

HENKILÖDATAN JÄLJILLÄ

Yksilöstä kertyvän tiedon kulku
ja käyttö digitaalisissa palveluissa

Riitta Vänskä ja Tiina Härkönen



© Sitra 2020

Sitran selvityksiä 168

**Henkilödatan jäljillä
– yksilöstä kertyvän tiedon kulku ja käyttö
digitaalisissa palveluissa**

Kirjoittajat: Asiantuntija Riitta Vänskä ja
johtava asiantuntija Tiina Härkönen Sitran
IHAN – Ihmislähtöinen datatalous -projektista.

Työryhmä: Tiina Härkönen, Maria Jalavisto,
Heli Parikka, Jaana Sinipuro ja Riitta Vänskä
Asiantuntijoina selvityksessä ovat toimineet
suomalainen ohjelmistoyritys Futurice sekä
matemaatikko Paul-Olivier Dehaye.

Toimitussihteeri: Kirsi Suomalainen, Sitra
Taitto: PunaMusta

ISBN 978-952-347-178-8 (PDF) www.sitra.fi
ISSN 1796-7112 (PDF) www.sitra.fi

SITRAN SELVITYKSIÄ -sarjassa julkaistaan
Sitran tulevaisuustyön ja kokeilujen tuloksia.

Sitran selvityksiä 168

Henkilödatan jäljillä – yksilöstä kertyvän tiedon kulku ja käyttö digitaalisissa palveluissa

Kesäkuu 2020

Sisällys

Esipuhe	2
Tiivistelmä	3
Sammanfattning	4
Summary	5
1 Johdanto	6
2 Ihmiset haluavat reilun datan palvelua ja yritykset reilua kilpailua	7
3 Digijälkiselvitys – näin yksilöstä kerätty data kulkee	9
4 Digijälkiselvityksen päähavainnot – läpinäkyvyys puuttuu, eikä tietosuojasetus riitä	23
5 Data liikkuu kolmansille osapuolille useita kanavia pitkin	25
6 Maksamme palveluista datalla, mutta miten käy yksityisyyden?	28
7 Digitaalisen mainonnan liiketoimintamallit kaipaavat tuuletusta	31
8 Suuria haasteita datapohjaisten kuluttajapalvelujen ja digitaalisen mainonnan toimijakentässä	33
9 Kuinka suojella yksityisyyttään – suosituksia jokaisen arkeen	35
10 Loikka reiluun datatalouteen – suosituksia yrityksille	37
11 Tulevaisuuden menestyjäpalvelut pohjautuvat luottamukseen	40
Lähteet	42
Sanasto	43
Liite 1. Selvityksen laajuus ja datan kulun seurantamenetelmä	45
Liite 2. Esimerkkejä testihenkilöiden digitaalisten palvelujen käytön yhteydessä löydettyistä kolmansista osapuolista	49
Liite 3. Palvelun tuottajille lähetetty kyselylomake	51

Esipuhe

Sitran reilun datatalouden IHAN-hanke käynnistettiin keväällä 2018. Samoihin aikoihin kaikissa EU-maissa alettiin soveltaa yleistä tietosuoja-asetusta, joka säätelee henkilötietojen käsittelyä. Asetus on maailman tiukin tietosuojaa koskeva sääntely.

Data on digitaalisesta toiminnasta syntyvä jälki, eikä sillä itsessään ole arvoa. Arvoa syntyy, kun dataa jalostetaan tiedoksi ja edelleen innovaatioiksi tai sen avulla kehitetään toimintaa. Datatalous puolestaan on talouden osa-alue, jossa liiketoimintamalli perustuu datan hyödyntämiseen ja käyttöön eri tavoin.

Monille meistä datatalous merkitsee hyviä ja ilmaisia palveluita ja olemme tyytyväisiä kohdennettujen mainosten aikaansaamaan ajansäästöön. Alustatalouden jättiläisten ympärille rakentunut digitaalinen mainonta on kuitenkin yksi ensimmäisistä datatalouden alueista, jonka lieveilmiöihin ja taustalla oleviin liiketoimintamalleihin on alettu kiinnittää yhä enemmän huomiota. Myös toimiala itse on huolestunut asenneilmapiirin ja lainsäätäjien aiheuttamasta paineesta. Pitkään menestynyt toimiala on muutoksessa ja edessä on toimintamallien uudistaminen.

Näemme Sitrassa datatalouden paitsi talouden osa-alueena myös ilmiönä, joka vasta hakee muotoaan. Ilmiö on aistein havaittava, mielenkiintoinen ja toistuva yhteiskunnan tapahtuma. Jotta ilmiöön voi tarttua, se tulee kuvata. Digitaalinen arkemme jättää monenlaisia jälkiä, joita olemme pyrkineet avaamaan meille kaikille tuttujen digitaalisten palveluiden avulla. Reilu datatalous on Sitran luoma käsite, jolla pyrimme kuvaamaan ilmiön tavoitetilaa. Reilu datatalous keskittyy luomaan palveluja ja dataan perustuvia tuotteita eettisesti. Se tuottaa arvoa kaikille: ihmisille, yrityksille ja yhteiskunnalle.

Tämän selvityksen tarkoituksena on analysoida ilmiötä sellaisena, kuin se näyttäytyy arjessamme erilaisten sovellusten kautta. Helppokäyttöisen palvelun taustalla on asioita, joita emme normaalisti näe tai joista emme ole edes kiinnostuneita. Haluamme valottaa yksilön käyttämien palvelujen taustalla olevien yritysten toimintaa ja taustalla vaikuttavia, teknisiä ja liiketaloudellisia, tekijöitä.

Innoittajana tähän tutkimukseen toimi Tietosuojavaltuutetun toimisto. MyData-yhteisön kautta taas löytyi osaja, jota ilman analyysit olisivat jääneet pintapuolisiksi.

Uskomme, että sanoittamalla vaikeasti ymmärrettäviä asioita ja hyödyntämällä tietosuoja-asetuksen meille antamia keinoja, pystymme muokkaamaan tulevaisuutta haluamaamme suuntaan. Tietosuoja-asetusta edelleen kehittämällä voimme mahdollistaa uusia, nykyistä kestävämpiä liiketoimintamalleja.

Reilun datatalouden rakentaminen on vasta alkumetreillä, ja onnistuu vain monipuolisen, yhteiskunnallisen keskustelun avulla.

JAANA SINIPURO

Projektijohtaja

Sitra

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. Verkkokäyttäjyymisestä 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 tarjoihtuun 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 liikehinnan takia. Vastuu yksityisyyden säilyttämisestä on tähän asti säilytetty 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 nykkykoneiston 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 plattformsekonomin 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. Ansvar 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.

Summary

According to a survey carried out by 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 capability to distinguish companies that use data sustainably from those who do not. A separate survey of companies revealed the business perspectives on the data economy and highlighted serious concerns among European companies about their competitive positions in the data economy in relation to major American and Chinese corporations.

Following on from these surveys, in the second half of 2019 Sitra carried out a trial with six individuals in Finland to investigate the flow of each person's data online. The Digitrail survey studied where people's data travels when they visit websites or log in to use 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 is holding the data. 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 created by companies working in and around digital advertising without the knowledge of consumers, and despite the large amount of data collected, they do not provide a true picture of the individual, although they influence the information offered to the individual. The General Data Protection Regulation (GDPR) only allows an individual 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 any information on the spread and exploitation of the data. As such, this online transaction cannot be considered fair.

The digital advertising business models that have emerged around the giants in the platform economy have been built in a fundamentally problematic way in terms of privacy. There is a very limited window of opportunity for users to assess the impact of giving consent for their data to be used when asked to do so upon visiting a site or downloading an application.

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 undergoing a transformation as a result of pressure from consumers and legislation, as well as movements within the field. So far, individuals have been responsible for guarding their privacy, and 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. As regards services used by children and young people, this presents a unique 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 a part of the fair data economy, which will create well-being for all involved.

1 Johdanto

Kiinnostus yksilöstä kerätyn datan liikkeisiin alustajattien palveluissa ja digitaalisen mainonnan verkostoissa sai alkunsa sarjasta kansainvälistä näkyvyyttä saaneita skandaaleja, joissa henkilöistä kerättyä dataa oli joko tietoisesti käytetty väärin, vuodettu ulos ihmisten tietämättä tai menetetty puutteellisen tietoturvan seurauksena. Näkyvin niistä oli vuoden 2018 Cambridge Analytica -skandaali, jossa miljoonien ihmisten Facebook-tietojen väitetään päätyneen vaalikampanjakäyttöön.

Sitran IHAN-hankkeessa selvitettiin tapahtumien vaikutusta kansalaisten asenteisiin datatalouden toimijoita ja palveluja kohtaan sekä yksilöiden ymmärryksen ja osaamisen tasoa datapohjaisista palveluista. Selvitys toteutettiin kyselytutkimuksena neljässä Euroopan maassa syksyllä 2018. Kysely paljasti selkeästi ihmisten huolen yksityisyydestään ja luottamuspuolan digitaalisten palvelujen tarjoajia kohtaan.

Datatalous koskettaa koko yhteiskuntaa, ja sen osapuolina ovat niin yksilöt, yritykset kuin muutkin organisaatiot. Sitra selvitti eurooppalaisten suur- ja pk-yritysten tietoisuutta, suhtautumista ja sitoutumista reilun datatalouden tarjoamiin liiketoimintamahdollisuuksiin keväällä 2019. Kysely osoitti, että lähes kolmasosa yrityksistä koki haasteena eurooppalaisten palvelujen syntyemiselle amerikkalaisten ja kiinalaisten toimijoiden kilpailun omilla säännöillään, ja yli

viidesosa näki merkittävänä haasteena yleisen tietosuojasetuksen (GDPR) asettamat vaatimukset ja muun vastaavan sääntelyn.

Kansalais- ja yritystutkimusten mukaan eniten dataa keräävien yritysten ja kuluttajien sekä eurooppalaisten yritysten edut ovat osittain ristiriitaisia. Merkittävä osa kyselyyn vastanneista eurooppalaistryksistä koki yksilöitä suojelevan lainsäädännön ongelmalliseksi ja pelikentän epäoikeudenmukaiseksi.

Mutta keitä ovat yksilödataa massoittain keräävät tahot, kuinka ne toimivat ja mitä datalla tehdään? Näihin kysymyksiin pyritään vastaamaan digijälkiselvityksessä. Se on kurkistus yksilöiden näkymättömissä kulkevan datan virtoihin. Selvityksen toteuttamiseen tarvittiin teknisiä ratkaisuja, rohkeita kansalaisia sekä tahto perehtyä monimutkaiseen ja läpinäkymättömään yksilödatan markkinaan.

Kuuden testihenkilön avulla toteutettu selvitys kertoo yksilöistä kerätyn datan liikkeistä valtavissa mainonnan ekosysteemeissä, markkinan erilaisista pelureista sekä siitä, kuinka hyvin kuluttajille annettu tieto datan käytöstä vastaa lainsäädäntöä ja todellista elämää. Asiantuntijoina selvityksessä olivat suomalainen ohjelmistoyritys Futurice sekä matemaatikko **Paul-Olivier Dehaye**, joka on yksi Cambridge Analytica -skandaalin paljastajista.

2 Ihmiset haluavat reilun datan palvelua ja yritykset reilua kilpailua

Ihmiset toivovat läpinäkyvyyttä yksilöistä kerättävän datan käytölle ja kykyä erottaa kestävä datankäyttö yritykset muista. Yrityksille suunnattu kysely valotti liiketoiminnan näkemyksiä datataloudesta ja toi esiin eurooppalaisten yritysten huolen datatalouden kilpailuasetelmasta.

Datatalouden kansalaiskysely paljasti luottamuspulan

Digitaalisten palvelujen käyttö -kyselyssä (Alankomaat, Ranska, Saksa ja Suomi) kartoitettiin ihmisten tietoisuutta datataloudesta, asenteita digitaalisia palveluntarjoajia kohtaan sekä aktiviteetteja oman datan suojaamiseksi. Kyselyyn vastasi 2000 henkilöä kustakin maasta.

Kansalaiskyselyn tärkeimmät havainnot voidaan tiivistää seuraavasti

- EU:n yleisen tietosuojasetuksen (GDPR) antamien oikeuksien käyttäminen on vasta alkuvaiheessa.
- Luottamuksen puute palveluntarjoajia kohtaan on pullonkaula.
- Tietovuodoilla on ollut vaikutusta ihmisten käyttäytymiseen.
- Kansalaiset toivovat läpinäkyvyyttä datan kululle ja käytölle.
- Reilut palvelut pitää voida tunnistaa.

Kysely osoitti, että vain pieni joukko suojaa omaa dataansa tai käyttää tietosuojasetuksen antamia oikeuksiaan. Yhdeksän prosenttia kyselyyn vastanneista oli pyytänyt

pääsyn palveluntarjoajan itsestään keräämiin tietoihin. Vastaajista 15 prosenttia oli lopettanut joidenkin palvelujen käytön tietovuotoihin liittyvien uutisten takia, kun taas 40 prosenttia kertoi luottamuksen puutteen estävän heitä käyttämästä digitaalisia palveluja. Kyselyn mukaan tärkein luottamusta lisäävä tekijä digitaalisissa palveluissa on datan käytön läpinäkyvyys. Vaikka luottamus on laskussa, vain 14 prosenttia vastaajista sanoi lukevansa palvelujen ja sovellusten käyttöehdot huolellisesti. Kysyttäessä tulevaisuuden mahdollisuuksista kyselyyn vastanneista 66 prosenttia koki digitaalisten palvelujen ”reilun datan merkin” erittäin tärkeäksi tai tärkeäksi (suomalaisista vastaajista 71 prosenttia).

Koska kansalaisten luottamuspula kohdistui datataloudessa toimiviin yrityksiin ja niiden tarjoamiin digitaalisiin palveluihin, syntyi tarve selvittää myös yrityskenttää. Tarkoituksena oli saada tietoa yritysten näkemyksistä sekä löytää kansalaisten ja yritysten tarpeita yhdistäviä tai erottavia tekijöitä.

Datatalouden yrityskysely toi esiin huolen kilpailuasetelmasta

REILUN DATATALOUDEN MALLISSA

- yritykset saavat käyttöönsä henkilödataa ihmisten suostumuksella ja vaihtavat sitä toistensa kanssa yhteisten sopimusten pohjalta
- ihmiset saavat dataansa vastaan erityisesti heille suunniteltuja palveluja
- yksilöistä kerätty data edustaa vain pientä osaa datatalouden kokonaisuudesta, sillä yritykset ostavat reiluja datatuotteita ja jakavat yhteistyössä myös paljon muuta dataa.

Sitra selvitti kyselytutkimuksella *Eurooppalaisten yritysten tulevaisuus datataloudessa* myös yritysten tietoisuutta, suhtautumista ja sitoutumista ”reilun datatalouden” mallin tarjoamiin liiketoimintamahdollisuuksiin neljässä EU-maassa. Reilun datatalouden kerrottiin tarkoittavan taloutta, jossa eri markkinatoimijat työskentelevät yhteisessä ympäristössä datan saannin ja käytettävyyden varmistamiseksi. Ne hyödyntävät dataa yhdessä ja luovat sen pohjalta uusi sovelluksia ja palveluja. Reilu datatalous vaatii dataverkostojen toimijoilta läpinäkyvää, yhteisiin sääntöihin perustuvaa datanvaihtoa muiden verkoston toimijoiden kanssa sekä ihmisen antamaa suostumusta datan keräämiseen ja käyttöön, silloin kun kyseessä on yksilödata.

Kysely toteutettiin keväällä 2019, ja sen tulokset perustuvat 1667 vastaukseen. Koh-

deryhmän muodostivat suuret sekä pienet ja keskisuuret yritykset Alankomaissa, Ranskassa, Saksassa ja Suomessa. Alle 10 henkilöä työllistävät yritykset eivät olleet mukana.

Yrityskyselytutkimuksen tärkeimmät havainnot

- Yrityksissä suhtauduttiin reilun datatalouden toimintaperiaatteisiin yleisellä tasolla melko positiivisesti, mutta sitoutuminen yksilön yksityisyyden kunnioittamiseen jopa asiakaskokemuksen kustannuksella koettiin haasteelliseksi.
 - Selkeimmän strategisen haasteen muodosti se, että vain 15 prosenttia vastaajista piti datan jakamista muiden tahojen kanssa hyvänä asiana.
 - Datatalouden tarjoamat mahdollisuudet hahmotettiin jo hyvin (kolmasosa vastaajista kertoo luoneensa siitä kilpailuetua), mutta digitaalisten liiketoimintamallien ymmärrys oli vielä alussa tai jäsentymättöä.
 - Yritykset näkivät reilun datatalouden vaatiman lainsäädännön (tietosuoja-asetus jne.) osin esteenä palvelujen syntymiselle.
 - Toisaalta yritykset, jotka panostivat merkittävästi tietosuoja-asetuksen toteuttamiseen käytännössä, saivat siitä myös selkeitä hyötyjä, koska se auttoi ymmärtämään yrityksen omia datavarantoja.
- Datatalouden markkinaa ei pidetty oikeudenmukaisena pelikenttänä. Yrityksissä (31 prosenttia vastaajista) koettiin, että teknologiajätit pelaavat omilla säännöillään. Kilpailu amerikkalaisten ja kiinalaisten yritysten kanssa nähtiin joko suurimmaksi (Alankomaat, Ranska ja Saksa) tai toiseksi suurimmaksi (Suomi) haasteeksi.

3 Digijälkiselvitys – näin yksilöstä kerätty data kulkee

Sitra selvitti yksilöstä kerätyn datan kulkua verkossa kuuden suomalaisen testihenkilön avulla loppuvuodesta 2019. Tavoitteena oli selvittää, minne testihenkilöiden dataa kulkee, kun he vierailevat verkkosivuilla tai käyttävät kirjautuneina digitaalisia palveluja.

Datan kerääjät toimivat eri rooleissa

Datan avulla verkkosivujen toimintaa ja digitaalisia palveluja voidaan optimoida niin, että niillä vierailu on käyttäjälle miellyttävää. Dataa keräämällä palvelu saadaan esimerkiksi muistamaan kävijä, jotta samoja kysymyksiä ei tarvitsi kysyä aina uudelleen tai vaikkapa kirjautua palveluun joka kerta erikseen.

Kohdennetun mainonnan lisäksi on lukuisia muita tärkeitä syitä kerätä ja hyödyntää dataa käyttäjistä. Verkkosivujen ja palvelujen kehittäminen kaiken kaikkiaan tapahtuu datan avulla. Kun ihmisten käyttäytymistä verkkopalvelussa seurataan, nähdään mikä toimii ja mikä taas ei. Data auttaa myös löytämään ongelmia palveluissa ja ratkaisemaan ne.

Verkkosivuja ja sovelluksia tekevät yritykset mahdollistavat ratkaisullaan datan keräämisen yritysasiakkailleen, jotta niillä puolestaan olisi mahdollisuus analysoida dataa ja saada tietoa omien nykyisten tai tulevien asiakkaidensa tarpeista. Yritykset käyttävät monipuolisia data-analytiikkatyövälineitä löytääkseen uusia tapoja palvella asiakkaitaan tai luoda uusia tuotteita. Ne hyödyntävät myös datan analysointiin erikoistuneita muita yrityksiä, jos niiden omat resurssit ovat riittämättömiä tai oikeaa osaamista ei löydy omasta organisaatiosta.

Kaikki datan kerääminen ei ole ongelmallista. Se on välttämättömän tärkeää, sillä ilman sitä nykyisen kaltaisia, edistyneitä, verkkopalveluja ei olisi. Digijälkiselvityksessä perehdytään joihinkin datan keräämisen ongelma-alueisiin ja tarkastellaan alustatoulouden suurimpia yrityksiä ja digitaalisen mainonnan toimijoita yksityiskohtaisemmin.

Seuranta käyntiin: kuusi testaajaa ja kuusi puhelinta

Selvityksessä seurattiin kahden viikon ajan testihenkilön käyttämän testipuhelimen kautta kulkevaa dataliikennettä. Testihenkilöiksi valittiin joukko eri-ikäisiä internetin palvelujen käyttäjiä, joilla olisi erilainen elämäntilanne. Heidän käyttämänsä palvelut olivat sekä Suomesta että muista maista. Palveluja tarjoavia yrityksiä pyydettiin myös vastaamaan testihenkilöiden tekemiin kirjallisiin, tietosuoja-asetuksen mukaisiin kysymyksiin.

Tutkimushenkilöitä käsitellään tässä selvityksessä anonymisti. Heitä kuvataan seuraavasti

1. Teini-ikäinen lukiolainen
2. Nuori yliopisto-opiskelija
3. Keski-ikäinen journalisti
4. Keski-ikäinen poliitikko
5. Johtavassa asemassa toimiva keski-ikäinen
6. Eläkeläinen

Selvityksessä etsittiin vastauksia seuraaviin kysymyksiin

- Mitä dataa ja kuinka paljon meistä kertyy eri palveluntarjoajille?
- Kenelle dataa kertyy ja kuka siitä hyötyy?
- Mihin henkilökohtaista dataamme käytetään?
- Mitä dataa kerätään kauttani myös kontakteistani?
- Miten meitä profiloidaan ja mihin profiileja käytetään?
- Mitä saamme vastineeksi datan keräämisestä?
- Millaista kauppaa datallamme käydään?
- Tapahtuuko datan kerääminen ja hyödyntäminen tietosuojasetuksen mukaisesti?

Selvityksen kysymyksiin haettiin vastauksia seuraavien menetelmien avulla

- Datan liikkumisen seurantaohjelma testihenkilöiden Android-testipuhelimissa
- Valittujen yritysten (14) tietosuojadokumenttaatio ja sen läpikäynti
- Yritysten vastaukset tietosuojasetuksen mukaisesti tiedusteluihin (8 kpl)
- Testihenkilöiden omien tulosten pohdinta.

Tarkempi kuvaus selvityksen laajuudesta ja datan kulun seurantamenetelmästä on liitteessä 1.

Selvityksessä analysoitiin kahden viikon ajalta kerättyä testihenkilöiden verkkoliikennedatata http-pyyntö-/vastauspakettien avulla. Sekä mobiilisovellukset että verkkosivut käyttävät Hypertext Transfer Protocol (HTTP) -nimistä tiedostonsiirtotapaa siirtäessään dataa paketteina internetissä oleville eri palvelimille. Paketit sisältävät varsinaista dataa sekä erilaista metadattaa, ja data voi olla hyvin erilaisissa muodoissa, kuten tekstiä, kuvaa, videota ja javascript-ohjelmia.

Selvityksessä keskityttiin seuraamaan datan kulkua mainonta- ja profilointiyritysten käyttöön. Todellisuudessa digitaalista mainosteknologiaa käyttäviä sivustoja oli todennäköisesti havaittua enemmän, sillä vain osassa mainonnan palvelimille lähete-

tyissä datoissa oli tieto siitä, mihin sivustoon pyyntö liittyi. Data voi olla epätarkkaa myös siksi, että osa mainontaan kytkeytyvistä yrityksistä tarjoaa myös muita palveluja, esimerkiksi analysoi verkkosivun käyttöä. Näitä käyttötapoja ei tässä selvityksessä pystytty erottamaan.

Mitä ovat kolmannen osapuolen toimijat?

Varsinaisen palvelun (ensimmäinen osapuoli) lisäksi dataa saattaa hyödyntää iso joukko muita toimijoita, eli ns. kolmannet osapuolet. Selvityksen avulla pyrittiin muodostamaan laajempi kuva siitä, minne dataa kulkeutuu palvelujen kautta.

Yksinkertaisimmassa tapauksessa sovelus tai verkkosivu kommunikoi vain palveluntarjoajan oman palvelimen eli ns. ensimmäisen osapuolen kanssa, eikä dataa siirtynyt muualle.

Kolmansilta osapuolilta dataa voidaan luovuttaa muille toimijoille, esimerkiksi datan kauppapaikoille. Myös näitä yhä kauempaa palvelun käyttäjästä olevia tahoja sanotaan usein kolmansiksi osapuoliksi. Tähän datan siirtymiseen tai yhdistämiseen ei päästy kiinni.

Verkkosivuja ja palveluja luotaessa käytetään usein valmista koodia, joilla toteutetaan erilaisia toiminnallisuuksia tai kytketään palvelu osaksi suurempaa verkostoa. Näiden koodinpätkien kautta tallentuu dataa kolmansien osapuolten palvelimille. Se ei vielä automaattisesti tarkoita, että dataa myytäisiin eteenpäin tai että se olisi muun kuin alkuperäisen palvelun käytössä, mutta verkkoliikennedatassa näkyvät yhteydennottoja muillekin kuin ensimmäisen osapuolen palvelimille.

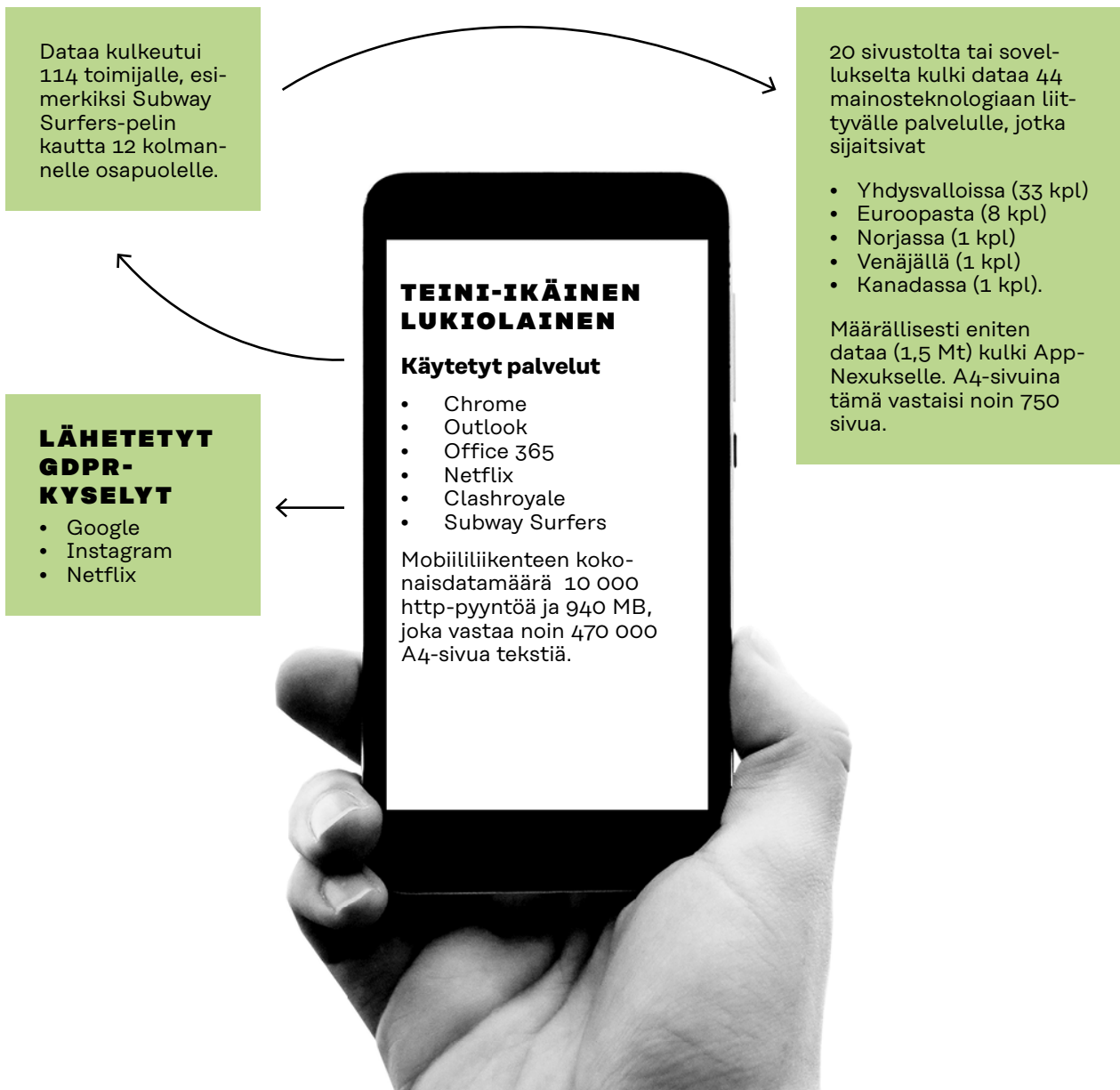
Jotkin näistä kolmannen osapuolen palveluista yhdistelevät käyttäjädataa eri sivustoilta. Jos käyttäjä vierailee useilla sivustoilla, joilla sama kolmannen osapuolen palvelu sijaitsee, voidaan kaikki käyttäjän tiedot yhdistää. Esimerkiksi Google, Facebook ja monet mainontaan liittyvät palvelut toimivat näin.

LUKIOLAINEN YLLÄTTYI KOLMANSIEN OSAPUOLTEN MÄÄRÄSTÄ

Lukiolaisen mielestä yksilöiden tulisi olla kiinnostuneempia datansa kulusta. Selvityksen tuloksissa testihenkilön yllätti kolmansien osapuolten määrä, vaikka hän tiesikin datansa kulkevan maailmalle eri palvelujen kautta. Datatalouden nykyistä systeemiä hän

ei kokenut haitalliseksi muuten kuin tilanteessa, jossa dataa käytetään väärin. Kuvitteellinen ”reilun datan merkki” ei herättäisi hänen mielestään luottamusta, koska merkin alkuperä ei olisi hänen tiedossaan.

LUKIOLAISEN DATAA KULKI 114 TOIMIJALLE



Käytettyjä palveluita:

20

Löydettyjä mainos- ja markkinointitoimijoita:

44

EU: 8

USA: 33

Norja: 1

Venäjä: 1

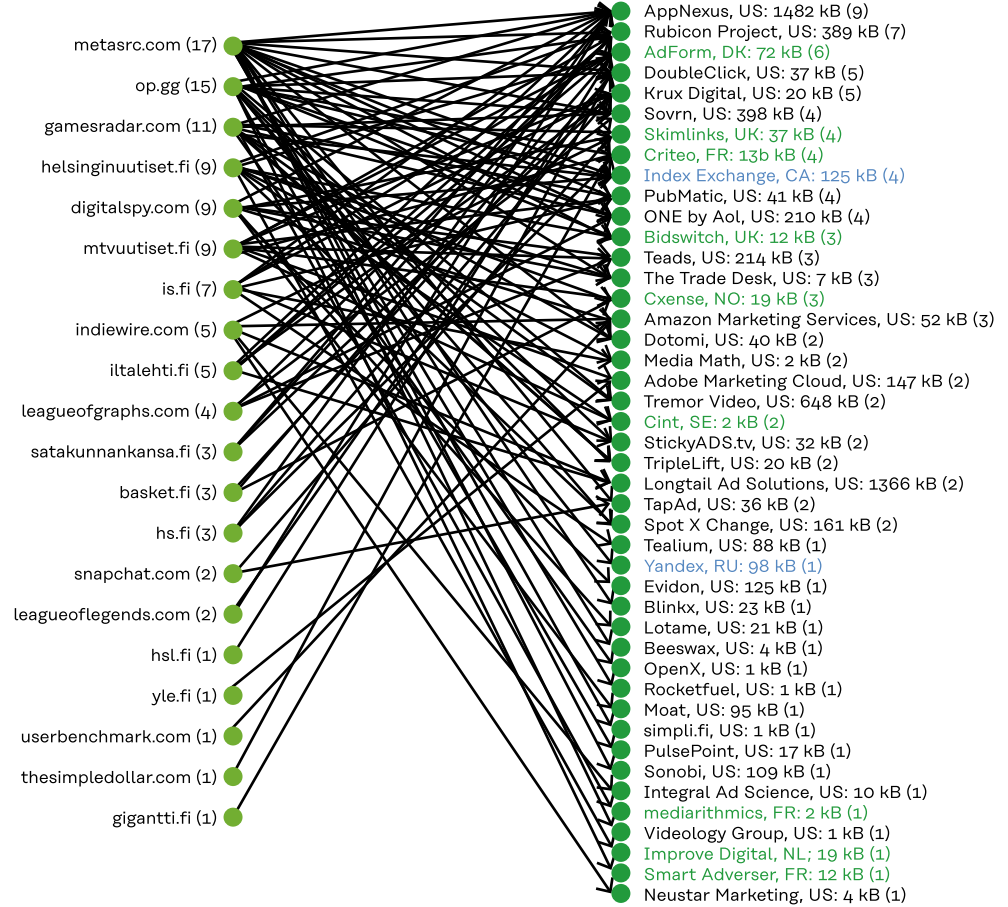
Kanada: 1

Eri toimijoille
menevän datan
kokonaismäärä
1800 https-pyyntöä,
6,2 MB, vastaa noin
3100 A4-sivua

DATAA KULKEUTUI 20 PALVELUN KAUITTA 44 MAINOS- JA MARKKINOINTITOIMIJALLE.

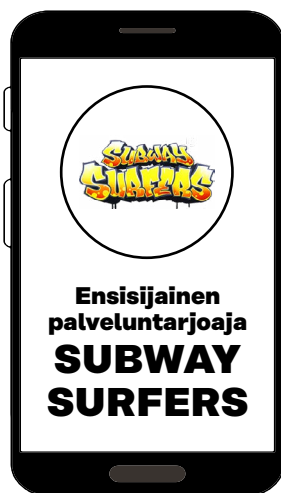
Verkkosivuja ja sovelluksia,
joita testihenkilö käytti

Mainontaan liittyviä toimijoita
Eurooppa/USA/Muut



SUBWAY SURFERS -PELISTÄ LÖYTYI KYMMENEN KOLMATTA OSAPUOLTA, JOISTA SEITSEMÄN OLI MAINONNAN TOIMIJOITA

Tarkempi selvitys eri toimijoista on liitteessä 2.



Kategoria

Tunnistetut kolmannet osapuolet

Ensijaisen palveluntuottajan palvelimet

Muut palvelun tuottamiseen tarvittavat osapuolet

Mainos-, markkinointi- ja profiloitintoimijat

Sosiaalinen media

Akamai
(palvelinylläpito)

Flurry Analytics
(sovellusoptimointi)

Moat

Chartboost

AdColony

Mintegral

ironSource

Tapjoy

Vungle

Facebook

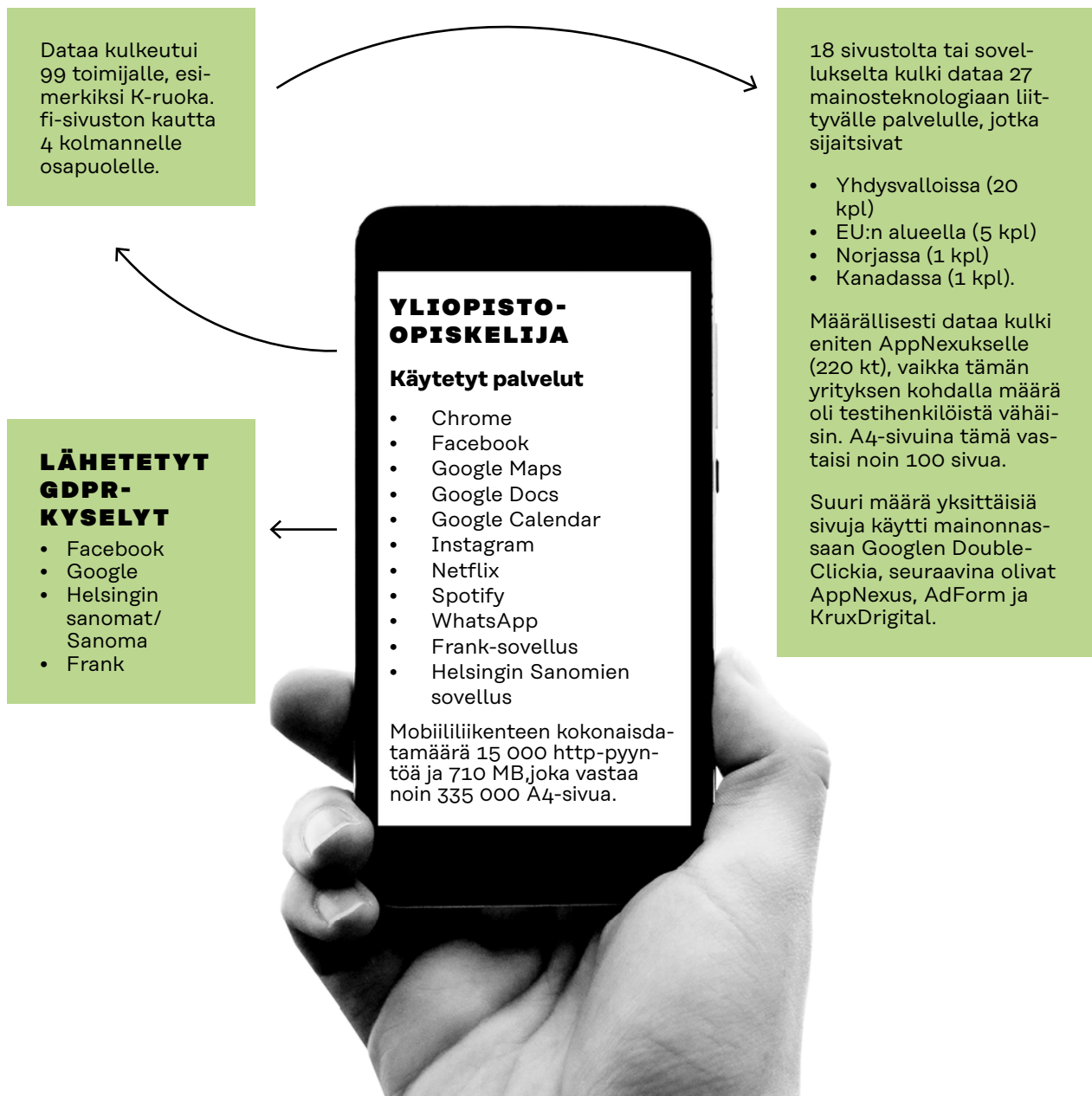
YLIOPISTO-OPISKELIJA KAVAHTI POLIITTISEN MANIPULOINNIN MAHDOLLISUUTTA

Yliopisto-opiskelija halusi osallistua selvitykseen, koska hän oli kiinnostunut datansa kulusta ja käyttötarkoituksista. Aihe kiinnosti häntä mediassa käydyin keskustelun pohjalta, mutta hän ei kuitenkaan tiennyt asiasta omasta mielestään riittävästi.

Ennen selvitystä testihenkilö ei lukenut evästekäytäntöjä koskaan ja koki dataansa

perustuvat mainokset hyödyllisiksi. Hänen kokemuksensa mukaan luottamusta kasvat-
taisi palvelujen käyttöehtojen läpinäkyvyys. Jos palveluntarjoajien toiminta olisi nykyistä läpinäkyvämpää, hän ei kokisi tietojensa jakamista haitallisena. Pahin mahdollinen skenaario opiskelijan mukaan olisi, että dataa käytettäisiin poliittiseen manipulointiin.

OPISKELIJAN DATAA KULKI 99 TOIMIJALLE



Käytettyjä palveluita:

18

Löydettyjä mainos- ja markkinointitoimijoita:

27

EU: 5

USA: 20

Norja: 1

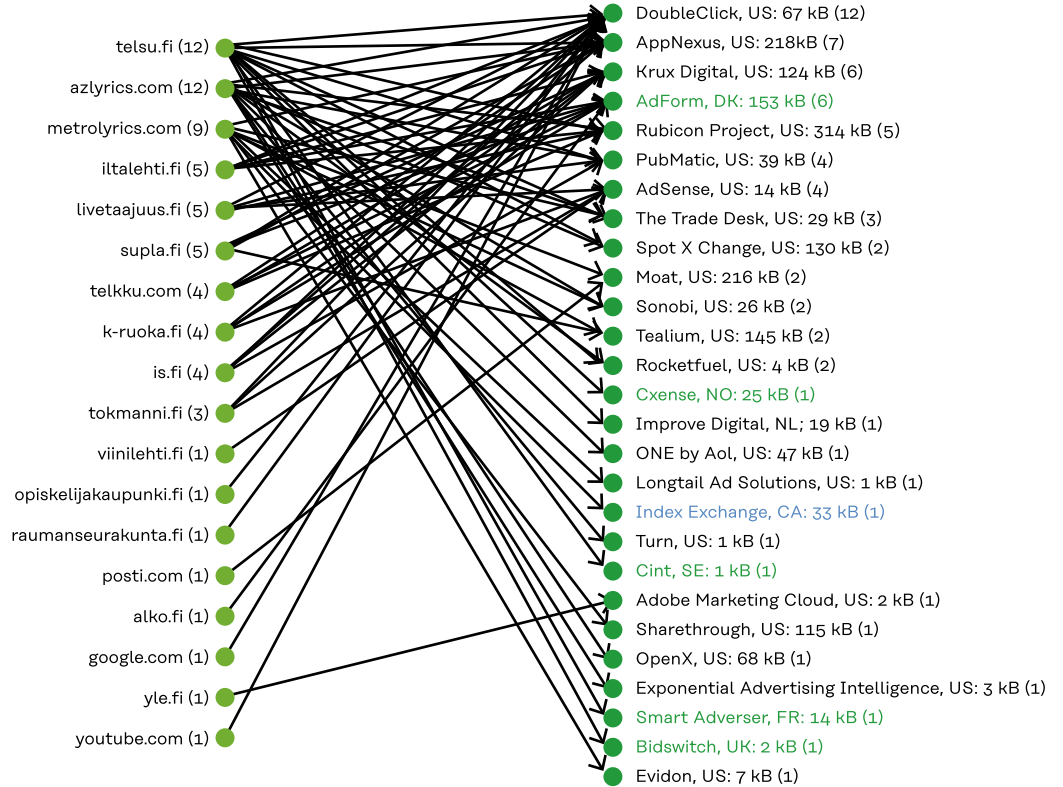
Kanada: 1

Eri toimijoille menevän datan kokonaismäärä 560 https-pyyntöä, 1,8 MB, vastaa noin 900 A4-sivua

DATAA KULKEUTUI 18 PALVELUN KAUTTA 27 MAINOS- JA MARKKINOINTITOIMIJALLE

Verkkosivuja ja sovelluksia, joita testihenkilö käytti

Mainontaan liittyviä toimijoita Eurooppa/USA/Muut



K-RUOKA.FI-PALVELUSTA LÖYTYI NELJÄ KOLMATTAA OSAPUOLTA, JOISTA KAKSI ON MARKKINOINTIYHTIÖITÄ



Kategoria

Tunnistetut kolmannet osapuolet

Ensijaisen palveluntuottajan palvelimet

Muut palvelun tuottamiseen tarvittavat osapuolet

Mainos-, markkinointi- ja profiloitintoimijat

giosg.com
(verkkosivun personointi)

feedbackly.com
(palautetyökalu)

AdForm

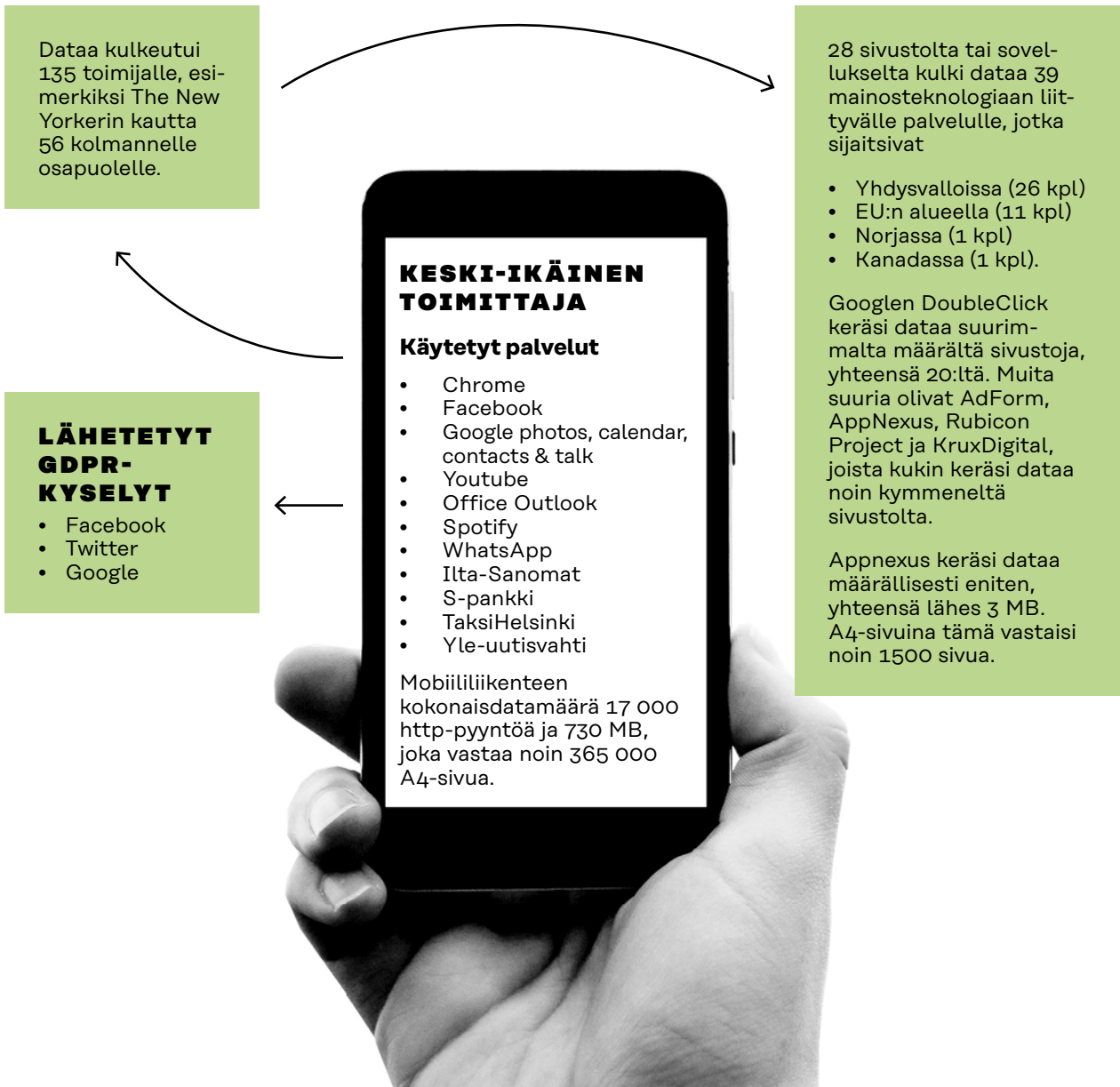
Krux Digital

TOIMITTAJA KYSEENALAISTAA MEDIAN ARVOLUPAUKSEN LUOTETTAVUUDESTA

Aihealue oli toimittajalle tuttu, koska hän oli kirjoittanut aiheesta, mutta selvitykseen osallistumisen kautta hän halusi saada konkreettista tietoa datansa kulusta ja sen jakamisesta. Hän oletti, että hänen datansa jää ensisijaiselle palveluntarjoajalle. Selvityksen tulokset muuttivat hänen suhtautumistaan dataa kerääviin yrityksiin ja perinteinen hyvän median arvolupaus luotettavuudesta kyseenalaistui.

Testihenkilön mukaan tavallinen ihminen ei pysty ymmärtämään datan kulkua monimutkaisessa verkostossa, vaan asian selittämiseen yleiskielellä tarvitaan asiantuntijaa. Selvitys vahvisti hänen olettamuksensa siitä, että data ei ole yksilön käsissä ja toisaalta tietosuoja-asetuksen mahdollistamiin kyselyihin ei saa vastauksia. ”Reilun datan merkki” olisi hänen mukaansa tarpeellinen.

TOIMITTAJAN DATAA KULKI 135 TOIMIJALLE



Käytettyjä palveluita:

28

Löydetyjä mainos- ja markkinointitoimijoita:

39

EU: 11

USA: 26

Norja: 1

Kanada: 1

Eri toimijoille
menevän datan
kokonaismäärä
1500 https-pyyntöä,
6,7 MB, vastaa noin
3400 A4-sivua

DATAA KULKEUTUI 28 PALVELUN KAUITTA 39 MAINOS- JA MARKKINOINTITOIMIJALLE

Verkkosivuja ja sovelluksia, joita testihenkilö käytti

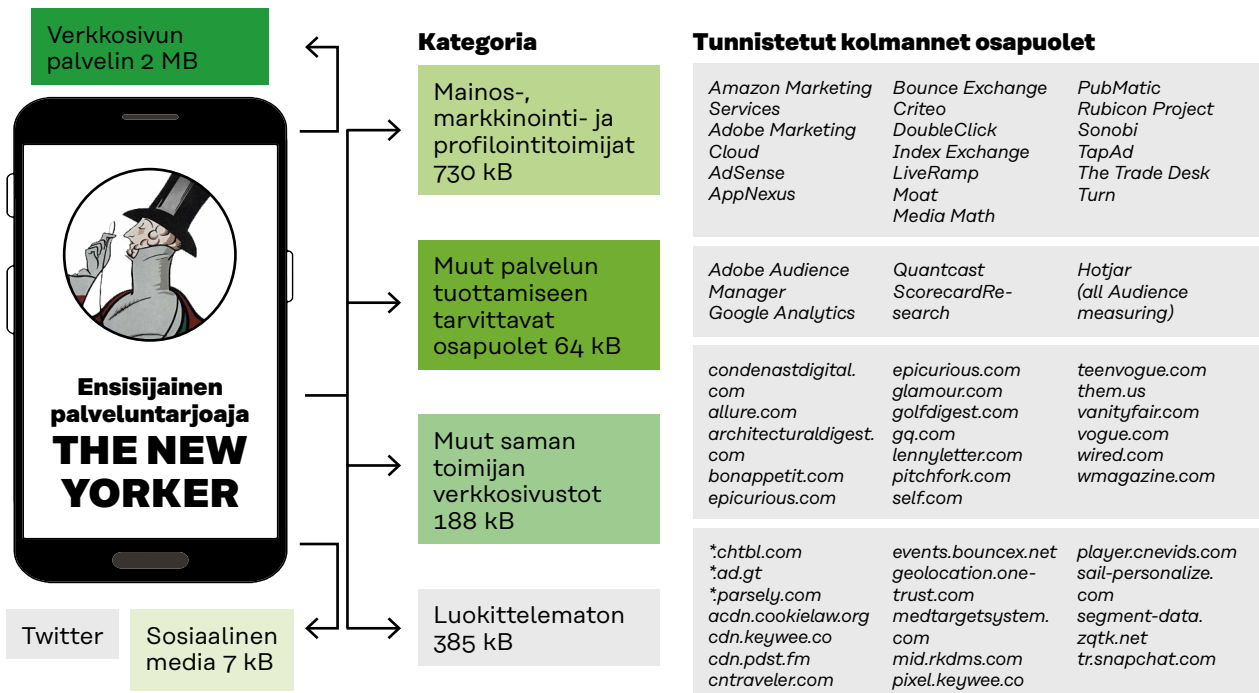
newyorker.com (18)
tori.fi (16)
spiegel.de (9)
suomi24.fi (8)
is.fi (8)
kauppalehti.fi (8)
hs.fi (8)
satakunnankansa.fi (6)
iltalehti.fi (5)
apu.fi (5)
aamulehti.fi (5)
iltasanomat.fi (4)
toa.st (4)
talouselama.fi (3)
tiekku.fi (3)
snapchat.com (2)
latimes.com (2)
hameensanommat.fi (2)
hsl.fi (2)
facebook.com (1)
taloustutkimus.fi (1)
yle.fi (1)
linkedin.com (1)
luontoliitto.fi (1)
stockmann.com (1)
google.com (1)
issuu.com (1)
youtube.com (1)

Mainontaan liittyviä toimijoita Eurooppa/USA/Muut

DoubleClick, US: 770 kB (20)
AdForm, DK: 235 kB (11)
KruX Digital, US: 80 kB (9)
AppNexus, US: 2847 kB (9)
Rubicon Project, US: 271 kB (8)
AdSense, US: 113 kB (6)
Criteo, FR: 108 kB (5)
PubMatic, US: 34 kB (4)
The Trade Desk, US: 45 kB (4)
Smart Adserver, FR: 168 kB (4)
StickyADS.tv, US: 9 kB (3)
Tealium, US: 254 kB (3)
Teads, US: 230 kB (3)
Amazon Marketing Services, US: 98 kB (3)
Adobe Marketing Cloud, US: 49 kB (3)
Moat, US: 488 kB (2)
Index Exchange, CA: 52 kB (2)
Celtra, US: 234 kB (2)
Cxense, NO: 8 kB (2)
Turn, US: 2 kB (2)
Bing, US: 13 kB (2)
Media Math, US: 69 kB (2)
TapAd, US: 10 kB (2)
Bidswitch, UK: 2 kB (1)
Cint, SE: 2 kB (1)
LiveRamp, US: 0 kB (1)
Skimlinks, UK: 0 kB (1)
Adality, DE: 2 kB (1)
Sonobi, US: 1 kB (1)
Nativo, US: 91 kB (1)
Spot X Change, US: 3 kB (1)
Bounce Exchange, US: 100 kB (1)
Bidtheatre, SE: 1 kB (1)
Media Innovation Group, UK: 1 kB (1)
Integral Ad Science, US: 75 kB (1)
Yieldlab, DE: 5 kB (1)
Improve Digital, NL: 107 kB (1)
Longtail Ad Solutions, US: 260 kB (1)
ONE by AOL, US: 2 kB (1)

THE NEW YORKER -PALVELUSTA LÖYTYY 56 KOLMATTAA OSAPUOLTA, JOISTA 22 ON MARKKINOINTIYHTIÖITÄ

Tarkempi selvitys eri toimijoista on liitteessä 2.

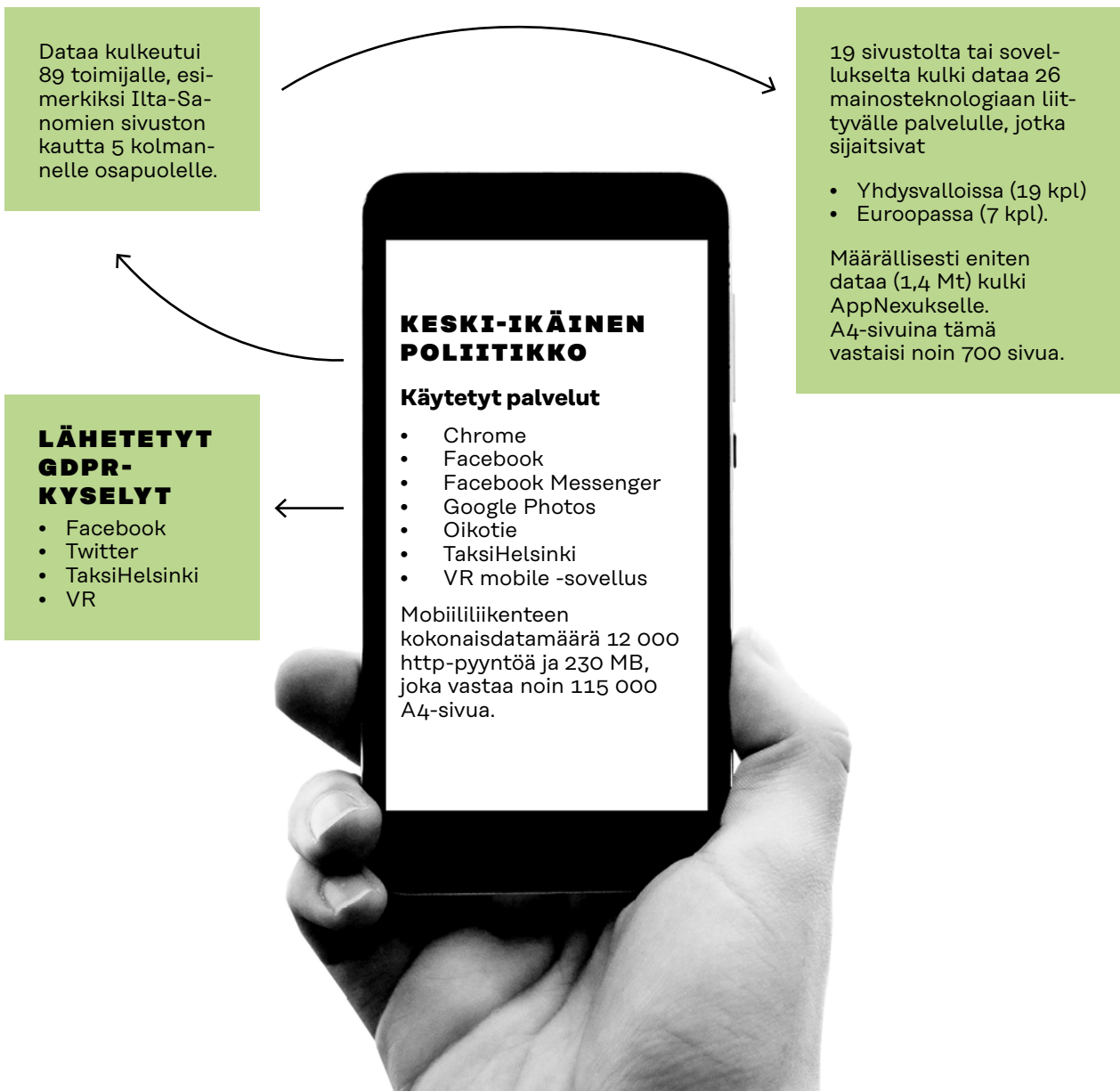


POLIITIKKO NÄKEE DATATALOUDEN AVOIMUUDEN JA LÄPINÄKYVYYDEN YHTEISKUNNALLISENA HAASTEENA

Testihenkilö kertoi, ettei ajattele dataansa arjessaan, vaikka tietääkin datatalouden toimintasääntöjen olevan merkityksellisiä sekä yksilö- että yhteiskunnallisella tasolla. Hän kertoi olevansa tietoinen kohdenne-tuista mainoksista, mutta olleensa tietämätön siitä, miten dataa huutokaupataan tai kuinka kohdentaminen käytännössä tapahtuu.

Testihenkilön mielestä yksilöiden tulisi ymmärtää enemmän datataloudesta, ja myös yhteiskunnan pitäisi miettiä, miten saisimme datatalouden avoimemmaksi ja näkyvämmäksi. Hän kertoi jatkossa tutustuvansa aiempaa tarkemmin digitaalisten palvelujen evästeisiin ja yrittävänsä estää niiden käyttöä mahdollisimman paljon. Hän kertoi myös kiinnittävänsä erityistä huomiota paikannus-tietojensa käyttöön.

POLIITIKON DATAA KULKI 89 TOIMIJALLE



Käytettyjä palveluita:

19

Löydettyjä mainos- ja markkinointitoimijoita:

26

EU: 7

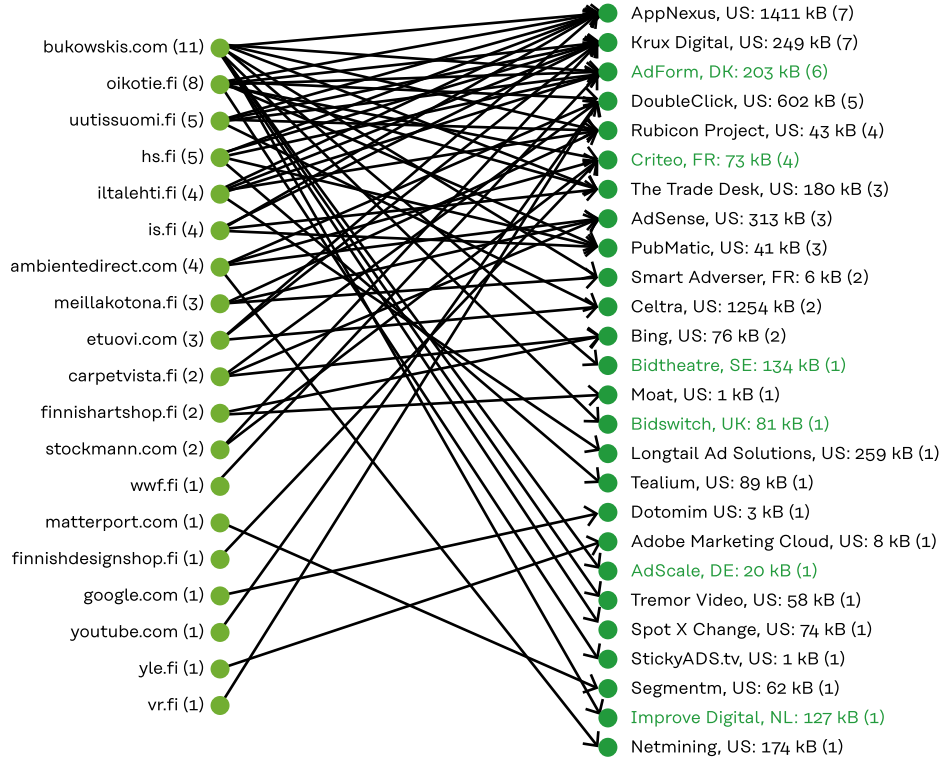
USA: 19

Eri toimijoille menevä datan kokonaismäärä:
1500 https-pyyntöä,
5,4 MB, vastaa noin
2700 A4-sivua

DATAA KULKEUTUI 19 PALVELUN KAUTTA 26 MAINOS- JA MARKKINOINTITOIMIJALLE

Verkkosivuja ja sovelluksia,
joita testihenkilö käytti

Mainontaan liittyviä toimijoita
Eurooppa/USA/Muut



ILTA-SANOMAT-PALVELUSTA LÖYTYI 6 KOLMATTAA OSAPUOLTA, JOISTA 4 ON MARKKINOINTIYHTIÖITÄ



Kategoria

Tunnistetut kolmannet osapuolet

Ensijaisen palveluntuottajan palvelimet

Muut palvelun tuottamiseen tarvittavat osapuolet

Akamai
(palvelinylläpito)

Chartbeat
(kävijämittaust)

Mainos-, markkinointi- ja profiointitoimijat

AdForm

AppNexus

Pliing.com

glimr.io

ESIMERKKISOVELLUS: TWITTER

Twitterin dataa nähtiin kulkevan palvelun omille palvelimille osoitteisiin twitter.com ja twimg.com. Lisäksi dataa kulki Google Analytics -palveluun. Se on yleisin verkkosivujen ja sovellusten käyttämä käyttäjäanalytiikkapalvelu. Google Analytics mainitaan kolmantena osapuolena myös Twitterin tietosuojaselosteessa.

JOHTAJA ANTAISI DATANSA YRITYKSEN KÄYTTÖÖN PALVELUA VASTAAN

Testihenkilö osallistui selvitykseen, koska hän oli utelias ottamaan selvää siitä, millaisia positiivisia tai negatiivisia puolia on datan luovuttamisessa yritysten käyttöön. Hän ei nähnyt datan luovuttamista tai kohdennettua mainontaa ongelmana. ”Reilun datan merk-

kiä” parempana vaihtoehtona testihenkilö pitäisi sitä, että käyttäjät voisivat valita joko datan luovuttamisen tai palvelun käytöstä maksamisen. Henkilökohtaisesti hän mieluummin antaisi datansa yritysten käyttöön palvelua vastaan.

JOHTAJAN DATAA KULKEUTUI 112 TOIMIJALLE

Dataa kulkeutui 112 toimijalle. Garmin-sovelluksesta ei löytynyt yhtään kolmatta osapuolta, mikä oli selvityksen mukaan yritysten joukossa harvinaista.

22 sivustolta tai sovellukselta kulki dataa 37 mainosteknologiaan liittyvälle palvelulle, jotka sijaitsivat

- Yhdysvalloissa (27 kpl)
- Euroopassa (9 kpl)
- Kanadassa (1 kpl).

Määrällisesti eniten dataa (2,2 Mt) kulki AdFormille. A4-sivuina tämä vastaisi noin 1100 sivua.

Useimmin verkkosivuilla käytetty mainosteknologia-yhtiö oli AppNexus (11 sivua).

LÄHETETYT GDPR-KYSELYT

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

KESKI-IKÄINEN JOHTAJA

Käytetyt palvelut

- Chrome
- Facebook
- Twitter
- Garmin
- Google Docs
- Google Maps
- TrainingPeaks

Mobiililiikenteen kokonaisdatamäärä 21 000 http-pyyntöä ja 980 MB, joka vastaa noin 490 000 A4-sivua.



Käytettyjä palveluita:

22

Löydettyjä mainos- ja markkinointitoimijoita:

37

EU: 9

USA: 27

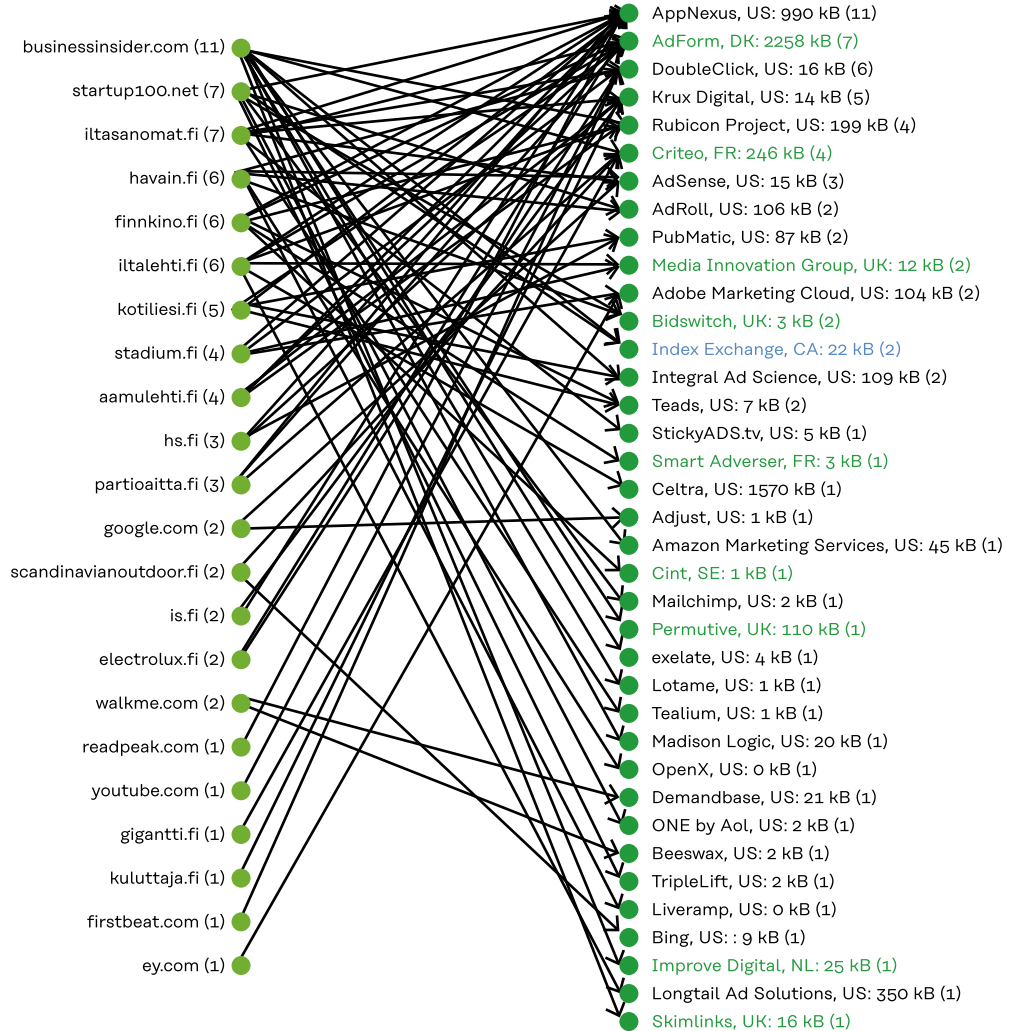
Kanada: 1

Eri toimijoille menevä datan kokonaismäärä:
800 https-pyyntöä,
6,2 MB, vastaa noin
3100 A4-sivua

DATAA KULKEUTUI 22 PALVELUN KAUTTA 37 MAINOS- JA MARKKINOINTITOIMIJALLE

Verkkosivuja ja sovelluksia,
joita testihenkilö käytti

Mainontaan liittyviä toimijoita
Eurooppa/USA/Muut



GARMIN-LIIKUNTASOVELLUKSESTA EI LÖYTYNYT KOLMANSIA OSAPUOLIA



Kategoria

Tunnistetut kolmannet osapuolet

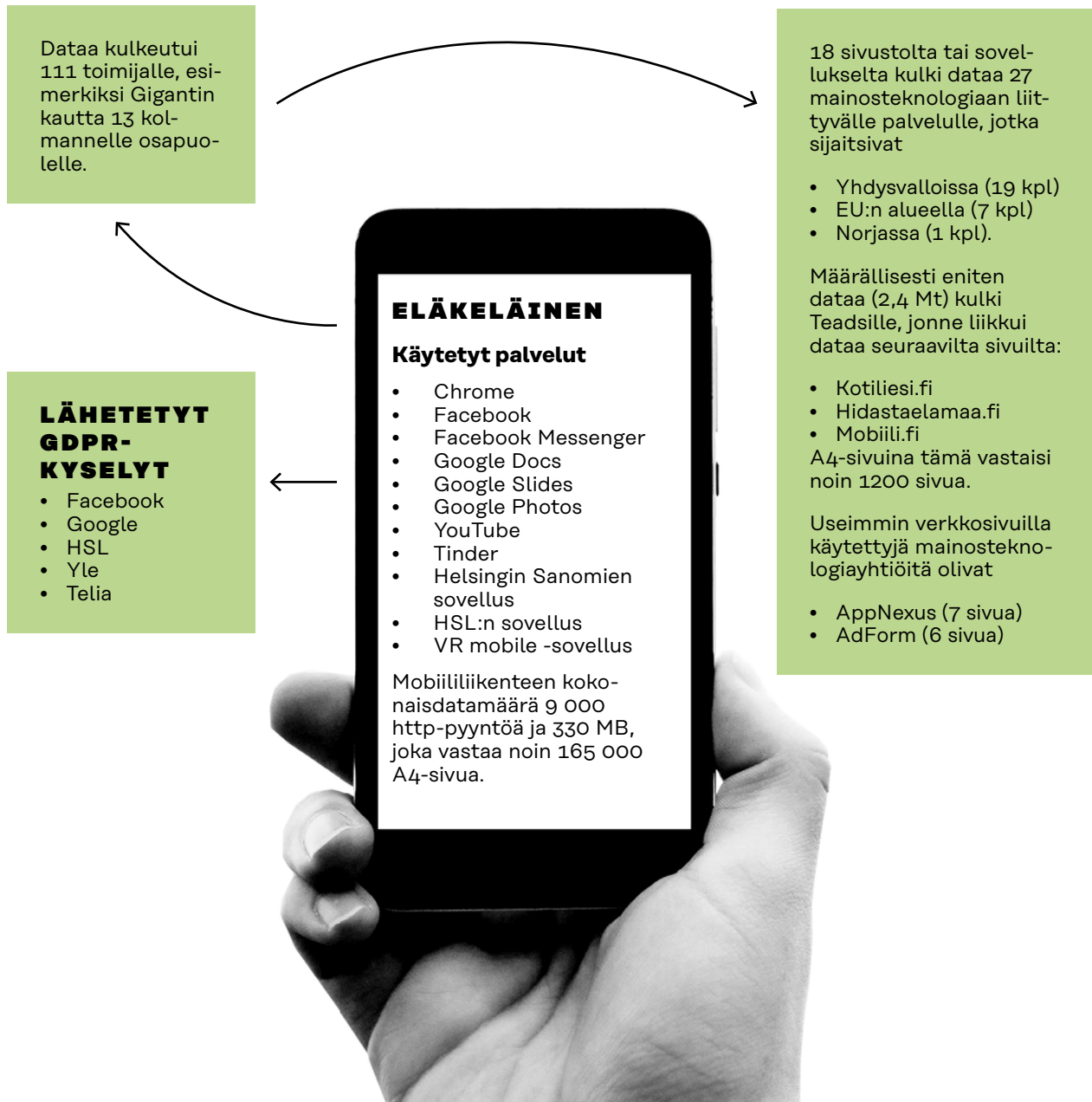
Ensisijaisen palveluntuottajan palvelimet

ELÄKELÄINEN TIETÄÄ OSALLISTUVANSA "MAAILMAN SUURIMPAAN VALHEESEEN" KLIKATESSAAN "KYLLÄ"

Testihenkilö halusi selvityksen yhteydessä oppia konkreettisesti omasta digijäljestään, koska tietää maksavansa digipalveluista datallaan. Hän kertoi tietävänsä, että dataa käytetään riippuvuutta aiheuttavien sovellusten rakentamiseen, mutta on antanut luvan datansa käytölle. Hyväksyessään palvelujen

käyttöehdot niitä lukematta hän kokee osallistuvansa ”maailman suurimpaan valheeseen”. Testihenkilö kertoi käyttävänsä Googlen hakukonetta paljon ja koki hyötyvänsä siitä. Googlen kuvapalvelun hän koki erityisen arvokkaaksi.

ELÄKELÄISEN DATAA KULKEUTUI 111 TOIMIJALLE



Käytettyjä palveluita:

18

Löydettyjä mainos- ja markkinointitoimijoita:

27

EU: 7

USA: 19

Norja: 1

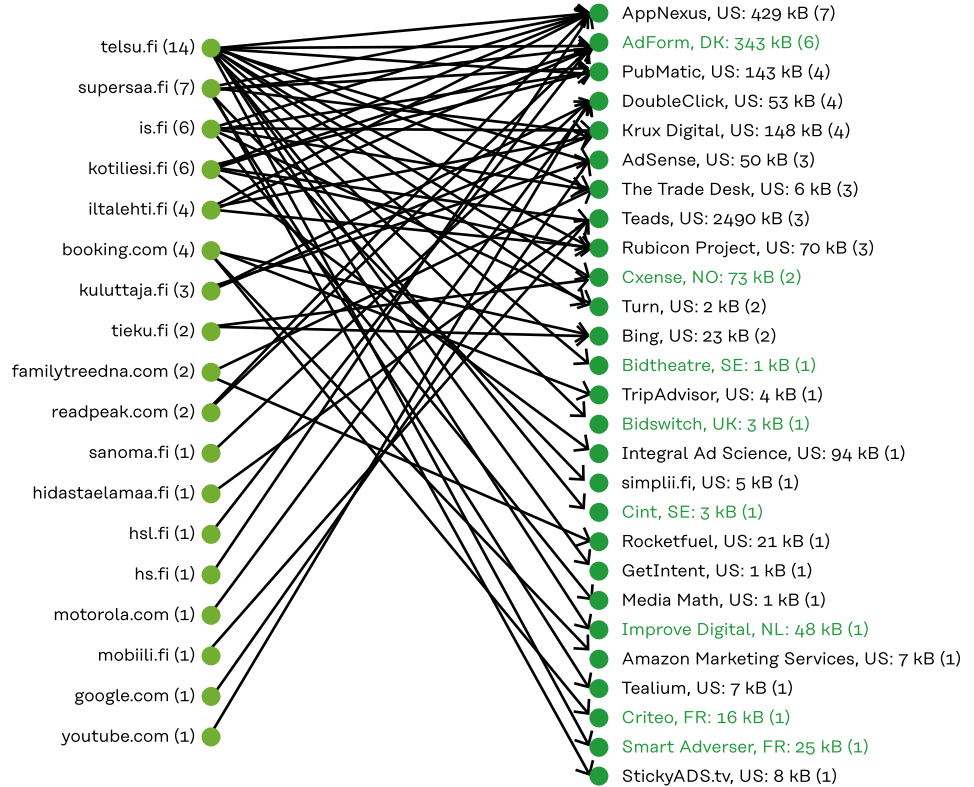
Eri toimijoille menevä datan kokonaismäärä:

500 https-pyyntöä,
4,0 MB, vastaa noin
2000 A4-sivua

DATAA KULKEUTUI 18 PALVELUN KAUTTA 27 MAINOS- JA MARKKINOINTITOIMIJALLE

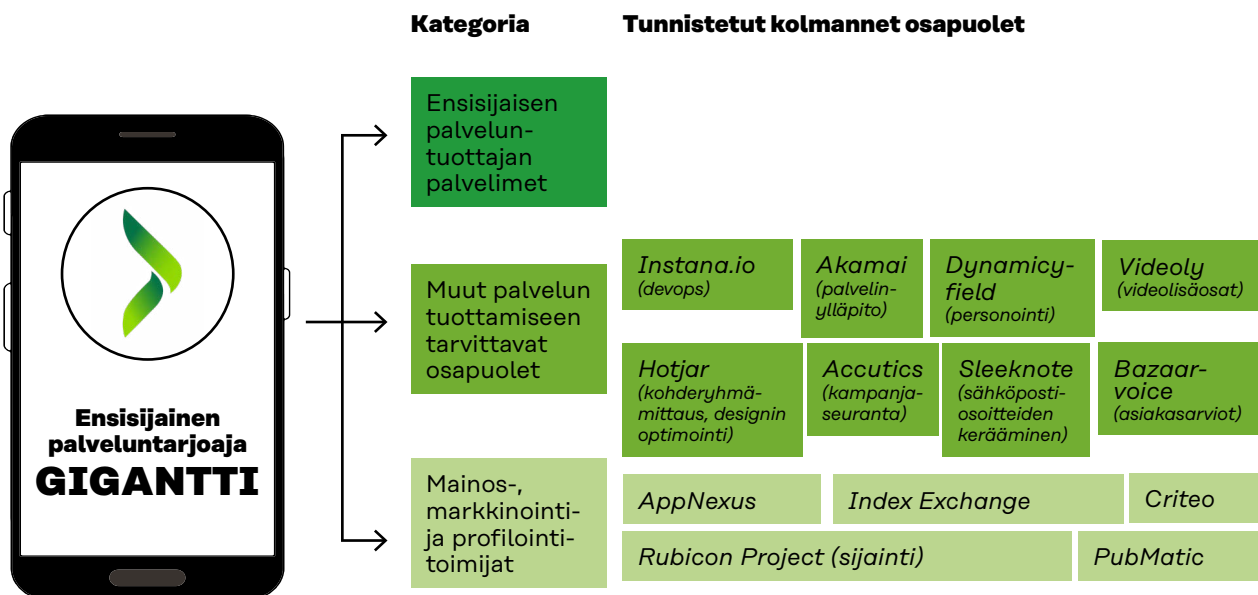
Verkkosivuja ja sovelluksia,
joita testihenkilö käytti

Mainontaan liittyviä toimijoita
Eurooppa/USA/Muut



GIGANTIN VERKKOSIVUILTA LÖYTYI 13 KOLMATTAA OSAPUOLTA, JOISTA 5 ON MARKKINOINTIYHTIÖITÄ

Tarkempi kooste eri toimijoista on liitteessä 2.



4 Digijälkiselvityksen päähavainnot – läpinäkyvyys puuttuu, eikä tietosuoja-asetus riitä

Ihmisen on mahdotonta tietää, mitä dataa hänestä on kertynyt ja kenellä dataa on. Yksilödataa rikastetaan datankulun eri vaiheissa yksilöstä muodostettavaa profiilia varten. Profiilit muodostetaan kuluttajien tietämättä, eivätkä ne laajamittaisesta datan keräämisestä huolimatta vastaa todellisuutta. Tietosuoja-asetus mahdollistaa vain rajoitetun näkyvyyden omaan dataan.

Kansalaisten ei ole mahdollista selvittää datansa liikkeitä

Selvityksen mukaan dataa liikkui usealle kolmannelle osapuolelle, ja suurin osa datasta meni amerikkalaisille yrityksille. Testihenkilöt tiesivät hyväksyneensä datan keruun samalla, kun ottivat palvelut käyttöön, mutta yllättyivät heille tuntemattomien osapuolten määrästä. Testihenkilöt eivät olleet lukeneet, mitä palvelujen ehdot ja evästeselosteet sisälsivät.

Suurin testihenkilöiden datasta löydetty määrä oli 56 erilaista kolmatta osapuolta yhdellä verkkosivulla, ja monien tutkimusten sekä palveluntarjoajien omien ilmoitusten mukaan määrät ovat huomattavasti tätäkin suurempia. Yksittäisiä datalähetyksiä eli ns. http-paketteja tutkittaessa ilmeni, että yhteensä 15 prosenttia datapaketeista meni digitaalisen mainonnan toimijoille.

Tällä hetkellä kuluttajapalveluissa hallitseva alustatalouden malli perustuu siihen, että suuret alustayritykset keräävät käyttäjistään mahdollisimman paljon dataa ja seuraavat käyttäjiä myös ”omien palvelujensa” ulkopuolella. Perusoletuksena on, että käyttäjistä kertynyt tieto siirtyy automaattisesti yritysten käyttöön.

Tietosuoja-asetus ei suojaa yksilön oikeuksia riittävästi

Suurin selvityksessä paljastunut tietosuoja-asetuksen ongelma liittyy laajaan ja läpinäkymättömään datatalouden ekosysteemiin. Verkkopalvelujen käyttäjä ei pysty hallitsemaan datansa kulkua, koska hän ei tiedä, missä hänen datansa on.

Vuonna 2018 Euroopassa voimaan astunut tietosuoja-asetus antaa yksilöille melko kapean näkymän henkilödatansa käytöstä. Yksilön on pääosin mahdollista saada tietoa vain ensisijaisen palveluntarjoajan keräämän datan osalta. Ensisijainen palveluntarjoaja kuitenkin luovuttaa dataa lukemattomille kolmansille osapuolille, joiden olemassaolosta on vaikea saada tietoa. Verkkopalvelujen käyttäjä ei tunne kolmansia osapuolia eikä voi kohdistaa asetuksen mahdollistamia toimia niihin. Palvelun käyttäjä ei voi esimerkiksi tarkistaa, perustuuko hänestä muodostettu profiili oikeisiin tietoihin. Käyttäjän on luotettava yleistäviin, vaikeasti luettaviin eväste- ja tietosuojaselosteisiin sekä pitkiin käyttöehtoihin.

Yksityiskohtaisiin tietopyyntöihin on vaikea saada vastausta

Sitran selvityksessä testihenkilöt lähettivät käyttämilleen yrityksille tietosuoja-asetuksen mukaisen pyynnön saada kopion omasta datastaan. Useat palvelut tarjoavat mahdollisuuden ladata käyttäjätiedot itsepalveluna palvelun sisällä olevalla automaattisella toiminnolla. Tätä kautta saatu tieto datan kerääntymisestä, profiloinnista ja datan käytöstä osoitettiin pintapuoliseksi. Lisäksi testihenkilöt lähettivät yrityksiin erillisen sähköpostin tai lähestyivät yrityksiä tietosuojaan liittyvän asiakaspalvelun lomakkeella ja pyysivät dataansa liittyviä tarkennuksia. Nämä tarkennukset tietopyynnöt oli valmistellut Paul-Oliver Dehaye. Pyynnöt olivat yksityiskohtaisia, ja niissä kysyttiin mm. sitä, millaista dataa palvelut keräävät käyttäjistä, miten ne profiloivat käyttäjiä ja miten palvelut kertovat datan käyttöön liittyvistä kolmansista osapuolista. Palvelun tuottajille lähetetty kyselylomake on liitteessä 3.

Testihenkilöiden oli vaikeaa tai lähes mahdotonta saada palveluntarjoajalta vastauksia yksityiskohtaisiin tietopyyntöihinsä. Yritykset eivät vastanneet tietosuoja-asetuksen oikeuttamiin tiedusteluihin datan alkuperästä ja kolmansista osapuolista. Samoin profilointiin liittyvät vastaukset olivat hyvin yleisluontoisia eivätkä auttaneet kysyjän omien tietojen arvioinnissa. Esimerkiksi yksi testihenkilö lähetti kolme sähköpostia Twitterille, mutta sai vain hankalasti tulkittavia vastauksia, joiden perusteella ei selvinnyt, mikä olisi ollut seuraava askel vastausten saamiseksi. Monissa EU/ETA-maissa tietosuojaviranomaisten antamat tietosuojasakot ovat tyypillisesti koskeneet muun muassa rekisteröityjen tietopyyntöihin vastaamatta jättämistä.

Palveluihin kertyy yksilödataa rikastetaan ja jalostetaan datankulun eri vaiheissa, ja yksilöstä muodostetaan profiili, jonka vaikutus näkyy mm. siinä, millaisia viestejä ja mainoksia hänelle näytetään. Lähes kaikki selvityksessä tutkitut ensimmäisen osapuolen palvelut käyttävät erilaisia datapohjaisesti määriteltyjä

ryhmäprofiileja palvelujen personoinnissa. Epäselväksi kuitenkin jäi, miten profiilit oli muodostettu tai käytettiin niitä muuhun kuin mainontaan. Myös profiilien ylläpito ja käytön pituus jäivät hämärän peittoon.

Profilointi mahdollistaa myös mielipidevaikuttamisen ja vaikuttaa palvelujen hinnoitteluun. Koska yritykset ja muut organisaatiot käyttävät profiileja lukuisiin eri tarkoituksiin, olisi tärkeää, että yksilöstä muodostettu profiili vastaisi todellisuutta. Tämä on erityisen tärkeää tulevaisuutta ajatellen, sillä profiilidataa voitaisiin yksilön suostumuksella hyödyntää mainontaa kriittisemmillä alueilla, kuten terveyden ja hyvinvoinnin palveluissa.

Koska testihenkilöt eivät saaneet selkeää tietoa heistä muodostetuista profiileista, tutkimuksen tekijät selvittivät MyDataAppNexuksen kautta, millaisia profiileja heistä itsestään on muodostunut. He totesivat, että datan myyjien luomat profiilit eivät vastanneet todellisuutta. AppNexus on nykyään Xandr ja näyttää profiilitiedot [verkkosivullaan](#). Toisaalta vain harva datan myyjä edes tarjoaa käyttäjälle mahdollisuutta tutustua profiiliinsa.

Mitä paremmin käyttäjä on asioista perillä, sitä helpompi on ”kysyä oikeita asioita”. Vastuun ei kuitenkaan pitäisi olla kansalaisella vaan yrityksillä.

Digitaalisen mainonnan toimintaperiaatteilla on laajempi yhteiskunnallinen vaikutus

Digitaalisen mainonnan ympäristö erilaisine riippuvuussuhteineen on niin monimutkainen, että alan asiantuntijoidenkin on vaikea hahmottaa sitä. Yksittäisten palvelujen takana on laaja verkosto tuntemattomia toimijoita, joille ensisijaiset palvelut, esimerkiksi kaupalliset mediat, ovat kanava kerätä raaka-ainetta, dataa, suurin mahdollinen määrä. Tällä on vaikutusta kaikkien yritysten toimintaan myös toimialan ulkopuolella yhtäältä kuluttajien yksityisyyden ja luottamuksen kannalta, ja toisaalta arvokkaan datan keräytymisenä harvojen käsiin.

5 Data liikkuu kolmansille osapuolille useita kanavia pitkin

Datasta luopumisen vastineena pidetään ilmaisia palveluja. Palvelujen todellista hintaa ei voi hahmottaa, koska tietoa datan leviämisestä ja käytöstä on mahdotonta selvittää. Vaihtokauppa ei siten ole reilu.

Yksilöistä kerättävä data on sekä tietoisesti luovutettuja yhteys- ja henkilötietoja että usein tiedostamatta, verkkokäyttäytymisen ja laitteiden liikkeiden vuoksi syntyviä jälkiä. Data kulkee digitaalisen mainonnan arvoketjuissa yhä kauemmas kuluttajasta, monikeroksiseen verkostoon, ”mustaan laatikkoon”.

Valtaosan yksilöiden tuottamasta datasta keräävät suuret amerikkalaiset alustayritykset, esim. Google, Facebook ja Amazon. Pelikenttää laajentavat myös suuret kiinalaiset alustayritykset, joista esimerkkeinä Alibaba, Baidu ja TikTok. Euroopassa datatalouden alustamenestyjiä on vain kourallinen, joista tunnetuimpia ovat musiikkipalvelu Spotify ja verkkokauppa Zalando. Afrikan merkittävin alustatoimija on Naspers.

Suuriin datankeräajiin kuuluvat myös peliyritykset, jotka ovat erityisesti lasten ja nuorten suosiossa. Tunnettuja peliyrityksiä ovat muun muassa amerikkalainen EA, ranskalainen Ubisoft ja suomalaista alkuperää oleva, kiinalaisen Tencentin enemmistöomistuksessa oleva, Supercell.

Pääasiallinen syy yksilöiden seurantaan on käyttäytymiseen pohjautuva mainonta. Henkilölle näytetään verkkokäyttäytymisen perusteella mainoksia asioista, joista hän on mahdollisesti ollut kiinnostunut lyhyellä aikavälillä. Markkinointi on ihmisiin vaikuttamista, mutta kuinka pitkälle yritykset

voivat mennä vaikuttamisessa ja mitkä menetelmät ovat sallittuja vaikuttamisessa?

Henkilö- ja käyttäytymisdataa voi siirtyä kolmansille osapuolille vähintään neljää eri väylää pitkin

- datahuutokaupat
- verkkosivut sekä muut palvelut ja sovellukset (tämä oli Digijälkiselvityksen ydin)
- yksilödataa keskenään yhdistelevät yritykset (cookie syncing platform)
- palvelujen rakentamisessa käytetty valmis koodi.

Princetonin yliopiston tutkimuksen mukaan eniten kolmansia osapuolia on uutissivustoilla ja vähiten hallituksen, yliopistojen ja kansalaisjärjestöjen sivustoilla.

Datahuutokaupassa mainostaja löytää mainospaikan

Norjan tietosuojavaltuutetun toimiston tilaamaan *The Great Data Race* -raportin mukaan osa verkkosivuista saa tulonsa käyttäjille näytettävistä mainoksista. Jo ennen kuin avaa verkkosivun, yksilön data myydään mainostajien huutokaupassa, 1/5 sekunnissa. Korkeimman tarjouksen tehnyt saa näyttää mainoksensa.

Huutokauppatoimijat saavat verkkosivun käyttäjän perustiedot. Näitä tietoja yhdistellen käyttäjistä olemassa olevaan tietoon,

julkisista rekistereistä saatuun tietoon sekä datakauppiailta ostettuun, käyttäjään liittyvään tietoon.

Datakauppiat eli ns. databrokerit ostavat ja keräävät kuluttajien dataa eri lähteistä, yhdistelevät ja paketoivat sitä ja myyvät jalostettuina datatuotteina ja -palveluina yrityksille. Tekoälyn avulla kukin toimija laskee hinnan, jonka se on valmis maksamaan käyttäjän näkemästä mainoksesta verkkosivulla. Suurin tarjous voittaa ja käyttäjä näkee mainoksen verkkosivun avautuessa. Jotta verkkosivun käyttäjä voisi ymmärtää, kuinka hänen dataansa käytetään, hänen tulisi lukea kaikkien huutokauppatoimijoiden tietosuojaselosteet.

Jo vuonna 2015 myytiin joka sekunti 1,3 miljoonan käyttäjän dataa huutokaupassa. Myyntitapahtumia oli 12 kertaa enemmän kuin New Yorkin pörssissä. Suurimmilla amerikkalaisilla toimijoilla, kuten Facebookilla, Yahoolla, Googlella ja Microsoftilla on omat huutokaupansa.

Verkkosivujen ja muiden palvelujen ja sovellusten kautta kulkeutuva data

Kun henkilö käyttää verkkopalveluja, on lähes aina perusolettamuksena, että yksilöstä syntynyt data on palveluntarjoajan ja hänen yhteistyökumppaneidensa käytössä. Jos henkilö hyväksyy palvelun käyttöehdot, hänestä kertynyt data voi esimerkiksi The New Yorkerin tapauksessa levitä sadoille eri toimijoille. Sitran selvityksessä löydettiin 56 eri toimijaa, joille dataa siirtyi (näistä toimijoista 18 liittyi mainontaan, markkinointiin ja profilointiin). The New Yorkerin tietosuojaselosteessa lueteltiin kuitenkin peräti 280 eri yhteistyökumppania.

Yksilöiden tietoja yhdistellään eri lähteistä

Digijälkiselvityksessä The New Yorkerin yhteistyökumppaneista löytyi kolmas osapuoli DoubleClick.net, jonka Google omistaa. Se on aktiivinen toimija dataa yhdistelevällä alustalla (cookie syncing platform). Tämä toimija löytyi lähes kaikkien testihenkilöiden dataa keräävistä tahoista. Digitaalisia palveluja käytettäessä syntyy useita tunnisteita (ID), jotka sisältävät tietoa yksilön digitaalisesta käyttäytymisestä. Osa toimijoista, erityisesti saman konsernin yritykset, jakaa keskenään heille syntynyttä tietoa käyttäjistä juuri dataa yhdistelevän alustan avulla ja saa siten kokonaisvaltaisemman kuvan käyttäjistään.

Datan kerääminen sovellusten rakentamiseen käytettävän valmiin koodin kautta

Consumer Reports Digital Labin toimittaja Kaveh Waddell nostaa artikkelissaan esiin uhan siitä, kuinka erilaisten ulkopuolelta otettujen koodinpätkien kautta voidaan kerätä käyttäjätietoa. Sovelluksia rakennettaessa on tapana käyttää valmista koodia toteuttamaan osaa toiminnallisuuksista (esim. chat-toiminnon lisääminen sovellukseen). Palvelun käyttäjä ei voi tietää olevansa tekemisissä sekä sovelluksen kehittäneen yrityksen että koodin kautta myös useamman eri organisaation kanssa. On siis mahdollista, että lainatun koodinpätkän kautta käyttäjän arkaluontoisikin tietoja päätyy muille yrityksille. Suhde on näin muodostettu useaan tiedonvälittäjään tai markkinointiyritykseen, joista käyttäjä ei ole luultavasti koskaan kuullut. Tämä tapahtuu ilman käyttäjän suostumusta ja joskus jopa ilman sovelluksen kehittäjän valvontaa. Myös muut datan keräämisen keinot ovat usein tavallisten käyttäjien näkymättömissä.

HAVAINNOLLISTAVA ESIMERKKI*

Yleisiä tapoja, joilla henkilön data voi levitä kolmansille osapuolille

*Esimerkki on luotu yhdistelemällä Sitran selvityksen tuloksia sekä seuraavia tutkimuksia ja artikkeleita:

Datatsynet 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.



Henkilö saapuu newyorker.com-sivustolle.



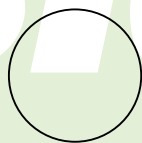
Häntä pyydetään hyväksymään evästeet.



Henkilö hyväksyy evästeet oletusasetuksineen.

Keksi?

0,2 SEKUNNIN HUUTOKAUPPA



Käynnistyy huutokauppa, jossa yritykset kisaavat henkilölle näkyvästä mainostilasta sivustolla.



Yritykset saavat dataa henkilöstä sivustolta ja sitä yhdistetään dataan muista lähteistä.



Algoritmit laskevat mainospaikalle hinnan kerätyn datan perusteella.



Huutokauppa on ohi 0,2 sekunnissa. Sivustolle ilmestyy voittajan kohdennettu mainos.

Myyty!

56

Sivustolle päästyään **henkilön dataan pääsee välittömästi käsiksi 56 yritystä.**

Useissa palveluissa hyödynnetään valmiita koodia toteuttamaan jokin toiminnallisuus (esim. kartta). Lainattu koodinpätkä voi myös kerätä dataa eri yrityksille.



Osa yrityksistä alkaa seurata henkilön käyttäytymistä lehden verkkosivuilla.

Ainakin **18 kyseisistä yrityksistä on keskittynyt profilointiin.**

Dataa yhdistelevillä **cookie syncing -alustoilla toimivat yritykset** voivat muodostaa kokonaisvaltaisen profiilin henkilöstä.

Evästeselosteiden mukaan **jopa 280 yritystä voi halutessaan saada dataa henkilöstä.**

? 280

6 Maksamme palveluista datalla, mutta miten käy yksityisyyden?

Alustatalouden jättiläisten ja digitaalisen mainonnan liiketoimintamallit on lähtökohtaisesti rakennettu yksityisyyden kannalta ongelmallisiksi. Myös palvelujen käyttäjien mahdollisuudet arvioida suostumuksen vaikutuksia ovat rajalliset suostumuksen hetkellä.

Mainosalustoille kuluttaja on datan lähde, jonka avulla tuotetaan arvoa. Asiakassuhde kuluttajaan on ensisijaisella palvelun tarjoajalla. Mainostaja puolestaan pyrkii muodostamaan asiakassuhteen tai vahvistamaan olemassa olevia asiakassuhteita.

Datalouden varhaisessa vaiheessa liikkeellä olleet teknologia- ja datajätit ovat muodostaneet kuluttajapalveluissa alustatalouden määrääviä markkina-asemia, sillä palvelut ovat saavuttaneet suuret käyttäjämäärät ja paitsi ilahduttavat myös helpottavat ihmisten arkea.

Menestyneimmät alustapalvelut ovat keskittyneet harvojen konsernien käsiin, ja niiden toimintamallit vaikeuttavat tasapuolista kilpailua. Se vaikuttaa negatiivisesti esim. eurooppalaisten yritysten kykyyn luoda uusia palveluja.

Vaikka suurimpien alustapalvelujen käyttäjämäärät ovat valtavia ja edelleen kasvussa, on jo selvästi nähtävissä, että luottamuksen puute digitaalisia palveluja tarjoavia yrityksiä kohtaan rajoittaa palvelujen käyttöä. Suurimmat datan kerääjät keskittyvät puolustamaan valta-asemaansa. Jos dataa kerätäisiin ja käytettäisiin aidosti yksilöiden suostumuksella, ihmiset voisivat

myös olla halukkaampia luovuttamaan dataansa.

Yksityisyyden hallinnan haasteita

Vaikka ihmiset eivät tunne hallitsevansa henkilökohtaisia tietojaan ja yksityisyyttään, he eivät näe vaihtoehtoja tietojen luovuttamiselle, koska haluavat turvata palvelujen saannin vastaisuudessakin. Hallinnan tunnetta haetaan säätelemällä tietoisesti itse annetun tiedon määrää, oikeellisuutta ja laatua. Tämä vääristää profiileja entisestään eikä toimi pitkällä tähtäimellä. Käyttäjälle näytettävät viestit tai mainokset eivät ehkä vastaakaan hänen todellisia tarpeitaan.

Näin toimien yksilöt menettävät niitä etuja, joita palvelujen takana olevilla algoritmeilla on pyritty tarjoamaan. Olisi järkevää, että yksilöt voisivat itse antaa luvan datan käytölle ja jakamiselle, jolloin tehdyt profiilit vastaisivat paremmin todellisuutta ja persoonintiin tähtäävät algoritmit olisivat hyödyllisempiä.

Yksityisyyden säilyttäminen digitaalisissa palveluissa vaatisi äärimmäistä aktiivisuutta. Yksilöiden odotetaan kykenevän monimutkaiseen arviointiin henkilökohtaisen datan

käytössä ja antamaan luvan sen käyttöön digitaalisten palvelujen tarjoajille vain silloin, kun hyödyt ovat suuremmat kuin haitat. Käytännössä tätä analyysiä on mahdotonta tehdä, koska tietoa ei ole riittävästi sen enempää itse datasta kuin datatalouden toimintamalleistakaan. Erityisesti lasten ja nuorten osalta vaatimus hyötyjen ja haittojen vertaamisesta on liikaa vaadittu. Lapset ovat kuitenkin tärkeä kohderyhmä mainostajille, jotka pyrkivät vaikuttamaan myös seuraavan sukupolven kulutustottumuksiin. OECD nostaa toimintaohjeessaan esiin huolen lasten datan keräämisestä pelien ja verkkoon yhdistettyjen lelujen kautta, koska näissä dataa kerätään usein käyttäjän tietämättä.

Yksityisyyden haasteita ovat esimerkiksi seuraavat (mukailtu: Lehtiniemi, T. & Korttinen, Y.):

1. Koska palvelun asettamien ehtojen hyväksymisen hetki on henkilötietojen keruun aloitusajankohta, yksityishenkilöiden on vaikea arvioida kaikkia tulevia haittoja ja hyötyjä. Vaikka välittömät haitat olisivat merkityksettömiä, pitkän ajan haittoja voi kehittyä vähitellen ajan mittaan.
2. Käyttäjien on hyväksyttävä ehdot täydellisesti voidakseen käyttää palvelua.
3. Usein tietoja keräävät tahot yhdistävät henkilötietoja eri henkilöiltä ja konteksteista, jolloin data-analyysin avulla voidaan löytää uutta tietoa.
4. Henkilötietojen odottamaton siirtyminen uusille osapuolille on yksityishenkilöille pääosin läpinäkymätöntä, mikä tekee merkityksellisestä päätöksenteosta vaikeampaa.

Tilanne käyttäjien hyväksymien ehtojen osalta on muuttumassa ja laki edellyttää, että käyttäjän on pystyttävä käyttämään palvelua, vaikka kieltäytyisikin datan keräämisestä. Käytänteet tarkentuvat ennakkotapausten kautta ja maiden tietosuojavaltuutetun toimistot ovat tärkeässä asemassa määrittelynsään ennakkotapausten kautta tarkempia sääntöjä.

EU-tuomioistuimen evästeiden käyttöä koskevasta päätöksestä ns. Planet 49 -tapauksessa todetaan, että evästeitä koskevaa suostumusta ei voi pätevästi antaa valmiiksi rastitetulla ruudulla, vaan suostumuksen antaminen edellyttää käyttäjältä aktiivisia toimenpiteitä. Suomen tietosuojavaltuutetun toimisto on yhdessä ennakkotapauspäätöksessään linjannut, mitä vaatimus aktiivisin toimenpitein annettavasta suostuksesta edellyttää eväste-ehdoilta. Päätöksessään Suomen tietosuojavaltuutetun toimisto on todennut muun muassa: vaikka käyttäjä ei tekisikään selainasetuksiinsa muutoksia tai jättäisi jonkin muun toimen toteuttamatta, se ei tarkoita, että käyttäjä olisi antanut suostumuksensa evästeiden tallentamiseen ja käyttöön.

Euroopassa ja esimerkiksi Brasiliassa yksilöitä on pyritty suojaamaan lainsäädännön avulla. World Economic Forumin artikkelin mukaan on myös muita keinoja datan suojaamiseen. Osa toimenpiteistä liittyy datan suojaamiseksi tehtyyn lakiin (Eurooppa ja Brasilia), osa datan lokalisoinnin vaatimukseen (Venäjä ja Intia), osa datan kryptauksen heikentämiseen (Australia, Kuuba, Marokko) ja osa tietojen säilyttämiseen (Kolumbia, Italia, Etiopia).

Alustayhtiöiden palvelut ovat ihmisille tärkeitä

Suurten alustayhtiöiden ja digitaalisen mainonnan koneiston kehitys nykyiseen muotoonsa on tapahtunut pikkukuhiljaa kuluttajien suosimien massapalvelujen kasvaessa. Suuret kuluttajille suunnatut alustapalvelut ovat tuoneet sekä iloa että hyötyä elämään. Kuluttajien saamaa hyötyä on vaikea mitata rahallisesti, mutta yhdysvaltalaisen yliopiston Massachusetts Institute of Technologyn (MIT) 65 000 henkilön tutkimuksessa on pyritty hahmottelemaan näiden tuotteiden arvoa loppukäyttäjille.

MIT:n tutkijat ovat peräkkäisinä vuosina kysyneet suoraan käyttäjiltä, paljonko rahaa heille tulisi antaa, jotta he suostuisivat luopumaan jostakin digitaalisesta palvelus-

taan. Vastaajat Yhdysvalloissa ovat halunneet 40–50 dollaria kuukaudessa luopuakseen Facebookista. Euroopassa ihmiset taas halusivat 59 euroa kuukaudessa kännyköiden karttapalveluista ja peräti 536 euroa kuukaudessa pikaviestisovellus WhatsAppista, sillä sen katsotaan olevan merkittävä väline kommunikoitaessa perheen ja ystävien kanssa. Lukuisia tuotteita arvioitaessa, ihmiset valitsivat kaikkein arvokkaimmaksi palveluksi hakukoneet. Hakukoneista luopumisen korvaukseksi he olisivat vuoden 2017 kyselyssä halunneet peräti 17 530 dollaria vuodessa.

Yksityisyyden rapautuminen on tapahtunut kuin huomaamatta, koska ihmiset saavat suosittuja palveluja dataansa vastaan. Suuret alustapalvelut ja digitaalinen mainonta ovat myös auttaneet kehittämään ja ylläpitämään kaikille avointa, ilmaista internetiä. Hyvät puolet ovat tähän asti riittäneet kattamaan yksityisyyden menetykseen liittyvät haitat. Kansalaisaktivismi, ihmisten verkkokäyttämisen muutokset ja medianäkyvyys kuitenkin viestivät, että yrityksille sallitun vaikuttamisen ja henkilökohtaisten tietojen keräämisen raja on tullut vastaan.

7 Digitaalisen mainonnan liiketoimintamallit kaipaavat tuuletusta

Suosituimmat kuluttajien alustapalvelut ovat rakentuneet vuosien aikana ja tuoneet vanavedessään suurten hyötyjen lisäksi haittoja. Yksityisyys on muodostunut markkinan hankalimmaksi kysymykseksi. Näiden alustapalvelujen ympärille rakentunut digitaalisen mainonnan kenttä on murroksessa.

Dataa kerätään, koska se on arvokasta raaka-ainetta palvelujen kehittämiseen ja uusien palvelujen luomiseen. Digitaalisen mainonnan ja data-analytiikan toimikentässä on valtava määrä yrityksiä, jotka saavat tulonsa joko yksilödataa keräämällä, sitä muutoin käsittelemällä, etsimällä, tallentamalla, yhdistämällä, jalostamalla tai analysoimalla – tai tekemällä kaikkea edellä mainittua. Datan avulla saadaan myös mainostilaa tarjoavat yritykset kohtaamaan organisaatiot, jotka haluavat mainostaa. Jokainen toimija saa oman osansa datan muodostamasta arvosta, koska datan paras ominaisuus on teknisesti lähes rajaton mahdollisuus hyödyntää samaa ja edelleen rikastettua dataa raaka-aineena yhä uudestaan ansaintaketjun eri osissa. Data voi monipuolistua ja kasvaa lisää arvoa jokaisen toimijan hallussa.

Internetin kohdennetun mainonnan vahva kasvu on kestänyt jo yli 20 vuotta. Tänä aikana toimialalle ja sen ympärille on kasvanut niin merkittävä liiketoiminta, että käytäntöjä ja ansaintamalleja on vaikea muuttaa. Digitoimistot, datan kerääjät, jalostajat ja analysoijat, sisältöjen julkaisijat

jne. ympäri maailman ovat rakentaneet oman liiketoimintansa alustatalouden liiketoimintamallin varaan, jossa yksilöiden käyttäytymisestä kerätty data tulee maksimoida mahdollisimman tarkkaa kohdistamista varten. Kuluttajamarkkinointia tekevät yritykset ovat joutuneet hyväksymään roolin heiltä pääosin näkymättömissä olevan datan hyödyntäjinä saaden vastineeksi mainospaikkoja. Samalla ne ovat hyväksyneet osansa koneiston mahdollistajina. Epäkohtaan ollaan heräämässä, ja keskustelu yksityisyydestä ja datan hallinnasta on alkanut niin toimialan sisällä kuin mainostajien ja toimialan välilläkin.

Salliiko tietosuoja-asetus vapaamatkustajia?

Eurooppalaiset yritykset ovat investoineet valtavia määriä rahaa ja aikaa noudattaakseen EU:n vuonna 2018 voimaan tullutta tietosuoja-asetusta. Niillä on Sitran yrityskselynsinkin mukaan huoli epätasaisesta kilpailuasetelmasta suhteessa amerikkalaisiin ja kiinalaisiin alustayhtiöihin. Nykytilanteessa suuret alustayritykset ja niiden ympärille rakentuneet suljetut ekosysteemit korjaavat

eurooppalaisista kerätyn datan liiketoimintahyödyt. Koska osa suurten alustojen ympärillä toimivista digitaalisen mainonnan ja datan keräämisen yrityksistä jää muille

paljon työtä aiheuttavan tietosuojasetuksen ulottumattomiin, tilanteesta on haittaa paitsi kuluttajille myös eurooppalaisten yritysten kilpailukyvyille.

SITRAN YRITYSKYSELYN (EUROOPPALAISTEN YRITYSTEN TULEVAISUUS DATATALOUDESSA) MUKAAN DATAN KÄYTTÖTAVAT VAIHTELEVAT YRITYSTEN KESKEN SUURESTI.

- Kaikki kyselyyn vastanneet yritykset hyödyntävät käyttäjädataa palvelun tuottamiseksi tai kehittämiseksi.
- Kaikki kyselyyn vastanneet yritykset käyttävät dataa asiakasanalytiikkaan.
- Lähes kaikki yritykset personoivat palvelunsa käyttäjäkohtaisesti.
- Kaikki suomalaiset toimijat kertoivat käyttävänsä henkilötietoa suoramarkkinointiin.
- Suuret toimijat pyrkivät ymmärtämään käyttäjiään mahdollisimman laajasti, jotta pystyisivät kohdentamaan mainontaa tarkasti käyttäjän kiinnostuksen ja elämäntilanteen mukaan. Kerätystä datasta voidaan jalostaa myös uusia tuotteita, joita yritykset myyvät eteenpäin. Nämä eivät välttämättä enää sisällä henkilökohtaista dataa, vaan voivat sisältää esimerkiksi aggregoitua tietoa käyttäjistä, esimerkkinä Telian myymä tieto ihmisjoukkojen liikkeistä.

8 Suuria haasteita datapohjaisten kuluttajapalvelujen ja digitaalisen mainonnan toimijakentässä

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 säilytetty kuluttajalle.

Tähän asti digitaalisen mainonnan liiketoiminta on pitkälti perustunut käyttäjien koneille kirjattuihin pieniin ohjelmanpätkiin eli evästeisiin, joiden avulla yksilöiden käyttäytymistä verkkosivuilla on seurattu. Erityisesti kolmannen osapuolen evästeet, joiden suhteen käyttäjällä ei ole näkyvyyttä tai toisaalta myöskään suoraa suhdetta toimijoihin, ovat osoittautuneet ongelmalliseksi.

Kuluttajien vaatimukset yksityisyyden suhteen ovat vahvistuneet, joten selainryityksetkin ovat joutuneet muokkaamaan tuotteitaan. Myös regulaatio on osaltaan vaikuttanut suuntaukseen. Applen Safari lopetti ensimmäisenä kolmannen osapuolen evästeet vuonna 2017, ja seuraavaksi niistä luopui Firefox vuonna 2018. Toimialaorganisaatio Interactive Advertising Bureau Europen (IAB Europe) mukaan selainten käytöstä jo 30 prosenttia on ilman kolmannen osapuolen evästeitä. Kun Googlen Chrome, joka edustaa noin 65:tä prosenttia käytöstä, on myös luopumassa kolmannen osapuolen evästeistä, niiden rooli mainosten tarjoamisessa käytännössä päättyy. Tämä on aiheuttanut markkinassa epävarmuutta, sillä yritys ei ole julkistanut tarkkaa aikataulua.

Evästeiden jälkeistä aikaa suunnitellaan jo kuumeisesti, sillä kuluttajien seurannasta ei edelleenkään olla tinkimässä.

Yritysten omien datavarantojen merkitys kasvaa

Alusta- ja teknologiayritykset tekevät medioissa näyttäviä lupauksia sitoutumisesta yksityisyyden suojelemiseen, mutta hankkivat yhä monipuolisempaa dataosaamista markkinoilta ja kehittävät uusia ratkaisuja, joilla pyrkivät hallitsemaan yhä suurempaa osaa yksilöiden digitaalisten palvelujen käytöstä. Kannettavien tietokoneiden ja puhelimien rinnalla ovat jo erilaiset puettavat mittarit (esim. urheilukellot), älytelevi-siot, kodinvalvontalaitteet ja ääniassistentit. Jokaisessa erillisessä laitteessa ja niiden sovelluksissa on omat tapansa kerätä dataa: ääntä, kuvaa, ulottuvuuksia, sijainti- ja reittitietoa, sydämen sykettä ja ovulaation hetkiä. Palvelujen käyttäjien syöttämien tietojen lisäksi kerätään tietoisien vaikuttamisen ulkopuolella olevaa käyttäytymis-, terveys- ja muuta tietoa. Mitä laajemmin dataa kerätään yhdelle toimijalle, sitä suuremmaksi yrityksen datavaranto kasvaa ja

sitä enemmän sillä on kykyä ymmärtää yksilöitä ja luoda uusia palveluja.

Alustatalouden jättiläisten ja digitaalisen mainonnan yritysten lisäksi muutkin yritykset pyrkivät luomaan arvoa datan avulla, mutta usein eri toimintamallilla. Kuluttajien jatkuvan seurannan sijaan yritykset pyrkivät yhä syvällisempään asiakassuhteeseen paitsi toimimalla vuorovaikutuksessa ihmisten kanssa sosiaalisessa mediassa myös keräämällä dataa esimerkiksi rakentamalla omia, suljettuja sovelluksiaan. Sovellus voi olla joko koko palvelun ydin tai se voi olla tapa muuten sitouttaa asiakasta yritykseen. Kun sovellus vaatii erillistä lataamista ja sisään kirjautumista, yritykset joutuvat punnitsemaan oman brändinsä vetovoimaa, palvelun hyödyllisyyttä ja asiakassuhteen merkitystä kuluttajalle. Asiakasymmärrys on olennaisen tärkeää asiakassuhteen ja palvelujen kehittämiseksi, joten omien datavarantojen kasvattaminen on tärkeää koko liiketoiminnan kannalta. Erillisten sovellusten rakentaminen mahdollistaa yrityksille tavan tarjoilla palvelu ja samalla kerryttää datavarantojaan, omilla ehdoillaan.

Toimijakenttä on jatkuvassa muutoksessa

Digitaalisen mainonnan ja siihen liittyvän data-analytiikan toimijakenttä koostuu pääosin rajatusta määrästä digitaalisten palvelujen konserneja, jotka kilpailevat toistensa kanssa. Yksi näistä on suuria kuluttajadatamääriä käsittelevä AppNexus (Xandr), jonka taustalla on mediajätti Warner, jonka puolestaan omistaa amerikkalainen telejätti AT&T. Tässä esimerkissä on nähtävissä, kuinka erilaiset yritykset kietoutuvat toisiinsa ja millä tavoin kuluttajien dataa voisi koota eri palveluista. Esimerkiksi teletietojen yhdistäminen muiden sovellus-

ten tuottamaan dataan on houkutteleva vaihtoehto luoda uutta liiketoimintaa.

Alustatalouden, digitaalisen mainonnan ja data-analytiikan toimijoiden kenttää venytetään laidoilta, kun ”perinteisiltä” toimialoilta tulevat yritykset hakevat uusia kasvualueita ja rooliaan datataloudessa ja haastavat vakiintuneita yhtiöitä. Digitaalisen mainonnan toimijoilla on myös huoli siitä, kuinka data ja valta keskittyvät muutamille teknologiajättiläisille. Niiden näkemyksen mukaan datan kierto mainonnan ekosysteemeissä on tärkeää, jotta yksilöistä kerätty data ei päädy vain Facebookin ja Googlen kaltaisten toimijoiden hyödynnettäväksi suljetuissa ekosysteemeissä (Interactive Advertising Bureau: The Socioeconomic Impact of Internet Tracking, February 2020).

Markkinan toimijoiden määrä ja roolit ovat jatkuvassa muutoksessa kentälle tulevien uusien pelureiden ja suurten yritysten tekemien yritysostojen vuoksi. Suurten markkinatoimijoiden arvioidaan edelleen pyrkivän ostamaan pienempiä yrityksiä ja näin kartuttamaan osaamistaan. Tähän kilpajuoksuun eurooppalaisilla yrityksillä on jo todella kova kiire.

Kuluttajapalvelujen dataliiketoiminta on laajaa ja on parhaimmillaan luonut miljardein dollarien menestystarinoita, mutta datatalouden potentiaalia on ulosmittaamatta erityisesti yritysten välisessä liiketoiminnassa. Koska datatalous on vielä kehityksensä alussa, on toistaiseksi nähty verrattain vähän kuluttajamarkkinan ulkopuolella olevia esimerkkejä yritysten välisistä datatalouden menestystarinoista. Yritysten ymmärryksen ja osaamisen kasvaessa datatalouden mahdollisuudet saadaan hyödynnettyä kaikilla toimialoilla, kaikissa yrityskokoluokissa.

9 Kuinka suojella yksityisyyttään – suosituksia jokaisen arkeen

Datatalouden pelisääntöjen ja mahdollisuuksien ymmärtämisestä tulisi tehdä kansalaistaito. Jokaisen kansalaisen tulisi ymmärtää myös oikeutensa, erityisesti yksityisyyden suhteen. Yksityisyyden kunnioittamisen tulisi ulottua koko arkeen, koska dataa kerätään monenlaisten älylaitteiden kautta.

Yksilön pitää saada vaikuttaa datansa käyttöön. Hyväksymällä käyttöehtojen oletusasetukset, henkilö saattaa antaa yritykselle vallan kerätä mahdollisimman paljon tietoa sekä palvelua käytettäessä että sen ulkopuolella.

Jotta osaaminen datataloudesta kasvaisi, tulisi jokaisen yksilön olla aktiivinen toimija ja vaatia yksityisyyden säilyttämisen oikeuksiaan.

Näillä keinoilla kykenet suojelemaan yksityisyyttäsi verkossa:

- Lue tarkoin kaikki yksityisyysasetukset ja kiinnitä huomiota siihen, minne dataa voi liikkua ensisijaiselta palveluntarjoajalta.
- Lue tietosuojaselosteissa mainittujen yhteistyökumppaneiden tietosuojaselosteet, jotta ymmärrät, mitä ne tekevät saamallaan datalla.
- Tutustu selaimesi asetuksiin kunnolla.
- Käytä useampaa selainta eri tarkoituksiin ja käytä selaimia, jotka on lähtökohtaisesti suunniteltu suojelemaan yksityisyyttä (esim. Brave).
- Käytä eri hakukoneita eri tarkoituksiin. Yksi hakukone, vaikkapa Google, voi olla työasioita varten ja toinen, vaikkapa Duckduckgo, muuta elämää varten.
- Asenna selaimeen mainonnan estäjä (ad blocker). Se estää selaimesi ja mainospal-

velimen yhteyden. Joissain selaimissa sellainen on valmiiksi, mutta tarvitset juuri sinun käyttöösi soveltuvan estäjän.

Valitse se sen mukaan, mitä eniten verkossa käytät: jos olet YouTuben suurkuluttaja, valitse siihen optimoitu mainosten estäjä. Jos pidät Facebookista, valitse estäjä sen mukaan. Mainonnan estäjän voi ladata verkosta. Princetonin yliopiston tutkimuksen mukaan Firefox-selaimen kolmansien osapuolten evästeiden estäminen on toimiva keino. Samoin GhosteryPrivacy-selaimen käyttö vähentää huomattavasti kolmansia osapuolia.

- Asenna järeitä yksityisyydensuojaohjelmistoja (VPN) kaikille laitteille. VPN:n käyttöönotto edellyttää hieman enemmän vaivannäköä. Se estää laitteesi ip-osoitteen välittämisen ulkopuolisille salaamalla verkkoyhteydessä olevan laitteen (puhelin, tabletti tai tietokone) verkkoyhteyden. Tällöin internetoperaattori ei myöskään pääse käsiksi liikenteeseesi. VPN:n voi hankkia siihen erikoistuneilta yrityksiltä, ja olet saattanut käyttää sellaista työntajasi kautta, sillä se suojaaa myös yritysten tietoja etätöissä

Tietosuoja-asetus antaa sinulle oikeuden saada lisätietoa kerätystä datasta.

- Pyydä palveluntarjoajalta tietosuoja-asetuksen mahdollistamalla tavalla tietoa sinusta kertyneestä datasta, kolmansista osapuolista, profiloinnistasi sekä paikoista, jonne dataasi on luovutettu.
- Ota yhteys tietosuojavaltuutettuun, mikäli palveluntarjoaja ei toimita dataasi tai koet, että sitä on käytetty väärin.
- Voit etsiä sinusta muodostettuja profileja myös esim. [Xandr-palvelun](#) kautta.

Edellä mainituilla keinoilla voit suojata yksityisyyttäsi paremmin verkossa, mutta yksilöistä kerätään tietoa myös selainten, sosiaalisen median ja hakukoneiden ulkopuolella. Erilaiset IoT- ja älylaitteet saattavat jakaa keräämäänsä dataa kolmansille osapuolille. Yksityisyyden turvaaminen kannattaa huomioida siis myös laajemmin, ei pelkästään puhelimen ja tietokoneen käytön kautta.

USEAT VIRALLISET TAHOT ANTAVAT OHJEISTUKSIA YKSITYISYYDEN SUOJAAMISEEN.

[Sitran digiprofiilitesti](#) auttaa ymmärtämään paremmin kuluttajille näkyvimmän datatalouden osa-alueen toimintaperiaatteita. Testin avulla saa myös tietoa siitä, kuinka omaa dataa voi suojata paremmin.

Tanskalainen European Data Ethics Forum on tehnyt kattavat [ohjeet mahdollisuuksista suojata yksityisyyttä](#).

European Interactive Digital Advertising Alliance on tehnyt [oppaan selainkäyttöön perustuvasta verkkomainonnasta ja yksityisyydensuojasta](#) verkossa. Sivustolta löytyy tietoa selainkäyttöön perustuvan mainonnan toimintaperiaatteista, [evästeistä](#) ja keinoista suojata yksityisyyttä internetissä.

10 Loikka reiluun datatalouteen – suosituksia yrityksille

Asiakaskokemus ja yritysvastuu ovat eurooppalaisten yritysten mahdollisuus profiloitua reilun datatalouden tekijöiksi. Alustatalouden pelisääntöjen kopioimisen sijaan tulisi hakea uudenlaisia liiketoimintamalleja, joissa dataa jaetaan verkostoissa yksilöiden luvalla.

Jokainen yritys on datavarantojensa ja tietosuojakypsyytensä suhteen erilaisessa vaiheessa. Yritys, joka käyttää yksilödataa reilun datatalouden periaatteiden mukaisesti, arvioi jatkuvasti omaa toimintaansa eettisyyden, datanhallinnan ja sääntöjen noudattamisen mukaan. Se kiinnittää erityistä huomiota myös osoitusvelvollisuuteen (accountability). Osoitusvelvollisuus voidaan viedä tasolle, jolla on yritykselle merkittävää lisäarvoa. Yrityksellä tulisi olla kyky tuottaa sen kaltaista aineistoa, josta syntyy parhaimmillaan jopa kilpailuetua. Vastuullisuuden kannalta helposti haltuun otettavissa oleva materiaali, joka tuo yhtiön johdolle, vastuullisuustoiminnoille ja asiakkaille selkeää tietoa, auttaa yritystä kehittämään toimintaansa edelleen.

Käyttöehdot ovat osa asiakaskokemusta

Suosituimpien digitaalisten palvelujen käyttöehdot ovat valitettavan usein yli 10 000 sanan pituisia ja sisällöltään sekavia lukuisine linkkeineen yhä uusiin teksteihin. On selvää, ettei niitä ole tarkoitettu kuluttajan helposti luettaviksi. Kuluttajien tarpeet jäävät helposti huomioimatta, kun palvelut saavat tätä kautta laajat oikeudet datan hyödyntämiseen. Pitkällä tähtäimellä tämä ei ole yritystenkään etujen mukaista.

Yrityksestä riippuen EU:n tietosuoja-asetus 2018 oli joko vain irrallinen projekti yrityksessä tai mahdollisuus arvioida omia datavarantojaan tai jotain näiden väliltä. Kuluttajille se on näkynyt internetin käytön hankaloitumisena lukuisten evästekäytäntöjen ja tietosuojaselosteiden muodossa, eivätkä kaikki ole ymmärtäneet, miksi näin tapahtui. Tietosuoja ei ole vielä näkyvä osa organisaatioiden arjessa, vaikka sitä pitäisi ajatella jatkuvana kehittämisenä. Eurooppalaisilla yrityksillä olisi mahdollisuus erottautua reiluina toimijoina, joilla on käytössään paitsi datan käytön eettiset ohjeet ja läpinäkyvä raportointi myös selkeät ja helpot eväste-, tietosuoja- ja käyttöehdot, jotka huomioivat nykyistä paremmin kuluttajien toiveet yksityisyyden suhteen.

Yritysten tulisi kehittää asiakasdatavaran tojaan oman toimintansa lähtökohdista ja jopa pitäytyä tarpeettomasta datan keräämisestä. Tietosuojakäytäntöjen ymmärrettävyyden tulisi panostaa selvästi nykyistä enemmän. Käyttöehtoja kannattaisi kehittää nimenomaan asiakkaan näkökulmasta sekä tarjota työkaluja, joilla asiakkaat voivat hallita dataansa helposti.

Data osaksi yritysvastuuta

Datan kerääminen yksilöistä jatkaa vahvaa kasvuaan, mutta toiminnan läpinäkyvämyys aiheuttaa laajasti huolta kuluttajien ja

mainostajien keskuudessa. Koska data on noussut maailman arvokkaimmaksi liiketoiminnan raaka-aineeksi, sitä tulisi muiden resurssien tavoin tarkastella myös yritys vastuun näkökulmasta, ja datan vastuullisen käytön tulisi nykyistä selkeämmin olla osa yritysten vastuullista toimintaa. Käytännössä tämä tarkoittaa läpinäkyvyyttä ja selkeyttä datan keräämisessä, hyödyntämisessä ja raportoinnissa.

Suuren mainostajia edustavan kansainvälisen World Federation of Advertisers (WFA) -organisaation mukaan 82 prosenttia sen jäsenyritysten työntekijöistä harkitsisi yrityksen jättämistä, mikäli yrityksen datankäyttö ei olisi eettistä. Alustajättiläisten, suurten teknologiayritysten ja digitaalisen mainonnan yritysten olisi kannattavaa huomioida tämä, sillä jäsenenä on valtavia kansainvälisiä organisaatioita ja WFA:n mukaan ne edustavat myös 90:ää prosenttia globaaleista markkinointiviestinnän panoksista

Vastuullisesti dataa hyödyntävät edelläkävijäyritykset eivät jätä vastuuta yksityisyyden suojelusta asiakkailleen, vaan hakevat keinoja ylittää lain vaatima minimitaso ja erottua muista. Ne etsivät perinteisestä poikkeavia liiketoimintamalleja ja rakentavat palvelut alusta alkaen asiakaskokemus huomioiden sisältämään kuluttajien toiveet myös yksityisyyden kunnioittamisesta. Kuluttajille annetaan enemmän liikkumavaraa datan keräämisen asteen valinnassa, ja toisaalta vastuullisuus näkyy myös siinä, että kerätty data varmasti hyödynnetään ja tuo arvoa asiakkaiden arvostamina, uusina palveluina.

Vastuullinen datan käyttö liittyy myös sen jakamiseen, kuitenkin siten, että se tapahtuu asiakkaiden suostumuksella. Paljon dataa omaavat organisaatiot voisivat jakaa omista datavarannoistaan muille mahdollistaen näin innovaatioita ja luoden sitä kautta kasvua myös ympäristöönsä. Reilussa datataloudessa palvelun käyttäjän suostumus on keskeinen, ja se tulee huomioida liiketoiminnan suunnittelussa, sopimusehdoissa, teknologia- ja kumppanivalinnoissa.

Datatalouden pelisäännöt uusiksi

Alustajättien vuosien varrella kasvattamaa etumatkaa, joka on syntynyt niiden ympärille rakentunein datan keräämisen ja jalostamisen keinoin, on vaikea kuroa kiinni. Niiden keräämän datan määrä sekä henkilödatan monipuolisuus ovat nykyisessä alustatalouden mallissa tavallisten yritysten saavuttamattomissa, ja Euroopassa on jääty kehityksessä jälkeen. Mikäli eurooppalaiset yritykset haluavat osansa datataloudesta, niiden on löydettävä vaihtoehtoisia tapoja toimia.

Yrityksen, joka etsii aktiivisesti uutta liiketoimintaa datapohjaisilla palveluilla, kannattaa aloittaa arvioimalla omia liiketoimintamallejaan ja laatimalla datastrategia. Se on konkreettinen keino kuvata organisaation omia dataa, kartoittaa puutteita sekä ottaa kantaa esimerkiksi datan jakamiseen. Parhaimmillaan se vie kohti uusia datatalouden kasvumahdollisuuksia ja auttaa selviämään tiukoissakin tilanteissa.

Alustatalouden liiketoimintamalleista poikkeavat datakumppanuudet ja verkostomaiset datan jakamisen mallit voisivat olla keino vauhdittaa eurooppalaista datataloutta. Jakamalla ja yhdistelemällä erilaisia dataa muiden organisaatioiden kanssa voitaisiin saada uusia näkymiä entistä älykkäämmiin tarpeisiin kohdistettujen palvelujen luomiseksi sekä kuluttaja- että yritysmarkkinoille. Tähän tarvitaan läpinäkyvät, selkeät ja kaikille osapuolille yhteiset pelisäännöt.

Digimainonnan panosten arvoeroosio ja vaihtoehtoiset toimintamallit

Digitaalisen mainonnan teknologioiden ja automaation lisäämisen uskottiin takavuosina poistavan välikädet mainonnan arvoketuista. Isossa-Britanniassa mainostajia edustavan organisaation ISBA:n ja konsulttiyritys PwC:n kaksi vuotta kestänyt kohdenetun mainonnan tutkimushanke (Programmatic Supply Chain Transparency Study 2020) osoitti, että vaikka vanhat välikädet

ovat kadonneet, ohjelmalliseen mainontaan käytettyjä rahoja menee uudelle välikäsien kerrokselle. Mainonnan hankintaketjun joka vaiheessa tapahtuu arvoeroosiota, jonka vuoksi vain puolet mainostajien käyttämistä rahoista päätyy mainospaikkojen omistajille eli sisältöjen julkaisijoille ja medioille, joilla on ensisijainen suhde kuluttajaan. Sisältöä tuottavat tahot häviävät digitaalisen mainonnan arvoketjuissa rahaa lukemattomille tuntemattomille toimijoille.

Medioiden lisäksi häviäjiä ovat myös mainostajat, joiden on vaikea luotettavasti todentaa digitaalisten mainostensa panos-tuotos-suhdetta. Mainostajan kannalta rahoista katoaa merkittävä osa jo ennen kuin brändin viesti edes saavuttaa halutun median, ja jopa 15 prosenttia koko panoksesta menee tahoille, jotka ovat täysin jäljittämättömissä. Tämä on selkeä osoitus siitä, että ohjelmallisen eli automatisoidun mainonnan markkina on kehittynyt liian monimutkaiseksi. Ongelma on erityisen hankala siksi, että digitaalisesta mainonnasta 90 prosenttia tapahtuu tänä päivänä ohjelmallisesti.

Mainostavat yritykset voisivat pohtia omia markkinointikäytäntöjään ja haastaa sekä omia työntekijöitään että markkinoinnin tukena käytettyjä toimistoja etsimään uudenlaisia, yhä kestävämpiä toimintamalleja. Kontekstuaalisen digimainonnan, jossa mainoksia tarjotaan esimerkiksi hakukonehaun tai verkkosivun sisällön pohjalta,

edelleen kehittäminen voisi olla yksi vaihtoehto. Jokainen digitaalisen mainonnan kampanja tai toimenpide tulisi arvioida yksilön koko suostumusketjun kannalta, ei vain mainostavan yrityksen tai mainospaikka tarjoavan, ensisijaisen palvelun osalta. Yritysten tulisi myös hyödyntää huomattavasti nykyistä tehokkaammin suhdettaan asiakkaaseen ja hakea vastavuoroisuutta etäällä tapahtuvan seurannan sijaan. Osa yrityksistä toimii aktiivisesti sosiaalisessa mediassa ja on näin suoraan yhteydessä asiakkaisiinsa, mutta dataa ei saada tehokkaasti toimimaan yrityksen hyväksi. Asiakastietojärjestelmän, verkkosivujen ja sovellusten tuottama data on yritykselle ainutlaatuista. Se tarjoaa lisämahdollisuuden syvälliseen ja jatkuvaan vuorovaikutukseen asiakkaiden kanssa, ja väylän palvelujen kehittämiseen.

Digitaalisen mainonnan yritysten sekä niiden varassa olevien medioiden ja mainostajien tulisi käynnistää laaja arviointi koko toimialan toimintamalleista ja etsiä tulevaisuuden vaihtoehtoja. Ensisijaisina kehittämiskohteina voisivat olla toimitusketjujen avoimuuden lisääminen yritysten ja kuluttajien suuntaan sekä yhteisten, toimialarajat ylittävien toimintatapojen etsiminen. Riittävän kunnianhimoisella itsesääntelyllä voitaisiin mahdollisesti välttää viranomaisten ja sääntelyelimien otteiden koveneminen ja tervehdyttää toimialaa kokonaisuutena.

11 Tulevaisuuden menestyjäpalvelut pohjautuvat luottamukseen

KUVITELLAAN HETKI DIGITAALISTA ARKEAMME VIIDEN VUODEN KULUTTUA.

”Olet tuntenut olosi viime aikoina vähäenergiseksi ja todennut elämäntaparemontin olevan paikallaan. Ystäväsi kertoi uudesta Elä paremmin -sovelluksesta. Palveluun tutustuessasi huomaat sillä olevan Reilu™-leiman. Nyt tiedät palvelun käyttävän henkilökohtaista dataasi eettisesti ja turvallisesti. Annat palvelulle luvan yhdistää lähikauppasi ostosdatan, Kanta-palvelun terveystiedot ja Sports Trackerin keräämän datan. Maksua tehdessäsi huomaat vakuutusyhtiösi tarjoavan palvelun sinulle ilmaiseksi. On vuosi 2025, ja maailmassa on reilusti toimiva datatalous.”

Tähän me pyrimme reilun datatalouden hankkeessa. Tavoitteena on tehdä elämästä helppoa ja mukavaa ja hyödyntää digitalisaation mahdollisuuksia niin, ettei yksityisyytemme vaarannu. Alustatalouden vanhasta mallista on kyettävä siirtymään hajautettuihin ja avoimiin liiketoimintaekosysteemiin, jotka pohjautuvat toimiviin datamarkkinoihin ja kuluttajien luottamukseen. Yksilön arvot muokkaavat yritysten toimintaympäristöä.

Reilun datatalouden arvopohja määräytyy seuraavien väittämien kautta:

- Luottamus digitaalisiin palveluihin rakentuu vaikutusmahdollisuudesta (itseäänmäärittäminen) ja hallinnan tunteesta (läpinäkyvyys) itseä koskevan tiedon käyttöön
- Ihmislähtöisen, reilun datatalouden lähtökohta on yksilön mahdollisuus vaikuttaa datansa hyödyntämiseen osana elinvoimaista digitaalista ekosysteemiä
- Reilun datatalouden ekosysteemeissä yhteiset pelisäännöt ja läpinäkyvyys määrittävät datan jakamista myös organisaatioiden välillä. Reilu datatalous tuottaa arvoa kaikille.

Sitra-laki määrittelee Sitran tavoitteen seuraavasti:

Sitran tavoitteena on edistää Suomen vakaata ja tasapainoista kehitystä, talouden määrällistä ja laadullista kasvua sekä kansainvälistä kilpailukykyä ja yhteistyötä toimimalla erityisesti sellaisten hankkeiden toteuttamiseksi, jotka vaikuttavat kansantalouden voimavarojen käyttöä tehostavasti tai tutkimuksen ja koulutuksen tasoa kohottavasti taikka jotka selvittävät tulevaisuuden kehitysvaihtoehtoja.

Datataloudessa pärjääminen on Suomen kilpailukykyyn kannalta ensiarvoisen tärkeää ja uskomme vahvasti, että Suomi voi näyttää esimerkkiä. Voimme luoda menestyviä palveluita, jotka pohjautuvat luottamukseen. Uudet datamarkkinat ovat tasavertaiset kaikenkokoisille yrityksille. Datamarkkinoilla dataa jaetaan saumattomasti, läpinäkyvästi ja luvitettuna eri toimijoiden välillä. Reilusta datataloudesta hyötyvät kaikki. Yksilöt saavat entistä kohdennetumpia palveluja, yritykset kasvua innovaatioista ja yhteiskunta lisääntyntä hyvinvointia. Reilu datatalous tarjoaa Suomelle ja Euroopalle

kilpailuedun globaaleilla datatalouden markkinoilla.

Tulevaisuus otetaan haltuun sekä liiketoimintakyvykkyksiä että teknologiaa kehittämällä. Tämä edellyttää uusien liiketoimintaekosysteemien syntymistä, innovatiivisia liiketoimintamalleja ja uudenlaisia palveluita. Palveluiden tuottaminen edellyttää myös, että data otetaan aktiiviseksi tuotantotekijäksi.

Reilu datatalous vaatii yhteistyötä niin päätöksentekijöiden kuin yritysten ja kansalaisjärjestöjenkin välillä. Edelläkävijöiden on kyettävä innostamaan ja vakuuttamaan muita. Tuotamme Sitrassa jatkossakin uutta tietoa, pyrimme tunnistamaan haasteita ja

kertomaan keinoista haasteiden ratkaisemiseksi. Haluamme vauhdittaa yhteiskunnallista keskustelua, levittää tietoisuutta, törmäyttää ja verkottaa eri toimijoita. Tärkeässä roolissa ovat käytännön kokeilut ja pilotit uusien toimintamallien testaamiseksi. Uusien toimintamallien levittäminen ja vakiinnuttaminen vaatii, että yksilöt ja yritykset hyötyvät niistä aidosti ja toimintaympäristö on suosiollinen uudelle toimintatavalle. Poliittisen ja hallinnollisen perustan rakentaminen muutokselle on yksi Sitran peruskeinoista, mutta reilun datatalouden muokkaamisessa yksilöiden rooli muutoksen tekijöinä on ensiarvoisen tärkeä.

Lähteet

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., Kortensniemi, Y. 2017. [Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach](#).

OECD Policy Note, 2020. [Growing Up Online. Addressing the Needs of Children in the Digital Environment](#)

Sitra 2019. [Digitaalisten palveluiden käyttö. Kyselytutkimus neljässä maassa](#).

Sitra 2019. [Eurooppalaisten yritysten tulevaisuus datataloudessa](#).

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 \(2020\)](#)

Muita hyödyllisiä linkkejä

Forbrukerrådet Norge 2020. [Out of Control - How consumers are exploited by the online advertising industry 2020](#).

Sitra 2019. [Datatalouden sääntökirja](#). (tilattavissa)

Sitra's Working paper 2020. [35 Proposals to make the European data strategy to work](#).

Sanasto

ADTECH (ADVERTISING TECHNOLOGY): Suora käänös: mainontaan liittyvä teknologia. Kyseessä on laaja määritelmä, jolla ei ole kunnollista suomenkielistä vastinetta. Laajasti ottaen AdTech sisältää niin digitaalisia työkaluja kuin analytiikkaakin, mutta keskusteluissa usein viitataan kompleksiseen ekosysteemiin, jossa on erilaisia toimijoita, jotka käyttävät dataa mm. mainonnan tehokkaaseen kohdentamiseen.

DATA BROKER: Yritys, jonka liiketoiminta perustuu datan keräämiseen ja yhdistelyyn ja siitä syntyvien datatuotteiden myyntiin.

DATATALOUS: Datatalous on talouden osa-alue, jossa liiketoimintamalli perustuu tiedon hyödyntämiseen ja käyttöön eri tavoin.

GDPR: Asetus (EU) 2016/679 eli Euroopan unionin uusi tietosuojasetus (GDPR) säätelee yksilön, yrityksen tai organisaation tekemää henkilötietojen käsittelyä EU:ssa. Asetus on erittäin tärkeä keino vahvistaa digiajan yksilöiden perusoikeuksia ja helpottaa liiketoimintaa selkeyttämällä sääntöjä, jotka koskevat yrityksiä ja julkisia toimijoita digitaalisilla sisämarkkinoilla. Asetus tuli voimaan 24.5.2016, ja sitä on sovellettu 25.5.2018 alkaen

HENKILÖTIETO: Kaikki sellainen tieto, jolla voidaan tunnistaa ja yksilöidä henkilöitä. Näitä tietoja ovat esimerkiksi nimi, osoite, sähköpostiosoite, henkilötunnus tai muut tunnistetiedot sekä verkkotunnistetiedot.

KOLMANNET OSAPUOLET (THIRD-PARTY DATA COMPANIES): Yrityksiä, jotka esim. keräävät, yhdistelevät, rikastavat ja myyvät dataa ihmisistä, mutta joilla ei kuitenkaan ole suoraa yhteyttä kuluttajaan. Esimerkiksi verkkosivu, jota kuluttaja käyttää, on ensimmäinen osapuoli ja sivulla mainontaa myyvä taho kolmas osapuoli.

MOBIILISOVELLUS TAI MOBIILIAPPLIKAATIO: Ohjelmisto, joka on suunniteltu toimimaan mobiililaitteilla, kuten esimerkiksi matkapuhelimilla, tableteilla tai älykelloilla. Mobiilisovelluksilla on lukemattomia erilaisia käyttötarkoituksia uutisista peleihin, tai pörssikursseista kuvankäsittelyyn.

OHJELMALLINEN MAINONTA: Automaattinen, datapohjainen mainonnan ostamisen ja myymisen tapa, joka perustuu algoritmeihin ja jonka yleisin muoto on digitaalinen huuto-kauppa. Programmatic Advertising.

OMADATA: Omadata eli MyData on henkilötietojen hallinnan ja käsittelyn periaate, jonka mukaan ihmisillä on oltava mahdollisuus hallita, hyödyntää ja luovuttaa eteenpäin heistä kerättäviä henkilötietoja (esim. teletunnistetiedot, terveystiedot, kuten tiedot yksilön geeniperimästä, energiatiedot, ostotiedot, liikkumistiedot, taloustiedot, ja verkkopalveluihin tallentuva tieto).

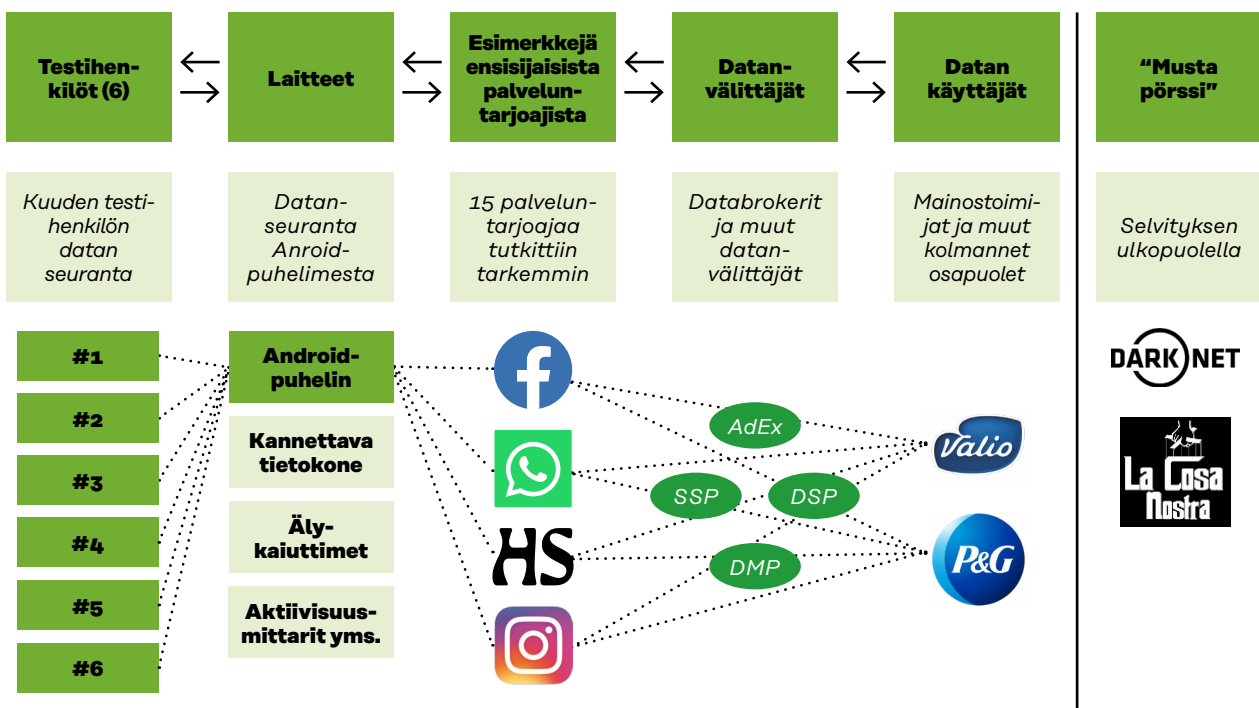
MONITOROINTISOVELLUS: Applikaatio, joka rakennettiin digijälkiselvitysprojektissa ja asennettiin testihenkilöiden testipuhelimiin. Sillä seurattiin testihenkilöiden käyttämiä palveluja.

VPN (VIRTUAL PRIVATE NETWORK): Tässä projektista testihenkilöiden data ohjattiin kulkemaan suojatun VPN-yhteyden kautta pilvessä olevalle palvelimelle, josta se tallennettiin.

Liite 1. Selvityksen laajuus ja datan kulun seurantamenetelmä

Henkilökohtaiseen dataan liittyvä datatalouden ekosysteemi on erittäin laaja ja kompleksinen. Suorien datan kerääjien lisäksi erilaisia osapuolia on paljon, ja harmaiden alueiden takia koko ekosysteemin kuvaaminen on hyvin haastava tehtävä.

KUVA 1. DIGITAALISEN MAINONNAN EKOSYSTEEMI ON MONIMUTKAINEN



Datan kulkua seurattiin niin ensimmäisen kuin kolmannen osapuolen toimijoille.

Tässä selvityksessä tunnistettiin erilaisia kolmansia osapuolia WebXRay-työkalun (<https://webxray.eu/>) tarjoaman ilmaisen tietokannan avulla. Se on tutkijoiden luoma ja käyttämä tietokanta, jonka avulla on mahdollista tunnistaa ja luokitella yleisimpiä verkon kolmansia osapuolia, jotka liittyvät mainontaan, markkinointiin ja profilointiin. On kuitenkin huomattava, että datan keräämiseen liittyvä ekosysteemi on jatkuvassa muutoksessa, joten tämäkään työkalu ei välttämättä tunnista kaikkia toimijoita vaan osa palveluista jouduttiin tunnistamaan käsityönä.

Datan kulun seurannan menetelmä

TESTIPUHELIMET

Kaikki testihenkilöt käyttivät samanlaisia Android-puhelimia. Testihenkilöitä ohjeistettiin käyttämään sensitiivisimpiä palveluja, kuten pankkipalveluja, muilla laitteilla kuin testipuhelimilla. Testiä varten tarvittiin Android-puhelimet, koska puhelimiin tehtiin monitorointisovel-

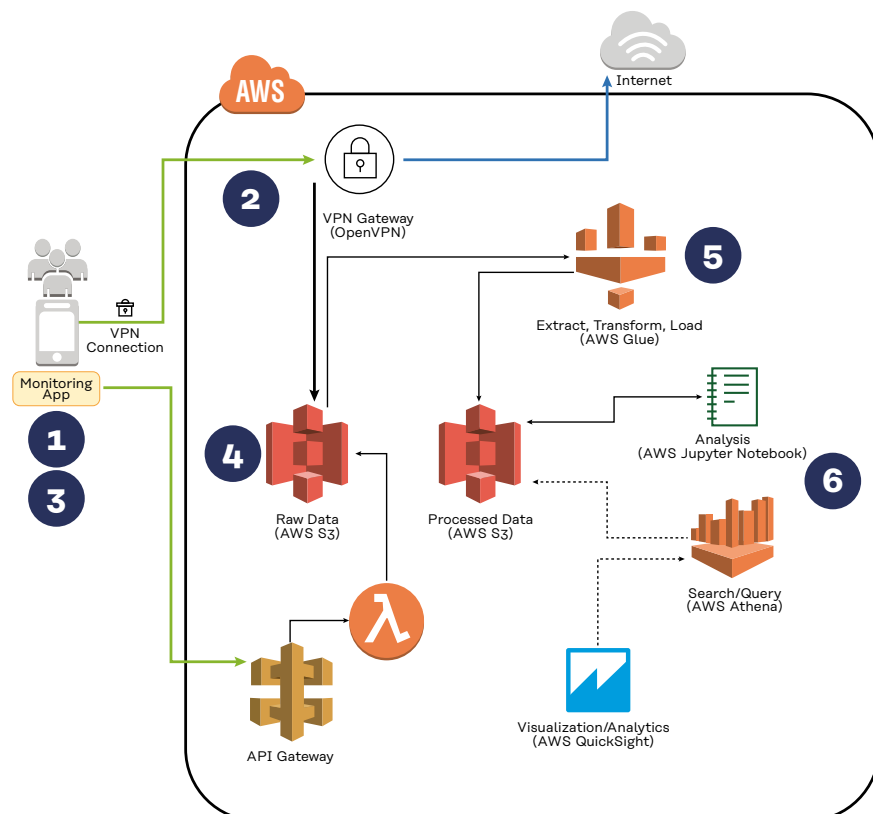
lus, joka seurasi, mitä ohjelmia testihenkilöt käyttivät. Applen iPhone-puhelimiin tätä ei teknisesti voitu tehdä. Puhelimien merkki oli Motorola ja malli Moto e5 plus. Puhtaan Androidin käyttö mahdollisti puhelimien ns. roottaamisen, jotta niihin pystyttiin vaihtamaan proxy-asetukset. Android-puhelimeissa ja -tableteissa on mahdollista ottaa käyttöön sovelluskehittäjien asetukset eli ylimääräinen asetusvalikko, joka vie käyttäjän perusasetuksia syvemmälle (roottaaminen). Tämä mahdollistaa esimerkiksi puhelinvalmistajien omien sovellusten poiston ja muita vastaavia operaatioita, jotka eivät välttämättä muuten onnistuisi valmistajien omien ohjelmistosuojausten vuoksi.

MONITOROINTISOVELLUS

Käytännössä analyysivaiheessa huomattiin, että monitorointisovelluksen keräämällä datalla on rajoituksia. Ohjelma erotti erilliset aktiiviset sovellukset. Kuitenkin suurin osa liikenteestä tuli Google Chrome -selaimesta, varsinkin kun testihenkilöitä kannustettiin käyttämään eri sovelluksia selaimella, jotta suojaetuimpien sovellusten, kuten Facebookin, data saataisiin käyttöön. Chrome-datasta monitorointiohjelma ei pystynyt erottelemaan esimerkiksi sitä, mille verkkosivulle mikin liikennepaketti kuului. Lisäksi osa verkkoliikenteestä liittyi taustalla pyöriiviin sovelluksiin, mutta näitä monitorointiohjelma ei pystynyt erottamaan aktiivisen sovelluksen datasta. Käytännössä suurin osa verkkoliikenteen analysoinnista perustuu seuraavassa kappaleessa kuvattuun kerättyyn liikennedataan ja sen suoraan analysointiin.

LIIKENNETATAN KERÄÄMISEEN LIITTYVÄ INFRASTRUKTUURI

KUVA 2. PUHELIMIIN ASENNETTIIN MONITOROINTISOVELLUS JA VPN-SOVELLUS JA VAIHDETTIIN PROXY-ASETUKSET VERKKOLIIKENTEEN SSL-SALAUKSEN OHITTAMISEKSI



Puhelimiin asennettiin monitorointisovellus ja VPN-sovellus ja vaihdettiin proxy-asetukset verkkoliikenteen ssl-salauksen ohittamiseksi (1). Liikennedata ohjattiin VPN-sovelluksen avulla kulkemaan VPN-tunnelia pitkin (2) ja Open VPN -serverin läpi. Open VPN -serveri oli sijoitettu AWS-pilvipalveluun. Open VPN -serverillä liikennedata tallennettiin AWS:n S3-ympäristöön (4). Tallentamisella ei häiritse itse dataliikennettä, vaan liikenne jatkoj Open VPN -serverin jälkeen muuttumattomana internetissä olevaan kohteeseensa.

Myös monitorointiohjelman data tallennettiin S3-palveluun. Näiden lisäksi analyysissä käytettiin ulkopuolisia aineistoja, kuten WebXRay-tietokantaa erilaisista datatalouden kolmansista osapuolista. Myös tämä data tallennettiin S3-palveluun. Nämä datajoukot yhdistettiin liikennedatan kanssa, ja saatu datajoukko puhdistettiin (5) ja tallennettiin uudestaan data-analyysia varten (6).

LIIKENNEDATAN MUOTO

Verkkoliikennedata koostuu HTTP-kyselyistä. Kyselyiden perusmuoto on aina sama: käyttäjän laite lähettää johonkin ip-osoitteeseen pyynnön (request) ja saa siihen kohteessa olevalta palvelimelta vastauksen (response). Näitä kutsutaan myös paketeiksi. Web-liikenteessä on käytössä kahden version HTTP-kyselyitä, vanhempaa HTTP/1.1-versiota ja uudempaa HTTP/2.0-versiota. Protokollissa on jonkin verran eroa, ja ne määräävät raamit sille, miten dataa siirretään ja millaista dataa ja metadattaa kyselyihin voi liittyä. Itse paketit voivat kuitenkin olla hyvin erilaisia ja sisältää hyvin erilaista dataa. Pakettien sisältönä voi olla myös erilaisia esim. Javascript-ohjelmia, joita voidaan myös käyttää datan kuljettamiseen. Tärkeimmät tässä selvityksessä käytetyt verkkoliikenteessä kulkevat tiedot ovat

1. paketin authority/host, joka kertoo mihin domain-osoitteeseen paketti on lähetetty
2. paketin datan tyyppi, eli millaista dataa kulkee
3. paketin queryString
4. paketin data, joskin se on useimmiten salattua/koodattua
5. paketin sisältämät evästeet
6. viittausinformaatio, eli onko jokin toinen palvelu pyytänyt lähettämään dataa myös muulle serverille kuin sen omille (tätä tietoa ei ole kaikissa paketeissa).

DATAN KERÄYSJAKSO JA DATAN KATTAVUUS

Selvityksessä analysoitiin kahden viikon ajalta kerätty testihenkilöiden verkkoliikennedata. Aineisto sisälsi noin 84 000 HTTP-pyyntö-vastauspakettia. Eri henkilöiden tuottamat pakettimäärät vaihtelivat 9 000:n ja 21 000:n välillä.

TESTIHENKILÖIDEN DATAN KULKEUTUMINEN KOLMANSILLE OSAPUOLILLE

Jokaisen testihenkilön osalta tarkasteltiin tilannetta tarkemmin kahdesta kulmasta.

1) Kultakin testihenkilöltä analysoitiin yhdestä kolmeen hänen käyttämäänsä sovellusta tai verkkosivua. Näiden palvelujen käytön aikaisesta liikenteestä etsittiin erilaisia kolmansia osapuolia ja tunnistettiin ne WebXRay-tietokannan sekä manuaalisen tarkistuksen avulla. Palvelut pyrittiin valitsemaan mahdollisimman monipuolisesti, jotta tulisi näkyviin, miten erilaisia kolmansia osapuolia eri sovellukset käyttävät. Tyypillisesti palvelut käyttivät useita kolmansia osapuolia, ja nämä toimijat esitellään tarkemmin.

2) Kunkin henkilön datasta etsittiin WebXRay-tietokannan avulla mainontaan liittyvät toimijat ja selvitettiin, mitkä verkkosivustot tai sovellukset käyttävät näitä kolmansia osapuolia. Tässä keskityttiin mainostajiin, koska heidän roolinsa on usein selkein.

Perinteisesti mainontaan liittyvät toimijat voivat ostaa kohderyhmätietoa muilta toimijoilta sekä kerätä itse dataa käyttäjistä. Dataa kerätään ja jaetaan eri toimijoiden kesken asettamalla käyttäjän koneelle lyhyitä ohjelmia (skriptejä) tai ns. kolmannen osapuolen evästeitä. Näiden avulla mainontaan liittyvät toimijat voivat tunnistaa samat käyttäjät eri verkkosivuilla ja rakentaa heidän käyttäytymisestään profileja. Tässä selvityksessä selvitettiin, mille sivuille mainontaan liittyvät toimijat olivat asettaneet evästeitä.

GOOGLE JA FACEBOOK

Google ja Facebook ovat teknologiajättiläisiä ja superalustoja, jotka pystyvät keräämään dataa käyttäjistä sekä omien palvelujensa että muiden sivustojen kautta ja myymään tämän tiedon avulla mainoksia. Esimerkiksi Facebook kerää tietoa tykkäysnappien ja evästeiden avulla. Vuonna 2019 Facebookilla oli 2,4 miljardia kuukausittaista käyttäjää ja Googlella 1,5 miljardia Gmail-sähköpostin käyttäjää. Ne eivät jaa suoraan tietoa käyttäjistään mainostajien kanssa vaan muodostavat datansa avulla mainostajille enemmän tai vähemmän tarkkoja kohderyhmiä: esim. jalkapallosta kiinnostuneet, Kalliossa asuvat tai keski-ikäiset naiset.

Facebook ja Google keräävät perinteisesti tietoa käyttäjistä asettamiensa kolmannen osapuolen evästeiden avulla. Käyttäjätietojen keräys on kuitenkin murroksessa. GDPR vaatii, että käyttäjältä pyydetään suostumus ennen kolmannen osapuolen evästeiden asettamista. Selaimista Safari ja Firefox ovat muuttuneet siten, että ne estävät oletusasetuksena kolmannen osapuolen evästeet kokonaan ja jopa rajoittavat ensimmäisen osapuolen evästeiden säilytysaikaa. Myös Googlen Chrome-selain tulee Googlen ilmoituksen mukaan antamaan käyttäjille uusia mahdollisuuksia hallita kolmansien osapuolten evästeitä.

Tämän kiertämiseksi ainakin Facebook ja Google ovat ryhtyneet keräämään dataa myös sivustojen itse asettamilla eli ensimmäisen osapuolen pikseleillä ja evästeillä. Tätä varten liikennedatasta etsittiin nyt ensin suorat Googlen ja Facebookin kolmannen osapuolen evästeet. Lisäksi etsittiin tietoa siitä, milloin Google tai Facebook asettaa evästeitä ensimmäisen osapuolen evästeiden seassa. Nämä evästeet tunnistettiin pelkästään evästeiden nimistä eli Facebookin tapauksessa tunnuksista ”_fbp” ja ”_fbc” ja Googlen tapauksessa ”__gads” ja ”__gac”. On epätodennäköistä, että muut palvelut käyttäisivät näitä nimiä omissa evästeissään.

MITÄ VERKKOLIIKENNEPAKETTIESSÄ SISÄLLÄ TARKALLEEN ON?

Vaikka selvityksessä verkkoliikenteen päälinjojen salaus purettiin Virtual Private Network (VPN) -tunnelin avulla, jolloin nähtiin pakettien sisään, niiden sisällä data kulki eri tavoin salattuna. Yritykset haluavat pitää itsellään sekä datan sisältämän liiketoimintatiedon (datatoukon yritysten pääomaa) että käyttäjiin kytkeytyvän tiedon, eli sen päätyminen kilpailijoiden tai muiden uteliaiden käsiin halutaan estää. Toisaalta tässä selvityksessä näimme pääosin, mihin verkko-osoitteisiin ja kuinka paljon dataa kulki, missä muodossa se oli sekä oliko http-kutsuilla jokin ”lähettäjä sivusto”. Emme kuitenkaan tiedä yksityiskohtaisesti, mitä tietoa lähetettiin.

Liite 2. Esimerkkejä testihenkilöiden digitaalisten palvelujen käytön yhteydessä löydetyistä kolmansista osapuolista

ADFORM on yksi suurimmista mainosteknologiaan liittyvistä yhtiöistä. AdForm ei ole amerikkalainen, kuten muut suuret alan toimijat, vaan kotoisin Tanskasta.

APPNEXUS (nyk. Xandr) on jättiläismäinen mainosteknologiayritys, jonka omistaa amerikkalainen suuryritys AT&T. Se tarjoaa [portaalin](#), jonka kautta yksittäiset käyttäjät voivat tarkistaa, millaisiin segmentteihin heitä on luokiteltu AppNexuksen datassa.

CONDENAST-KONSERNIN lehdille, kuten [vogue.comille](#) ja [wired.comille](#), lähteneissä paketeissa oli sisältönä yksi parametri, sama tunniste. Vastauksena tullut paketti tallensi näiden kolmantena osapuolena olevien lehtien evästeiden käyttäjän puhelimeen.

CHARBEAT analysoi sovelluksen käyttäjien toimintaa. Tietosuojaselosteen mukaan ennen datan tallentamista ip-osoitteesta poistetaan viimeinen oktetti, joten sitä ei voi yhdistää käyttäjiin. Lisäksi yritys ilmoittaa, ettei se säilö muutakaan tietoa, josta käyttäjän voisi tunnistaa.

DOUBLECLICK JA ADSENSE ovat molemmat Google-mainontaan liittyviä yhtiöitä. Kummankin tietosuojaselosteena toimii Googlen yleinen tietosuojaseloste. AdSensen sivulla mainitaan, että palvelua käyttävien sivustojen pitäisi mainita Käyttäjilleen, että ”third party vendors, including Google, use cookies to serve ads based on user’s prior visits to your website or other websites”. Tässä sanamuoto on hyvin heikko, ja siinä halutaan korostaa, että palvelu on vain osa monista toimijoista.

FEEDBAKLY on suomalainen yritys, joka kerää tietoa verkkosivusta ja optimoi asiakaspolkuja.

FLURRY ANALYTICS on analytiikkapalvelu, jonka tarkoituksena sanotaan olevan sovelluksen optimointi. Sivuston perusteella ei ole helppo ymmärtää, mitä dataa se yksilöstä käyttää ja miten sitä yhdistetään. Lisäksi palvelun tietosuojalinkki vie Verizon Median tietosuojakäytäntöön. Verizon Media on osa jättiläismäistä Verizon Communications yhtiötä, joka mm. omistaa Yahoona. Linkitetty tietosuojakäytäntö ei yksilöi Flurry Analyticsin käyttämää dataa, ja siellä kerrotaan, että dataa jaetaan Verizon median ja ”luotettujen partnereiden” kesken.

HOTJAR on työkalu, jolla analysoidaan käyttäjien polkuja verkkosivustolla. Se voi mitata esimerkiksi klikkauksia tai sivun vieritystä. Työkalu voi myös antaa palautetta siitä, kuinka

sivua voi parantaa. Palvelun tietosuojakuvauksen mukaan henkilökohtaista dataa myydä ulkopuolisille.

MOAT analysoi mainonnan vaikuttavuutta ja sitä, miten ja missä mainokset ovat näkyneet (attention analytics). Se on osa Oracle Cloudia. Oraclen tietosuojakäytäntö sisältää paljon linkkejä, eikä sieltä ole helppo löytää mainintaa siitä, miten Moatin dataa käytetään.

TAPJOY on mainonnan myyntiin liittyvä palvelu. Kerätty liikennedata paljastaa, että se hyödyntää ns. tracking ID:tä, joka on laite- ja käyttäjäkohtainen tunniste. Tämä tekee TapJoylle mahdolliseksi yhdistää dataa eri palveluista.

Liite 3. Palvelun tuottajille lähetetty kyselylomake

TO WHOM IT MAY CONCERN:

I am hereby requesting access and portability of my personal data according to 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; 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 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.

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

SITRAN SELVITYKSIÄ 168

Sitran selvityksiä -sarjassa julkaistaan Sitran tulevaisuustyön ja kokeilujen tuloksia.

ISBN 978-952-347-178-8 (PDF) www.sitra.fi

ISSN 1796-7112 (PDF) www.sitra.fi


SITRA.FI

Itämerenkatu 11–13

PL 160

00181 Helsinki

P. 0294 618 991

 @SitraFund