SITRA STUDIES 169

AUGUST 2020

ON THE TRAIL OF PERSONAL DATA

The flow and use of data collected from individuals using digital services

Riitta Vänskä and Tiina Härkönen

Sitra studies 169

On the trail of personal data – the flow and use of data collected from individuals using digital services

Authors: Riitta Vänskä, Specialist and Tiina Härkönen, Leading Specialist from Sitra's IHAN – Human-driven data economy project.

Working group: Tiina Härkönen, Maria Jalavisto, Heli Parikka, Jaana Sinipuro and Riitta Vänskä
The experts consulted for the report were Futurice, a Finnish software company, and mathematician Paul-Olivier Dehaye.

Subeditor: Kirsi Suomalainen, Sitra
Layout: PunaMusta

ISBN 978-952-347-179-5 (PDF) www.sitra.fi
ISSN 1796-7112 (PDF) www.sitra.fi

SITRA STUDIES is a publication series which focuses on the conclusions and outcomes of Sitra's future-oriented work.

Sitra studies 169
On the trail of personal data – the flow and use of data collected from individuals using digital services
August 2020

Sisällys

Foreword

Sitra's fair data economy project IHAN was launched in the spring of 2018. At about the same time, all EU countries began enforcing the General Data Protection Regulation, which regulates the processing of personal data and is the most rigorous data protection regulation in the world.

Data refers to a trace generated by digital activities and does not have any value in itself. Data acquires value when it is enriched into information and further into innovations, or when it is used to develop operations. Data is a central concept in the data economy, which refers to an area of the economy whose business model is based on the diverse use of data.

For many of us, the data economy means good-quality services that we enjoy free of charge, and we feel grateful for the time we save due to targeted advertisements. However, the digital advertising built round the giants of the platform economy is one of the first areas of the data economy whose side effects and background business models have begun to attract increasing attention. The industry itself has become concerned about the pressure caused by prevailing general sentiments as well as by legislators. The industry is undergoing a change and a reform of its operating models.

At Sitra, we see the data economy not only as part of the economy but also as a phenomenon that is only beginning to take shape. This phenomenon is a perceivable, interesting and recurring societal event. To tackle it we need to describe it. Our digital everyday lives leave many types of traces that we have tried to examine by looking at digital services that are familiar to us all. 'Fair data economy' is a concept created by Sitra, which aims at describing the desired state of the phenomenon. It is a sector of the economy focused on creating data-driven products and services ethically. It creates value for all: people, companies and society.

The aim of this report is to analyse the phenomenon as it appears in our everyday lives through various applications. There are things behind the easy-to-use services which we do not normally see and which we are not even interested in. We want to shed light on the activities of the companies whose services individuals use, as well as on the underlying technical and financial factors at play.

The inspiration for this report came from the Finnish Office of the Data Protection Ombudsman. Through the MyData community, we also found a specialist without whom the analyses of this report would have remained superficial at best.

We believe that by putting difficult-to-understand matters into words and by utilising the means given to us by the General Data Protection Regulation, we are able to shape our future and move in the direction we wish. By further developing the GDPR, it will be possible to enable new business models that are more sustainable than the current ones.

Building the fair data economy is in the early stages of its development and can only succeed through multi-voiced social discussion.

JAANA SINIPURO
Project Director
Sitra

Summary

According to a survey related to Sitra's IHAN project on the fair data economy, people would like more transparency over the use of the data collected from individuals and the ways to identify companies that use data in an ethical manner from those that do not. A separate survey of companies revealed the business perspectives on the data economy and highlighted serious concerns among European companies concerning their competitive positions in the data economy in relation to major American and Chinese corporations.

Following on from these surveys, Sitra worked with six test people in Finland to investigate the flow of an individual's data online at the end of the year 2019. The Digitrail survey studied where people's data travels when they visit websites or log on to digital services.

The results of the Digitrail survey show that it is impossible for people to know what data has been collected about them and who holds it. The individual data arising from online behaviour is refined at various stages in the data flow to create a profile of the individual. Profiles are generated by companies working in and around digital advertising unbeknown to consumers, and despite the large amount of data collected, profiles do not provide a true picture of the individual, although they influence the information offered to the individual. The General Data Protection Regulation only permits a person to gain limited access to their data. Free services are considered adequate recompense for handing over data. However, the true price of these services cannot be judged, because it is impossible to find information on the spread and exploitation of the data. As such, this online transaction cannot be considered fair.

When it comes to privacy, the digital advertising business models that have emerged around the giants in the platform economy have been built in a fundamentally problematic way. Users have a limited opportunity to evaluate the impacts of the consent they give for the use of data when they are asked for it.

The most popular platform services have been built up over the course of many years, providing significant benefits as well as drawbacks. The field of platform companies, digital advertising and data analytics is under transformation due to pressure from consumers and legislation, as well as from trends within the field. So far, individuals have been responsible for guarding their privacy, while privacy has become the most complex issue for the platform companies and digital advertising market. It is essential for consumers to form an understanding of the ground rules in the market. In terms of services used by children and young people, this presents a particular set of challenges.

The respect for privacy should extend to the customer experience and corporate responsibility, as these aspects provide European companies with an opportunity to establish themselves as fair data economy operators, thereby gaining a competitive advantage. Instead of replicating the old ground rules for the platform economy and supporting the existing digital advertising machinery, it is important to seek new business models. The data economy offers enormous potential, and European companies have the opportunity to succeed with innovations enabled by new operating models. This could be realised by means such as sharing data between companies in data partnerships or data networks with ethically sustainable methods and with the individual's consent. The new services created in this way will be part of the fair data economy, which will create well-being for all involved.

Tiivistelmä

Sitran reilun datatalouden IHAN-hankkeeseen liittyvän kyselytutkimuksen mukaan ihmiset toivovat läpinäkyvyyttä yksilöistä kerättävän datan käytölle ja keinoja tunnistaa eettisesti reilulla tavalla dataa käyttävät yritykset muista. Yrityksille suunnattu kysely puolestaan valotti liiketoiminnan näkemyksiä datataloudesta ja toi esiin eurooppalaisten yritysten vakavan huolen datatalouden kilpailuasetelmasta suhteessa amerikkalaisiin ja kiinalaisiin suuryrityksiin.

Näiden kyselyiden jatkoksi Sitra selvitti yksilöstä kerätyn datan kulkua verkossa kuuden suomalaisen testihenkilön avulla loppuvuodesta 2019. Digijälkiselvityksessä tutkittiin, minne henkilöiden dataa kulkee, kun he vierailevat verkkosivuilla tai käyttävät kirjautuneina digitaalisia palveluja.

Digijälkiselvityksen tulosten perusteella ihmisten on mahdotonta tietää, mitä dataa heistä on kertynyt ja kenellä dataa on. Verkkokäyttäytymisestä syntyvää yksilödataa rikastetaan datankulun eri vaiheissa yksilöstä muodostettavaa profiilia varten. Digitaalisen mainonnan ympärillä toimivien yritysten synnyttämät profiilit muodostetaan kuluttajien tietämättä, eivätkä ne laajamittaisesta datan keräämisestä huolimatta vastaa todellisuutta, vaikka niillä on vaikutusta yksilöille tarjoiltuun tietoon. Yleinen tietosuoja-asetus mahdollistaa vain rajoitetun näkyvyyden omaan dataan. Datan luovuttamisen vastineena pidetään ilmaisia palveluja. Palvelujen todellista hintaa ei kuitenkaan voi hahmottaa, koska tietoa datan leviämisestä ja käytöstä on mahdotonta selvittää. Siten vaihtokauppaa ei voi pitää reiluna.

Alustatalouden jättiläisten ympärille muodostuneet digitaalisen mainonnan liiketoimintamallit on lähtökohtaisesti rakennettu yksityisyyden kannalta ongelmallisiksi. Myös palvelujen käyttäjien mahdollisuudet arvioida datan käyttöön liittyvän suostumuksen vaikutuksia ovat rajalliset suostumuksen hetkellä.

Suosituimmat alustapalvelut ovat rakentuneet vuosien aikana ja tuoneet vanavedessään suurten hyötyjen lisäksi haittoja. Alustayhtiöiden, digitaalisen mainonnan ja data-analytiikan kenttä on murroksessa sekä kuluttajilta ja lainsäädännöstä tulevan paineen että kentän sisäisen liikehdinnän takia. Vastuu yksityisyyden säilyttämisestä on tähän asti sälytetty yksilölle, ja yksityisyydestä onkin muodostunut alustayhtiöiden ja digitaalisen mainonnan markkinan hankalin kysymys. Markkinan pelisääntöjen ymmärtäminen olisi kuluttajille ensiarvoisen tärkeää. Lasten ja nuorten käyttämien palvelujen osalta se muodostaa aivan erityiset haasteensa.

Yksityisyyden kunnioittaminen tulisi ulottaa asiakaskokemukseen ja yritysvastuuseen, sillä niiden kautta eurooppalaisilla yrityksillä olisi mahdollisuus profiloitua reilun datatalouden tekijöiksi ja saada siitä kilpailuetua. Alustatalouden vanhojen pelisääntöjen kopioimisen ja digimainonnan nykykoneiston tukemisen sijaan tulisi hakea uudenlaisia liiketoimintamalleja. Datatalouden potentiaali on valtava ja eurooppalaisilla yrityksillä olisi mahdollisuus menestyä uusien toimintamallien mahdollistamilla innovaatioilla. Niihin voitaisiin päästä esimerkiksi jakamalla dataa yritysten kesken datakumppanuuksissa tai dataverkostoissa, eettisesti kestävin keinoin, yksilön luvalla. Näin syntyvät uudet palvelut ovat reilua datataloutta, joka synnyttää hyvinvointia kaikille osapuolille.

Sammanfattning

Enligt en enkät i Sitras projekt IHAN om rättvis dataekonomi önskar människor transparens i hur data som insamlas om individer används och färdigheter att skilja på företag som använder data på ett hållbart sätt och andra företag. En enkät riktad mot företag belyste å sin sida affärsverksamhetens synpunkter på dataekonomi och lyfte fram europeiska företags allvarliga oro om konkurrenskonstellationen inom dataekonomi i förhållande till amerikanska och kinesiska storföretag.

Som en fortsättning på dessa enkäter utredde Sitra i slutet av 2019 hur data som insamlas om en individ rör sig på nätet med hjälp av en finländsk testperson. I utredningen om digitala spår studerade man vart data om personer sänds, när personerna besöker webbplatser eller använder digitala tjänster som inloggade användare.

Enligt resultaten i utredningen om digitala spår är det omöjligt för människor att veta vilka data som insamlats om dem och vem som innehar dessa data. Individuella data som uppstår genom nätbeteende berikas i olika skeden av färden för en profil som skapas om individen. Profiler som företag verksamma inom digital annonsering genererat skapas utan konsumenternas vetskap, och trots omfattande datainsamling motsvarar de inte verkligheten, även om de inverkar på information som erbjuds individerna. Den allmänna dataskyddsförordningen möjliggör endast begränsad insyn i egna data. Gratis tjänster anses vara utbytet för överlämning av data. Det går emellertid inte att skapa sig en uppfattning om tjänsternas verkliga pris, eftersom det är omöjligt att utreda hur data sprids och används. Således kan byteshandeln inte anses vara rättvis.

Affärsmodeller för digital annonsering, som bildats omkring plattformsekonomins jättar, har i regel byggts upp så att de är problematiska med tanke på integriteten. Även möjligheterna att bedöma påverkan av samtycke till användning av tjänsten är begränsade för dem som använder tjänsterna då de ger sitt samtycke.

De populäraste plattformstjänsterna har byggts under flera år och förutom stora fördelar även fört med sig nackdelar i kölvattnet. Fältet för plattformsföretag, digital annonsering och dataanalys befinner sig i omvälvning på grund av det tryck som kommer från konsumenter, lagstiftning samt interna rörelser på fältet. Ansvaret för att bevara integriteten har hittills legat hos individen, och integriteten har blivit den svåraste frågan för plattformsbolagen och den digitala annonsmarknaden. Det vore av största vikt för konsumenter att förstå marknadens spelregler. När det gäller tjänster som används av barn och ungdomar skapar detta särskilda utmaningar.

Respekten för integritet borde sträckas till kundupplevelsen och företagsansvaret, eftersom europeiska företag via dessa skulle ha möjlighet att profilera sig som skapare av en rättvis dataekonomi och få konkurrensfördelar. I stället för att kopiera gamla spelregler för plattformsekonomi och stötta det nuvarande maskineriet inom digital annonsering borde man söka nya slags affärsmodeller. Dataekonomins potential är enorm och europeiska företag har en möjlighet till framgång med hjälp av innovationer som möjliggörs av nya operativa modeller. Dessa kan uppnås till exempel genom att dela data mellan företag inom datapartnerskap eller datanätverk, med etiskt hållbara metoder, med individens tillstånd. Tjänster som uppkommer på detta sätt utgör en rättvis dataekonomi som skapar välfärd för alla parter.

1 Introduction

Interest in the movement of data collected on individuals through the services and digital advertising networks of platform giants originated from a series of internationally infamous scandals, in which data collected from individuals had either been deliberately misused, had leaked out without people's knowledge, or was lost as a result of inadequate security measures. The most prominent of these was the 2018 Cambridge Analytica scandal, in which the Facebook data of millions of people was allegedly used in an election campaign.

Sitra's IHAN project examined the impact of these events on people's attitudes to data economy operators and services and charted their understanding and knowledge of data-based services. The study was carried out as a survey in four European countries in the autumn 2018. The survey clearly revealed people's concerns about their privacy and their lack of trust in digital service providers.

The data economy affects society as a whole and involves individuals and businesses as well as other organisations. In spring 2019, Sitra examined large, small and medium-sized European companies' awareness of and attitude and commitment to business opportunities offered by the fair data economy. The survey showed that nearly a third of the companies felt that the fact that American and Chinese operators played by their own rules constituted a challenge for European companies, and more than a fifth considered the requirements of the General Data Protection Regulation (GDPR) and other similar regulations to be a major challenge.

According to public and business surveys, the interests of those businesses that collect the most data are partly contradictory to those of consumers and European companies. A significant number of the European companies that responded to the survey found that legislation that protects individuals is problematic and that the playing field is unfair.

But who are the companies that collect individual data in large quantities, how do they operate, and what do they do with the data? We aim to answer these questions in the digitrail survey to provide a glimpse of the data flows that normally remain unseen by people. Technical solutions, courageous citizens and the will to explore the complex and opaque market for personal data were needed to carry out the survey.

The survey, conducted with the help of six test subjects, illustrates the movements of data collected from individuals in huge advertising ecosystems, introduces the different players in the market and examines how well the information given to consumers about the use of their data accords with legislation and reality. The experts consulted for the report were Futurice, a Finnish software company, and mathematician **Paul-Olivier Dehaye**, one of the people who exposed the Cambridge Analytica scandal.

2 People want fair use of personal data and companies want fair competition

People want greater transparency on how the data collected from individuals is used, and to be able to distinguish companies that use data sustainably from those who do not. A survey for companies revealed the business perspectives on the data economy and highlighted serious concerns among European companies about their competitive positions within it.

A survey of citizens on the data economy revealed a lack of trust

The use of digital services survey (Finland, France, Germany and the Netherlands) measured people's grasp of the data economy, their attitudes towards the digital service providers and their activities to protect their personal data. Two thousand people from each country responded to the survey.

The most important findings of the survey can be summarised as follows:

- The application of the rights granted by the General Data Protection Regulation (GDPR) is only at an early stage.
- Lack of trust in the service providers is a bottleneck.
- Data breaches have had an impact on people's behaviour.
- People would like more transparency in the flow and use of data.
- Fair data services should be easy to recognise.

The survey showed that only a small number of people protect their data or exercise the rights granted by the GDPR. Nine per cent of respondents had requested access to the data collected about them by a service provider. Fifteen per cent of respondents had stopped using some services due to news reports about data leakage, while 40 per cent said a lack of trust had prevented them from using digital services. According to the survey, the most important trust-building factor in digital services is the transparency of data usage. While trust is falling, only 14 per cent of the respondents said that they read the terms of use for services and applications carefully. When asked about future opportunities, 66 per cent of respondents felt that the "fair data label" of digital services was very important or important (71 per cent of the Finnish respondents).

Because people's lack of trust concerned companies operating in the data economy and the digital services they provide, it was necessary to explore the business field too. The aim was to gain an insight into the views of businesses and to detect the factors that link or separate the needs of citizens and businesses.

The data economy survey for companies raised concerns about the competitive setup

IN A FAIR DATA ECONOMY MODEL

- companies have access to personal data with people's consent, and are allowed to share it with each other on the basis of common agreements
- people receive services specifically designed for them in exchange for their data
- data collected from individuals represents only a small part of the data economy, as companies buy fair data products and share much more data cooperatively.

With The future of European companies in the data economy survey, Sitra studied the awareness, attitude, and commitment of the companies to the business opportunities offered by the "fair data economy" model in four EU countries. A fair data economy was defined as an economy where different market actors operate in a common environment to ensure data collection and usability. Together they make good use of the data and develop new applications and services based on them. A fair data economy requires transparent data sharing between the actors based on common rules and, in the case of personal data, people's consent to collect and use the data.

The survey was conducted in spring 2019, and the findings are based on 1,667 responses. The target group consisted of large and small and medium-sized enterprises in Finland, France, Germany and the Netherlands. Companies employing fewer than 10 people were not included.

Key findings of the companies' survey:

- Generally, companies took a fairly positive approach to the principles of the fair data economy, but the commitment to respecting individual privacy, even at the expense of the customer experience, was seen as challenging.
- The most obvious strategic challenge was that only 15 per cent of respondents considered data sharing with others to be a positive thing.
- The opportunities offered by the data economy were already well understood (a third of respondents said that they generated a competitive advantage from it), but the understanding of digital business models was still incipient and unorganised.
- Companies considered the legislation required by the fair data economy (GDPR, etc.) as a partial obstacle to service creation.
- On the other hand, the companies that invested significantly in the practical implementation of the GDPR also benefited from it, as it helped them understand their own data resources.

The marketplace of the data economy was not considered to be a level playing field. Thirty-one per cent of respondents felt that technology giants play by their own rules. Competition with American and Chinese companies was seen as either the biggest (France, Germany and the Netherlands) or the second biggest (Finland) challenge.

3 The Digitrail survey – the flow of individual data

Sitra studied the online flow of individual data collected from six Finnish test subjects at the end of 2019. The aim was to find out where the data of the test subjects flows when they visit websites or are logged on to digital services.

Data collectors operate in different roles

Data can be used to optimise the functions and digital services of websites to provide an agreeable user experience. For example, collecting data enables the service to remember the visitor so that the same questions are repeated, and the user does not need to log in to the service each time.

In addition to targeted advertising, there are a number of other important reasons for collecting and using user data. Overall, the data is used in the development of websites and services. What works and what does not can be discerned when monitoring people's behaviour when using online services. Data also helps in detecting and solving problems in the services.

Website and application development companies enable data collection for their corporate customers, who can analyse the data and obtain information about the needs of their current and future customers. Companies use versatile data analytics tools to find new ways to serve their customers or create new products. They also use the services of companies specialising in data analysis, if their own resources are insufficient or the required skills are not found in their own organisation.

Not all data collection is problematic. Data collection is essential because without it we would not have the highly developed online services we have now. The Digitrail survey examines some of the problem areas of data collection and looks in more detail at the largest companies in the platform economy and in digital advertising.

The monitoring begins: Six test subjects and six mobile phones

In the survey, the data flow passing through the test subjects' mobiles was monitored over a two week period. A number of online service users of different ages and with different life situations were selected as test subjects. The services they used were from various countries. The companies providing the services were also asked to answer the test subjects' written questions in accordance with the GDPR.

The test subjects are treated anonymously in this report. They are described as follows:

1. Upper secondary school teenager
2. Young university student
3. Middle-aged journalist
4. Middle-aged politician
5. Middle-aged person in a managerial position
6. Retiree

The study sought answers to the following questions:

- What kind of data and how much of it is accumulated by different service providers?
- Who the data is collected for and who benefits from it?
- What is our personal data used for?
- What data about my contacts is collected via me?
- How are we profiled and what the profiles are used for?
- What do we get in return for data collection?
- How is our data traded?
- Is data collected and used in line with the GDPR?

The following methods were used to find answers to the questions posed in the study:

- A data flow monitoring application in the test subjects' Android test mobile phones.
- Reading through the data protection documentation of 14 selected companies.
- Analysing companies' responses to the queries made in accordance with the GDPR (8 pcs).
- Test subjects' thoughts on their results.

A more detailed description of the scope of the study and of the data flow monitoring method is provided in Annex 1.

In the study, the test subjects' network traffic data was collected over a period of two weeks by using http request/response packets. Both mobile applications and web pages use a file transfer method called Hypertext Transfer Protocol (HTTP) to transfer data as packets to different servers on the internet. The packets contain actual data and several types of metadata. The data can be in very different formats, such as text, image, video, and javascript programs.

The study focused on monitoring the data flow for advertising and profiling companies. In reality, it is likely that there were more than the detected websites using digital advertising technology, as only some of the data sent to the advertising servers contained information about which site the request was related to. Data can also be inaccurate because some advertising-related companies provide other services as well, such as website use analysis. Such uses could not be distinguished in this study.

What are third-party actors?

In addition to the actual service provider (first party), data may be used by a large number of other actors, i.e. the so-called third parties. The study aimed to provide a broader picture of where data flows from the services.

In the simplest case, the application or web page communicated only with the service provider's own server, i.e. the so-called first party, and data was not diverted elsewhere.

Third parties may hand over the data to other actors, such as data marketplaces. These actors, who are still further away from the service, are also often referred to as third parties. The data flowing to these parties could not be traced.

When creating websites and services, a ready-made code is often used to implement different functionalities or to integrate the service into a larger network. These code snippets store data on the third-party servers. It does not automatically mean that the data is being sold or used other than by the original service, but traffic data shows communications to servers other than those of the first party.

Some of these third-party services combine user data from different websites. If a user visits multiple sites where the same third-party service is located, all the user's data can be merged. Google, Facebook, and many advertising-related services work this way, for instance.

UPPER SECONDARY SCHOOL STUDENT SURPRISED BY THE NUMBER OF THIRD PARTIES

The student concerned believes that people should take more of an interest in the flow of their data. He was surprised by the number of third parties involved, although aware that the data was being sent out into the world via various services. The test subject did not consider the current data economy system harmful, except in a situation where data is misused. In his opinion, the imaginary ”fair data symbol” would not inspire confidence, because you would not know the source of the symbol.

THE UPPER SECONDARY SCHOOL STUDENT'S DATA WAS TRANSMITTED TO 114 ACTORS

Data was delivered to a total of 114 companies. For instance, the Subway Surfers game delivered data to 12 third parties

Data was delivered via 20 websites or applications to 44 advertising technology-related services located in

- United States (33)
- Europe (8)
- Norway (1)
- Russia (1)
- Canada (1)

The greatest amount of data (1.5 MB) was sent to AppNexus. This corresponds to about 750 A4 pages of text.

TEENAGER UPPER SECONDARY SCHOOL STUDENT

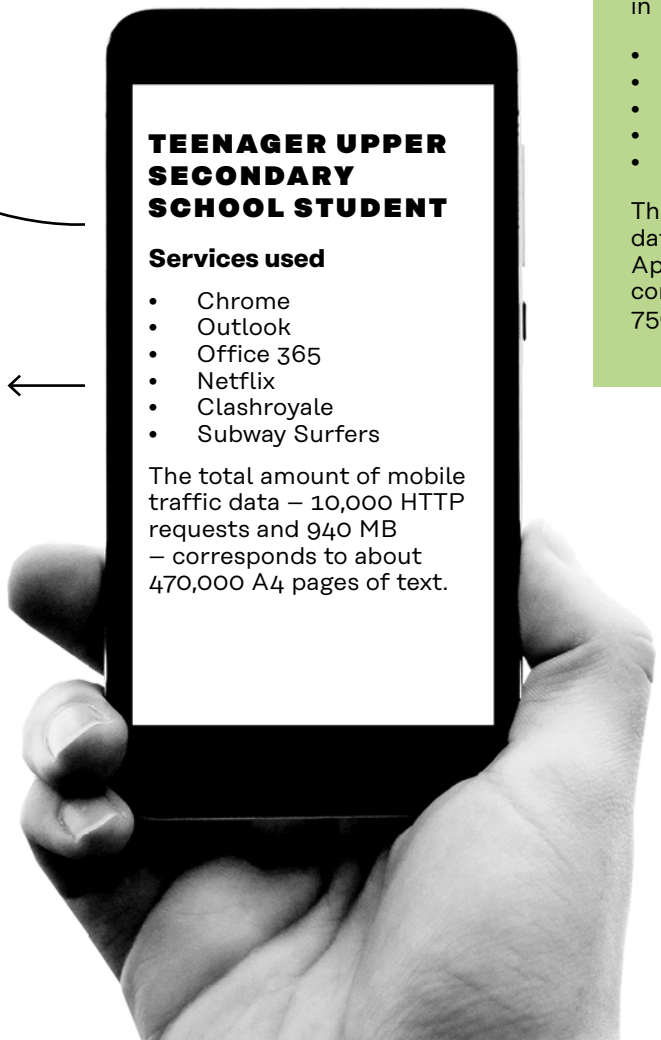
Services used

- Chrome
- Outlook
- Office 365
- Netflix
- Clashroyale
- Subway Surfers

The total amount of mobile traffic data – 10,000 HTTP requests and 940 MB – corresponds to about 470,000 A4 pages of text.

GDPR QUERIES SENT TO

- Google
- Instagram
- Netflix



Used services:

20

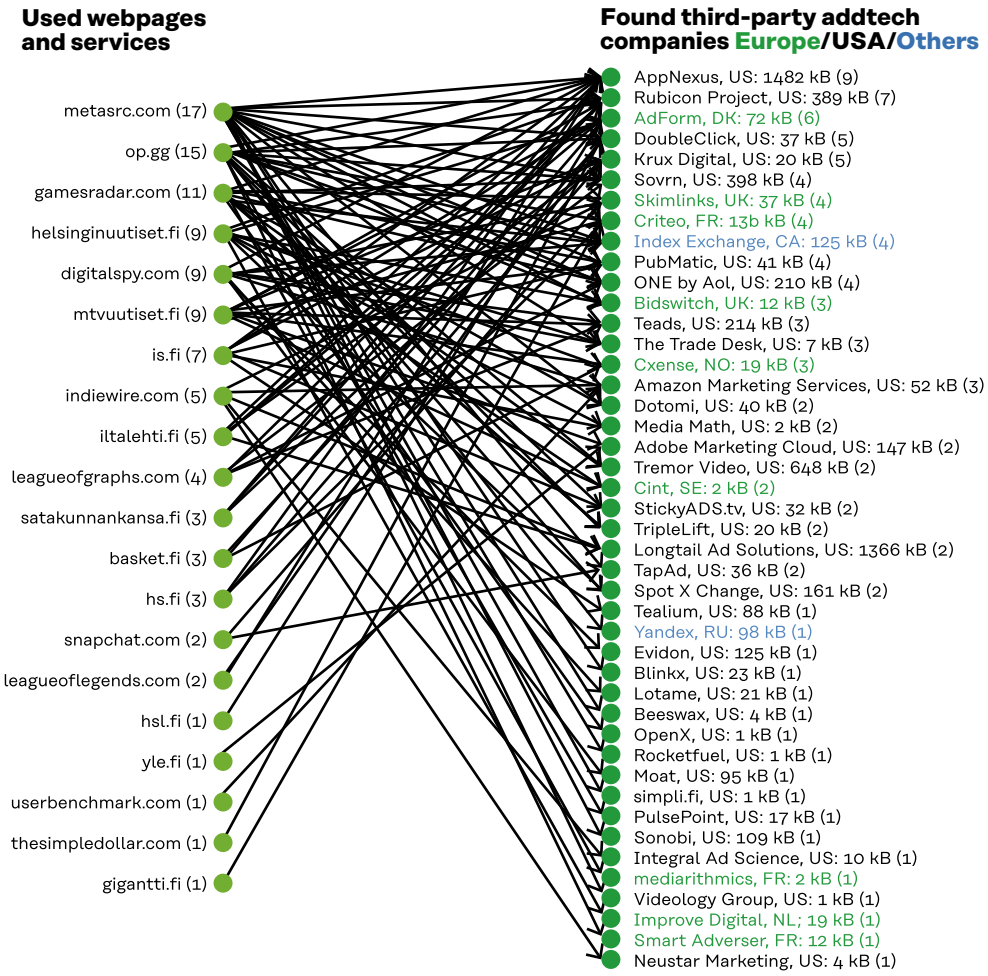
Total number of AdTech / marketing companies identified:

44

EU: 8
USA: 33
Norway: 1
Russia: 1
Canada: 1

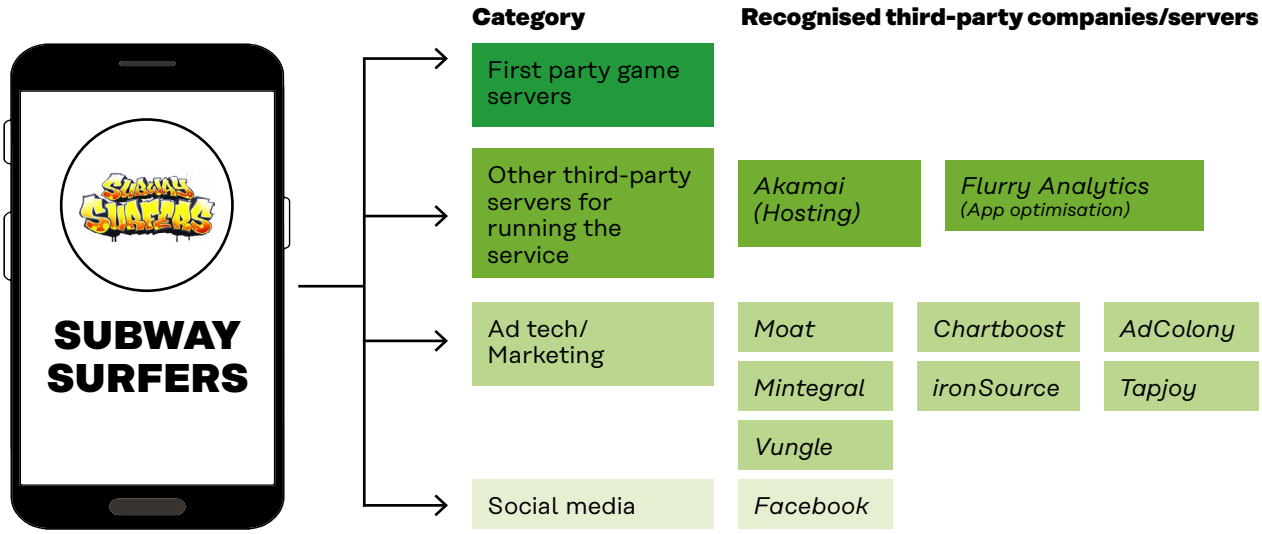
Total data to these companies:
1800 requests,
6.2 MB ~ 3100 pages

DATA WAS SENT THROUGH 20 SERVICES TO 44 ADVERTISING AND MARKETING COMPANIES



10 THIRD-PARTY ACTORS WERE FOUND TO BE LINKED TO SUBWAY SURFERS. 7 OF THEM WERE ADVERTISING COMPANIES

Annex 2 gives a more detailed description of the different actors involved.



UNIVERSITY STUDENT WARY OF THE POSSIBILITY OF POLITICAL MANIPULATION

The university student wanted to participate in the study because she was interested in the flow and use of her data. The student was interested in the issue because of the debate around it in the media but felt she did not know enough about it.

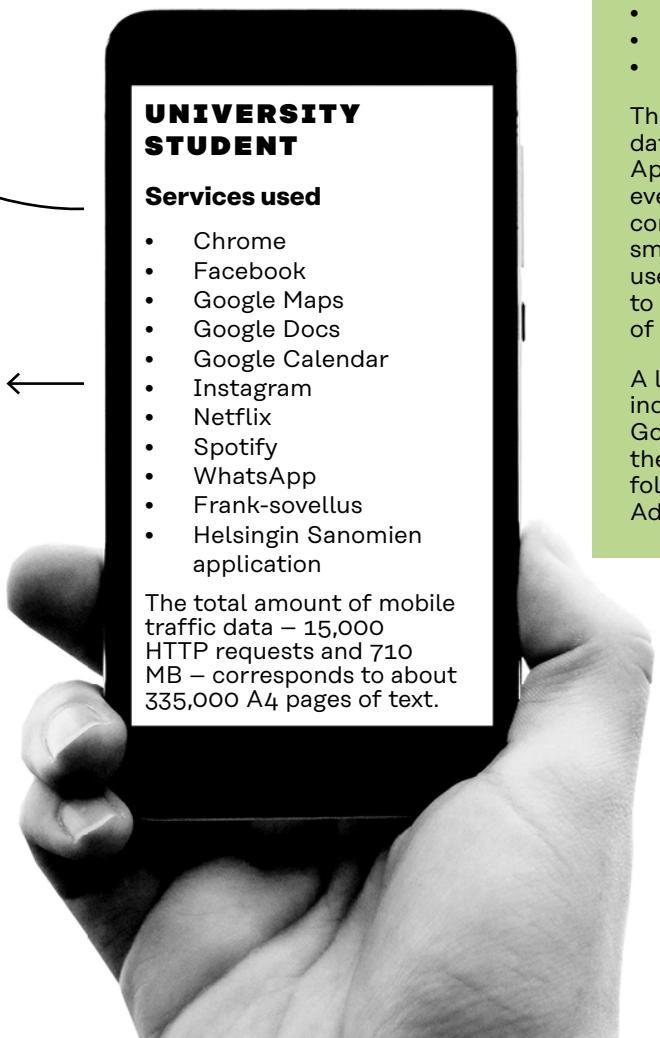
Prior to the study, the subject had never read cookie policies and found ads based on their data useful. In her experience, transparency about the terms of use would increase users' confidence in the service. If service providers were more transparent, she would not feel that the sharing of her information would be harmful. According to the student, the worst-case scenario would be the use of her data for political manipulation.

THE UNIVERSITY STUDENT'S DATA WAS TRANSMITTED TO 99 COMPANIES

Data was delivered to a total of 99 companies. For instance, the K-ruoka.fi website delivered data to four third parties.

GDPR QUERIES SENT TO

- Facebook
- Google
- Helsingin sanomat/ Sanoma
- Frank



A total of 18 websites or applications delivered data to 27 advertising technology-related services located in

- United States (20)
- EU (5)
- Norway (1)
- Canada (1)

The greatest amount of data was delivered to AppNexus (220 KB), even though this company had the smallest number of users. This corresponds to about 100 A4 pages of text.

A large number of individual pages used Google's DoubleClick in their advertising, followed by AppNexus, AdForm and KruxDrigital

Used services:

18

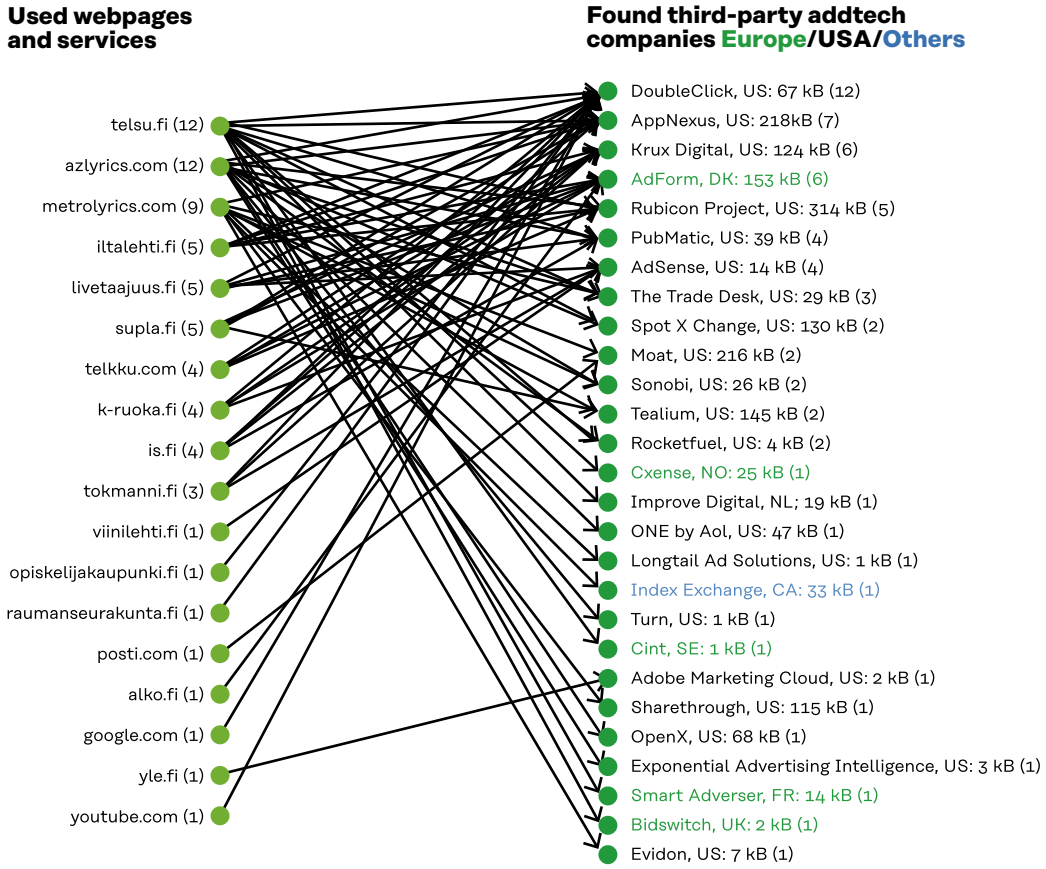
Total number of AdTech / marketing companies identified:

27

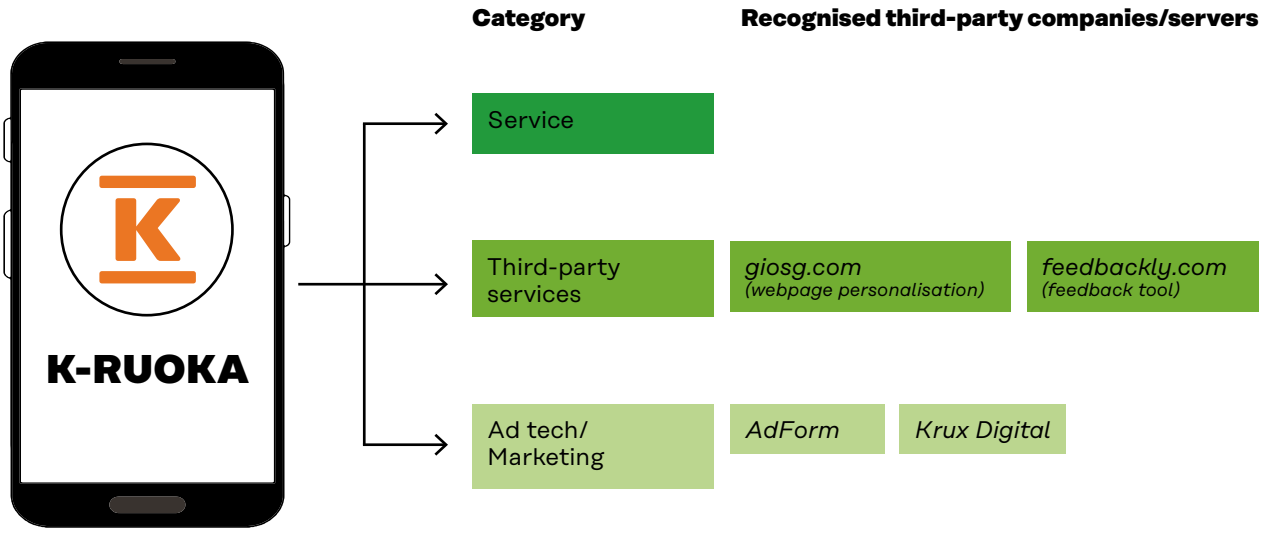
EU: 5
USA: 20
Norway: 1
Canada: 1

Total data to these companies:
560 requests,
1.8 MB ~ 900 pages

DATA WAS PASSED THROUGH 18 SERVICES TO 27 ADVERTISING AND MARKETING COMPANIES



4 THIRD-PARTY ACTORS WERE FOUND TO BE LINKED TO THE K-RUOKA.FI SERVICE. 2 OF THESE WERE MARKETING COMPANIES



JOURNALIST QUESTIONS THE MEDIA'S VALUE PROPOSITION ON RELIABILITY

The subject, a journalist, was familiar with the topic because she had written about it. However, by participating in the study, she was able to get specific information about the flow of her data and how it was shared. She assumed that the data would remain with the primary service provider. The results of the study changed her attitude towards data collection companies, leading her to question the media's value proposition on reliability.

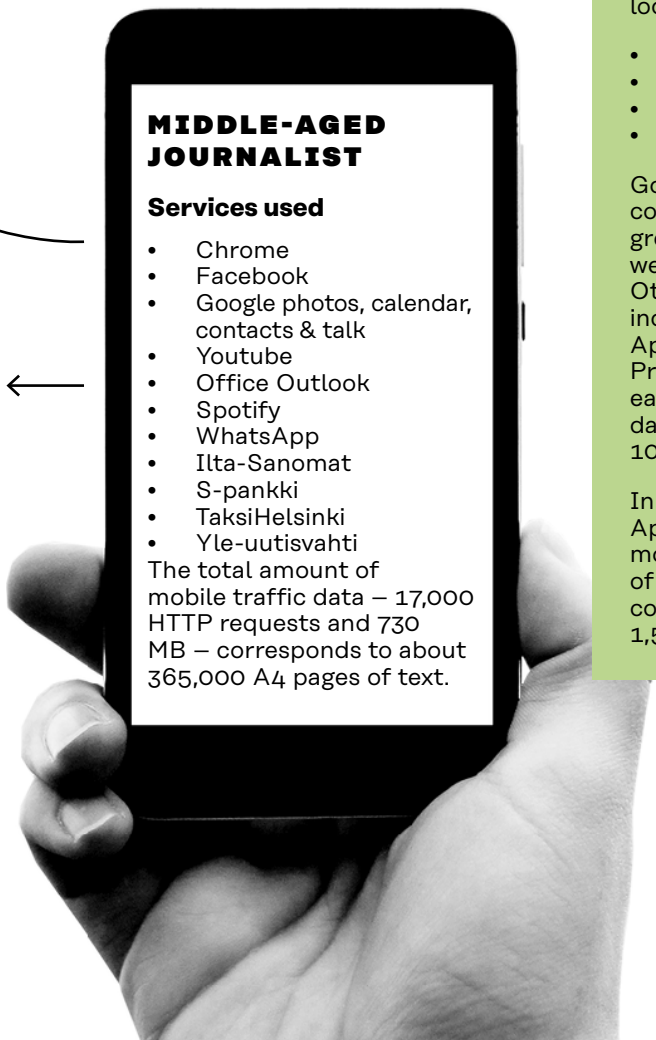
According to the subject, it is impossible for a layman to understand data flow in a complex network. An expert is needed to explain the subject in plain language. The study confirmed the subject's assumption that data is not in the hands of the individual and that companies did not answer the queries in accordance with the GDPR. In her view, a "fair data label" is needed.

THE JOURNALIST'S DATA WAS SENT TO 135 COMPANIES

Data was delivered to a total of 135 actors. For instance, The New Yorker delivered data to 56 third parties.

GDPR QUERIES SENT TO

- Facebook
- Twitter
- Google



MIDDLE-AGED JOURNALIST

Services used

- Chrome
- Facebook
- Google photos, calendar, contacts & talk
- Youtube
- Office Outlook
- Spotify
- WhatsApp
- Ilta-Sanomat
- S-pankki
- TaksiHelsinki
- Yle-uutisvahti

The total amount of mobile traffic data – 17,000 HTTP requests and 730 MB – corresponds to about 365,000 A4 pages of text.

A total of 28 websites or applications delivered data to 39 advertising technology-related services, which were located in

- United States (26)
- EU (11)
- Norway (1)
- Canada (1)

Google's DoubleClick collected data from the greatest number of websites, a total of 20. Other major collectors included AdForm, AppNexus, Rubicon Project and KruxDigital, each of which collected data from approximately 10 websites.

In quantitative terms, AppNexus collected the most data, with a total of nearly 3 MB. This corresponds to about 1,500 A4 pages of text.

Used services:

28

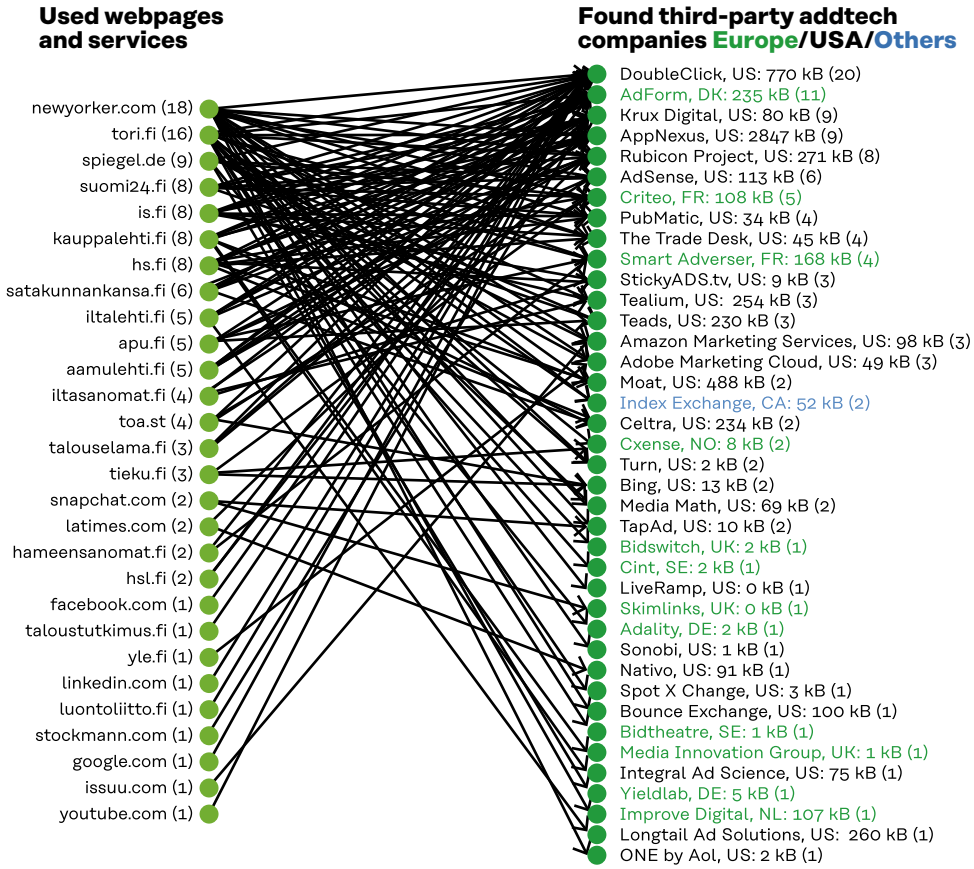
Total number of AdTech / marketing companies identified:

39

EU: 11
USA: 26
Norway: 1
Canada: 1

Total data to these companies:
1500 requests,
6.7 MB ~ 3400 pages

DATA WAS SENT THROUGH 28 SERVICES TO 39 ADVERTISING AND MARKETING COMPANIES



56 THIRD-PARTY BODIES WERE FOUND TO BE LINKED TO THE NEW YORKER'S SERVICE. 22 OF THESE WERE MARKETING COMPANIES

Annex 2 gives a more detailed description of the different actors involved.

Category	Recognised third-party companies/servers		
Ad tech/ Marketing 730 kB	*Amazon Marketing Services* *Adobe Marketing Cloud* *AdSense* *AppNexus*	*Bounce Exchange* *Criteo* *DoubleClick* *Index Exchange* *LiveRamp* *Moat* *Media Math*	*PubMatic* *Rubicon Project* *Sonobi* *TapAd* *The Trade Desk* *Turn*
Other third-party servers for running the service 64 kB	*Adobe Audience Manager* *Google Analytics*	*Quantcast* *ScorecardRe-search*	*Hotjar (all Audience measuring)*
Other same publisher's webpages 188 kB	*condenastdigital.com* *allure.com* *architecturaldigest.com* *bonappetit.com* *epicurious.com*	*glamour.com* *golfdigest.com* *gq.com* *lennyletter.com* *pitchfork.com* *self.com* *teenvogue.com*	*them.us* *vanityfair.com* *vogue.com* *wired.com* *wmagazine.com*
Uncategorised 385 kB	**.chtbl.com* **.ad.gt* **.parsely.com* *acdn.cookielaw.org* *cdn.keywee.co* *cdn.pdst.fm* *cntraveler.com*	*events.bouncex.net* *geolocation.one-trust.com* *medtargetsystem.com* *mid.rkdms.com* *pixel.keywee.co*	*player.cnevids.com* *sail-personalize.com* *segment-data.zqtk.net* *tr.snapchat.com*

POLITICIAN SEES THE TRANSPARENCY OF THE DATA ECONOMY AS A SOCIETAL CHALLENGE

The subject, a politician, said that she does not think about her data in day-to-day life, although she knows that the operating rules in the data economy are relevant to private individuals and society. The subject said she was aware of targeted ads but was unaware of how data is auctioned or how targeting happens in practice.

The subject believes that individuals should understand the data economy better, and that society should also think about how to make the data economy more transparent and visible. She said that in the future she would read digital service cookie policies more carefully and will try to prevent their use as much as possible. She also said that she would pay particular attention to the use of her location information.

THE POLITICIAN'S DATA WAS SENT TO 89 COMPANIES

Data was delivered to a total of 89 actors. For instance, the Ilta-Sanomat website delivered data to five third parties

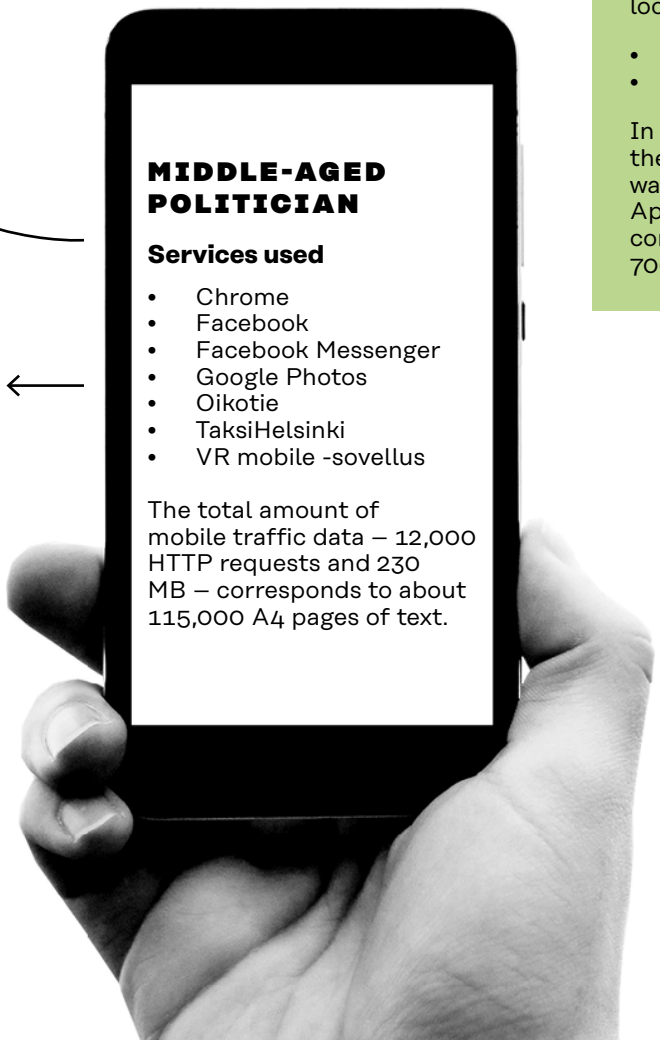
A total of 19 websites or applications delivered data to 26 advertising technology-related services, which were located in

- United States (19)
- Europe (7)

In quantitative terms, the most data (1.4 MB) was delivered to AppNexus. This corresponds to about 700 A4 pages of text.

GDPR QUERIES SENT TO

- Facebook
- Twitter
- TaksiHelsinki
- VR



Used services:

19

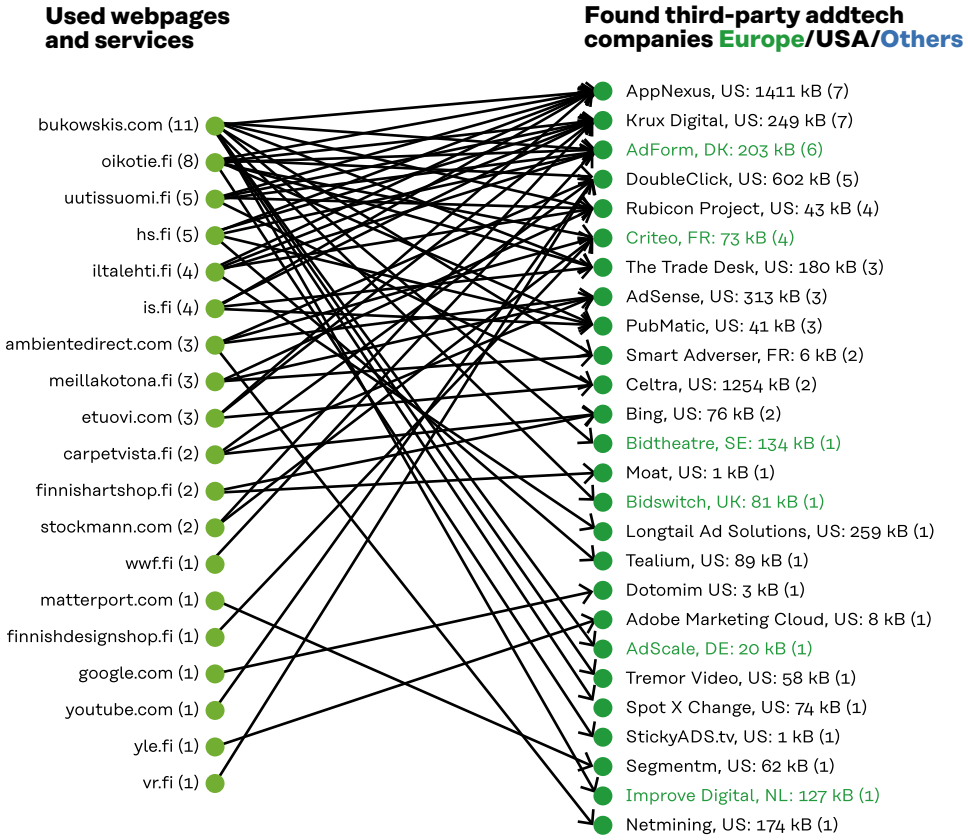
Total number of AdTech / marketing companies identified:

26

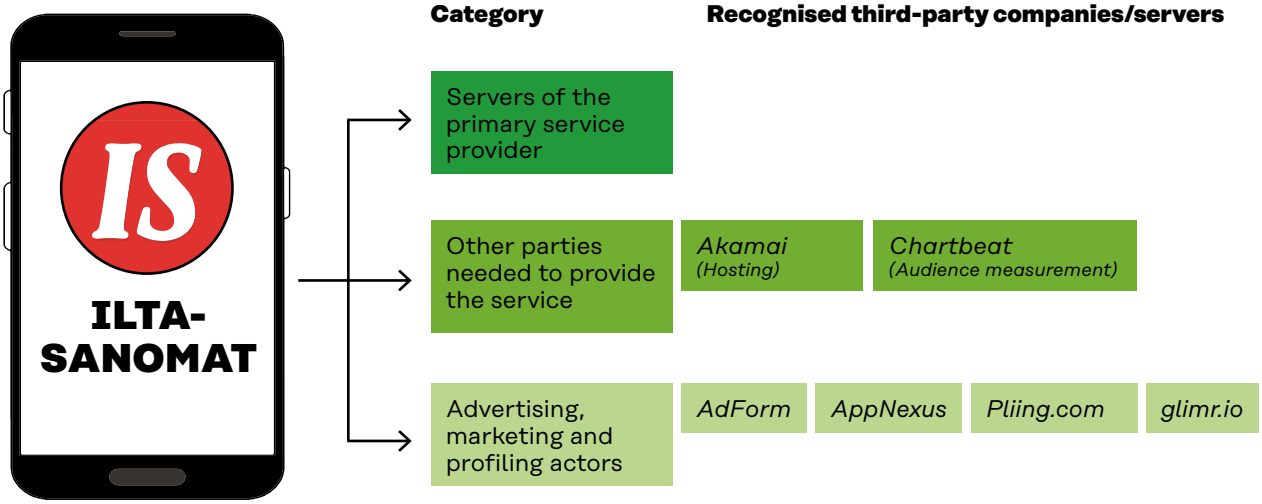
EU: 7
USA: 19

Total data to these companies:
1500 requests,
5.4 MB ~ 2700 pages

DATA WAS SENT THROUGH 19 SERVICES TO 26 ADVERTISING AND MARKETING COMPANIES



6 THIRD-PARTY ACTORS WERE FOUND TO BE LINKED TO THE ILTA-SANOMAT SERVICE. 4 OF THESE WERE MARKETING COMPANIES



EXAMPLE APPLICATION: TWITTER

Data on Twitter was delivered to its own servers in twitter.com and twimg.com. Data was also delivered to Google Analytics, which is the most common user analytics service used by websites and applications. Google Analytics is mentioned as a third party in Twitter's Privacy Policy.

SENIOR MANAGER WOULD MAKE OWN DATA AVAILABLE TO COMPANIES IN RETURN FOR THE SERVICE

The subject, a senior manager, took part in the study because he was curious to find out the positive or negative aspects of making his data available to companies. He did not consider data disclosure or targeted advertising to be a problem. He thinks that having users choose whether to grant access to their data or to pay for services would be a better alternative to the ”fair data label”. Personally, he would rather make his own data available to companies in exchange for services.

THE SENIOR MANAGER'S DATA WAS SENT TO 112 COMPANIES

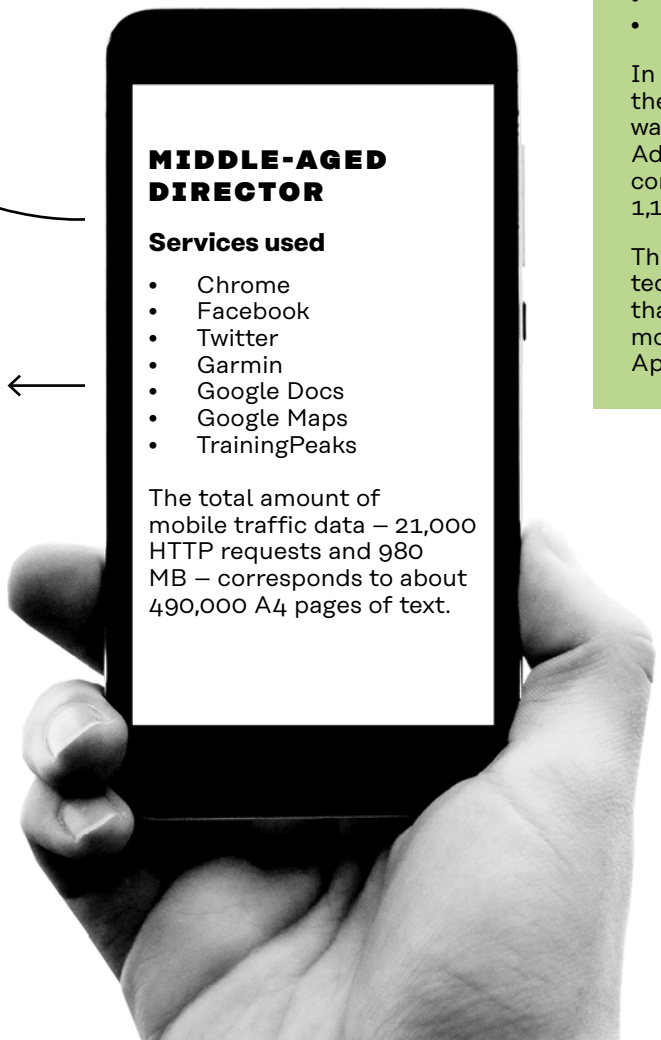
Data was delivered to 112 actors. The Garmin application included no third parties, which was rare among the companies in the report.

A total of 22 websites or applications delivered data to 37 advertising technology-related services, which were located in

- United States (27)
- Europe (9)
- Canada (1)

In quantitative terms, the most data (2.2 MB) was delivered to AdForm. This corresponds to about 1,100 A4 pages of text.

The advertising technology company that the web pages used most often was AppNexus (11 pages).



GDPR QUERIES SENT TO

- Facebook
- Twitter
- Google
- Garmin
- TrainingPeaks
- Helsingin Sanomat/ Sanoma

Services used:

22

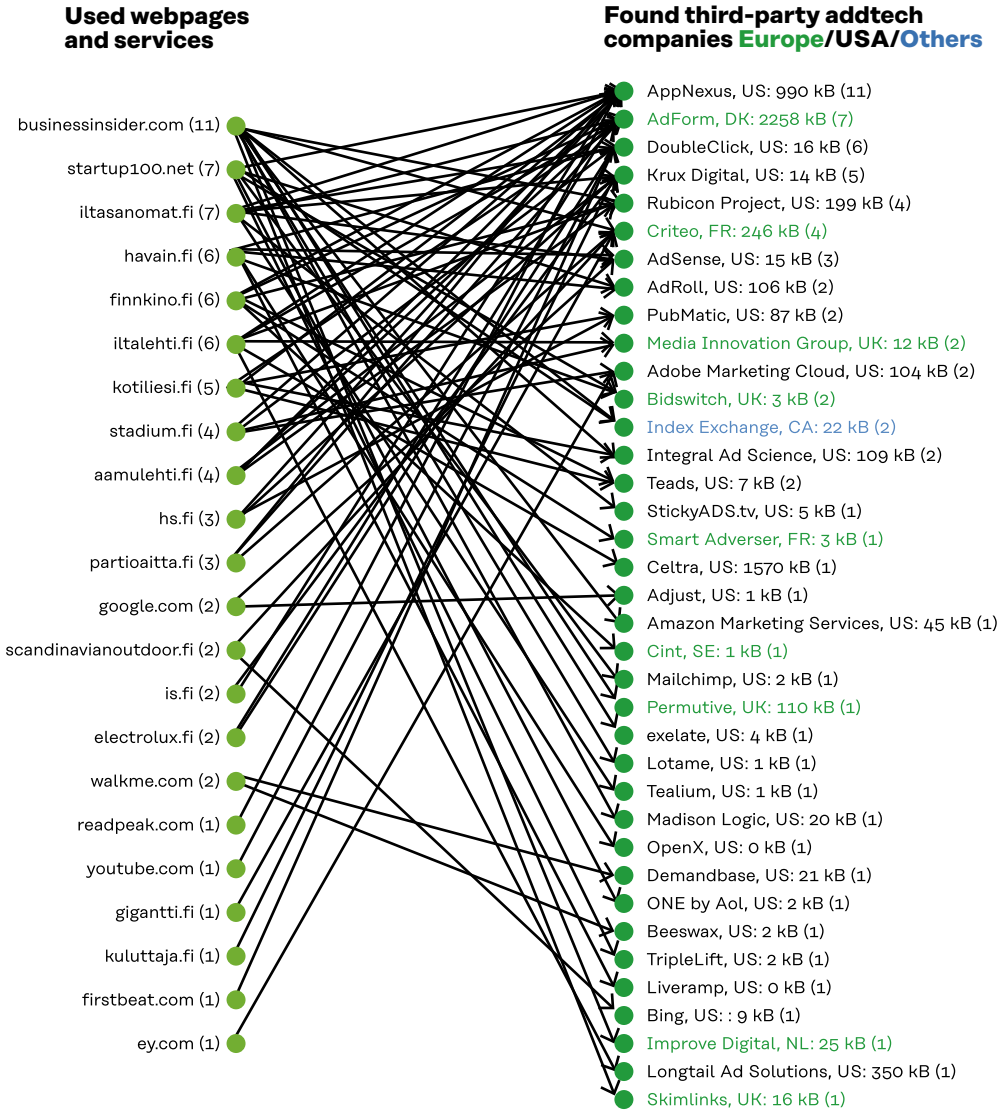
Total number of AdTech / marketing companies identified:

37

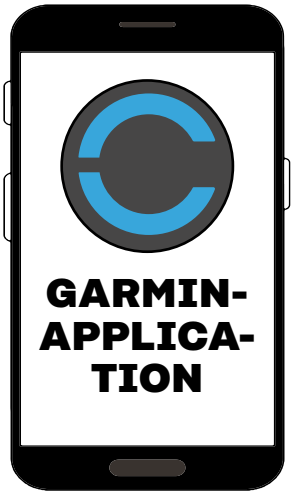
EU: 9
USA: 27
Canada: 1

Total data to these companies:
800 requests,
6.2 MB ~ 3100 pages

DATA WAS SENT THROUGH 22 SERVICES TO 37 ADVERTISING AND MARKETING COMPANIES



NO THIRD PARTIES WERE FOUND IN THE GARMIN SPORTS APP



Category	Recognised third-party companies/servers
Servers of the primary service provider	

RETIREE AWARE OF PARTICIPATING IN "THE WORLD'S BIGGEST LIE" WHEN CLICKING "I AGREE"

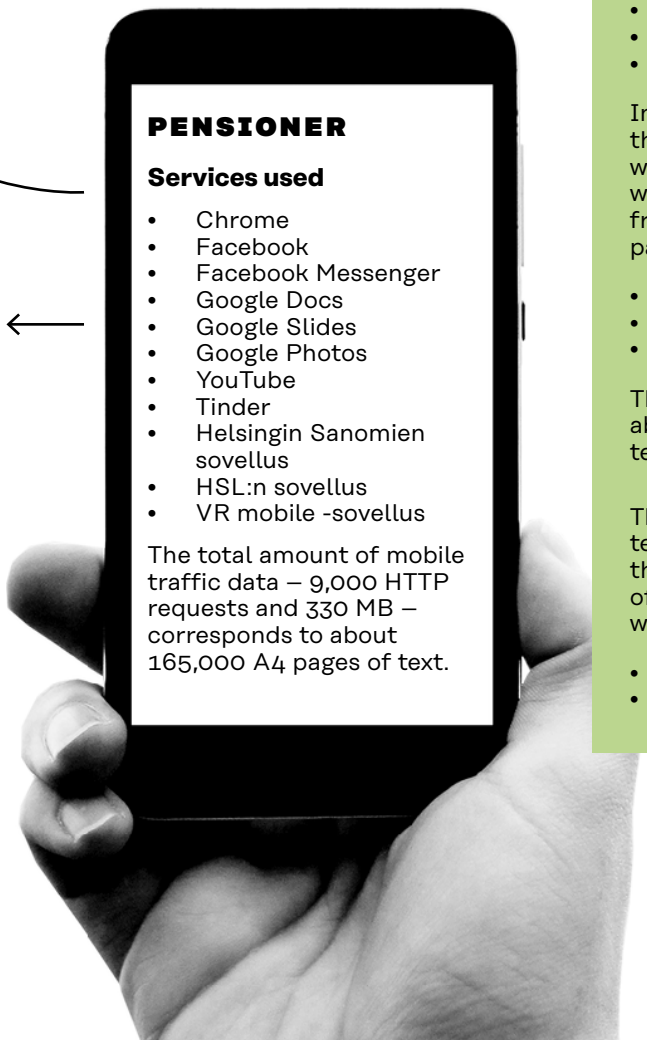
Through the study, the subject, a retiree, wanted to learn concretely about her own digital trails, because she knows that she is paying for digital services with her data. The subject said she knew that data is used to build addictive applications but has nevertheless authorised the use of her data. By accepting the terms of services without reading them, the subject feels that she is part of "the world's biggest lie". The subject said that she used Google's search engine a lot and felt she benefited from it. She considered the Google Photos service to be particularly valuable.

THE PENSIONER'S DATA WAS DELIVERED TO 111 COMPANIES

Data was delivered to a total of 111 actors. For instance, the Gigantti website delivered data to 13 third parties.

GDPR QUERIES SENT TO

- Facebook
- Google
- HSL
- Yle
- Telia



PENSIONER

Services used

- Chrome
- Facebook
- Facebook Messenger
- Google Docs
- Google Slides
- Google Photos
- YouTube
- Tinder
- Helsingin Sanomien sovellus
- HSL:n sovellus
- VR mobile -sovellus

The total amount of mobile traffic data – 9,000 HTTP requests and 330 MB – corresponds to about 165,000 A4 pages of text.

18 websites or applications delivered data to 27 advertising technology-related services, which were located in

- United States (19)
- in the EU (7)
- Norway (1)

In quantitative terms, the most data (2.4 MB) was delivered to Teads, which received data from the following pages:

- Kotiliesi.fi
- Hidastaelamaa.fi
- Mobiili.fi

This corresponds to about 1,200 A4 pages of text.

The advertising technology companies that were used most often on the web pages were

- AppNexus (7 pages)
- AdForm (6 pages)

Used service:

18

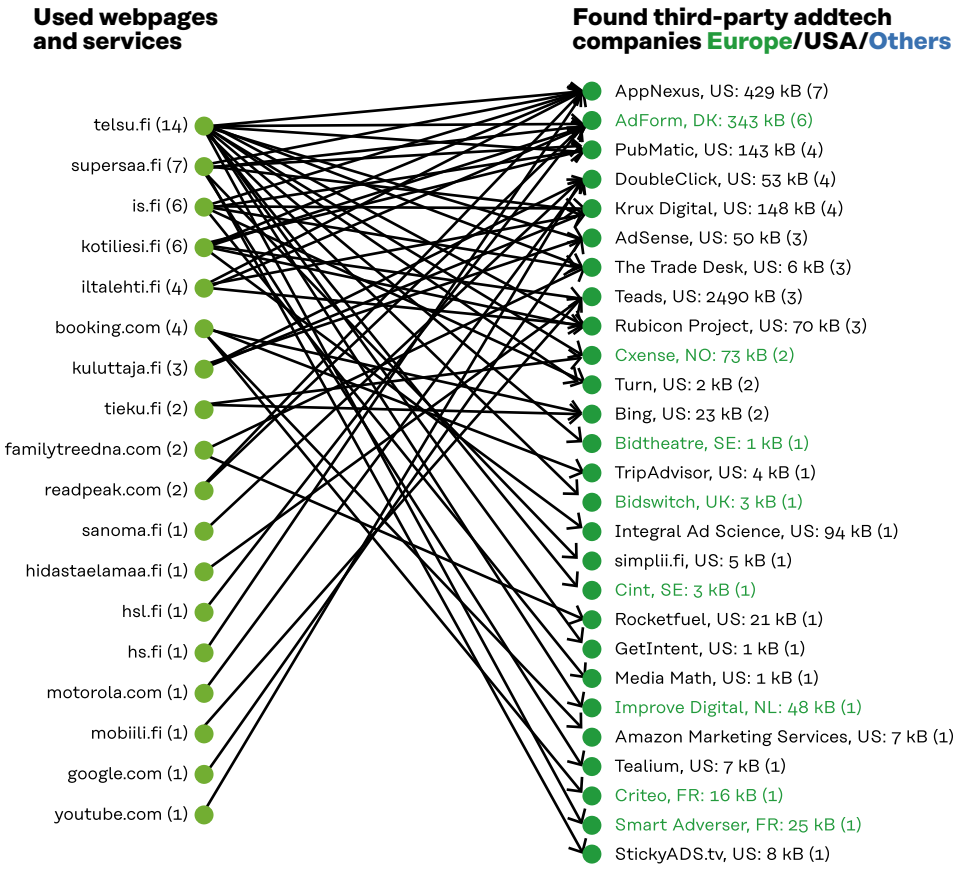
Total number of AdTech / marketing companies identified:

27

EU: 7
USA: 19
Norway: 1

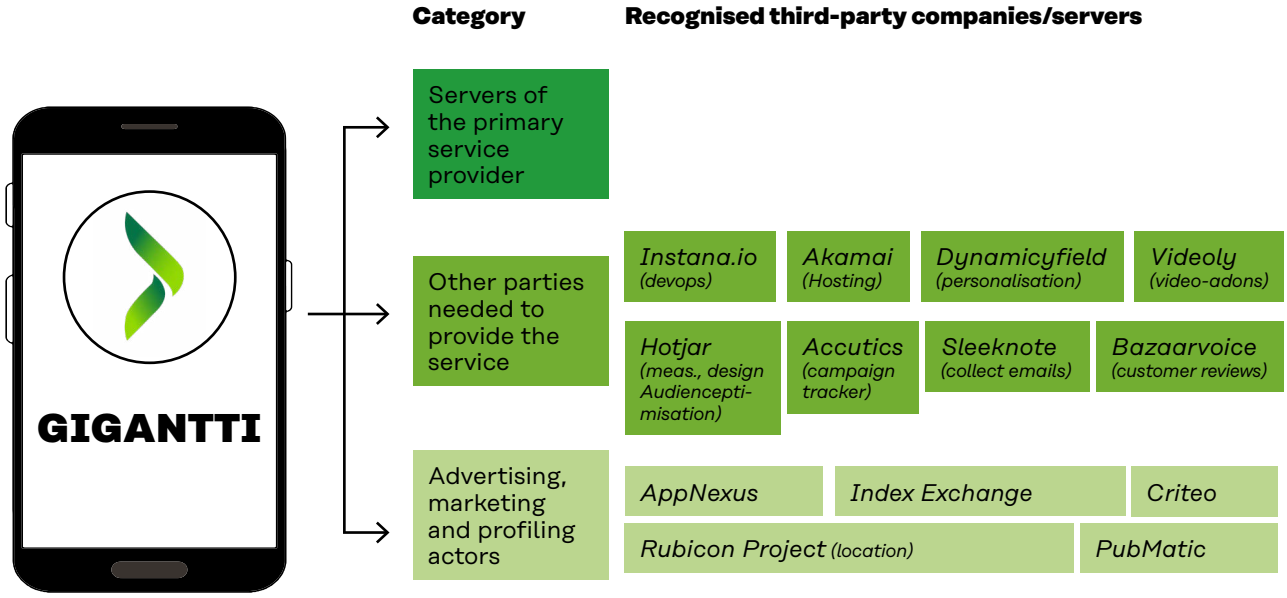
Total data to these companies:
500 requests,
4.0 MB ~ 2000 pages

DATA WAS SENT THROUGH 18 SERVICES TO 27 ADVERTISING AND MARKETING COMPANIES



13 THIRD-PARTY ACTORS WERE FOUND TO BE LINKED TO THE GIGANTTI SERVICE. 5 OF THESE WERE MARKETING COMPANIES

Annex 2 gives a more detailed description of the different actors involved.



4 Main observations of the Digitrail survey – lack of transparency, insufficient data protection regulation

It is impossible for people to know what data has been collected from them and who has it. A person's data is enriched at various stages of the data flow to create a profile of them. Profiles are formed without consumers' knowledge and, despite the extent of data collection, they do not reflect reality. The General Data Protection Regulation permits individuals to gain only limited access to their data.

It is not possible for individuals to find out how their data is circulated

According to the survey, data was sent to several third parties, mainly US companies. The test subjects knew they had accepted data collection when using the services, but were surprised by the number of third parties of which they were unaware. The test subjects had not read the terms of the services or the cookie policies.

The largest number of individual third parties on a single web page found in the test subjects' data was 56 and, according to many studies and service providers' own reports, the number can be considerably higher. The examination of individual data transmissions, or HTTP packets, revealed that altogether 15 per cent of the data packets were sent to digital advertising bodies.

At present, in consumer services the dominant model of the platform economy is based on large platform companies collecting as much data as possible from their users and monitoring users outside their 'own services'. The basic assumption is that the information accrued from users is automatically made available to companies.

The GDPR does not adequately protect the rights of the individual

The main problem with the General Data Protection Regulation, the study found, relates to the extensive and opaque ecosystem of the data economy. Online service users cannot control their data flow because they do not know where their data is.

The GDPR, which entered into force in Europe in 2018, provides individuals with a fairly narrow view of the use of their personal data. For the most part, an individual can only obtain information on the data collected by the primary service provider. However, the primary service provider discloses data to countless third parties whose identity is hard to verify. The user of online services is not familiar with these third parties and cannot therefore target them with the measures provided by the GDPR. A service user cannot, for instance, check whether the profile created about them is based on accurate information. The user must rely on sweeping, difficult-to-read cookie and privacy policies and lengthy terms of use.

It is difficult to get responses to detailed information requests

In Sitra's survey, the test subjects sent a GDPR-based request to the companies they used to obtain a copy of their own data. Several services allow their users to download their own user data using an automated function within the service. The information on data accumulation, profiling and the use of data obtained in this manner proved to be superficial. In addition, the test subjects requested clarifications on their data by sending a separate email message to the companies or by approaching them using a data protection-related customer service form. These more detailed requests for information had been prepared by Paul-Oliver Dehaye. The detailed requests asked, among other things, which kinds of data the services collected from users, how the services profiled their users, and how they informed users about the third parties involved in the use of the data. The questionnaire sent to the service providers is included in Appendix 3.

It was difficult or nearly impossible for the test subjects to get a response to their detailed requests for information from the service provider. The companies did not respond to the enquiries on the origin of the data and third parties to which the users were entitled on the basis of the GDPR. Similarly, the profiling-related responses were very general and did not help the person making the query assess their own data. For example, one of the test subjects sent three email messages to Twitter but only received responses that were difficult to interpret and did not explain how they could obtain more answers. In many EU/EEA countries, the data protection fines imposed by data protection authorities have typically been related to such things as the failure to respond to requests for information from data subjects.

Individual data accumulated in the services is enriched and processed at various stages of the data flow. The data is also used to create a profile of individuals, the effect of which is reflected, for example, in the types of messages and advertisements displayed to them. Nearly all of the first-party services examined in the survey personalise their services by using various group profiles that are determined based on the data. However, it remained unclear how these profiles were formed and whether they were used for purposes other than advertising. The maintenance of profiles and the duration of their use also remained vague.

Profiling also enables opinion formation and has an impact on the pricing of services. Since profiles are used by companies and other organisations for a wide range of purposes, it would be important for the people's profiles to reflect reality. This will be particularly important in the future, since profile data could be used, with the individual's consent, in more critical areas than advertising, such as health and well-being services.

Because the test subjects did not receive clear information about their profiles, the authors of the study used MyDataAppNexus to examine their own profiles. They discovered that the profiles created by data vendors did not reflect reality. AppNexus now goes by the name Xandr and displays profile information on its website. It is worth bearing in mind that only a few data vendors provide the user with the opportunity to access their own profile.

The more aware and well informed the users are about the issue, the easier it is for them to ask the right questions. However, responsibility for this should not rest with the individual, but with companies.

The operating principles of digital advertising have a wider social impact

The environment of digital advertising with its varying interdependencies is so complex that even experts in the field struggle to understand it. Individual services are backed by an extensive network of unknown actors, for whom primary services, such as commercial media, provide a channel for collecting as much raw material – data – as possible. This has an impact on the activities of all companies, including those outside the industry, when it comes to consumer privacy and trust, as well as in terms of how valuable data ends up in the hands of a few.

5 Data is delivered to third parties via multiple channels

Free services are considered adequate pay back for handing over data. The true price of these services cannot be judged because it is impossible to find any information on the circulation and use of the data. The exchange is therefore not fair.

Data collected from individuals contains both consciously disclosed contact information and personal data, and unknowingly created traces of online behaviour and location data. In the value chains of digital advertising, data flows further and further away from the consumer into a multi-layered network, a "black box".

The majority of the data produced by individuals is collected by large American platform companies, such as Google, Facebook and Amazon. The field is further expanded by big Chinese platform companies, such as Alibaba, Baidu and TikTok. In Europe, there are only a handful of data economy success stories, the most well-known of which are the music service Spotify and online store Zalando. The most prominent African platform operator is Naspers.

The heavyweight data collectors also include gaming companies, which are particularly popular with children and young people. Well-known gaming companies include the American EA, French Ubisoft, and Supercell, a company of Finnish origin in which the Chinese Tencent now has a majority stake.

The main reason for monitoring individuals is behavioural advertising. A person is shown ads about things they may have recently been interested in based on their online behaviour. Marketing is based on influence, but how far can companies go when it comes to influencing people and what methods does it permit?

Personal and behavioural data can be delivered to third parties via at least four different channels

- data auctions
- websites and other services and applications (this formed the core of the Digitrail survey)
- companies that combine individuals' data (cookie syncing platform)
- finished code used in building services.

A Princeton University study found that news sites contain the largest number of third parties, while government, university and NGO sites contain the smallest.

In a data auction, the advertiser finds a place to advertise

According to the report The Great Data Race commissioned by the Norwegian Data Protection Authority, some websites receive their income from advertisements displayed to users. Even before you open a webpage, your data will be sold at an advertisers' auction in a fifth of a second. Advertisers with the highest bids get to display their ads.

Auction operators receive basic information about website users. This information is

added to existing information on users, information from public registers and information related to users purchased from data vendors.

Data vendors, or data brokers, buy and collect consumer data from different sources, combine and package it and then sell it to businesses as enriched data products and services. Artificial intelligence allows each actor to calculate the price it is willing to pay for the advertisement the users see on a web page. The highest bid wins and users see the advertisement on opening a web page. To understand how their data is deployed, web page users should read the privacy policies of all auction operators.

Already in 2015, the data of 1.3 million users was auctioned every second. The number of sales events exceeded that of the New York Stock Exchange twelvefold. Major American companies, such as Facebook, Yahoo, Google and Microsoft, run their own auctions.

Data transmitted through websites and other services and applications

When people use online services, the basic assumption is almost without fail that the data generated from individuals is available to service providers and their partners. If users agree to the terms of use of the service, the accumulated data may, in the case of The New Yorker, for example, spread to hundreds of different operators. Sitra's investigation discovered 56 different operators to which data was transferred (18 of them were linked to advertising, marketing and profiling). However, The New Yorker's privacy policy listed as many as 280 different partners.

People's data is combined using different sources

The Digitrail survey found that The New Yorker's collaborators included a third party, DoubleClick.net, owned by Google. This is an active operator in the data combining platform (cookie syncing platform) and was included in the list of data collectors of almost all the survey's test subjects. The use of digital services generates several identifiers (IDs) that contain information about people's digital behaviour. Some actors, particularly companies belonging to the same group, share the generated user data with each other through a data aggregation platform, thus gaining a more comprehensive view of their users.

Collecting data through a ready-made code used for building applications

Kaveh Waddell, the deputy editor of Consumer Reports Digital Lab, has written of his concern that user data can be collected through various imported code snippets. When building applications, it is usual to use ready-made code when including certain functionalities (such as when adding a chat function to an application). The service users cannot know that they are involved with several different organisations via both the company that developed the application and via the code. It is therefore possible that the borrowed snippets will even deliver sensitive information about users to other companies. This forms a relationship with several mediators or marketing companies that the users have probably never even heard of. All this is done without the users' consent and sometimes even without the application developers' supervision. Other means of data collection are also often invisible to ordinary users.

A HYPOTHETICAL EXAMPLE*

Common ways in which a person's data can spread to third parties

*This example has been created by combining results from Sitra's study and the following research and articles:

Datatilsynet Norge 2015. The Great Data Race. How commercial utilisation of personal data challenges privacy.

Waddell, K. 2020. Some developers don't know what their apps do with your data. Here's why most apps use off the shelf code and some of it can be risky.

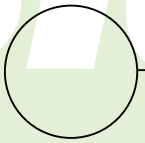
Englehardt, S., Narayanan, A. 2016. Online Tracking: a 1-million-site Measurement and Analysis, Princeton University.

A person arrives at newyorker.com

They are asked to accept the use of cookies.

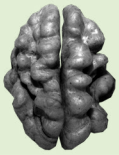
The person gives concent to the use of cookies at default settings.

THE 0.2 SECOND AUCTION



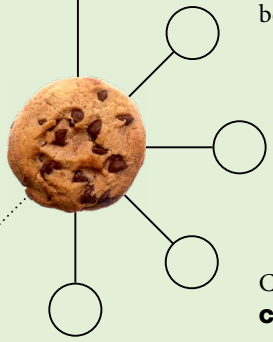
An auction where companies bid for the adspace visible to the person, is held immediately.

Participants get data about the person from the target website and combine it with data from other sources.



Algorithms determine a value for the ad space based on the gathered data.

The auction is over in 0.2 seconds. The winning bidder's ad is shown to the person visiting the website.



Once on the site **the persons data is immediately accessible by 56 companies.**

Some companies begin tracking the person's behaviour on the website.

At least **18 of said companies are focused on profiling.**

Companies that are a part of data combining **cookie syncing platforms** can form a comprehensive profile of the person.

According to the site's cookie policy, **as many as 280 companies can get information about the person at will.**

Many online services utilise pre-written code to provide some features (e.g. maps). Borrowed code can also collect data for different companies.

6 We pay for services with data, but what happens to our privacy?

When it comes to privacy, the business models of the giants of the platform economy and digital advertising have been built in a fundamentally problematic way. Users have a limited opportunity to evaluate the effects of their consent when they are asked for it.

For advertising platforms, the consumer is a source of data which is used to generate value. The primary service provider is the party that has a customer relationship with the user. Advertisers strive to create customer relationships or strengthen existing ones.

The technology and data giants that managed to climb onto the data economy bandwagon early on have occupied dominant market positions in consumer services. Their services have attracted a great number of users, and they not only provide entertainment but also make people's daily lives easier.

The most successful platform services are concentrated in the hands of only a few groups, and their operating models hamper fair competition. This has a negative impact, for instance on European companies' ability to create new services.

Although the number of users on the largest platform services is enormous and still increasing, it is already becoming apparent that people's lack of trust in digital service providers restricts the use of digital services. The largest data collectors focus on safeguarding their dominant position. If the data was collected and used with people's genuine consent, they might also be more willing to disclose their data.

Challenges in privacy management

Although people do not feel that they control their personal data and privacy, they do not see any alternative to disclosing their data in situations where they want to ensure that they can continue to access services. A sense of control is sought by consciously regulating the amount, accuracy and quality of the information disclosed. This further distorts user profiles and does not work in the long run. Messages or ads that are displayed to users may not meet their actual needs.

As a result, people lose the benefits that the algorithms behind the services seek to provide. It would be sensible to let the individuals themselves authorise the use and sharing of their data. This way, user profiles would be more in line with reality and personalisation algorithms would be more useful.

Maintaining your privacy when using digital services requires extreme effort. People are expected to be able to carry out

complex assessments of the use of their personal data and to allow it to be used by digital service providers only when the advantages outweigh the disadvantages. In practice, conducting this sort of analysis is impossible because the information about the data itself and the operating models of the data economy are far from sufficient. Requiring people to weigh up the pros and cons is too much to ask, especially when it comes to children and young people. And yet, children are an important target group for advertisers that seek to influence the next-generation's consumer habits. In its code of conduct, the OECD raises concerns about the collection of children's data through games and online toys, as data is often amassed without the user's knowledge.

Problems related to privacy include the following (adapted from Lehtiniemi, T. & Kortesniemi, Y.):

1. Given that the collection of personal data starts when the user accepts the terms and conditions set by the service, it is difficult for individuals to assess all future advantages and disadvantages. Even if immediate harm is negligible, long-term disadvantages may develop gradually over time.
2. Users have to accept the terms and conditions in full in order to use the service.
3. Those who collect data often aggregate the personal data of different people and contexts, so that new information can be found by conducting a data analysis.
4. The unexpected transfer of personal data to new parties is largely unclear to individuals, making meaningful decision-making more difficult.

When it comes to user-accepted terms and conditions, the situation is problematic and the law requires that users be able to use the service, even if they refuse the collection of their data. These practices will be refined through precedents, and the offices of the National Data Protection Ombudsmen will play an important role in making use of them to define more precise rules.

In the decision on the use of cookies issued by the Court of Justice of the European Communities in the so-called Planet 49 case, it is stated that the consent for cookies cannot be validly given through a pre-ticked box. Instead, users must give their active consent. In one of its precedent decisions, the Office of the Finnish Data Protection Ombudsman has outlined what the requirement for consent by active measures requires from cookie policies. In its decision, the Finnish Data Protection Ombudsman's Office stated that even if users do not alter their browser settings or fail to take any other action, it does not mean that they have given their consent to the storage and use of cookies.

In Europe and in Brazil, for example, efforts have been made to protect individuals through legislation. According to an article by the World Economic Forum, there are also other ways to protect your data. Some of the measures relate to data protection acts (Europe and Brazil), the requirements of data localisation (Russia and India) and the weakening of data encryption (Australia, Cuba, Morocco), while others relate to data retention (Colombia, Italy, Ethiopia).

People find the services of platform companies important

The large platform companies and the digital advertising machinery have developed to their current form gradually, in parallel with the growth of consumer-friendly mass services. Large consumer-oriented platform services have provided people with both joy and benefits. It is difficult to measure the benefits consumers have received financially, but a study of 65,000 people conducted by the Massachusetts Institute of Technology (MIT) attempted to outline the value of these products for end-users.

Over consecutive years, MIT researchers have asked service users how much money

they think they should be given for them to agree to giving up one of their digital services. Respondents in the United States have wanted $40–50 a month to give up Facebook. In Europe, people wanted €59 a month for mobile phone map services and as much as €536 per month for the instant messaging application WhatsApp, which is seen as an important tool for communicating with family and friends. When reviewing numerous products, people selected search engines as the most important service. In the 2017 survey, people wanted as much as $17,530 a year for giving up search engines.

The erosion of privacy has taken place as if by stealth, since people receive popular services in exchange of their data. Large platform services and digital advertising have also helped to develop and maintain a free-of-charge internet that is open to all. So far, the advantages have been enough to outweigh the disadvantages of privacy loss. However, civil activism, changes in people's online behaviour and media exposure signal that the limit has been reached when it comes to influencing carried out by companies and the collection of personal data.

7 The business models of digital advertising need to be reconsidered

The most popular consumer platform services have been built over the course of many years, providing significant benefits as well as drawbacks. Privacy has become the most problematic issue in the market. The field of digital advertising built around these platform services is in transition.

Data is collected because it provides valuable raw material for developing services and creating new ones. The domain of digital advertising and data analytics includes an enormous number of companies that generate income either by collecting personal data, or by processing, searching, storing, combining, enriching or analysing it – or all of the above. Data helps to bring together companies that offer advertising space and organisations that want to advertise. Each company receives its share of the value created by data, because the best feature of data is its almost unlimited technical ability to utilise the same and further enrich data as a raw material in different parts of the business revenue chain over and over again. Data can diversify and increase in value in the possession of every actor.

The strong growth of targeted internet advertising has already lasted for over 20 years. During this period, the industry and its neighbouring fields have become such a dominant business that it is difficult to change its practices and revenue models. Throughout the world, digital agencies, data collectors, data enrichers and analysts, content publishers and others have based their own operations on the business model of the platform economy, where data collected on people's behaviour requires maximisation for optimal targeting. Consumer marketing companies have had to accept their role as utilisers of data that is mostly invisible to them, receiving advertising space in return. They have also accepted their role as the enablers of the whole juggernaut. The problem is starting to become apparent, and the debate on privacy and data management has taken off not only within the industry but also between advertisers and the industry

Does the GDPR allow free riders?

European companies have invested much time and money to comply with the EU's Data Protection Regulation, which entered into force in 2018. Companies are worried about their uneven competitive position in relation to American and Chinese platform companies, a sentiment ascertained by a Sitra business survey. In the current situation, large consumer platform companies

and the closed ecosystems built around them reap the business benefits of data collected on Europeans. Since some of the digital advertising and data collection companies operating around large platforms cannot be reached by the labour-intensive data protection regulation, the situation is detrimental not only to consumers but also to the competitiveness of European companies.

ACCORDING TO THE SITRA BUSINESS SURVEY THE FUTURE OF EUROPEAN COMPANIES IN THE DATA ECONOMY, THE WAYS DATA IS USED VARIES GREATLY BETWEEN COMPANIES.

- All companies that responded to the survey utilise user data to provide or develop their service.
- All companies that responded to the survey use data for customer analytics.
- Almost all companies personalise their services according to the user.
- All Finnish actors said that they use personal data for direct marketing.
- Large actors strive to gain as comprehensive an understanding of their users as possible so that they can accurately target their advertising according to users' interests and situation in life. The data collected can be refined into new products, which the companies will then sell to others. These products may no longer contain personal data, but they may include aggregated information about users, such as the data on crowd movements sold by Telia.

8 Major challenges in the field of data-driven consumer services and digital advertising

The field of digital advertising and data analytics is being transformed due to pressure from consumers and legislation, as well as changes within the field. The task of protecting privacy has until now been shouldered by consumers.

Until now, the business of digital advertising has been largely based on small program snippets recorded on the users' devices known as cookies, used to monitor the behaviour of website visitors. Third-party cookies have turned out to be particularly problematic, since they do not provide users with visibility concerning or a direct relationship with operators.

Consumers' demands for privacy have increased, so browser companies have had to modify their products. Regulation has also contributed to this trend. Apple Safari was the first browser to give up third-party cookies in 2017, and next they were abandoned by Firefox in 2018. According to the industry organisation Interactive Advertising Bureau Europe (IAB Europe), 30 per cent of browser usage is already carried out without third-party cookies. As Google Chrome, which represents about 65 per cent of overall usage, is also about to give up using third-party cookies, the role of cookies in offering advertisements will practically end. As the company has not announced any exact schedule for doing this, the market has shown signs of uncertainty. Plans are already frantically being made for the post-cookie era, since consumer monitoring is still not something that companies are willing to give up.

Growing importance of companies' own data resources

Platform and technology companies make grand promises in the media about their commitment to protecting privacy, while also acquiring increasingly diverse data expertise from the market and developing new solutions to manage a bigger part of individuals' digital service use. Alongside laptops and mobile phones, various wearable meters (e.g. sports watches), smart TVs, home surveillance devices and voice assistants have already become available. Each individual device and its applications have their own ways of collecting data: sound, images, dimensions, location and route information, as well as data on heart rates and periods of ovulation. In addition to the information provided by service users, the devices collect behavioural and health data and other information which lie beyond any conscious influence. The more extensive the data collected for one operator is, the bigger its data pool becomes and the greater its ability to understand individuals and create new services.

In addition to platform economy giants and digital advertising companies, other companies also strive to create value through data, but often through a different operating model. Instead of continuously monitoring consumers, companies are seeking to create increasingly in-depth customer relationships, not only by interacting with people on social media, but also by collecting data, such as by building their own closed applications. An application can either be the core of the entire service or it can provide a way of otherwise committing the customer to the company. When the application needs to be separately downloaded or when it requires a login, companies have to consider and weigh the attractiveness of their brand, the usefulness of their service and the importance of the customer relationship to the consumer. Customer understanding is crucial for developing customer loyalty and services, which means that increasing the company's own data pool is important for the entire business operation. Building separate applications allows companies to have their service available while they accumulate their data resources on their own terms.

The field of actors is constantly changing

The field of actors of digital advertising and related data analytics mainly consists of a limited number of digital service groups that compete with each other. One of these is AppNexus (Xandr), a major consumer data processor and a division of the media giant Warner, which in turn is owned by the American telecommunications giant AT&T. This example shows how different companies intertwine and how consumer data could be gathered from different services. For instance, combining telecommunications data with data generated by other applications provides an attractive option for creating new business.

The field of actors of the platform economy, digital advertising and data analytics expands as companies from "traditional" industries seek new growth areas and their role in the data economy and as they challenge established companies. Digital advertising companies are also concerned at the concentration of data and power in the hands of a few technology giants. They find it important that data circulates in advertising ecosystems, so that the data collected on individuals does not end up in closed ecosystems, only to be utilised by the likes of Facebook and Google (Interactive Advertising Bureau: The Socioeconomic Impact of Internet Tracking, February 2020).

The numbers and roles of actors in the market are constantly changing, as new players enter the field and large companies make business acquisitions. Large market players are expected to continue trying to buy up smaller companies and thus increase their expertise. European companies are busy trying to join this race.

The data business of consumer services is substantial and has managed to create success stories worth billions of dollars, but the data economy still has a lot of untapped potential, particularly in B2B. Because the data economy is in the early stages of its development, there are still relatively few examples of business-to-business success stories outside the consumer market. As companies' understanding and competence increase, they will be able to exploit the potential of the data economy across all industries and in businesses of all sizes.

9 How to protect your privacy – recommendations for everyday life

Understanding the rules and opportunities of the data economy should be a civic skill. Everyone should also know their rights, especially in terms of privacy. Respect for privacy should feature in all aspects of everyday life, as data is collected through a wide range of smart devices.

People must be able to influence how their data gets used. By accepting the default settings of the terms of use, people may be giving companies the right to collect as much information as possible both when using the services and outside them.

To increase knowledge of the data economy, everyone should be an active agent and demand that their rights to privacy be preserved.

You can protect your privacy online by using the following methods:

- Carefully read all privacy settings and pay attention to where your data can be sent from your primary service provider.
- Read the privacy policies of the partners mentioned in a company's Privacy Policy to understand what they do with the data they receive.
- Properly familiarise yourself with your browser settings.
- Use different browsers for different purposes and use browsers that have been designed to protect your privacy to begin with (e.g. Brave).
- Use different search engines for different purposes. You can use one search engine, say, Google, for work and another, for example Duckduckgo, for other purposes.
- Install an ad blocker in your browser. It prevents the connection between your browser and an ad server. Some browsers have one preinstalled, but you will need a blocker that is right for you. Choose your blocker based on the online service you use the most: If you are a heavy user of YouTube, choose an ad blocker that is optimised for that particular service. If you like Facebook, choose a Facebook-optimised blocker. Ad blockers can be downloaded online. According to a study conducted at Princeton University, the Firefox browser's third-party cookie blocker works well. Similarly, using the GhosteryPrivacy browser considerably reduces the number of third parties.
- Install powerful privacy software (VPNs) on all devices. Setting up a VPN requires a little more effort. It prevents your device's IP address from being transmitted to third parties by encrypting the network connection of your network-connected device (phone, tablet or computer). This means that your internet operator will not be able to access your traffic either. You can purchase a VPN from specialised companies, and you may have already used one through your employer, since VPNs are often used in protecting business data for remote working.

The GDPR gives you the right to acquire more information about the data collected.

- The data protection regulation entitles you to ask your service provider for information about the data accumulated about you, third parties, profiling and the places where your data has been disclosed.
- Contact the Data Protection Ombudsman if your service provider does not send you your data or if you feel that it has been misused.
- You can also look for profiles made of you, for example, by using the Xandr service.

The above methods allow you to better protect your online privacy, but information about individuals is also collected outside of browsers, social media, and search engines. Various IoT and smart devices may share the data they collect with third parties. So it is a good idea to pay more attention to protecting your privacy, and not just through using your phone and computer.

SEVERAL OFFICIAL BODIES PROVIDE GUIDELINES ON HOW YOU CAN PROTECT YOUR PRIVACY.

Sitra's Digital Profile Test helps you to better understand the operating principles of that part of the data economy that is the most evident to consumers. The test also provides information on how you can better protect your own data.

The Danish European Data Ethics Forum has prepared extensive guidelines on the possibilities of protecting one's privacy.

The European Interactive Digital Advertising Alliance has written a guide to browser-based online advertising and online privacy. The website contains details about the operating principles of browser-based advertising, cookies, and ways of protecting your privacy online.

10 A leap into the fair data economy – recommendations for companies

Customer experience and corporate responsibility provide European companies with the opportunity of becoming recognised as the heralds of the fair data economy. Instead of replicating the ground rules for the platform economy, it is important to try to find new business models in which data is shared within networks with people's permission.

Each company is at a different stage of maturity in terms of how it uses data pools and handles data protection. A company that uses individual data in accordance with the principles of the fair data economy continuously evaluates its own operations according to their ethicality, data management and rule compliance. It also pays particular attention to accountability. Accountability can reach a level that provides companies with significant added value. They should therefore have the ability to produce content that at best can be used to produce a competitive advantage. Material that produces clear information for the company's management, sustainability measures and customers and that is easily adoptable in terms of responsibility helps the company to further develop its operations.

Terms of Use are part of the customer experience

Unfortunately, the terms of use of the most popular digital services are often more than 10,000 words long and confusing, with links to numerous new texts. It is clear that they are not intended to be easily read by the consumer. It is easy to overlook the needs of consumers, as this allows services to acquire extensive rights to use their data. However, in the long term, this is not in the interests of companies either.

Depending on the company, the EU's Data Protection Regulation of 2018 was treated either as a one-off project within the company or as an opportunity to evaluate its own data resources – or as something in between. For consumers, it has meant increasing difficulties in using the internet that have been brought about by cookie policies and privacy statements, in addition to which not all have understood why these changes have even taken place. Data protection is not yet a visible part of the everyday life of organisations, even though it should be regarded as a continuous development. European companies now have the opportunity to stand out as fair players who not only adhere to ethical guidelines and exercise transparent reporting on data usage, but who also have clear and easy-to-read cookie policies, privacy statements and terms of use that take better account of consumers' privacy preferences.

Companies should develop their customer data pools on the basis of their own operations and go as far as to refrain from unnecessary data collection. More attention should be paid to the comprehensibility of data protection policies. It would be a good idea to develop terms of use specifically from the customer's point of view and to provide tools that allow customers to easily manage their data.

Making data part of corporate social responsibility

The collection of data from individuals continues to increase steadily, but the opacity of the activity causes widespread concern among consumers and advertisers. As data has become the world's most valuable resource for business, it should also be viewed in terms of corporate responsibility, and the responsible use of data should become a more pronounced part of responsible business operations. In practice, this means transparency and clarity in data collection, use and reporting.

According to the mammoth World Federation of Advertisers (WFA), 82 per cent of employees in its member companies would consider leaving the company if the company's data usage was unethical. Platform giants, large technology companies and digital advertising companies should take this into account, as the WFA members are enormous international organisations and, according to the WFA, they also represent 90% of the investments in global marketing communications.

Pioneering companies that use data responsibly do not leave the responsibility for privacy protection to their customers, but seek ways to stand out and exceed the minimum requirements imposed by law. They look for business models that differ from the norm and build services from scratch, taking into account the customer experience and acknowledging consumers' desire for privacy. Consumers are given more choice in selecting the degree of data collection, and responsibility is also reflected in the fact that the collected data is definitely put to use and that it brings value through new services that customers appreciate.

The responsible use of data is also related to its sharing, as long as it is carried out with the customers' consent. Organisations that have a lot of data could share some of it with others, thus enabling innovations and creating growth around them. In the fair data economy, the consent of service users is central and must be taken into account in business planning, terms of agreement as well as technology and partner choices.

Revamping the rules of the data economy

The head start which the platform giants have gained over the years and which has been created by collecting and processing data that have accumulated around the companies is difficult to catch up. In the current model of the platform economy, the amount of data as well as the diversity of personal data the giants have collected are inaccessible to ordinary companies, and Europe is lagging behind. If European companies want their share of the data economy, they must find alternative ways of doing things.

A company that is actively looking for new business opportunities through data-driven services should start by evaluating its own business models and developing a data strategy. This provides a concrete way of describing the organisation's own data, identifying its shortcomings and taking a stand on, for example, data sharing. At best, this will lead to new data economy-related growth opportunities and help to cope even in tough situations.

Data partnerships and network-based data sharing models that differ from the business models of the platform economy could be a way to accelerate the European data economy. Sharing and combining different types of data with other organisations could provide a means for gaining new prospects in a more intelligent way to create services that target the needs of both the consumer market and the business market. This requires rules that are transparent, clear and common to all parties.

Value erosion of digital advertising input and alternative operating models

Some years ago, the increase in digital advertising technologies and automation was

believed to remove middlemen from the advertising value chains. In the UK, the two-year study on targeted advertising (Programmatic Supply Chain Transparency Study 2020) by ISBA, an organisation representing advertisers, and the consulting firm PwC showed that, although the old intermediaries have disappeared, the money spent on programmatic advertising flows to a new layer of middlemen. Every stage of the advertising supply chain includes value erosion, which means that only half of the money used by advertisers ends up with the owners of advertising spaces, i.e. the content publishers and media that have a primary relationship with the consumer. In the digital advertising value chains, content producers lose money to countless unknown actors.

In addition to media, the losing side also includes advertisers who find it difficult to reliably verify the input-output ratio of their digital ads. From the advertiser's point of view, a significant portion of money is lost even before the brand's message reaches the desired medium, and up to 15 per cent of the total input goes to entirely untraceable entities. This is a clear indication that the market for systematic or automated advertising has become too complex. The problem is particularly difficult because today 90 per cent of digital advertising is carried out programmatically.

Advertising companies could reflect on their own marketing practices and challenge both their own employees and the agencies used in marketing to discover new, more sustainable operating models. One option could be to further develop contextual digital advertising, where advertisements are offered, for example, on the basis of a search engine search or web page content. Every digital advertising campaign or measure should be assessed for the entire consent chain of the individual, not just for the advertising company or the primary service providing the advertising space. Companies should also make much more effective use of their relationship with the customer and seek reciprocity rather than remote monitoring. Some companies are active on social media and are therefore directly connected to their customers, but the data cannot be effectively put to work for the company. The data generated by a customer information system, website and applications provides a unique opportunity for the company. It offers an additional opportunity for in-depth and continuous interaction with customers and a way of developing services.

Digital advertising companies as well as the media and advertisers that depend on them should launch a broad assessment of industry-wide operating models and look for future options. Primary targets for development could include increasing the transparency of supply chains for businesses and consumers and seeking common cross-industry approaches. Sufficiently ambitious self-regulation could potentially prevent public authorities and regulatory bodies from employing increasingly stringent approaches and help to reorganise the industry as a whole.

11 In future, successful services are based on trust

LET US IMAGINE A MOMENT IN OUR DIGITAL EVERYDAY LIFE IN FIVE YEARS' TIME.

"You have recently felt like you have no energy and realised that you need to change your lifestyle. Your friend has told you about the new Live Better application. As you get to know the service, you notice that it has a Fair™ label. Now you know that the service uses your personal data ethically and in a secure manner. You give the service permission to combine your local store's shopping data, the health data in your national digital health service and data collected by Sports Tracker. When making a payment, you notice that your insurance company offers you the service for free. It is the year 2025, and the world has a well-functioning data economy."

This is what we are working on in the fair data economy project. The point is to make life easy and comfortable and to take advantage of the opportunities of digitalisation without compromising our privacy. The old model of the platform economy must be able to move towards decentralised and transparent business ecosystems that are based on a functioning data market and consumer trust. The values of individuals shape the operating environment of businesses.

The value base of the fair data economy is determined by the following statements::

- Trust in digital services is based on the ability to have an influence (self-determination) and on a sense of control (transparency) in the use of personal data
- The starting point for a human-oriented, fair data economy is the ability of the individual to influence the utilisation of their data as part of a dynamic digital ecosystem
- In fair data ecosystems, the sharing of data between organisations is also determined by common rules and transparency. The fair data economy creates value for everyone.

The Sitra Act defines Sitra's objective as follows:
Sitra aims to promote the stable and balanced development of Finland, the quantitative and qualitative growth of the economy and international competitiveness and cooperation, in particular by working to realise projects that contribute to improving the efficiency of the use of national resources or to raising the level of research and education or that explore future development options.

Making the most of the data economy is essential for Finland's competitiveness, and we firmly believe that this is where Finland can set an example. We can create successful services that are based on trust. The new data market is equal for companies of all sizes. In the data market, data is shared seamlessly, transparently and with permission between different actors. Everyone benefits from the fair data economy. Individuals will receive more tailored services, businesses will grow due to innovations and society will become more prosperous. The fair data economy provides Europe with a

competitive advantage in the global data economy market.

We can take control of the future by developing both business capabilities and technology. This requires the creation of new business ecosystems, innovative business models and new types of services. The provision of services also requires that data be taken as an active production factor.

The fair data economy requires cooperation not only between decision-makers but also between companies and NGOs. The forerunners must be able to inspire and convince others. At Sitra, we will continue to produce new information, aim to recognise challenges and spread information about the ways in which the challenges can be solved. We want to stimulate social debate, spread awareness, and bring different actors together. Practical experiments and pilots that test new operating models play a significant role. The dissemination and consolidation of new operating models requires that individuals and companies genuinely benefit from them and that the operating environment is favourable to such approaches. Setting a political and administrative foundation for change is one of Sitra's basic methods to achieve this, but in shaping the fair data economy individuals assume a crucial role as the agents of change.

Sources

Brynjolfsson, E., Collisa A., Eggersc, F. 2019. Using massive online choice experiments to measure changes in well-being. Sloan School of Management, Massachusetts Institute of Technology.

Datatilsynet Norge 2015. The Great Data Race. How commercial utilisation of personal data challenges privacy.

Englehardt, S., Narayanan, A. 2016. Online Tracking: a 1-million-site Measurement and Analysis, Princeton University.

Interactive Advertising Bureau IAB, Europe 2020. A Guide to the Post Third-Party Cookie Era.

ISBA/PwC 2020. Programmatic Supply Chain Transparency Study.

Lehtiniemi, T., Kortesniemi, Y. 2017. Can the obstacles to privacy self-management be over-come? Exploring the consent intermediary approach.

OECD Policy Note, 2020. Growing Up Online. Addressing the Needs of Children in the Digital Environment

Sitra 2019. The use of digital services report.

Sitra 2019. The future of European companies in data economy.

Waddell, K. 2020. Some developers don't know what their apps do with your data. Here's why most apps use off the shelf code and some of it can be risky.

World Economic Forum, Suga 2020. How to restore trust in data.

World Federation of Advertisers. Survey: Data ethics (2020)

Other useful links

Forbrukerrådet Norge 2020. Out of Control - How consumers are exploited by the online advertising industry 2020.

Sitra 2019. Rulebook for a fair data economy.

Sitra's Working paper 2020. 35 Proposals to make the European data strategy to work.

Glossary

ADTECH (ADVERTISING TECHNOLOGY): AdTech refers to advertising-related technology and is defined very broadly. Generally speaking, AdTech includes both digital tools and analytics, but in discussions it is often used to refer to a complex ecosystem with various actors who use data for, for instance, efficient targeted advertising.

DATA BROKER: A company whose business is based on the collection and aggregation of data and selling the resulting data products.

DATA ECONOMY: The data economy refers to a part of the economy, the business model of which is based on the diverse use of data.

GDPR: Regulation (EU) 2016/679 or the new General Data Protection Regulation (GDPR) of the European Union regulates the processing of personal data by an individual, company or organisation in the EU. The regulation is highly important for strengthening the basic rights of individuals and facilitating business by clarifying the rules that apply to companies and public actors in the digital internal market. The regulation entered into force on 24 May 2016, and it has been applied from 25 May 2018.

MOBILE APPLICATION: A piece of software that has been designed to work on mobile devices, such as mobile phones, tablets or smart watches. Mobile applications are created for countless purposes ranging from news and games to stock exchange and image processing.

MONITORING APPLICATION: An application that was built for the Digitrail survey project and installed on the test mobiles of the test subjects. It was used to track the services used by these subjects.

MYDATA: MyData is a principle applied to the management and processing of personal data which says that people must have the possibility to manage, utilise and hand over the personal data that is collected about them (e.g. call detail records, health data, including information on the individual's genetic heritage, energy information, purchase data, location data, financial data and data stored in online services).

PERSONAL DATA: All data that can be used to identify and specify a person. This data includes names, addresses, email addresses, personal identification numbers and other identifiers, including online identifiers.

PROGRAMMATIC ADVERTISING: An automatic, data-driven and algorithm-based means of buying and selling advertisements which usually takes the form of a digital auction.

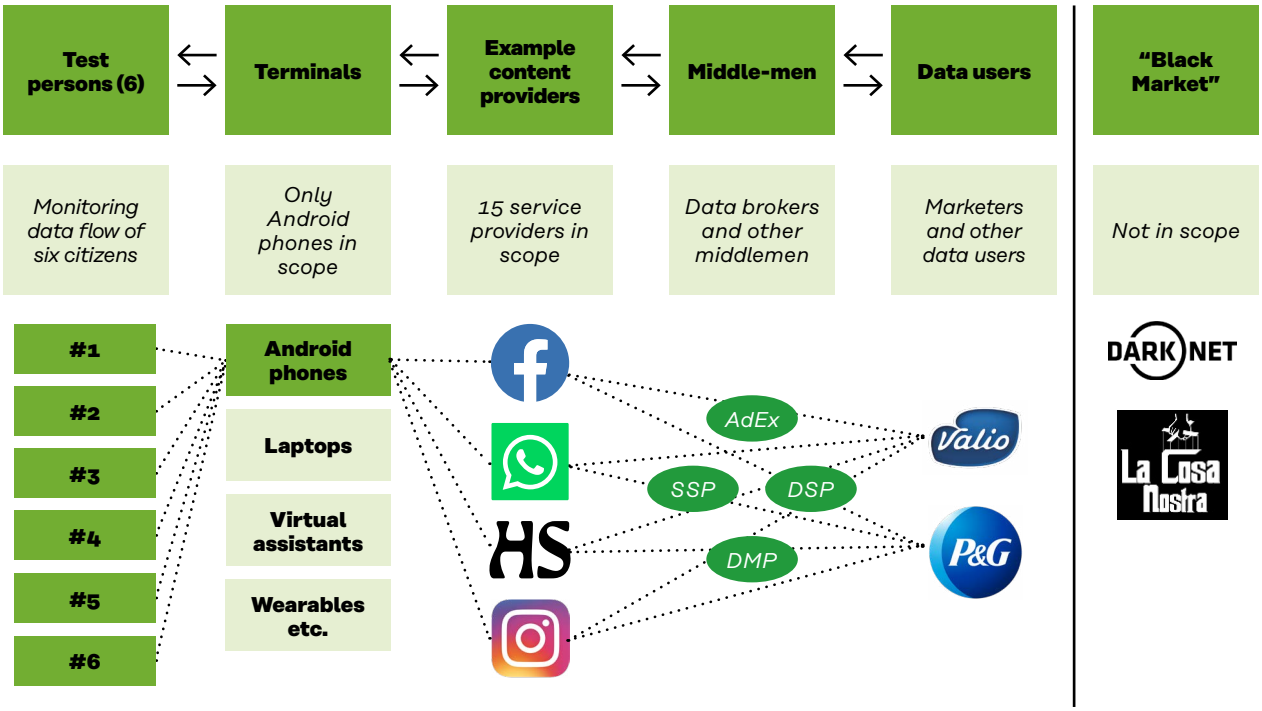
THIRD-PARTY DATA COMPANIES: Companies that, for instance, collect, aggregate, enrich and sell people-related data but that do not have a direct connection to the consumer. For example, a web page used by a consumer is the first party, while the third party data company sells advertisements on the page.

VPN (VIRTUAL PRIVATE NETWORK): In this project, the test subjects' data was directed through a protected VPN connection to a cloud-based server from where it was then stored.

Appendix 1. The scope of the report and the data tracking method

The data economy ecosystem related to personal data is vast and complex. In addition to actual data collectors, there are a number of different parties and, due to grey areas, describing the entire ecosystem is very challenging.

FIGURE 1. THE ECOSYSTEM OF DIGITAL ADVERTISING IS HIGHLY COMPLEX



The flow of data was tracked for both first party and third-party actors.

In this report, the various third parties were identified using a free-of-charge database provided by the WebXRay tool. The database, created and used by researchers, enables the identification and categorisation of the most common online third parties related to advertising, marketing and profiling. It needs to be noted, however, that the data collection-related ecosystem is in a constant state of flux, so even this tool may not recognise all actors. Therefore, some of the services were identified manually.

Data tracking method

TEST MOBILES

All test subjects used a similar Android phone. The test subjects were instructed to use the most sensitive services, such as bank services, on devices other than their test telephone.

The test required Android mobiles due to the monitoring application that tracked the programs the subjects were using. Adding this application to an Apple iPhone was technically

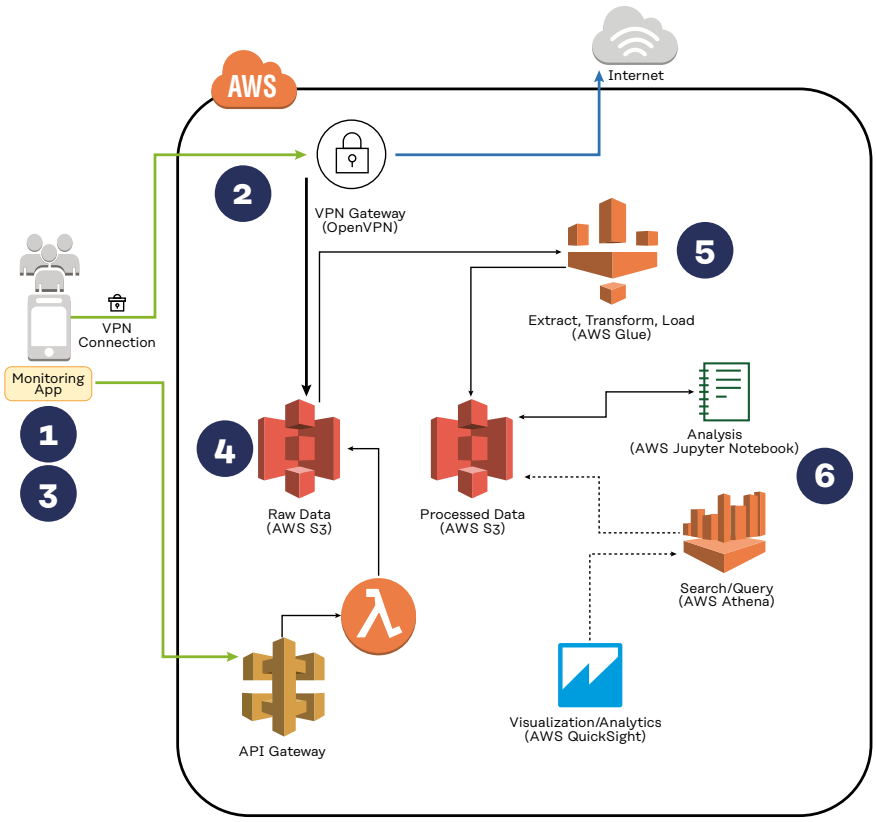
unfeasible. The make of the mobiles was Motorola Moto e5 Plus. Using a clean Android enabled the rooting of the mobiles, which in turn enabled the changing of the proxy settings. On Android mobiles and tablets it is possible to enable application developer settings, an additional settings menu, which allows the user to access settings beyond the basic settings (rooting). This makes it possible to, for instance, remove the telephone manufacturer's own applications and other similar operations, which might not otherwise be possible due to the manufacturers' own software security.

MONITORING APPLICATION

During the analysis phase it was discovered that there were certain limitations related to the data that the monitoring application collected. The program recognised the individual active applications. However, the majority of the traffic came from the Google Chrome browser, due particularly to the fact that the test subjects were encouraged to use the different applications through the browser to gather data from the most well-protected applications, such as Facebook. The monitoring program was unable to process the Chrome data to specify, for example, the website to which each of the traffic packet belonged. In addition, part of the online traffic was related to the applications running in the background, and the monitoring program could not distinguish these from the data of the active application. In practice, most of the online traffic analysis is based on traffic data and its analysis, as described in the next paragraph.

INFRASTRUCTURE RELATED TO THE COLLECTION OF TRAFFIC DATA

FIGURE 2. THE MOBILES WERE EQUIPPED WITH A MONITORING APPLICATION AND A VPN APPLICATION, IN ADDITION TO WHICH THEIR PROXY SETTINGS WERE CHANGED TO BY-PASS THE SSL ENCRYPTION OF ONLINE TRAFFIC



The mobiles were equipped with a monitoring application and a VPN application, in addition to which their proxy settings were changed to by-pass the SSL encryption of online traffic (1). A VPN application was used to direct the traffic data via a VPN tunnel (2) and through an Open VPN server. The Open VPN server was located in the AWS cloud service. On the Open VPN server, the traffic data was stored in the S3 environment of the AWS (4). The storing did not interfere with the data traffic itself. Instead, after the Open VPN server, the traffic continued unchanged to its online destination.

The data of the monitoring program was also stored in the S3 service. In addition to these, the analysis used external materials, such as the WebXRay database to provide information on the various third parties of the data economy. This data was stored in the S3 service as well. These data sets were combined with the traffic data, and the resulting dataset was cleaned (5) and store again for the data analysis (6).

FORMAT OF THE TRAFFIC DATA

Online traffic data consists of HTTP queries. The format of the queries is always the same: the device of the user sends a request to an IP address and receives a response from the destination server. These are also called packets. Web traffic uses two HTTP query versions: the older HTTP/1.1 version and the more recent HTTP/2.0. The protocols differ in some respects, and they determine the framework for data transfers and query-related data and metadata. The packets themselves can be very dissimilar and contain highly varying data. The contents of the packets may involve various JavaScript programs that can be used to transfer data. In this report, the most important data delivered in the online traffic includes

1. packet authority/host, which indicates the domain address to which the packet has been sent
2. packet data type, i.e. the type of data delivered
3. packet queryString
4. packet data, although it is usually encrypted/coded
5. cookies included in the packet
6. reference information, i.e. whether some other service has asked to send data to a server other than its own (this information is not included in every packet).

DATA COLLECTION PERIOD AND DATA COVERAGE

The report analysed the test subjects' online traffic data that was collected in the course of two weeks. The material included approximately 84,000 HTTP request/response packets. The number of packets ranged from 9,000 to 21,000 between individuals.

DELIVERY OF THE TEST SUBJECTS' DATA TO THIRD PARTIES

The situation of each test subject was assessed from two viewpoints.

1) One to three applications or websites per test subject were analysed. The traffic that occurred during the use of these services was examined for third parties, which were identified with the help of the WebXRay database as well as manually. The services chosen were as diverse as possible in order to show the variety of third parties used by different applications. The services typically used several third-party data companies, and these actors will be presented in more detail below.

2) Each person's data was scanned for advertising-related actors using the WebXRay database. The aim was to discover which websites or applications used these third-party

companies. Here, the examination focussed on advertisers, since their role is often the most important.

Traditionally, advertising-related actors can both buy target audience data from other actors and collect data on the users themselves. The data is collected and shared between the different actors by installing small programs (scripts) or so-called third-party cookies on the user's device. They enable the advertising-related actors to recognise the same users on different web pages and to compile profiles according to their behaviour. This report sought to find out the pages on which the advertising-related actors had placed their cookies.

GOOGLE AND FACEBOOK

Google and Facebook are technology giants and super platforms that are able to collect data on users through both their own services and those of other websites and that use this data to sell advertisements. Facebook, for instance, collects information through its Like buttons and cookies. In 2019, Facebook had 2.4 billion monthly users, and 1.5 billion people used Google's Gmail electronic mail. They do not share their user data with advertisers directly but use the data to create more or less accurate target groups – such as football fans, people living in a certain part of town or middle-aged women – which they then offer to their advertisers.

Facebook and Google traditionally collect information on their users using third party cookies. However, the collection of user data is changing. GDPR requires that the user be asked consent before installing any third-party cookies. The Safari and Firefox browsers now block third party cookies by default and even restrict the storage time of first party cookies. According to Google, the company's Chrome browser will provide the users with new opportunities to manage third party cookies in the future.

To get around this issue, Facebook and Google have begun to collect data through pixels and cookies installed by web sites, i.e. the first parties, themselves. This is why the traffic data was first examined for Google's and Facebook's direct third-party cookies. In addition, the examination sought to find out when Google or Facebook installed cookies as part of the first party cookies. These cookies were identified based on the names of the cookies, which included the identifier "_fbp" and "_fbc" for Facebook and "__gads" and "__gac" for Google. It is unlikely that other services would use these names in their own cookies.

WHAT EXACTLY IS INSIDE ONLINE TRAFFIC PACKETS?

Even though the topmost encryption layer of online traffic was decrypted using the Virtual Private Network (VPN) tunnel to see inside the packets, the data they contained was also encrypted in several ways. Companies want to keep both the business data contained in the data (capital of the data economy companies) and user-related information to themselves to stop it from reaching the hands of competitors or other prying eyes. This report let us discover the web addresses to which the data was delivered, the amount and format of the data delivered, and whether the HTTP requests had a "sender site". However, we do not know in detail what the data sent included.

Appendix 2. Examples of the third-party data companies discovered during the test subjects' use of digital services

ADFORM is one of the largest advertising technology-related companies. AdForm is not American, like the rest of the major actors in the field, but Danish.

APPNEXUS (currently Xandr) is a gigantic advertising technology company owned by the large-scale American company AT&T. It provides a portal through which individual users can check the segments into which they have been categorised in the AppNexuksen data.

The packets sent to the magazines of the **CondeNast group**, such as **vogue.com** and **wired.com**, contained one parameter, the same identifier. The packet that was received in response stored a cookie of these third-party magazines on the user's telephone.

CHARBEAT analysed the activities of the application users. According to the Privacy Statement, the final octet is removed from the IP address before storing the data, meaning it cannot be connected to the users. In addition, the company states that it does not store any other information that could be used to identify the user, either.

Both **DOUBLECLICK AND ADSENSE** are companies that are linked to Google advertising. Google's Privacy Statement acts as the privacy statement of both companies. The pages of AdSense mention that the websites using the service should let their users know that "third party vendors, including Google, use cookies to serve ads based on user's prior visits to your website or other websites". The wording is very weak, and it aims to highlight the fact that the service only forms a part one of many actors.

FEEDBAKLY is a Finnish company which collects data on the web page and optimises customer journeys.

FLURRY ANALYTICS is an analytics service which claims to be designed for application optimisation. The website makes it difficult to understand what person-related data it uses and how the data is aggregated. In addition, the Privacy Statement link directs to the Privacy Statement policy of Verizon Media. Verizon Media is part of the gigantic Verizon Communications company, which owns, for instance, Yahoo. The Privacy Statement linked does not specify the data used by Flurry Analytics, and it says that the data is shared between Verizon media and "trusted partners".

HOTJAR is a tool used for analysing user paths on a website. It can used to gauge clicks or page scrolls. The tool can also give feedback on how to improve a page. According to the Privacy Policy of the service, the users' personal data will not be sold to anyone.

MOAT analyses the impact of advertising as well as how and where the advertisements have been displayed (attention analytics). It is part of Oracle Cloud. Oracle's Privacy Policy contains many links and it is not easy to find any mention of how the Moat data is used.

TAPJOY is a service related to the selling of advertising. The traffic data collected reveals that it utilises a so-called tracking ID, which is a device- and user-specific identifier. This enables TapJoy to aggregate data from different services.

Appendix 3. Request for information sent to service providers

TO WHOM IT MAY CONCERN:

I am hereby requesting access and portability of my personal data in accordance with both Articles 15 GDPR and Article 22 GDPR. Additionally, I am also using specific provisions in the GDPR to ask for more information on the processing of my personal data.

Please note that it is not legal to require data subjects to use an in-house form (see, for instance, UK Information Commissioner's Office: 'Subject Access Code of Practice' (9 June 2017) p 13 and Information Commissioner's Office: 'Guide to the GDPR: Right to access' (22 May 2019), stating that 'even if you have a form, you should note that a subject access request is valid if it is submitted by any means, so you will still need to comply with any requests you receive in a letter, a standard email or verbally […] although you may invite individuals to use a form, you must make it clear that it is not compulsory').

A. UNDER THE SCOPE OF ARTICLE 15 GDPR:

1. Please confirm as to whether or not you are processing personal data concerning me.
2. Please provide me with a copy of all the personal data concerning me that you or one of your subprocessors holds and that falls under the scope of Article 15 GDPR, **excluding data that is included in the standard data request.** This includes for instance - solely for the purpose of providing here a nonexhaustive list of examples - any data derived about me, such as opinions, inferences, settings, segments, audiences and preferences. (Note that opinions, inferences and the like are considered personal data. See Case C-434/16 Peter Nowak v Data Protection Commissioner [2017] ECLI:EU:C:2017:994, 34.) For data that is available to the controller in machine readable format, it must be provided to me in that form in accordance with the principle of fairness and provision of data protection by design.
    a. Any information held by Facebook about my profile being added to Custom Audiences by advertisers, including IDs of the Custom audiences.
    b. Any browsing information tracked by Facebook on third party websites or apps, through tools such as Facebook SDK or Facebook Pixel.
    c. Any replacement ID, as presented in the 2012 Audit of Facebook by the Irish DPC, at 1.9.2.2.4 of http://www.europe-v-facebook.org/ODPC_Review.pdf
    d. Any trace of notifications sent to my devices.
3. Please confirm for how long each category of personal data is stored, or the criteria used to make this decision, in accordance with the storage limitation principle and Article 15(1)(d).
4. Any **third parties to whom data has been disclosed**, named with contact details in accordance with Article 15(1)(c). Please note that there is an explicit obligation to name recipients in third countries (Art 15(1)(c)) and that the European data protection regulators have stated that by default, controllers should name precise recipients and not "categories" of recipients. If they do choose to name categories, they must justify why this is fair, and be specific, naming "the type of recipient (i.e. by reference to the activities it carries

out), the industry, sector and sub-sector and the location of the recipients. (Article 29 Working Party, 'Guidelines on Transparency under Regulation 2016/679' WP260 rev.01, 11 April 2018) Please note that in the case of any transferred data processed on the basis of consent, there is no option to just name categories of recipients without invalidating that legal basis (Article 29 Working Party, 'Guidelines on Consent under Regulation 2016/679' (WP259 rev.01, 10 April 2018) 13). Make sure as well that it is possible from the data provided to match granular data points to the named recipients, for instance in cases where data dumps would use partner IDs to refer to recipients.

5. If any data was not collected, observed or inferred from me directly, please provide precise information about **the source of that data,** including the name and contact email of the data controller(s) in question ("from which source the personal data originate", Article 14(2)(f)/15(1)(g)). Make sure as well that it is possible from the data provided to match granular data points to the named sources, for instance in cases where data dumps would use partner IDs to refer to sources.
6. Please confirm where my personal data is physically stored (including backups) and at the very least **whether it has exited the EU at any stage (if so, please also detail the legal grounds and safeguards for such data transfers).**
7. Should you seek to restrict the scope of your response under Article 15(4), please provide me with a detailed assessment in writing of the balancing you have chosen, such as a Data Protection Impact Assessment.

B. UNDER ARTICLE 20 GDPR:

For data falling within the right to data portability (GDPR, art 20), which includes all data I have provided, **excluding data that is included in the standard data request**, and data which have been indirectly observed about me (Article 29 Working Party, Guidelines on the Right to Data Portability (WP 242), 13 December 2016, 8), and where lawful bases for processing include consent or contract, I wish to have that data:

1. **sent to me in commonly used, structured, machine-readable format**, such as a CSV file. A PDF is not a machine-readable format (Article 29 Working Party, 'Guidelines on Transparency under Regulation 2016/679' WP260 rev.01, 11 April 2018).
2. accompanied with an **intelligible description of all variables.** This would include a detailed listing of names of partners, if those are referred through identifiers in my personal data.

Note that the data falling under the scope of Article 20 will most likely partially overlap with the data obtained under Article 15, but the format of the result is more restricted under Article 20.

C. UNDER EITHER ARTICLES 13 OR 14 GDPR:

1. Please inform me of all **processing purposes and the lawful basis for those purposes by category of personal data.** This list must be broken down by purpose (Art 15(1)(a) GDPR), lawful basis aligned to purposes, and categories of data concerned aligned to purposes and lawful bases. Separate lists where these three factors do not correspond are not acceptable (Article 29 Working Party, 'Guidelines on Transparency under Regulation 2016/679' (WP260 rev.01, 11 April 2018), page 35.). A table may be the best way to display this information.

2. Please inform me of the **specified legitimate interest** where legitimate interest is relied upon (Article 14(2)(b)).

D. UNDER ARTICLE 22 GDPR:

1. Please confirm whether or not you make any automated decisions (within the meaning of Article 22, GDPR). If the answer is yes, please provide meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for me. (Article 15(1)(h))

E. UNDER ARTICLE 26 GDPR:

1. Please provide the **identity of all joint controllers** of my personal data, as well as the essence of you contracts with them (Article 26). Please note that the definition of "processing" (Art 4(2)) includes such operations as "disclosure by transmission, dissemination or otherwise making available". Therefore, in all cases where data is transmitted to another data controller who will process the data under their own purpose, the transfer itself must have been done under joint controllership provisions.

F. IN ACCORDANCE WITH ARTICLE 46:

1. Where relevant, please provide me with a list of applicable safeguards as well as information about the relevant data (as referred to in Article 46(2) GDPR).

G. IF YOUR ORGANISATION CONSIDERS ME A CONTROLLER FOR WHOM YOU PROCESS PERSONAL DATA

1. Furthermore, if your business considers me the controller of any personal data for which your business acts as processor, please provide me **with all the data you process on my behalf in machine readable format** in accordance with your obligation to respect my to determination of the means and purposes of processing.

I do understand that according to Article 11 GDPR, you might need additional information to identify me for the purpose of this request. I have included some of that information below. However, should this not be sufficient to identify me, please provide to me (in accordance with the principle of fairness and the second part of Art 11(2) GDPR), with a definite list of the required information in order to identify me.

Please note: You may only use the following information for the purposes of identifying me and responding to my request:

Name:	YOUR NAME
E-mail:	YOUR EMAIL
User name:	USER NAME OR SIMILAR IF USED

I ask you to provide the requested information to me without undue delay and in any event within one month. According to Article 15(3) GDPR, you have to answer this request without cost to me.

Yours sincerely,

YOUR NAME

SITRA

SITRA STUDIES 169
Sitra studies is a publication series which focuses on the conclusions and outcomes of Sitra's future-oriented work.

ISBN 978-952-347-179-5 (PDF) www.sitra.fi
ISSN 1796-7112 (PDF) www.sitra.fi

SITRA.FI
Itämerenkatu 11–13
PO Box 160
FI-00181 Helsinki
Tel +358 294 618 991
@SitraFund