply computer vision and machine learning techniques to the data to extract meaningful logs from them. This tool, thus, carries more privacy risks—though the screenshots are encrypted, in the case of a data breach on the university servers, participants' personal data, including sensitive information, such as banking details, can be leaked.

In general, screen recording tools are non-intrusive since, similarly to smartphone loggers, they function in the background. However, depending on their implementation, they might use a lot of smartphone resources, thus affecting the **user experience**. We suggest that

Table 2. Overview of different approaches for tracking online media use for mobile devices.

Approach	Types of information	Technical complexity	User experience	Availability	Available tools
Smartphone log	visited URLs only, no content; can get other behavior data, e.g. calls log.	high	medium (can be highly intrusive depending on the implementation of a specific tool)	yes, but no browsing tracking functionality support (e.g. MobileDNA, not open source)	MobileDNA (Van Damme et al., 2019), the tool is not open source and does not track which URLs were visited
Proxy	URLs + some content (including limited in-app browsing)	high	low (difficult to set up, potentially intrusive)	yes, not academic (e.g. Charles Proxy),	None
Standalone browser/news app	Content, but only that accessed through this app/browser	medium	medium (highly intrusive)	no (outdated)	None
Browser extension	content, but only that accessed through the browser where the extension is installed	medium	medium (highly intrusive)	no (prototype only)	None
Screen-capturing	All the content including in-app browsing	medium; high for data processing	medium (can be highly intrusive depending on the implementation of a specific tool)	yes, for Android (including open source)	Screenomics (Reeves et al., 2019); unnamed screen recorder (Krieter, 2019)

the screen recording approach is currently the most potent one for mobile tracking. First, it is the only approach that allows the acquirement of a comprehensive overview of users' mobile media consumption, including in-app browsing. The second major advantage is that there are currently two functioning screen recording-based tools developed by academic researchers, one of which is available open-source. Both of the screen recording-based tools developed by academics have their advantages and disadvantages. We suggest that the integrated privacy implementation of the tool developed by Krieter (2019a), as well as the fact that it is available open-source, are major advantages over the tool developed by Reeves et al. (2021). Still, before deploying the tool from Krieter (2019a) for actual data collection, we recommend extensively testing it, including trials with a small sample of participants, to examine how it affects the functionality of smartphones with different configurations and in order to avoid usability issues.

Recommendations and Limitations

In this section, we provide recommendations for communication researchers interested in using user-centric tracking data. First, we suggest what researchers should focus on when selecting a tracking approach or tool for their studies; second, we discuss the combination of mobile and desktop tracking; third, we put forward suggestions for the further development of online tracking and, finally, outline the limitations of using in-house-developed tracking tools and briefly discuss alternative solutions.

Selection of a Suitable Tracking Tool

The first important point researchers need to consider when selecting a tracking tool is the aim of the study. For scholars interested specifically in Facebook media consumption, *Eule* (Haim & Nienierza, 2019) is the most suitable tool. Those interested in a broader set of websites should opt for other—non-website-specific—desktop tracking tools. A second selection criteria is the availability of the tool. As of now, there is no “out-of-the-box” solution available open-source. Therefore, a central question is whether researchers can get access to it (i.e., via personal or institutional agreements with the developers) or have the capacity to revise the tool. If access to an existing tool is not possible, researchers might opt for developing their own tool. In this case the choice between screen-scraping and a proxy might be made depending on the technical knowledge and recommendations of the researchers themselves or IT specialists hired to develop the tool.

Based on our evaluation, we suggest that the most promising approach for mobile tracking is screen recording. It allows for the most comprehensive capture of media usage data on mobiles, and in addition, there are two fully functioning screen recording-based, academic-developed tools available (one of them (Krieter, 2019b) open-source). However, we believe that for specific purposes, approaches other than screen recording might also be suitable. For example, for communication scholars interested in general mobile device usage (i.e., calling patterns, screen time, etc.), smartphone loggers might be of higher relevance than screen recorders. The reason for that is that the data from the loggers come in a format that is easier to analyze (i.e., one can get the usage metadata directly rather than extracting them from the screenshots or videos and filtering out relevant information). In addition, researchers interested in the usage of specific apps and/or those wishing to conduct experimental studies (i.e., examining how changes in the layout of a given news app affect users’ news consumption patterns) might wish to opt for a standalone browser or app approach instead.

Combination of Desktop and Mobile Tracking

In the ideal combination scenario, the desktop and mobile tracking tools used would be based on the same technical approach. This would allow the building of a universal back-end infrastructure around both tools and would make the structures of the data very similar or identical, subsequently, simplifying the data processing. There are two approaches that could combine mobile and desktop tracking: a transparent proxy (once there is a way to decrypt HTTPS traffic for mobiles outside of the home network) and browser extensions (in the case that more mobile browsers add extension support). At the moment, however, mobile tracking tools based on these approaches do not exist and, due to approach-specific obstacles (i.e., a lack of support for extensions on mobile browsers), they might not be developed in the future either. Therefore, researchers willing to combine desktop and mobile tracking will have to deal with tools based on two different approaches. We suggest that to get both mobile and desktop tracking data, it is currently best to use screen recording accompanied by the desktop tracking tool that is the most suitable for researchers' needs (or at least the one that is easily available to them).

In the cases where the use of a screen-recording approach is not possible for a given research project, we suggest combining desktop tracking with the gathering of mobile digital trace data (i.e., asking participants to share their mobile browser histories with the researchers).

Further Development of Tracking Tools in the Academic Community

Despite the efforts and the existence of several promising approaches, so far, there is no "out-of-the-box" tracking tool that is available to the academic community. Therefore, we list features that are crucial for researchers who are faced with the task of developing their own tracking tool. First, it is crucial that the actual content that participants have seen is collected and that the collection of content from social media is not fundamentally excluded. Second, comprehensive privacy features must be implemented. Here we suggest implementing, first, dedicated anonymized log-in credentials for the end-users. This is necessary to protect their privacy and allows tracking data to be easily merged with survey data without compromising users' identities. In addition, we suggest the inclusion of the technical requirements for both black- and whitelists. This will allow researchers to customize the tool for different research purposes. Implementing a privacy button that allows users to temporarily disable the tracking tool will give users additional control over their data and might positively affect their willingness to participate. However, this can increase the awareness of being observed and result in behavioral adjustment. This shows that there is a need for future studies on how tracking tools and different (privacy) features affect participants' behavior and their willingness to participate in a tracking study. Finally, we strongly argue the case for making the tools available open-source to the academic community in order to promote scientific research, similarly to the way open access data and analytic tools do (e.g., Dienlin et al., 2020). In addition, this will prevent the investment of resources in the development of tools that already exist and promote cooperation for the maintenance of previously developed tools. The problem regarding updates, thus, could be solved by creating and maintaining long-term research infrastructures within which the tools would be regularly updated so as to remain compatible with the new versions of browsers and operating systems on both mobile and desktop devices. Unfortunately, the creation of open-access tools is not always regarded as equally important a research output as, for instance, journal articles, and their developers are not always cited when someone else uses the tool. Thus, it is necessary to ensure that the original developers are properly credited during the subsequent use of the tools. We suggest that this can be partially addressed if tool maintenance is addressed and updated infrastructure is created and hosted, either at one of the university departments that has already developed its own tracking tool or by an academic institution that specifically focuses on research infrastructure (such as GESIS in Germany or FORS in Switzerland), and distributed under a license that allows free-of-charge use for research purposes if the original creators of the tool are properly credited.

Limitations of Academic Tracking and Alternatives

The development of in-house tracking tools can be costly and complex, and not all research teams have enough funding and/or technical capabilities to develop their own tracking tools or even to be able to set up the back-end infrastructure for an existing one. In these cases, we suggest using one of the alternatives. If there is enough funding, one can purchase tracking data from a private company, such as *comScore* or *TNS NIPO*. But this comes with several caveats: First, to the best of our knowledge, such companies only offer domain- or URL-level data. In addition, the companies only have data available for certain countries and within their panels, which might be a major disadvantage if researchers are interested in a specific population. As an alternative, researchers might opt for a form of digital trace data (Menchen-Trevino, 2013). Both alternatives are limited in terms of their application as they do not allow the examination of research questions related to content personalization and/or (incidental) news exposure on social media. Still, if the data provided by the alternative means allow the answering of the research questions that scholars have in mind, using such alternatives might be preferable due to the amount of resources that can be saved with them.

Paper Limitations

As in the present paper we focus on technical questions, we were only able to discuss ethical questions, as well as the question of participation rates, to a limited extent. Further, we mainly focused on the process of data collection and discussed data pre-processing, data encryption, and data storage only marginally. These steps are essential for these highly sensitive data, and a similar overview is needed for these questions. Finally, we largely described the technical implementation of the front-end for each of the approaches. Back-end infrastructure has not been discussed here in detail; however, researchers interested in implementing tracking should keep in mind that it is also important and can pose significant technical challenges.

Acknowledgements

We gratefully thank Mykola Makhortykh for his useful comments that really helped to improve the manuscript. Further, we would like to thank the editor and the anonymous reviewers for their excellent reviews and suggestions.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This paper has been written within the project “Reciprocal relations between populist radical-right attitudes and information behavior: a longitudinal study of attitude development in high-choice information environments” lead by Silke Adam (U of Bern) and Michaela Maier (U of Koblenz-Landau) and sponsored by the SNF (100001CL_182630/1) and DFG (MA 2244/9-1).

References

Adam, S., Maier, M., Aigenseer, V., Urman, A., & Christner, C. (2019, July 17–20). *WebTrack – Tracking users' online information behavior while screen-scraping content* [Poster presentation]. 5th International Conference on Computational Social Science (IC²S²), Amsterdam, Netherlands.

Apple Developer. (n.d.). App Extensions. Retrieved August 12, 2020, from https://developer.apple.com/app-extensions/

Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. H. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures*, *11*(3), 173–190. https://doi.org/10.1080/19312458.2017.1317337

Bodo, B., Helberger, N., Irion, K., Zuiderveen Borgesius, F., Moller, J., van der Velde, B., Bol, N., van Es, B., & de Vreese, C. H. (2017). Tackling the algorithmic control crisis: The technical, legal, and ethical challenges of research into algorithmic agents. *Yale Journal of Law and Technology*, *19*(1), 133–180. Available at: https://digitalcommons.law.yale.edu/yjolt/vol19/iss1/3

Bruns, A. (2013). Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship. *First Monday, 18*(10). https://doi.org/10.5210/fm.v18i10.4879

Clement, J. (2020, June 17). *Share of mobile internet traffic in global regions 2020*. Statista. https://www.statista.com/statistics/306528/share-of-mobile-internet-traffic-in-global-regions/

Coffey, S. (2001). Internet audience measurement: A practitioner's view. *Journal of Interactive Advertising*, *1*(2), 10–17. https://doi.org/10.1080/15252019.2001.10722047

Constantinides, M., & Dowell, J. (2016). User interface personalization in news apps. In F. Cena, M. Desmarais, D. Dicheva, & J. Zhang (Eds.), *Extended proceedings of the 24th Conference on User Modeling, Adaptation and Personalisation (UMAP '16),* Halifax, Canada.

Constantinides, M., Dowell, J., Johnson, D., & Malacria, S. (2015). Exploring mobile news reading interactions for news app personalisation. In S. Boring, E. Rukzio, H. Gellersen, & K. Hinckley (Eds.), *Proceedings of the 17th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct: Copenhagen, Denmark, August 24–27, 2015* (pp. 457–462). ACM.

De Vreese, C. H., & Neijens, P. (2016). Measuring media exposure in a changing communications environment. *Communication Methods and Measures*, *10*(2–3), 69–80. https://doi.org/10.1080/19312458.2016.1150441

Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., Lukito, J., Bier, L. M., Zhang, R., Johnson, B. K., Huskey, R., Schneider, F. M., Breuer, J., Parry, D. A., Vermeulen, I., Fisher, J. T., Banks, J., Weber, R., Ellis, D. A., & De Vreese, C. (2020). An agenda for open science in communication. Journal of Communication. Advance online publication. https://doi.org/10.1093/joc/jqz052

Flaxman, S. R., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(1), 298–320. https://doi.org/10.1093/poq/nfw006

Haim, M. (2019). *Browser Plug-in for Computational Observation at FBforschung.de* (Version 1.1.0) [Source code]. GitHub. https://github.com/MarHai/fbforschung

Haim, M., & Nienierza, A. (2019). Computational observation: Challenges and opportunities of automated observation within algorithmically curated media environments using a browser plug-in: Challenges and opportunities of automated observation within algorithmically curated media environments using a browser plug-in. *Computational Communication Research*, *1*(1), 79–102. https://doi.org/10.5117/CCR2019.1.004.HAIM

Jünger, J. (2018). Mapping the field of automated data collection on the web: Collection approaches, data types, and research logic. In C. Stützer, M. Welker, & M. Egger (Eds.), *Computational social science in the age of big data: Concepts, methodologies, tools, and applications* (pp. 104–130). Herbert von Halem.

Jürgens, P., Stark, B., & Magin, M. (2020). Two Half-Truths Make a Whole? On Bias in Self-Reports and Tracking Data. *Social Science Computer Review*, *38*(5), 600–615. https://doi.org/10.1177/0894439319831643

Karr, C. (2014). *Roxy-Proxy* [Source code]. GitHub. https://github.com/audaciouscode/Roxy-Proxy

Kleppe, M., & Otte, M. (2017). Analysing and understanding news consumption patterns by tracking online user behaviour with a multimodal research design. *Digital Scholarship in the Humanities*, *32*(2), 158–170. https://doi.org/10.1093/llc/fqx030

Krieter, P. (2019a). Can I record your screen? Mobile screen recordings as a long-term data source for user studies. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (MUM '19)*. ACM.

Krieter, P. (2019b). *Screenlogger-Android* [Source code]. GitHub. https://github.com/pkrieter/screenlogger-android

LiKamWa, R., Liu, X., Lane, N., & Zhong, L. (2013, June). MoodScope: Building a mood sensor from smartphone usage patterns. In H.-H. Chu (Eds.), *Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services* (pp. 389–402). ACM.

Liu, B. (2008). *Web data mining: Exploring hyperlinks, contents, and usage data*. *Data-centric systems and applications*. Springer.

Marres, N., & Weltevrede, E. (2013). Scraping the social? Issues in live social research. *Journal of Cultural Economy*, *6*(3), 313–335. https://doi.org/10.1080/17530350.2013.772070

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, *67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015

Menchen-Trevino, E. (2013). Collecting vertical trace data: Big possibilities and big challenges for multi-method research. *Policy & Internet*, *5*(3), 328–339. https://doi.org/10.1002/1944-2866.POI336

Menchen-Trevino, E., & Karr, C. (2012). Researching real-world web use with Roxy: Collecting observational web data with informed consent. *Journal of Information Technology & Politics*, *9*(3), 254–268. https://doi.org/10.1080/19331681.2012.664966

Möller, J., van de Velde, R. N., Merten, L., & Puschmann, C. (2020). Explaining online news engagement based on browsing behavior: Creatures of habit? *Social Science Computer Review*, *38*(5), 616–632. https://doi.org/10.1177/0894439319828012

Mosaic, M. (2018, June 22). *How to use Charles Proxy to view the data your mobile apps send & receive.* https://null-byte.wonderhowto.com/how-to/use-charles-proxy-view-data-your-mobile-apps-send-receive-0185364/

Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018). Networks of audience overlap in the consumption of digital news. *Journal of Communication*, *68*(1), 26–50. https://doi.org/10.1093/joc/jqx007

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143–150. https://doi.org/10.1186/s12874-018-0611-x

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D., & Nielsen, R. (2018). *Reuters Institute Digital News Report 2018*. Reuters Institute for the Study of Journalism. http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475

Newzoo. (n.d.). *Top countries by smartphone users*. Retrieved August 12, 2020 from https://newzoo.com/insights/rankings/top-countries-by-smartphone-penetration-and-users/

Prior, M. (2009). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, *73*(1), 130–143. https://doi.org/10.1093/poq/nfp002

Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., Chiatti, A., Cho, M., Roehrick, K., Yang, X., Gagneja, A., Brinberg, M., Muise, D., Lu, Y., Luo, M., Fitzgerald, A., & Yeykelis, L. (2021). Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. *Human–Computer Interaction*, *36*(2), 150–201. https://doi.org/10.1080/07370024.2019.1578652

Research Group for Media, Innovation and Communication Technologies. (2019). *MobileDNA* (Version 1.5.0). [Mobile app]. Google Play. https://play.google.com/store/apps/details?id=be.ugent.mobiledna&hl=en

Shepard, C., Rahmati, A., Tossell, C., Zhong, L., & Kortum, P. (2011). LiveLab: Measuring wireless networks and smartphone users in the field. *ACM SIGMETRICS Performance Evaluation Review*, *38*(3), 15–20. https://doi.org/10.1145/1925019.1925023

Statcounter GlobalStats. (n.d.). *Mobile browser market share worldwide: July 2019–July 2020.* Retrieved August 12, 2020, from https://gs.statcounter.com/browser-market-share/mobile/worldwide

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, *38*(5), 503–516. https://doi.org/10.1177/0894439319843669

Teufl, P., Zefferer, T., & Stromberger, C. (2013). Mobile device encryption systems. In J. Lech, L. J. Janczewski, H. B. Wolfe, & S. Shenoi (Eds.), *Security and privacy protection in information processing systems* (pp. 203–216). Springer.

Tossell, C. C., Kortum, P., Rahmati, A., Shepard, C., & Zhong, L. (2012). Characterizing web use on smartphones. In J. A. Konstan, E. H. Chi, & K. Höök (Eds.), *CHI 2012, it's the experience. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2769–2778). ACM.

Tossell, C. C., Kortum, P., Shepard, C., Barg-Walkow, L. H., Rahmati, A., & Zhong, L. (2012). A longitudinal study of emoticon use in text messaging from smartphones. *Computers in Human Behavior*, *28*(2), 659–663. https://doi.org/10.1016/j.chb.2011.11.012

Van Atteveldt, W., Bogaardt, L., van Hees, V., Loecherbach, F., Möller, J., Trilling, D., & Welbers, K. (2019). *Gathering mobile news consumption traces: An overview of possibilities and a prototype tool based on Google Takeout* [Conference presentation]. 69th International Communication Association Conference, Washington, USA.

Van Damme, K., Courtois, C., Verbrugge, K., & De Marez, L. (2015). What's APPening to news? A mixed-method audience-centred study on mobile news consumption. *Mobile Media & Communication*, *3*(2), 196–213. https://doi.org/10.1177/2050157914557691

Van Damme, K., Martens, M., van Leuven, S., Vanden Abeele, M., & Marez, L. de (2020). Mapping the mobile DNA of news. Understanding incidental and serendipitous mobile news consumption. *Digital Journalism*, *8*(1), 49–68. https://doi.org/10.1080/21670811.2019.1655461

Von Randow, K. (2020). *Charles* (Version 5.4.6) [Computer software]. https://www.charlesproxy.com/download/

Wagner, D., Rice, A., & Beresford, A. (2014). Device analyzer: Understanding smartphone usage. In I. Stojmenovic, Z. Cheng, & S. Guo (Eds.), *Mobile and ubiquitous systems: Computing, networking, and services* (pp. 195–208). Springer.

Wieland, M., Au, A.-M. in der, Keller, C., Brunk, S., Bettermann, T., Hagen, L. M., & Schlegel, T. (2018). Online behavior tracking in social sciences: Quality criteria and technical implementation. In C. Stützer, M. Welker, & M. Egger (Eds.), *Computational social science in the age of big data: Concepts, methodologies, tools, and applications* (pp. 131–160). Herbert von Halem.