

# Quantifying the digitisation of everyday lives

Measurement opportunities for large-scale surveys

CLS working paper number 2021/7

By Alex Papadopoulos

This paper was written in June 2019, and published as a working paper in March 2021 by the UCL Centre for Longitudinal Studies.

UCL Social Research Institute

University College London

20 Bedford Way

London WC1H 0AL

[www.cls.ucl.ac.uk](http://www.cls.ucl.ac.uk)

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Social Research Institute, University College London. It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit [www.cls.ucl.ac.uk](http://www.cls.ucl.ac.uk).

This document is available in alternative formats. Please contact the Centre for Longitudinal Studies.

Tel: +44 (0)20 7612 6875

Email: [clsfeedback@ucl.ac.uk](mailto:clsfeedback@ucl.ac.uk)

## Disclaimer

This working paper has not been subject to peer review.

CLS working papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Any opinions expressed here are those of the author(s) and not those of the UCL Centre for Longitudinal Studies (CLS), the UCL Social Research Institute, University College London, or the Economic and Social Research Council.

## How to cite this paper

Papadopoulos, A. (2021) *Quantifying the digitisation of everyday lives: Measurement opportunities for large-scale surveys*. CLS Working Paper 2021/7. London: UCL Centre for Longitudinal Studies.

# Contents

<b>INTRODUCTION</b>	<b>3</b>
<b>THE CHALLENGE</b>	<b>3</b>
<b>PASSIVE AND ACTIVE TRACKING</b>	<b>4</b>
<b>THE DATA SO FAR</b>	<b>4</b>
<b>RESEARCH APPROACH</b>	<b>7</b>
<b>FINDINGS</b>	<b>8</b>
<b>POTENTIAL PARTNERS</b>	<b>9</b>
<b>POSSIBLE TECHNOLOGICAL SOLUTIONS</b>	<b>9</b>
<b>BROWSER PLUG-INS – [PASSIVE, REQUIRES ONE-OFF INSTALLATION BY THE USER]</b>	<b>10</b>
<b>DESKTOP TIME TRACKING – [PASSIVE]</b>	<b>10</b>
<b>SELF-TRACKING SMARTPHONE APPLICATIONS – [PASSIVE]</b>	<b>11</b>
<b>MOUSE TRACKING – [ACTIVE/PASSIVE]</b>	<b>12</b>
<b>SESSION REPLAY – [ACTIVE/PASSIVE]</b>	<b>12</b>
<b>BROWSER FINGERPRINTING – [PASSIVE]</b>	<b>12</b>
<b>COOKIES (HTTP COOKIE) – [PASSIVE]</b>	<b>13</b>
<b>DEEP PACKET INSPECTION (DPI) – [PASSIVE]</b>	<b>14</b>
<b>MOST VIABLE SOLUTION: DEEP PACKET INSPECTION (DPI) OVER VIRTUAL PRIVATE NETWORK (VPN) – [PASSIVE]</b>	<b>15</b>
<b>CONCLUSIONS</b>	<b>16</b>
<b>REFERENCES AND APPENDIX</b>	<b>17</b>

# Introduction

The use of technology in people's everyday lives has dramatically increased during the past decade. The introduction of new and improved hardware (smartphones, wearables, increased internet speed), as well as the abundance of software for a variety of needs have established new ways of entertainment, participation in life events, communicating and many more. In the UK, 78% of the population owns a smartphone, a 61% increase since 2008. According to Ofcom's report<sup>1</sup>, people claim to spend a total of 24 hours per week online, more than twice the amount of time reported in 2007. Changes have been noted on the ways people use technology as well as the amount of time they spend using it. While in 2007 the desktop computer was the most popular household technology item with 69% of people having one at home, in 2018 with the introduction of smartphones and tablets, and streaming players (like iPlayer and Netflix) the number has decreased to 28% and users turned to mobile technologies. In comparison to other demographics, young adults (18-34) spend the largest percentage of time with digital devices. This dramatic shift to embed technology in most, if not all, aspects of everyday life raises multiple questions about the impact it has on people's lives. Plenty of studies have tried to touch on various aspects of technology use and its impact on physical and mental wellbeing, using self-reported scales like The Compulsive Internet Use Scale (CIUS)<sup>2</sup>, media use and attitudes, and personal tracking of screen time, but there are no recorded scientific studies that use accurate, live background data from users to measure online activity in a longitudinal manner. It is also worth mentioning that all available data of this kind come from commercial digital companies that collect and analyse them for specific purposes.

## The challenge

Painting a scientific picture of people's digital and online lives, is not only going to help uncover potential underlying pathologies, but it is also going to help realize the extent to which the digital affects the physical and vice-versa. Longitudinal studies using real time data from technology users will be able to accurately answer questions like:

- How do people use technology?
- What types of technologies do people use and to what extent?
- How much time do people spend online and how do they divide this time between different activities?
- What types of devices do people use and what do they use them for?
- What impact does technology use have on people?

There is no one way to collect and analyse information about digital activity and behaviour, with methodologies varying from interviews and self-reported questionnaires, to diary studies and website analytics. Self-reports of digital behaviour, though widely used, are subject to measurement error, particularly recall problems. In this report, we aim to identify robust, new measures of online activity including direct objective measures.

---

<sup>1</sup> Ofcom's Communications Market Report: <https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr-2018/report>

<sup>2</sup> The Compulsive Internet Use Scale (CIUS): <https://www.ncbi.nlm.nih.gov/pubmed/19072079>

## Passive and Active tracking

Different tracking methods are related to different tracking mechanisms. An example could be the distinction between passive and active tracking – passive being the tracking of an activity that happens without the person's active involvement (but not necessarily without their consent) as a background mechanism i.e. a step counter on a smartphone, and active tracking being the tracking of an activity that happens consciously with an effort from the user, such as a diary entry or a total calorie amount for a meal entry on a smartphone application.

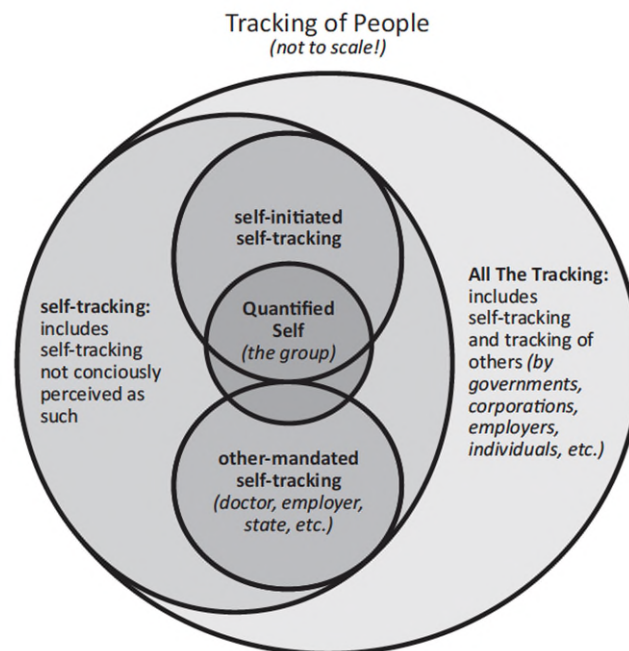


Figure 1.2 Taxonomy of types of tracking.  
Source: Whitney Boesel, *The Society Pages*.

Figure 1

Figure 1<sup>3</sup>, presents a Venn diagram that explains the various ways of tracking people's activities, that can also apply to digital activities. As it can be observed on the diagram, tracking of people can take various forms, from conscious self-tracking, to self-tracking that happens unknowingly and not perceived as such, such as personal diary keeping which is an activity that can serve other purposes as well as tracking. Finally, the tracking from external sources like doctors, employers, or the government, which can extend from health history records to marital status and vehicle registrations which are all forms of tracking. Related to the discussion about active and passive tracking above, Figure 1 breaks down a variety of active and passive tracking methods and their overlap. It is evident that much of the data that is collected from people comes from self-tracking in various forms, both in passive and active ways.

## The Data so far

Current studies have been consistently showing that the adoption of technology from people of all ages has been dramatically increasing, but at the same time it is constantly shifting

<sup>3</sup> The Society Pages: <https://thesocietypages.org/cyborgology/author/whitneyerinboesel/>

through different types of technologies. Technology use is not limited to one device or a specific activity, and user activity is constantly changing, as it can be seen from the table below:

**Table No.1 : Technological Activities**

<b>Category</b>	<b>Individual Activities included in category</b>
<b>E-mail</b>	Send/Receive Emails
<b>Communications</b>	Communicating via instant messaging Making voice calls Making video calls
<b>Transactions</b>	Online shopping (purchasing goods/services/tickets etc) Trading/auctions e.g. ebay
<b>Banking</b>	Banking
<b>Social Media</b>	Using Social Networking
<b>News</b>	Accessing News
<b>Information for work/school/college</b>	Finding/downloading information for work/business/school/college/university/homework
<b>Watch short video clips</b>	Watching short video clips (e.g. on YouTube, Dailymotion, Vimeo, or Facebook)
<b>Watch TV content</b>	Watching TV programmes or film content online (iPlayer, Netflix, Amazon etc.)
<b>Health</b>	Finding information on health related issues
<b>Radio/Audio services</b>	Listening to radio Streamed audio services (Spotify, Deezer, Apple Music etc.)
<b>Government Services</b>	Using local council/Government sites Finding information, completing processes such as tax returns etc.
<b>Games</b>	Playing games online or offline
<b>Remote Services</b>	Assessing files through cloud services (Dropbox, Google Drive etc) Remotely control TV services at home such as Sky+, SkyQ etc. Remotely control or monitor household appliances
<b>Upload/add content</b>	Uploading/Adding content to the internet e.g. photos, videos, blogposts

Of course, technology use is not limited in these categories, especially as the internet grows and expands creating new activities and ways of interaction between users (e.g. Virtual Reality (VR), cryptocurrency mining, online activism, mapping etc.) but it is important to acknowledge that the table above showcases a big chunk of activities that people engage with online.

As it was previously mentioned, online behaviour has been changing rapidly during the past decade and this can be observed both on the activities people engage online, but also on the types of devices that people use to access online content. As mentioned in the introduction above, people have been using smartphones and tablets a lot more and desktop PCs a lot less during the past decade, and new devices like smart speakers and smart watches were introduced and adapted successfully. Figure 2, provides a comparison of device use between 2008 and 2018, with the very recent addition of smart speakers, VR headsets, and Smart TV sets.

These figures vary between age groups and demographic populations.

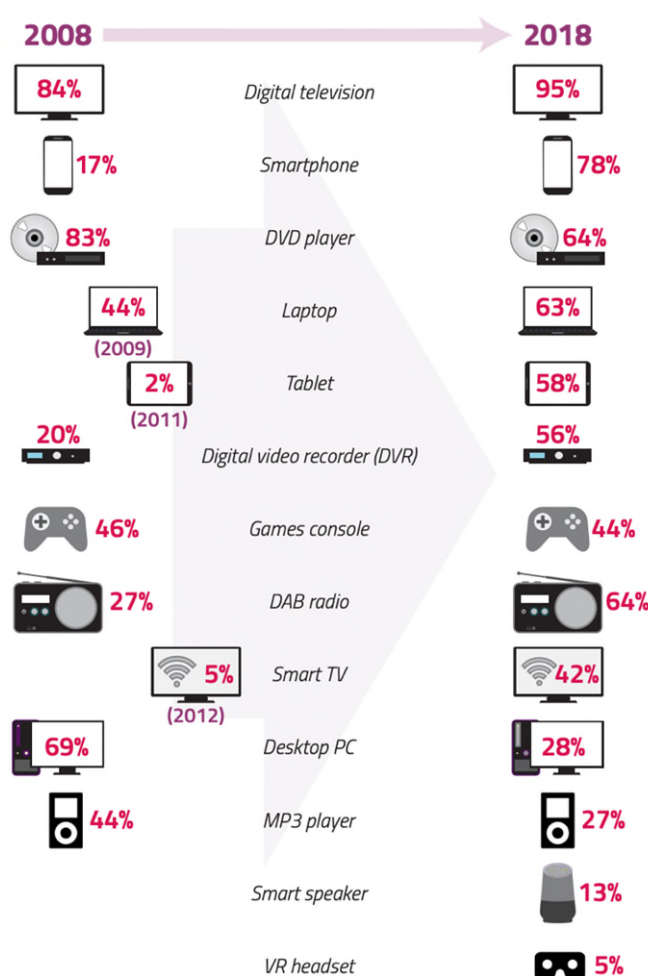


Figure 2

For example, although 70% of adults own and use a smartphone, between the ages 16-24 this increases to 95% but decreases significantly between the ages 55-64 (50%), with a further drop between the ages 65-74 (22%). Similar rates can be observed for other device types like desktop computers, tablets and smart TVs, as well as on the rates of people going online (88% in total, decreasing to 65% at the ages of 65-74).

As the population of people who use such technologies and the usage itself grows, new areas of research emerge both regarding the technological and the human factor. A big part of research is concerned with the rapid adoption of technology within the youth, examining implications on mental and physical health and general well-being related with technology use. Quantifying technology use and using it for social science studies has become a complicated matter as technology use is becoming an integral part of everyday life and it is becoming seemingly impossible to separate confounding variables. For this reason, this study aims to identify more specific

<sup>4</sup> [https://www.ofcom.org.uk/data/assets/pdf\\_file/0017/107360/Ofcom-Technology-Tracker-H2-2017-data-tables.pdf](https://www.ofcom.org.uk/data/assets/pdf_file/0017/107360/Ofcom-Technology-Tracker-H2-2017-data-tables.pdf)



and accurate methods of measurement in order to be implemented and aid future longitudinal studies. The methodology, findings, and conclusions will be discussed in the sections below.

## Research Approach

New and emerging technologies, as well as already existing methodologies for collecting and analysing user data regarding online behaviour, screen time, technology preferences and other technology use related information, could potentially aid national longitudinal studies with more specific and accurate data. Current data streams mostly come from participants' self-reports that are currently being collected through interviews, surveys, and diary studies.

With an aim to identify the most viable technology options for big scale studies, this review focuses on already existing work done by scientific and commercial parties. The segregation between academic and commercial research was applied mainly due to the differences in methodologies and goals of the studies at hand, as commercial companies are mainly interested in advertisement and data exploitation for profit, unlike their academic counterparts which are interested in purely scientific research.

The scientific literature review focuses on current and previous methodologies to measure screen-time and digital activities. The review of commercial activities relies on reports from big commercial organisations (i.e. Ofcom and Comscore), though consumerist nature of commercial research should be borne in mind. One of the goals of this study was to identify organisations and companies that could be potential partners in future research and data collection..

In terms of identifying existing research, the approach that was taken consisted of five steps:

- Current knowledge of methodologies/practices/technologies
- Organisations/companies/scientific institutes that these practices are used
- Types of measures and research areas that those measures apply to
- Potential different types of quantifiable measures that could be used for specific purposes
- Feasibility for nation-wide longitudinal studies

There are many examples for self-tracking devices and applications driven by the cultural phenomenon of self-tracking by technology, called “Quantified Self”, which incorporates wearable and other technologies to gain better knowledge about one’s self, with the goal of improving physical, mental, and/or emotional performance. The “Quantified Self” has created a lot of opportunities for the development and utilisation of a wide range of data collection technologies. Examples being the wide spread introduction of wearable technologies such as the Fitbit, the Apple Watch, and a variety of smartphone applications measuring exercise levels, tracking sleep, analysing screen time usage and many more. <sup>5</sup>

See also: [Neff, G. and Nafus, D., 2016. \*Self-tracking\*. MIT Press.](#)

[Lupton, D., 2016. \*The quantified self\*. John Wiley & Sons.](#)

Screen time usage was a central focus with the aim of identifying technologies that specifically aid its measurement accurately and in real time. Much of this review was focused around identifying previous scientific studies that measured screen time usage, as well as new

---

<sup>5</sup> Examples of quantified self applications can be found in this article: <https://josiahvorst.com/9-apps-tools-you-can-use-for-a-quantified-self-framework/>

technologies that could aid future studies measure this more accurately. It is important to note that so far, there are no published scientific articles that mention direct measurement of screen time usage.

It was hypothesized that academic organisations that deal with big data would be involved with research of this nature, and could potentially aid similar research in the future, but as far as we were able to ascertain this is not the case, nor could we identify the existence of data sets of this type (that are not coming from or aimed to commercial use). This was a major setback in the current review, as it would be a very useful asset to have, especially taking into account the differentiation of academic and commercial research.

Despite this setback, technical solutions such as wearable technologies, smartphone applications, web browser extensions, and home network solutions have been meticulously explored and collected in a spreadsheet (see appendix) as possible solutions. The most viable will be presented in the findings section below.

## Findings

Academic research on the development of tools that measure and classify technology use has been sparse and mostly based on self-report methodologies. Quite a few studies have been using tailored scales (e.g. the Online Cognition Scale [OCS] that measures problematic internet use, the Compulsive Internet Use Scale [CIUS] etc.). Much of the findings on academic research indicate that research on that field is currently in its primal stages, and is not involved with large data sets fed by real-time information of usage. Large academic institutes like the Oxford Internet Institute, are concerned with aspects of digital technology and young people, including gaming and screen time, but they too use self-reported methods for their studies. Appendix No.1 showcases some of the most relevant academic research that has been conducted on screen time and technology use, mainly in the social sciences.

In contrast, research in the commercial sector done by large statistical companies for advertisement and other commercial purposes has been heavily based on data streams coming directly from users, as well as from the information taken by interviews and surveys, and can give a lot of useful insights about how and how much people use technology. Companies like Ofcom (<https://www.ofcom.org.uk/home>), Ukom (<https://ukom.uk.net/>), Globalwebindex (<https://www.globalwebindex.com/>), and Comscore (<https://www.comscore.com/>) to name a few, have been developing extensive reports on online behaviour and trends regarding technology use among a variety of demographic groups.

The downside of research conducted by commercial companies, is that it is tailored and bound to specific products and services, using technologies like cookies and APIs that link back to the research funding provider. Wide-spread technologies for tracking online behaviour like cookies, and other types of analytics require direct communication with the source that provides the material the user sees, and it is generally connected with commercial uses. As an example, cookies from an e-commerce webpage can be controlled and observed only from the company that owns the site, and any advertisement or other organisations that the company collaborates with and gives them access, but not from any external sources.

## Potential Partners

The table below presents a list of the most useful organisations and key people that were identified from this research.

**Table No.2: Related Organisations**

Organisations	Link	Key People
<b>The Centre for Internet and Technology Addiction</b>	<a href="https://virtual-addiction.com/">https://virtual-addiction.com/</a>	Dr. David Greenfield Assistant Clinical Professor of Psychiatry, University of Connecticut School of Medicine
<b>Oxford Internet Institute</b>	<a href="https://www.oii.ox.ac.uk/">https://www.oii.ox.ac.uk/</a>	Professor Philip Howard, Professor Victoria Nash, Professor Andrew Przybylski
<b>Edison Research</b>	<a href="https://edisonresearch.com/">https://edisonresearch.com/</a>	Commercial organisation
<b>Emarketer</b>	<a href="https://www.emarketer.com/">https://www.emarketer.com/</a>	Commercial organisation
<b>Comscore</b>	<a href="https://www.comscore.com/">https://www.comscore.com/</a>	Commercial organisation
<b>Statista</b>	<a href="https://www.statista.com/api">https://www.statista.com/api</a>	Commercial organisation
<b>Ofcom</b>	<a href="https://www.ofcom.org.uk/home">https://www.ofcom.org.uk/home</a>	Commercial organisation
<b>Ukom</b>	<a href="https://ukom.uk.net/">https://ukom.uk.net/</a>	Commercial organisation
<b>Globalwebindex</b>	<a href="https://www.globalwebindex.com/">https://www.globalwebindex.com/</a>	Commercial organisation
<b>Marketwatch</b>	<a href="https://marketwatch.com">https://marketwatch.com</a>	Commercial organisation

## Possible Technological Solutions

This section provides some possible technological solutions that could potentially be used for research on digital habits and activity tracking. Most of these solutions include tracking of specific activities, time spent, and exact interactions with said technologies and can provide a variety of data for user's digital behaviour. Regarding the user involvement for tracking these activities, most of the proposed solutions are classified under "passive tracking" methodologies, which gives the advantage of not expecting from the user to keep track of every digital activity they do, especially when multiple devices come in the picture.

In terms of installation and use from participants, most of the proposed technologies require one-off remote installation or activation from the participant to enable data tracking. Additionally, certain solutions require the installation of specialized hardware (e.g. standalone DPI packages need to be installed in home networks) or the user to agree to provide access to data from specific devices (e.g. Screen Time and Digital Wellbeing are native apps for iOS and Android devices respectively).

There exist some smartphone applications that allow users to manually track the time they spend online or their online activities, but we have focused on passive measurement as this is likely to give greater accuracy and granularity than self-report.

Depending on the technology, different methods are used to track and access the data users are providing. Many of these solutions are web-based only, meaning that other applications on the same device, or multiple devices are not supported and data from those should be collected in different ways, where others are more unified and can provide information for more than one devices and applications. Web applications run on remote servers and can only be accessed by web browsers, so the data that is circulated are only accessible to the web browser and the server, and is not related to the physical device that is being used or other

applications running on that device, meaning that technologies that track web activity cannot generally track what happens outside the browser. Similarly, some applications are bound to user accounts rather than devices (mostly seen in smartphone applications), and can cover everything a user does from a specific account regardless of the physical device. In terms of the accuracy of these measures, due to the fact that they all feed from real-time data directly from the user, there is little room for inaccuracies, except in situations where the tracking is network-dependent and there are external faults in the network the user is on.

More on how these proposed technologies work will be discussed below, and the most viable solution for large scale social research will be presented at the end of this section.

### Browser plug-ins – [passive, requires one-off installation by the user]

Browser plug-ins allow users to keep track of how much time they spend in each of the websites they visit. Such plugins have the ability to provide the users with both detailed or aggregate reports of their online habits; i.e. how much time they spend in known social websites, how much time they spent on a specific one, or even how much time they spent on specific activities of a social website. Some of the existing plugins have the ability to allow the users to create an account and gather data from more than one of their browsers or devices as well as the option to view the reports online.

One of the downsides of browser plugins is that they can only capture data from websites but not from applications running outside the browser or even other extensions running in the same browser.

Examples: Web Activity Time Tracker, Trackr, Webtime Tracker, StayFocusd, TimeYourWeb

### Desktop time tracking – [passive]

Desktop time tracking software can be installed in Windows, Linux, OSX, or other operating systems (OS) and can gather information about the applications the user is using at any given time. There are a number of ways these applications can function but most work by using the OS' APIs to get access to the applications the user has open, and their window titles or other information.

Such applications have the added benefit of being able to track both online and offline applications, but lack in their abilities to finely track the user's behaviour inside complex applications such as browsers due to their ability to only track window titles or at best tab titles.

More advanced applications (usually seen in enterprise/corporate environments) might pair this tracking functionality with a local DPI/VPN to gain access to additional information from the user's network traffic.

Examples: RescueTime, SelfControl, ManicTime, Project Hamster

### Self-tracking smartphone applications – [passive]

Similar to desktop time tracking applications, both iOS 12 (screen time) and Android (digital wellbeing) operating systems come with pre-installed applications allowing the user to track how much time they spent on each application and present them with reports on daily/weekly/monthly usage.

In addition to the ability to track application usage, both OSes allow users to limit the amount of time they spent on each application by setting a limit on how many hours/minutes the user can have the application open per day.

Currently neither OS provides APIs to allow developers access to these data and instead only provide this information through their own applications and reports. However, data from these applications can be accessed and shared by the users, although there is no current framework or knowledge to the day, for accessing raw data externally. Another limitation of these screen-time measures is that they are highly aggregated.

Figures 3 and 4 below, provide a picture of what kind of data these applications can provide.

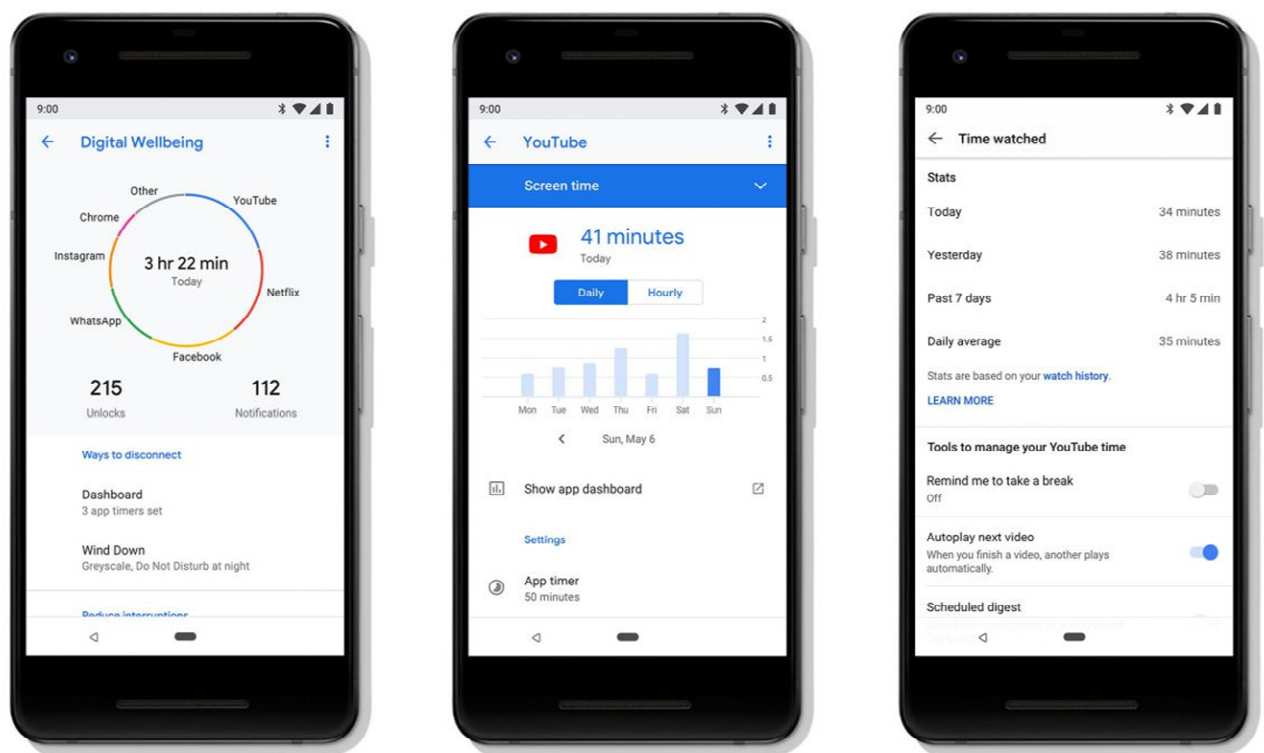


Figure 3: Digital Wellbeing by Android



Figure 4: Screen Time by iOS

### Mouse Tracking – [active/passive]

Mouse tracking (also known as cursor tracking) is the use of software to collect users' mouse cursor positions on the computer. This goal is to automatically gather richer information about what people are doing, typically to improve the design of an interface. Often this is done on the Web and can supplement eye tracking in some situations.

Mouse tracking can be performed natively from the website by installing specific javascript code, or by having the user install an external plug-in on their browser.

### Session Replay – [active/passive]

Session replay is the ability to replay a visitor's journey on a web site or within a web application. Replay can include the user's view (browser or screen output), user input (keyboard and mouse inputs), and logs of network events or console logs. It is supposed to help improve customer experience[1] and to identify obstacles in conversion processes on websites. However it can also be used to study web site usability and customer behaviour as well as handling customer service questions as the customer journey with all interactions can be replayed. Some organizations even use this capability to analyse fraudulent behaviour on websites.

### Browser Fingerprinting – [passive]

Browser fingerprinting is a powerful method that websites use to collect information about your browser type and version, as well as your operating system, active plugins, timezone, language, screen resolution and various other active settings.

These data points might seem generic at first and don't necessarily look tailored to identify one specific person. However, there's a significantly small chance for another user to have 100% matching browser information. Panopticlick found that only 1 in 286,777 other browsers will share the same fingerprint as another user.



Websites use the information provided by browsers to identify unique users and track their online behaviour. This process is therefore called “browser fingerprinting.”

### Cookies (HTTP cookie) – [passive]

Cookies were designed to be a reliable mechanism for websites to keep information in the user's browser and retrieve them later as the user navigates through the website, i.e. allowing the website to remember if a user has visited before, or keeping the their items in the shopping cart of an online store.

Since their original inception usage of cookies has changed and they now mainly hold information for the server-side part of the website (backend) to identify the user. The first time the user opens a page on a website, the backend will create a new random identifier (typically a string of random letters and numbers) for the user and store it on their browser in the form of a cookie. On every subsequent page the user opens on the same website (domain) the browser will include that cookie, with the user's identifier, allowing the backend to keep track of the user's actions in that website.

A special type of such cookies, namely tracking cookies are used to track users' web browsing habits across multiple websites (domains). These cookies usually contain more information about the user than just an identifier and will be sent not only to the website that originally created them but to their affiliate websites as well. They are commonly used for legitimate marketing and advertising purposes, but because they contain a history of the user's actions on multiple sites, they may be exploited or misused to track the user's behaviour.

Cookies can either have an expiration time, can be manually cleared by the user, or can even be manipulated or cleared by plugins and third party tools designed to keep the user safer from websites trying to track them.

## Deep Packet Inspection (DPI) – [passive]

Deep packet inspection (DPI) is a type of data processing that inspects in detail the data being

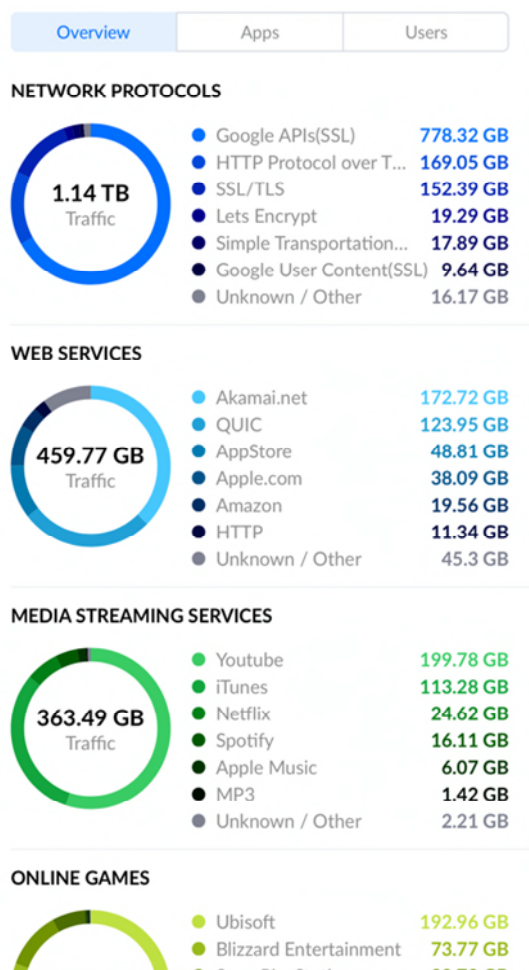


Figure 5

sent over a computer network, and usually takes action by blocking, re-routing, or logging it accordingly. DPI is often used to ensure that data is in the correct format, to check for malicious code, eavesdropping and internet censorship among other purposes. DPI keeps track of the types of data that it goes into the network stream and the quantity of it, and most User Interfaces designed for end-user usage can categorise data into groups of services that the user can see and interact with.

There are multiple ways to acquire packets for DPI. Using port mirroring which is a method of monitoring network traffic (the switch sends a copy of all network packets seen on one port (or an entire VLAN) to another port, where the packet can be analysed) is a very common way, as well as an optical splitter.

DPI is used in a wide range of applications, at the so-called "enterprise" level, in telecommunications service providers, and in governments, but it can also be used in smaller self-installed home networks to provide information about the activity that is being undertaken within the network at hand, providing information for all the devices connected and their whereabouts.

Most standalone DPI products cannot look into encrypted connections so the insights they can provide about the users' patterns on web-sites and applications that are only provided over SSL/TLS will be limited to access times and total size of traffic exchanged.

The downsides of DPIs is its inability to track how much time the user spent on an application, and the fact that it cannot track offline application usage as it relies on network traffic.

Figure 3 is taken from a home network installed with a Ubiquity router, and it provides a detailed overview of the services and applications that are used by the owners of the network, as well as the traffic for each one. A variety of settings can provide different measurements and categorisations.



## Most Viable Solution: Deep Packet Inspection (DPI) over Virtual Private Network (VPN) – [passive]

Most DPI solutions will require the installation of a hardware device in the user's network, this might not always be optimal or even feasible. An alternative to this would be to have the user to connect through a virtual private network (VPN – a network that extends a private network across a public network, and enables users to send and receive data across shared or public networks as if their computing devices were directly connected to the private network.) that will route all the user's online activity through a server that can perform deep packet inspection, anonymise, and log all the user's incoming and outgoing packets for further analysis.

This approach has been previously used by a number of wellbeing and health applications on mobile devices that routed the device's traffic through their own VPN servers and provided them with statistics about their website and application usage. (see also: <https://mysenseapp.com/>)

As previously mentioned, DPIs cannot peer into encrypted connections by default, thus reducing the amount of information we can get out of their interactions. The only way around this is for our VPN server to act as a "man in the middle" (MITM) which allows it to decrypt the data being sent from the website the user is accessing, and re-encrypt it with its own key/certificate, that the user must have explicitly installed on their device. This enables the VPN server to fully access all data that comes through it.

Even though the data gathered from DPIs only provides raw information about the traffic captured (protocol, URL, time, and content) post-processing them can yield a trove of valuable information and aggregated data that can be used to better understand the user's patterns and habits. It would be possible to categorise the interactions into application or website type based on the request's URL alone, to use the size and type of the content to deduce the length of the interaction, or to actually perform natural language processing, metadata extraction on text content, or even visually process images and video to better understand the user.

## Conclusions

There is a variety of possible solutions for tracking digital behaviour that can provide a complete picture of a person's technology habits and behaviours, and new technological solutions and frameworks emerge every day for this use. From external solutions like smartphone application installations and active self-tracking, to extensive passive data extrapolation from APIs, cookies, and home networks.

Both active and passive ways of tracking users' digital behaviours have several pros and cons in terms of ease of use, installation and activation issues, as well as user trust and consent complications. Although active tracking technologies offer more freedom to the users and keep them in control of their data, they can sometimes be difficult to use, time-consuming, and could potentially lead to drop out rates. On the other hand, passive tracking technologies, can generally be installed just once and then be used continuously for as long as the study requires, with no external or human factors interfering with the amount and accuracy of data.

That being said, there are a few things that need to be considered before the implementation of any of these technologies for the purposes of tracking human digital behaviour. First and foremost, the ethics of digital surveillance are highly complicated and regulated, especially with the introduction of the new EU Data Protection Regulation (GDPR) as of 2018. Aspects of data privacy and active consent to provide sensitive data (like browser history contents or similar data) should be highly considered. Second, although many of the proposed solutions include open source software and native applications that bear no additional costs from the devices themselves, the cost of hardware and software that could potentially be needed (sophisticated home networks, screen tracking software etc) should be taken into account. Finally, importance should be given to participants' levels of digital literacy, as some of the proposed technologies can be complicated to install or activate. It is also possible that in the near future, native applications like iOS' screen time will be replicated and become accessible for a wide variety of technologies including wearables, desktops, and smart TVs which will make the process of collecting such data easier for researchers.

# References and Appendix

## References (uncategorised)

Title	Author	Year Published	Source
<b>Adults' Media Use and Attitudes Report 2018 (Ofcom)</b>	Saville Rossiter-Base	2018	Ofcom
<b>From Databases to Dataspaces: A New Abstraction for Information Management</b>	Michael Franklin, Alon Halevy, David Maier	2005	ACM SIGMOD
<b>How much time do people spend online each day? (UKOM Insights)</b>	UKOM	2018	UKOM
<b>How much time do you spend on social media? Research says 142 minutes per day</b>	Saima Salim	2019	Digital Information World
<b>Lively Data, Social Fitness, and Biovalue: the Intersections of Health Self-Tracking and Social Media</b>	Deborah Lupton	2015	University of Canberra
<b>Ofcom's Communications Market Report</b>	Ofcom	2018	Ofcom
<b>People spend most of their waking hours staring at screens</b>	Quentin Fottrell	2018	MarketWatch
<b>Personal Tracking of Screen Time on Digital Devices</b>	Rooksby et. al	2016	University of Glasgow
<b>The Infinite Dial 2019</b>	Edison Research	2019	Edison Research

## Appendix No.1 : Related Academic Literature

Author(s)	Title	Journal	Year Published
<b>Amy Orben, Andrew K. Przybylski</b>	Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies	Psychological Science	2019
<b>Andrew K. Przybylski, Netta Weinstein</b>	A Large-Scale Test of the Goldilocks Hypothesis: Quantifying the Relations Between Digital-Screen Use and the Mental Well-Being of Adolescents	Psychological Science	2017
<b>C R Elder, C M Gullion, K L Funk, L L DeBar, N M Lindberg &amp; V J Stevens</b>	Impact of sleep, screen time, depression and stress on weight change in the intensive weight loss phase of the LIFE study	International Journal of Obesity	2012
<b>Cally A Davies, Corneel Vandelanotte, Mitch J Duncan et. Al</b>	Associations of physical activity and screen-time on health related quality of life in adults		

<b>Delfien Van Dyck</b> Email author, <b>Greet Cardon, Benedicte Deforche, Neville Owen, Katrien De Cocker, Katrien Wijndaele and Ilse De Bourdeaudhuij</b>	Socio-demographic, psychosocial and home-environmental attributes associated with adults' domestic screen time	BMC Public Health	2011
<b>Emmanuel Stamatakis, Mark Hamer and David W. Dunstan</b>	Screen-Based Entertainment Time, All-Cause Mortality, and Cardiovascular Events	Journal of the American College of Cardiology	2011
<b>Emmanuel Stamatakis, Kris Rogers, Ding Ding, David Berrigan, Josephine Chau, Mark Hamer and Adrian Bauman</b>	All-cause mortality effects of replacing sedentary time with physical activity and sleeping using an isotemporal substitution model: a prospective study of 201,129 mid-aged and older adults	International Journal of Behavioral Nutrition and Physical Activity	2015
<b>Ilaria Montagni, Elie Guichard, Claire Carpenet, Christophe Tzourio, Tobias Kurth</b>	Screen time exposure and reporting of headaches in young adults: A cross-sectional study	Cephalalgia, 36(11), 1020–1027.	2016
<b>Jennifer J. Otten Benjamin Littenberg Jean R. Harvey-Berino</b>	Relationship Between Self-report and an Objective Measure of Television-viewing Time in Adults	Obesity Journal	2010
<b>Katie E.GunnellabcdJennifer BrunetbefMathieuBélange rghi</b>	Out with the old, in with the new: Assessing change in screen time when measurement changes over time	Preventive Medicine Reports Volume 9, March 2018, Pages 37-41	2017
<b>Sarah M. Camhi,1 Molly E. Waring,2 Susan B. Sisson,3 Laura L. Hayman,4 and Aviva Must5</b>	Physical Activity and Screen Time in Metabolically Healthy Obese Phenotypes in Adolescents and Adults	Journal of Obesity	2013
<b>Shields, Margot; Tremblay, Mark S.</b>	Screen time among Canadian adults: A profile	Health Reports Vol. 19, Iss. 2,	2008
<b>Stephanie Schoeppe, Amanda L. Rebar, Camille E. Short, Stephanie Alley, Wendy</b>	How is adults' screen time behaviour influencing their views on screen time restrictions for children? A cross-sectional study	BMC Public Health	2015

**Van Lippevelde, Corneel  
Vandelanotte**

<b>Susan Christina Stephanie MistiLeyvaRD</b>	<b>B.Sisson, M.Shay, T.Broyles,</b>	Television-Viewing Time and Dietary Quality Among U.S. Children and Adults	American Journal of Preventive Medicine	2012
---	---	--	--	------