TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

TIETOVUODOT KOLMANSILLE OSAPUOLILLE SUOMALAISISSA TERVEYDENHUOLLON VERKKOPALVELUISSA

Sampsa Rauti

TURUN YLIOPISTON JULKAISUJA ANNALES UNIVERSITATIS TURKUENSIS
SARJA SER. F OSA TOM. 54 | TECHNICA INFORMATICA | TURKU 2025

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

TIETOVUODOT KOLMANSILLE OSAPUOLILLE SUOMALAISISSA TERVEYDENHUOLLON VERKKOPALVELUISSA

Sampsa Rauti

TURUN YLIOPISTON JULKAISUJA ANNALES UNIVERSITATIS TURKUENSIS
SARJA SER. F OSA TOM. 54 | TECHNICA INFORMATICA | TURKU 2025

Turun yliopisto

Teknillinen tiedekunta
Tietotekniikan laitos
Tietotekniikka
Teknologian tohtoriohjelma

Ohjaajat

Professori, FT Ville Leppänen
Turun yliopisto

Apulaisprofessori, TkT Tuomas Mäkilä
Turun yliopisto

Tarkastajat

Professori, TtT Kristina Mikkonen
Oulun yliopisto

Professori, TkT Kari Systä
Tampereen yliopisto

Vastaväittäjä

Professori, TkT Pasi Tyrväinen
Jyväskylän yliopisto

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck-järjestelmällä.

ISBN 978-952-02-0120-3 (Painettu)
ISBN 978-952-02-0121-0 (PDF)
ISSN 2736-9390 (Painettu)
ISSN 2736-9684 (Sähköinen)
Painosalama, Turku, 2025

Omistettu vanhemmilleni

TURUN YLIOPISTO
Teknillinen tiedekunta
Tietotekniikan laitos
Tietotekniikka
RAUTI, SAMPSA: Tietovuodot kolmansille osapuolille suomalaisissa terveydenhuollon verkkopalveluissa
Väitöskirja, 214 s.
Teknologian tohtoriohjelma
Huhtikuu 2025

TIIVISTELMÄ

Verkkosivustot ovat nykyään tärkeä työkalu monien kriittisten ja välttämättömien arkipäivän palvelujen tarjoamiseen. Verkkopalvelujen määrän kasvaessa ja käytön lisääntyessä yksityisyyskysymyksistäkin tulee entistä olennaisempia. Moderneissa verkkopalveluissa suuri huolenaihe ovat niihin sisällytettävät kolmannen osapuolen palvelut, kuten erilaiset analytiikkatyökalut, ja näiden palvelujen käyttäjistä keräämät tiedot. Tässä väitöstutkimuksessa käsitellään suomalaisten verkkopohjaisia terveyspalvelujen yksityisyyttä keskittyen niissä esiintyviin kolmannen osapuolen palveluihin ja näille vuotaviin henkilötietoihin. Koska palveluissa käsitellään esimerkiksi asiakkaiden lääkityksiin, sairauksiin ja hoitotoimenpiteisiin liittyviä tietoja, niitä ei saisi päätyä kolmansien osapuolien haltuun.

Väitöstutkimus pohjautuu pitkälti lukuisissa terveysalan verkkopalveluissa – verkkoapteekeissa, terveydenhuollon verkkopalveluissa ja erilaisilla mielenterveysalan toimijoiden sivustoilla – tehtyyn verkkoliikenneanalyysiin, jonka avulla selvitetään näistä palveluista kolmansille osapuolille vuotavia henkilötietoja. Tämän ohella tarkastellaan myös, tuleeko käyttäjä asianmukaisesti informoiduksi henkilötietojensa käsittelystä ja suostutellaanko häntä hyväksymään tiedonkeruu harhaanjohtavin keinoin. Verkkoliikenneanalyysin lisäksi väitöstutkimuksessa haastateltiin web-kehittäjiä tietovuotojen syiden selvittämiseksi ja ehkäisevien toimien löytämiseksi.

Tutkimus osoittaa, että suomalaisilla terveysalan sivustoilla on ollut vakavia tietovuotoja. Kolmansien osapuolten keräämien tietojen joukossa on mm. reseptilääkkeiden nimiä ja terveydenhuollon ajanvarauksiin liittyviä tietoja yhdistettyinä käyttäjän yksilöiviin henkilötietoihin kuten IP-osoitteeseen. Esimerkiksi tarkastelluista 163 verkkoapteekista 57 (35,0 %) vuoti reseptilääkkeiden nimiä kolmansille osapuolille. Vaikka tilanne on julkisen sektorin terveyspalveluissa parempi kuin yksityisellä sektorilla, on molemmilla runsaasti parannettavaa henkilötietojen käsittelyssä. Palvelujen käyttäjä puolestaan ei useinkaan tiedonkeruun kieltämälläkään voi estää näitä tietovuotoja, ja käyttäjän informointi henkilötietojen siirtymisestä kolmansille osapuolille on myös tavallisesti riittämätöntä. Löydökset antavat aihetta tarkemmalle regulaation noudattamiselle, auditoinneille, lokaalien analytiikkaratkaisujen suosimiselle ja huolellisemmalle kolmansien osapuolien valinnalle erityisesti kriittisten terveysalan verkkopalvelujen osalta.

ASIASANAT: Tietovuodot, yksityisyys, verkkopalvelut, kolmannet osapuolet

UNIVERSITY OF TURKU
Faculty of Technology
Department of Computing
Information and Communication Technology
RAUTI, SAMPSA: Data leaks to third parties in Finnish web-based health services
Doctoral dissertation, 214 pp.
Doctoral Programme in Technology
April 2025

ABSTRACT

Today, websites are an important tool for delivering critical and essential services to accomplish everyday tasks. As the number and use of web services grow, privacy issues become increasingly important. A significant concern in modern web services is the third-party tools used as parts of these websites, such as various web analytics tools, and the personal data these services collect. This dissertation discusses the privacy of Finnish web-based healthcare services, focusing on the third parties and the personal data leaks to them. Because these services process data concerning, for instance, prescription medicines, diseases, and treatments, they should not fall into the hands of third parties.

The current thesis is based on network traffic analysis conducted on several healthcare websites, such as online pharmacies, web-based health services and mental health websites. With this analysis, we study the types of personal data that may potentially leak to third parties from the inspected web services. We also examine whether users are appropriately informed of the data processing activities taking place on the studied websites and whether they are manipulated into accepting data collection through deceptive methods. Besides network traffic analysis, we carry out expert interviews to study the reasons for data leaks and the ways to prevent them.

The study shows that there have been serious data leaks on Finnish healthcare websites. Among the data collected by third parties were, for instance, prescription medicine names and information on appointment bookings linked to uniquely identifying personal data such as IP addresses. For example, 57 out of the 163 studied online pharmacies (35.0%) leaked prescription medicine names to third parties. Although the situation in public sector health services is better than in the private sector, both have much to improve in processing personal data. Users of health services often cannot prevent these data leaks, even by declining data collection, and they are not usually sufficiently informed about data transfers to third parties. The findings call attention to better compliance with existing regulation, auditing, the preference for local analytics solutions, and more careful selection of third parties, especially in the case of critical health services.

KEYWORDS: Data leaks, privacy, web services, third parties

Kiitokset

Kiitän Ville Leppästä paitsi väitöskirjatyöni ohjaamisesta, myös erinomaisena esimiehenä toimimisesta yli kymmenen vuoden ajan. Villen ansiosta olen päässyt osallistumaan lukuisiin haastaviin ja mielenkiintoisiin tehtäviin ja projekteihin. Merkittävin näistä lienee tämänkin väitöskirjan syntyyn johtanut Intimacy in Data-Driven Culture -projekti, jonka osalta olen saanut erittäin vapaat kädet toteuttaa tutkimusta itsenäisesti haluamallani tavalla. Ville on myös antanut minulle riittävät resurssit tutkimuksen tekemiseen ja tarvittaessa ohjannut projektia oikeaan suuntaan. Kiitokset myös Tuomas Mäkilälle väitöstyön ohjauksesta sekä hyvistä neuvoista ja mielenkiintoisista uusista näkökulmista työn loppuvaiheessa.

Kiitän Kristina Mikkosta ja Kari Systää väitöskirjan esitarkastajina toimimisesta. Heidän palautteensa auttoi merkittävästi työn laadun parantamisessa ja saattamisessa lopulliseen muotoonsa. Kiitokset myös Pasi Tyrväiselle, joka suostui vastaväittäjäkseni.

Olen kiitollinen myös Samuli Laatolle mainiosta tutkimusyhteistyöstä, joka on yksityisyys- ja tietoturva-aiheiden lisäksi usein liittynyt niinkin yllättäviin teemoihin kuin Pokémon GO -peliin tai digiteologiaan. Tämä on sekä laajentanut omaa tutkimuskenttääni että tarjonnut tervetullutta vaihtelua tavallisiin tutkimusrutiineihin.

Esitän kiitokset myös kaikille Agoran huoneessa 452A työskenteleville, työskennelleille tai siellä muuten vain vieraileville henkilöille. Jari-Matti Mäkelän ja Anne-Maarit Majanojan kanssa olemme käyneet monia mielenkiintoisia keskusteluja, ihmetelleet yliopisto-opetukseen liittyviä haasteita ja ratkaisseet monenlaisia käytännön ongelmia. Jari Lehto ja Jarko Papalitsas ovat merkittävästi keventäneet opetustaakkaani ja mahdollistaneet sen, että olen voinut työskennellä IDA-projektin parissa ja kirjoittaa tämän väitöskirjan. Kiitän Tapani Joelssonia ja Paavo Nevalaista tervetulleista keskeytyksistä ja tiedusteluista väitöskirjani valmistumiseen liittyen; ne ovat nyt tuottaneet tulosta. Kiitokset myös Johannes Holvitielle, Shohreh Hosseinzadehille, Sami Hyrynsalmelle ja Jukka Ruohoselle vuosien yhteistyöstä.

Tähän väitöskirjaan liittyvä tutkimus on kaikilta osin ollut erinomaisen tutkimusryhmän yhteistyön tulosta. Kiitän kaikkia IDA-tutkimusryhmän tutkimusavustajia ja projektitutkijoita; Timi Heinoa, Robin Carlssonia, Panu Puhtilaa, Sammani Rajapakshaa, Henna Lohea ja Esko Vuorista. Tämän tutkimusryhmän ohjaaminen on ollut tutkijanurani hienoin kokemus, ja yhdessä olemme tehneet yhteiskunnallisesti hyvin merkittäviä löytöjä. Suuret kiitokset myös Lauri Koivuselle teknisestä

avusta ja pitkästä tutkimusyhteistyöstä sekä Kalle Hjerppelle IDA-projektin rikastamisesta uusilla mielenkiintoisilla tutkimusaiheilla. Lisäksi kiitän Sini Mickelssonia erinomaisesta poikkitieteellisestä yhteistyöstä IDA-projektissa ja oikeustieteellisen näkökulman tuomisesta tietovuototutkimukseemme.

Lopuksi kiitän vanhempiani Siro ja Jukka-Pekka Rautia, jotka ovat kaikin tavoin tukeneet minua opiskeluni ja elämäni joka vaiheessa.

Turussa 2.4.2025
Sampsa Rauti

Sisällys

Kiitokset . . . . . **iii**

Sisällys . . . . . **v**

Artikkelit . . . . . **vii**

1 Johdanto **1**

2 Yksityisyys ja kolmannet osapuolet terveyspalveluissa . . **4**
2.1 Yksityisyys ja tietovuodot 4
2.2 Tietoturva 6
2.3 Web-analytiikka ja kolmannet osapuolet 8
2.3.1 Web-analytiikan perusteet 8
2.3.2 Kolmansille osapuolille lähetettäviä henkilötietoja . 11
2.3.3 Tiedonkeruu verkkosivulla 15
2.3.4 Analytiikkaan liittyvä aiempi tutkimus 16
2.4 Verkkoapteekit ja yksityisyys 18
2.5 Terveydenhuollon verkkopalvelut ja yksityisyys 20
2.6 Motivaatioita tutkimukselle 21

3 Tutkimustyön kuvaus **25**
3.1 Tutkimuskysymykset 25
3.2 Tutkimusasetelma, aineistot ja menetelmät 27
3.2.1 Aineistot 28
3.2.2 Verkkoliikenteen nauhoittaminen ja analyysi 29
3.2.3 Tapaustutkimus 31
3.2.4 Tilastollinen analyysi 31
3.2.5 Asiantuntijahaastattelut 32
3.2.6 Gioia-menetelmä 32
3.3 Eettiset kysymykset 33
3.4 Tutkimusprosessi ja julkaisut 34
3.5 Yhteenvedot alkuperäisjulkaisujen sisällöistä 36

3.5.1 Julkaisu I: Assessing discrepancies between network traffic and privacy policies of public sector web services . . . 36
3.5.2 Julkaisu II: Several Online Pharmacies Leak Sensitive Health Data to Third Parties . . . 39
3.5.3 Julkaisu III: Analyzing third-party data leaks on online pharmacy websites . . . 42
3.5.4 Julkaisu IV: Health data leaks to third parties in web-based health services . . . 45
3.5.5 Julkaisu V: Analysis of Third-Party Data Leaks on Finnish Mental Health Websites . . . 49
3.5.6 Julkaisu VI: Data Leaks to Third-Party Services on Medical Websites . . . 52
3.5.7 Julkaisu VII: Third-party data leaks on municipal websites . . . 56
3.5.8 Julkaisu VIII: Leaky Democracy: Third Parties in Voting Advice Applications . . . 58
3.5.9 Julkaisu IX: Insights from Developers: Understanding Third-Party Data Leaks in Web Environment . . . 62

4 Johtopäätökset . . . 67
4.1 Keskeisimmät löydökset . . . 67
4.2 Kontribuutiot . . . 67
4.2.1 Tutkimuskysymys TK1 . . . 68
4.2.2 Tutkimuskysymys TK2 . . . 71
4.2.3 Tutkimuskysymys TK3 . . . 72
4.2.4 Tutkimuskysymys TK4 . . . 74
4.3 Tutkimuksen rajoitukset . . . 78
4.4 Jatkotutkimusaiheet . . . 79

Lähteet . . . 81

Alkuperäisjulkaisut . . . 87

Alkuperäisjulkaisut

Tämä väitöskirja perustuu seuraaviin julkaisuihin, joihin viitataan jatkossa roomalaisilla numeroilla:

I Timi Heino, Robin Carlsson, Sampsa Rauti, Ville Leppänen. Assessing discrepancies between network traffic and privacy policies of public sector web services. Teoksessa: Proceedings of the 17th International Conference on Availability, Reliability and Security, 2022, ss. 1–6, ACM.

II Robin Carlsson, Sampsa Rauti, Sini Mickelsson, Tuomas Mäkilä, Timi Heino, Elina Pirjatanniemi, Ville Leppänen. Several online pharmacies leak sensitive health data to third parties. Teoksessa: Proceedings of World Conference on Information Systems and Technologies, 2023, ss. 164–175, Springer.

III Sampsa Rauti, Robin Carlsson, Sini Mickelsson, Tuomas Mäkilä, Timi Heino, Elina Pirjatanniemi, Ville Leppänen. Analyzing third-party data leaks on online pharmacy websites. Health & Technology, vol. 14, 2024, ss. 1–18.

IV Sampsa Rauti, Robin Carlsson, Samuli Laato, Timi Heino, Panu Puhtila, Ville Leppänen. Health data leaks to third parties in web-based health services. Teoksessa: Proceedings of the Annual Doctoral Symposium of Computer Science, 2024, ss. 117–121, CEUR workshop proceedings.

V Sampsa Rauti, Esko Vuorinen, Panu Puhtila, Robin Carlsson. Analysis of Third-Party Data Leaks on Finnish Mental Health Websites. Teoksessa: Proceedings of the 47th MIPRO ICT and Electronics Convention, 2024, ss. 1543–1548, IEEE.

VI Sampsa Rauti, Esko Vuorinen, Robin Carlsson, Panu Puhtila. Data Leaks to Third-Party Services on Medical Websites. Teoksessa: Proceedings of 16th International Conference on Security of Information and Networks, 2023, ss. 1–7, IEEE.

VII Sampsa Rauti, Robin Carlsson, Panu Puhtila, Ville Leppänen. Third-party data leaks on municipal websites. Teoksessa: International Congress on

Information and Communication Technology, 2024, ss. 599–610, Springer Nature.

VIII Timi Heino, Sampsa Rauti, Samuli Laato, Robin Carlsson, Ville Leppänen. Leaky Democracy: Third Parties in Voting Advice Applications. Teoksessa: Proceedings of 8th International Conference on Smart Computing and Communication, 2024, ss. 351–360, Springer.

IX Sampsa Rauti, Panu Puhtila, Samuli Laato, Tuomas Mäkilä. Insights from Developers: Preventing Third-Party Data Leaks in Web Environment. Lähetetty vertaisarviointiin.

Julkaisut on liitetty väitöskirjaan tekijänoikeuksien omistajien luvalla.

1 Johdanto

Verkkosivustot ovat tehokas työkalu monien välttämättömien arkipäivän palvelujen tarjoamiseen, kun asiakkailla on haasteita päästä hyödyntämään palveluja fyysisissä toimipisteissä. Erityisesti monet haavoittuvassa asemassa olevat ihmiset, kuten vakavasti sairaat ja syrjäseuduilla asuvat, voivat hyötyä merkittävästi sähköisistä palveluista [1]. Kun yhä suurempi osa väestöstä pääsee hyödyntämään verkkopalveluja, myös lainsäätäjät ovat ymmärtäneet sähköisten palvelujen tärkeyden. Suomessa säädettiin vuonna 2019 laki digitaalisten palveluiden tarjoamisesta (306/2019)[1]. Lain tarkoituksena on parantaa digitaalisten palveluiden saavutettavuutta, laatua sekä tietoturvallisuutta, jotta kaikki voivat hyödyntää niitä mahdollisimman tasapuolisesti. Viime vuosina COVID-19-pandemia on entisestään lisännyt verkkopalvelujen tarvetta ja kiihdyttänyt digitalisaatiokehitystä [2].

Verkkopalvelujen käytön lisääntyessä yksityisyyskysymyksistä tulee yhä olennaisempia [3]. Verkossa tarjottavista välttämättömistä palveluista tärkeitä esimerkkejä ovat apteekki- ja terveyspalvelut, joiden yksityisyyshaasteiden tarkasteluun tämä väitöstutkimus keskittyy. Näissä palveluissa käsitellään hyvin arkaluonteisia terveystietoja, kuten reseptilääketilauksia ja terveyspalvelujen ajanvarauksia. Apteekki- ja terveyspalveluita tarjoavissa toimipaikoissa on tavallisesti varsin selvää, että tietoja esimerkiksi asiakkaan lääkkeistä, sairauksista ja hoitotoimenpiteistä ei anneta kolmansille osapuolille, ellei tähän ole laillisia perusteita esimerkiksi asiakkaan suostumuksen tai laillisen velvoitteen nojalla. Asiakkaiden terveystietoja tulisi käyttää vain asiakkaan hyvän hoidon ja terveyden edistämiseen sekä apteekkien ja terveyspalvelun lakisääteisten velvoitteiden täyttämiseen. Apteekkien ja terveydenhuollon toimipaikoissa pyritään myös tilajärjestelyin huolehtimaan siitä, että asiointi onnistuu yksityisesti ja luottamuksellisesti.

Valitettavasti verkossa näitä yksityisyyden ja luottamuksellisuuden periaatteita ei kuitenkaan aina noudateta yhtä hyvin kuin fyysisessä maailmassa. Kolmannen osapuolen analytiikkapalveluista ja seurantamekanismeista on tullut verkossa hyvin suosittuja ja arkipäiväisiä [4, 5], jopa kriittisissä ja välttämättömissä sähköisissä palveluissa ja julkisen sektorin ylläpitämillä verkkosivuilla [6]. Analytiikan käyttöä perustellaan esimerkiksi käyttäjäkokemuksen parantamisella ja mainonnan kohdentamisella. Useissa tärkeissä sähköisissä palveluissa arkaluonteisia henkilötietoja saat-

[1] https://www.finlex.fi/fi/laki/alkup/2019/20190306

taakin asiakkaan ymmärtämättä ja huomaamatta välittyä kolmansille osapuolille [7], mikäli ohjelmistokehittäjät, sivuston omistajat ja tietosuojavastaavat eivät ole noudattaneet riittävää huolellisuutta sähköisten palvelujen suunnittelussa ja ylläpidossa. Myöskään verkkoapteekeissa ja terveydenhuollon sähköisissä palveluissa asiakkaan yksityisyys ei siis välttämättä aina ole itsestäänselvä asia, mikä antaa aihetta tarkastella näitä verkkopalveluja tarkemmin ja luo lähtökohdat tälle väitöstutkimukselle.

Tutkimuksessa tarkastellaan kolmansia osapuolia verkkoapteekeissa, erilaisissa verkkopohjaisissa terveyspalveluissa kuten lääkärikeskusten, terapiapalveluiden ja laboratorioiden verkkosivuilla, ja lopulta vielä mielenterveyteen liittyvissä verkkopalveluissa. Tarkastelun kohteena on erityisesti se, vuotaako sivuilta terveystietoja kolmansille osapuolille ja mitä tahoja nämä tietojen vastaanottajat ovat. Lisäksi tarkastellaan käyttäjän mahdollisuuksia vuotoihin vaikuttamiseen ja sitä, kuinka hyvin käyttäjä tulee informoiduksi siitä, mitä henkilö- ja terveystietoja hänestä luovutetaan ja mille osapuolille. Lisäksi apteekkien ja terveysalan verkkopalveluita vertaillaan yksityisyyshaasteiden osalta muihin kriittisiin verkkopalveluihin, kuten vaalikoneisiin ja kuntien verkkosivuihin. Väitöstutkimus sisältää myös kehittäjähaastatteluja, joiden avulla pyritään selvittämään kehittäjien näkemyksiä tutkimuksen kohteena olevien kolmannen osapuolen tietovuotojen syistä. Väitöstutkimuksessa tarkastellaan niin ikään ratkaisuja kriittisten verkkopalvelujen yksityisyyden parantamiseen lähinnä ohjelmistoteknisestä näkökulmasta. Tutkimus on tehty osana IDA – Intiimiys datavetoisessa kulttuurissa -akatemiahanketta (327397).

Tehtyjen verkkoliikenneanalyysien yhteydessä paljastui vakavia tietovuotoja, joissa arkaluonteisia henkilötietoja ja terveystietoja siirtyi kolmansien osapuolien haltuun, yleensä ilman käyttäjän asianmukaista suostumusta. Tieteellisten tulosten lisäksi tavoitteena oli myös vaikuttaa terveysalan verkkosivustojen tietoturvaan. Tämän vuoksi huolestuttavista löydöksistä ilmoitettiin tietosuojavaltuutetun toimistolle tai suoraan sivustoja ylläpitävien tahojen tietosuojavastaaville. Esimerkiksi suomalaisten verkkoapteekkien kohdalla onkin jo saavutettu tässä suhteessa suuria parannuksia kymmenien apteekkien verkkosivuilla, kun kolmannen osapuolen analytiikkapalvelujen määrää on merkittävästi vähennetty. Tutkimus siis pyrkii myös yhteiskunnalliseen vaikuttavuuteen tarkasteltujen verkkopalvelujen tietosuojan parantamisen kautta.

Väitöskirja koostuu tämän johdantoluvun lisäksi kolmesta luvusta:

- Luvussa 2 esitellään väitöstutkimuksen teemaan liittyvää teoreettista taustaa. Luvussa käydään ensin läpi verkkoapteekkeihin ja terveysalan sähköisiin palveluihin liittyvää aiempaa tutkimusta. Luku sisältää myös tutkimuksen kannalta oleellisia perustietoja tietoturvasta ja yksityisyydestä. Lisäksi käsitellään verkkopalveluissa usein esiintyviä kolmansia osapuolia, jotka tämän tutkimuksen puitteissa useimmiten ovat erilaisia seuranta- ja analytiikkapalveluja. Analytiikkapalvelujen osalta perehdytään myös niille yleisesti lähetettäviin hen-

kilötietoihin. Lopuksi teoriakatsauksen perusteella tunnistetaan vielä aukkoja aihepiirin tutkimuksessa ja perusteita jatkotutkimukselle, mikä luo pohjan väitöskirjan tutkimuskysymyksille.

- Luvussa 3 kuvataan tehtyä tutkimustyötä. Luvussa esitellään aluksi tutkimuskysymykset, ja käydään sitten läpi kerätyt aineistot ja käytetyt tutkimusmenetelmät sekä tutkimusprosessin vaiheet. Luku sisältää myös väitöstutkimukseen kuuluvien alkuperäisjulkaisujen yhteenvedot.
- Luvussa 4 esitetään johtopäätökset ja väitöstutkimuksen päätulokset. Tähän sisältyvät vastaukset tutkimuskysymyksiin, tutkimuksen tärkeimmät kontribuutiot, rajoitteet ja esiin nousevat jatkotutkimusaiheet.

2 Yksityisyys ja kolmannet osapuolet terveyspalveluissa

2.1 Yksityisyys ja tietovuodot

Yksityisyydellä (engl. privacy) viitataan tavallisesti yksilön oikeuteen päättää itseään koskevien tietojen keräämisestä, tallentamisesta, uudelleenkäytöstä ja luovuttamisesta kolmansille osapuolille [8, 9]. Henkilötietoja tallentavien ja käsittelevien ohjelmistojen ja verkkopalvelujen tulisi olla suunniteltu ja toteutettu tavalla, joka suojaa niiden henkilöiden yksityisyyttä, joiden tietoja käsitellään.

Yksityisyys on yhä tärkeämpää digitaalisessa maailmassa verkkopalvelujen ja kolmansien osapuolien kerätessä jatkuvasti tietoja käyttäjien toiminnasta, kuten sivustoilla vierailemisesta, hakukyselyistä ja ostohistoriasta. Verkkosivustot käyttävät evästeitä ja seurantateknologioita rakentaakseen käyttäjälle profiilin, jota voidaan sitten käyttää vaikkapa mainonnan kohdentamiseen. Tämän vuoksi käyttäjälle tulisi verkkopalveluissa tarjota selkeät ja helppokäyttöiset yksityisyysasetukset, jotka antavat hänelle mahdollisuuden hallita omien henkilötietojensa käsittelyä.

Vaikka yksityisyyden käsite vaihtelee paikasta ja kulttuurista riippuen [10], oikeutta yksityisyyteen pidetään yhtenä ihmisoikeuksista ja siitä määrätään ihmisoikeuksien yleismaailmallisen julistuksen 12. artiklassa. Tässä väitöskirjassa yksityisyyttä tarkastellaan eurooppalaisesta näkökulmasta. EU:n yleinen tietosuoja-asetus eli GDPR (General Data Protection Regulation) tuli voimaan EU:n jäsenvaltioissa 25. toukokuuta 2018. Tavoitteena oli yhdenmukaistaa yksityisyyden suojaa koskevia säännöksiä EU:ssa ja vastata digitaalisten palveluiden nykyisiin yksityisyysvaatimuksiin. GDPR edellyttää nykyaikaisten tietosuojakäytäntöjen käyttöönottoa digitaalisissa palveluissa ja ohjaa kunnioittamaan käyttäjien oikeutta yksityisyyteen [11].

Tietosuoja (engl. data security) määritellään tässä väitöstutkimuksessa lain takaamaksi yksityisyyden suojaksi. Se turvaa käyttäjän oikeuksien ja vapauksien toteutumisen, kun hänen henkilötietojaan käsitellään. Henkilötietoja tulee käsitellä vain silloin, kun niiden käsittelyyn on lainmukainen peruste. Samalla päämääränä on myös selvästi osoittaa käyttäjälle, milloin ja millä perusteilla henkilötietoja käsitellään. Yksi yleinen ja tämän väitöstutkimuksen kannalta olennainen peruste henkilötietojen käsittelylle on käyttäjän oma suostumus. GDPR:n mukaan käyttäjän tulee antaa suostumus ”vapaaehtoisesti, yksilöidysti, tietoisesti ja yksiselitteisesti”[12]. Suostumusta on tärkeää pyytää selkeällä ja yksinkertaisella kielellä, jotta käyttäjä varmasti

ymmärtää antamansa suostumuksen merkityksen. Yleisesti ottaen tiedonkeruuseen ja evästeiden käyttöön pyydetään suostumus evästebannerilla (engl. cookie consent banner), verkkosivuston kävijälle ensimmäisellä käyttökerralla näytettävällä ponnahdusikkunalla, jossa on vaihtoehdot suostumuksen antamiseen ja antamatta jättämiseen.

GDPR-asetuksen artiklassa 5 esitellään tietosuojaperiaatteita, joita digitaalisten palvelujen yksityisyyden osalta tulisi noudattaa[1]. Ensinnäkin käyttäjän henkilötietoja on käsiteltävä tiedon luottamuksellisuus ja eheys säilyttäen. Henkilötietojen käsittelyn on myös oltava lainmukaista, kohtuullista ja käyttäjälle läpinäkyvää. Tietoja on kerättävä ja käsiteltävä tiettyä selkeää ja laillista tarkoitusta varten. Kerättävän tiedon määrä olisi myös syytä minimoida; henkilötietoja on kerättävä vain tarpeellinen määrä käsittelyn tarkoitus huomioiden. Kerätyt tiedot tulee lisäksi säilyttää sellaisessa muodossa, että käyttäjä on tunnistettavissa ainoastaan tietojen käsittelyä varten tarpeellisen ajan. On huomionarvoista, että GDPR koskee myös EU-alueella toimivia ulkomaisia yrityksiä.

Henkilötietojen käsittelyn läpinäkyvyyden kannalta olennaista on käyttäjien asianmukainen informointi. Informoinnin keinoksi on pitkälti vakiintunut tietosuojaseloste (engl. privacy policy). Palvelun käyttäjän tulisi esimerkiksi saada hyvä kuva siitä, mitkä tahot hänen henkilötietojaan käsittelevät ja mitä tarkoitusta varten. Edelleen käyttäjän tulisi olla tietoinen siitä, jos hänestä siirretään kolmansille osapuolille erityisen arkaluonteisia henkilötietoja. GDPR määrittelee useita niin sanottuja erityisiä henkilötietoryhmiä, joihin kuuluvien henkilötietojen käsittely on lähtökohtaisesti kiellettyä[2]. Tällaisista tietoja ovat esimerkiksi henkilön terveyttä koskevat tiedot ja poliittiset mielipiteet.

Vaikka GDPR on tärkeä osa yksityisyyden parantamista digitaalisissa palveluissa, tämä työ ei analysoi tietovuotoja oikeustieteellisestä näkökulmasta, eikä useimmissa tapauksissa ota kantaa regulaatioon. Tässä väitöskirjassa tietovuotoja tarkastellaan enimmäkseen tekniseltä ja eettiseltä kannalta. Pyrkimyksenä on monessa tapauksessa osoittaa esimerkiksi, että tiettyjä tietoja siirtyy kolmansille osapuolille ottamatta suoraan kantaa tiedonsiirtojen laillisuuteen tai laittomuuteen. Sen sijaan voidaan useissa tapauksissa argumentoida, että eettisesti tällaiset tiedonsiirrot eivät ole hyväksyttäviä – aina silloinkaan, kun niille väitetysti on olemassa jokin oikeusperuste.

Tietovuodolla (engl. data leak) tarkoitetaan salassapidettävän tiedon vuotamista ulkopuolisille [13, 14]. Tietovuotoja pidetään usein tietomurtoina (engl. data breach), joissa ulkopuolinen henkilö saa luvattoman pääsyn esimerkiksi yrityksen tietoihin, tai tapauksina, joissa sisäpiiriläinen paljastaa salaisia tietoja julkisuuteen. On kuitenkin tärkeää huomata, että tietovuoto voi olla myös tahatonta tiedon vuotamista, johon

[1] https://gdpr-info.eu/art-5-gdpr/

[2] https://tietosuoja.fi/erityisten-henkilotietoryhmien-kasittely

mikään taho ei suoraan aktiivisesti ja tahallisesti hyökkäämällä tai tietoa vuotamalla vaikuta. Tästä hyvä esimerkki ovat monet tässä väitöstutkimuksessa käsiteltävät verkkopalvelujen tietovuodot kolmansille osapuolille. Tässä yhteydessä tietovuoto on siis tiedon leviämistä ulkopuolisille, ei niinkään jonkin tahon harjoittamaa tietoista ja aktiivista tiedon levittämistä. Kolmannet osapuolet toki tarjoavat palveluitaan ja keräävät tietoa tarkoituksellisesti esimerkiksi kohdennettua mainontaa varten. Tietovuotoja voidaan ehkäistä ja yksityisyyttä parantaa panostamalla vahvaan tietoturvaan, jota käsitellään seuraavaksi.

2.2 Tietoturva

Tietoturva (engl. information security) on resurssien suojaamista oikeudetonta toimintaa vastaan[3]. Tietoturvan saavuttamiseksi on erilaisia menettelyjä ja teknisistä ratkaisuja, jotka on tarkoitettu suojaamaan luottamuksellisia ja arkaluonteisia resursseja, kuten dataa ja suoritettavaa ohjelmakoodia. Näitä resursseja suojataan luvattomalta käytöltä, väärinkäytöltä, paljastumiselta, muokkaamiselta, poistamiselta tai erilaisilta häiriöiltä [15]. Esimerkiksi tietoturvavaatimusten määrittäminen, uhkaskenaarioiden kartoittaminen, tietoturva-auditoinnit ja tietoturvatestaus ovat monien tietoturvamenettelyiden olennaisia osia [16, 17]. Toisaalta on myös kehitetty runsaasti erilaisia teknisiä ratkaisuja tietoturvan parantamiseksi. Nämä menettelyt ja tekniset ratkaisut on suunnattu pitkälti ohjelmistokehittäjille, mutta omalta osaltaan myös ohjelmistojen ja palvelujen ylläpitäjät sekä loppukäyttäjätkin ovat vastuussa tietoturvasta.

Tietoturvan nähdään usein koostuvan kolmesta peruselementistä: luottamuksellisuudesta, eheydestä ja saavutettavuudesta [9]. Nämä esitetään usein kuvan 1 kaltaisena kolmiomallina, ns. CIA-kolmiona (engl. CIA triad). Tarkemmin näiden kolmen peruselementin merkitykset ovat datan suojaamisen kannalta seuraavat:

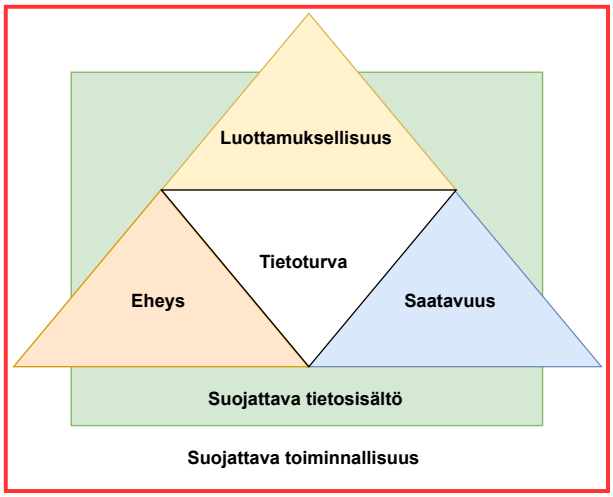
- *Luottamuksellisuudella* (engl. confidentiality) tarkoittaa datan suojaamista henkilöiltä tai tahoilta, joilla ei ole oikeuksia nähdä tai käsitellä sitä [18, 19]. Ohjelmiston käsittelemän, välittämän tai tallentaman datan tulisi siis näkyä vain niille, joilla on siihen käyttöoikeus. Luottamuksellisuus saavutetaan usein käyttämällä vahvaa salausta, mikä tekee tietojen tarkastelun oikeudettomille tahoille erittäin haasteelliseksi. Luottamuksellisuus on olennainen osa kriittisten ohjelmistojen tietoturvaa, koska se auttaa suojaamaan arkaluonteisia tietoja luvattomalta pääsyltä ja käytöltä.

- *Eheys* (engl. integrity) tarkoittaa datan täsmällisyyttä ja täydellisyyttä koko sen elinkaaren ajan [20, 21]. Eheys siis edellyttää, ettei mikään oikeudeton taho ole päässyt vääristelemään tai tuhoamaan tietoja. Ohjelmiston käsittelemään

[3] https://csrc.nist.gov/glossary/term/infosec

dataan tehdyillä muutoksilla voi olla vakavia seurauksia ja se voi aiheuttaa ohjelmiston toimintaan häiriöitä tai virheellisiä tuloksia. Datan eheyden varmistamiseksi ja luvattomien muutosten estämiseksi voidaan käyttää esimerkiksi tarkistussummia, digitaalisia allekirjoituksia ja pääsynvalvontaa.

- *Saatavuus* (engl. availability) viittaa siihen, että datan tulee olla käytettävissä niille, joilla on siihen käyttöoikeus [22, 21]. Käyttäjien tulee voida luottaa siihen, että ohjelmisto pystyy aina tarvittaessa toimimaan käyttötarkoitustaan vastaavalla tavalla. Kun palvelu ei ole saatavissa, tuottavuus heikentyy ja käyttäjät saattavat tulla tyytymättömiksi. Ohjelmistojen ja järjestelmien saatavuutta voidaan pyrkiä varmistamaan varajärjestelmällä, joka voidaan ottaa käyttöön ensisijaisen järjestelmän käytön häiriintyessä tai estyessä.



Kuva 1. Tietoturvan kolme peruselementtiä.

On syytä huomioida, että vaikka edellä puhutaan tietoturvan kolmesta peruselementistä dataan sovellettuina, tulisi ne yleisesti käsittää laajemmin moniin muihinkin resursseihin, kuten ohjelmistoihin ja palveluihin liittyvinä. Vaikkapa suoritettavan ohjelmakoodin eheys on datan eheyden lisäksi ensiarvoisen tärkeää [23]. Tässä väitöstutkimuksessa suojattavana resurssina tarkastellaan kuitenkin verkkopalvelun käyttäjän luovuttamaa dataa ja keskitytään lähinnä vain sen luottamuksellisuuteen; tarkemmin sanottuna kysymykseen arkaluonteisten terveystietojen mahdollisesta joutumisesta vääriin käsiin.

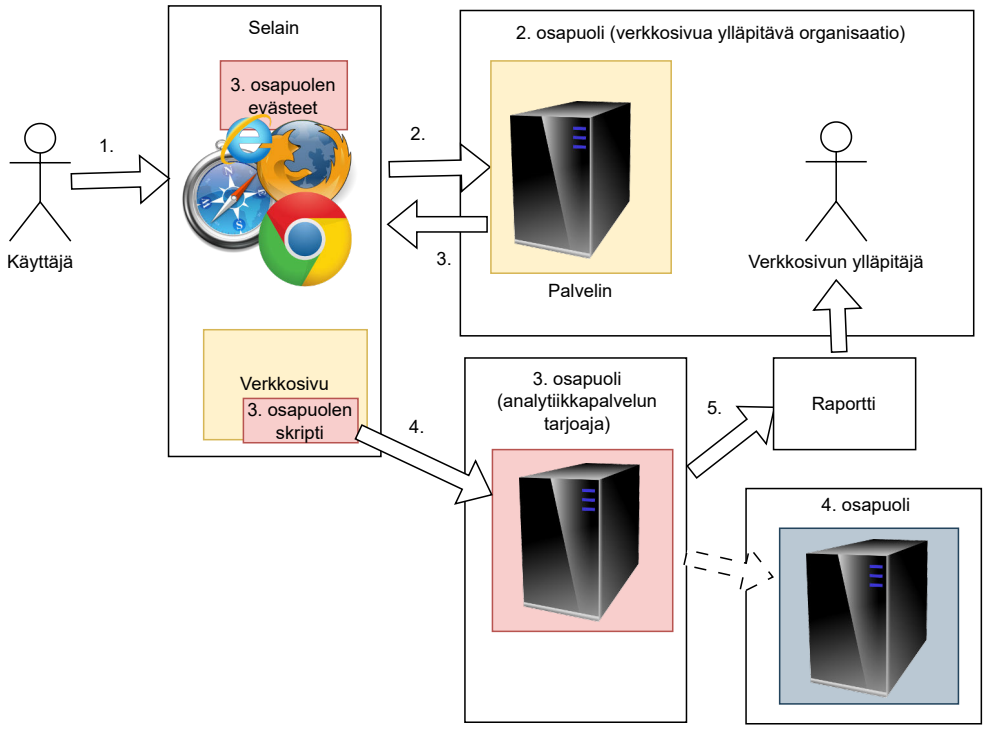
Tietoturvaa voidaan ohjelmistoteknisestä näkökulmasta ajatella pitkälti työkaluna vahvan yksityisyyden aikaansaamiseen. Asianmukaisten teknisten ja organisatoristen tietoturvatoimien käyttäminen johtaa myös käyttäjien parempaan yksityisyy-

teen. Erityisesti edellä käsitelty luottamuksellisuus on olennainen tietoturvan elementti vahvaa yksityisyyttä tavoiteltaessa [9].

2.3 Web-analytiikka ja kolmannet osapuolet

2.3.1 Web-analytiikan perusteet

Web-analytiikka (engl. web analytics) on verkkosivuston käyttäjäseurantaa: käytön mittaamista, datan keräämistä ja analysointia sekä raporttien laatimista sivuston käytön ymmärtämiseksi ja optimoimiseksi [24]. Analytiikkaratkaisut liittyvät läheisesti myös markkinoinnin tehostamiseen ja mainonnan kohdentamiseen [25]. Web-analytiikka auttaa yritystä tai organisaatiota pääsemään haluamiinsa tavoitteisiin keräämällä tietoa verkkosivuston vierailijoiden toiminnasta.



Kuva 2. Kolmannen osapuolen tiedonkeruu verkkosivulla.

Kuvassa 2 on korkealla tasolla kuvattu analytiikan toiminta. Nuolet kuvaavat datan liikkumista. Käyttäjä saapuu toisen osapuolen tarjoamalle verkkosivulle tai palveluun, ja verkkosivu ladataan toisen osapuolen palvelimelta käyttäjän selaimeen. Osana verkkosivua on kolmannen osapuolen analytiikkaskripti, jota suoritetaan käyttäjän selaimella verkkopalvelua käytettäessä ja joka kerää tietoa käyttäjän toimista sivustolla. Kuten kuvasta voidaan nähdä, tämä tieto on tarkoitettu hyödyttämään

toista osapuolta erilaisten raporttien muodossa, mutta samalla se päätyy kolmannelle osapuolelle. Kolmas osapuoli seuraa käyttäjää ja tunnistaa tämän myös *seurantaevästeiden* (engl. tracking cookies, käyttäjän laitteelle tallennettuja pieniä tekstitiedostoja) avulla. Kun kolmas osapuoli kerää ja säilyttää dataa, syntyy aina riski siitä, että data tavalla tai toisella päätyy myös eteenpäin ns. neljännelle osapuolelle. Toisen osapuolen verkkopalvelu näkyy kuvassa keltaisella, kolmannelle osapuolelle kuuluvat elementit taas punaisella.

Käytännössä sivuston käyttäjien toimia siis seurataan ja mitataan sivustoon upotetuilla pienillä JavaScript-koodinpätkillä, jotka aktivoituvat käyttäjän saapuessa sivulle ja alkavat lähettää analytiikan kerääjälle tietoja esimerkiksi käyttäjän laitteesta, selaimesta ja sivustolla suoritetuista toimista käyttäjän vuorovaikuttaessa sivun ja sen HTML-elementtien kanssa.

Vaikkapa vierailijan saapuminen etusivulle, tietyn tuotteen katselu tai tuotteen tilaaminen ovat asioita, jotka usein tunnistetaan kävijän selaimessa ja niihin liittyvä data otetaan talteen. Yleisesti web-analytiikassa onkin määritetty sivuston toiminnan kannalta yrityksen tärkeimmät tavoitteet, joita erityisesti tulisi seurata. Näitä tavoitteita sanotaan myös *konversioiksi* (engl. conversions) [26]; tavallisia esimerkkejä ovat vaikkapa sivustolle rekisteröityminen tai tuotteen ostotapahtuma. Käyttäjän yksityisyyden kannalta voikin pitää perin ongelmallisena, että juuri näihin sivuston olennaisimpiin toimintoihin usein myös liittyy käyttäjän luovuttamaa tai hänen toimintansa tuloksena syntyvää mahdollisesti arkaluonteista tietoa (kuten tieto tilatusta tuotteesta). Konversiot siis antavat yritykselle kaikkein arvokkainta tietoa, koska ne ovat linkki sivuston toiminnallisuuden ja liiketoiminnan välillä, mutta samalla ne myös usein maksavat asiakkaalle hänen yksityisyytensä, kun tiedot lähetetään analytiikan tarjoavan kolmannen osapuolen palvelimelle.

Analytiikkaratkaisuilla tuotettavien raporttien pohjana ovat dimensiot ja mittarit [27]. *Dimensioita* eli ulottuvuuksia (engl. dimensions) ovat esimerkiksi tietyn sivun nimi tai URL-osoite, tuotteen nimi, ja käyttäjän sijaintitieto. *Mittarit* (engl. meters) puolestaan ovat dimensioihin yhdistettäviä lukuarvoja, esimerkiksi sivun katselukerrat, sivulla vietetty aika tai tietyn tuotteen myynnit. Vaikkapa eniten ostetuista tuotteista on näiden tietojen perusteella helppo laatia raportti.

Analytiikalla voidaan myös mallintaa sitä, miten käyttäjät kulkevat sivustolla kohti jotakin konversiota, kuten tuotteen ostoa [28]. Läheskään kaikki kävijät eivät nimittäin ainakaan ensimmäisellä käyntikerralla osta mitään, ja voidaankin esimerkiksi tarkkailla, kuinka moni käy vain etusivulla, kuinka moni etenee katselemaan jotakin tuotetta, kuinka moni lisää tuotteen ostoskoriin ja kuinka moni lopulta ostaa tuotteen. Tämä on sivuston houkuttelevuuden ja käytettävyyden kehittämisen kannalta varsin hyödyllistä tietoa, eikä sivuston kehittäjä välttämättä ajattele, että tässä yhteydessä kerätään myös toisinaan hyvinkin arkaluonteista tietoa siitä polusta, jota yksittäinen käyttäjä seuraa sivustolla.

Voidaan ajatella, että yritykselle suoraan hyödyllisten dimensioiden ohella analytiikkapalvelut keräävät omaan käyttöönsä myös eräänlaisia ”piilodimensioita”, kuten monenlaista teknistä tietoa, jolla voidaan tunnistaa ja profiloida myös yksittäistä sivujen kävijää [29]. Yksittäistä asiakasta ja esimerkiksi tämän ostokäyttäytymistä seuraamalla Googlen kaltaiset suuret teknologiayritykset voivat kohdentaa mainontaa juuri tietylle käyttäjälle sopivaksi. Tässä on luonnollisesti se hyvä puoli, että käyttäjän näkemät mainokset ovat häntä kiinnostavia, mutta samalla analytiikkapalveluilla on usein tiedossaan sellaisiakin käyttäjien henkilötietoja, jotka eivät monen käyttäjän mielestä välttämättä kuuluisi niille. Tällainen seuranta on aina potentiaalinen uhka käyttäjän yksityisyydelle, erityisesti kun sitä harjoittavat kolmannet osapuolet.

Vuonna 2013 kävi ilmi, että Yhdysvaltojen kansallisella turvallisuusvirastolla (NSA) on suora pääsy useiden suurimmista teknologiayritysten, kuten Googlen, Microsoftin ja Metan palvelimille. Yhdysvaltain viranomaiset voivat näin ollen reaaliaikaisesti nähdä ja ottaa talteen esimerkiksi kaiken Googlen keräämän datan[4]. Käytännössä tiedonkeruun kohteina ovat olleet erityisesti ulkomaalaiset, myös siis esimerkiksi pohjoismaiset Googlen ja muiden suurten teknologiayritysten analytiikkapalveluita hyödyntävien sivustojen vierailijat [30].

Web-analytiikkaa voikin käyttää verkkosivuilla myös sellaisten lokaalisti asennettavien työkalujen avulla, jotka säilyttävät datan hallinnan verkkosivun ylläpitäjällä, eivätkä luovuta sitä millekään kolmannelle osapuolelle. Yksi tällainen ratkaisu on Matomo, joka on ilmainen avoimen lähdekoodin käyttäjäseurantaohjelmisto [3, 4]. Erityisesti terveyspalvelujen kaltaisten välttämättömien verkkopalvelujen soisi ottavan käyttöön enemmän tällaisia ratkaisuja globaalien suuryritysten ylläpitämien analytiikkapalvelujen sijaan.

On myös syytä huomata, että kaikki kolmannet osapuolet, jotka vastaanottavat dataa, eivät välttämättä harjoita käyttäjäseurantaa tai hyödynnä vastaanottamiaan henkilötietoja. Monien yritysten ja organisaatioiden sivuilla on esimerkiksi tavoitettavuuden helpottamiseksi chat-palveluita, jotka eivät sinänsä ole analytiikkapalveluita, mutta jotka kuitenkin sijaitsevat kolmannen osapuolen palvelimilla ja periaatteessa voivat tallentaa tietoja käyttäjistä. Koska valtaosa kolmansista osapuolista kuitenkin on – tämänkin väitöstutkimuksen löydösten perusteella – analytiikkapalveluja ja koska arkaluonteisten tietojen päätyminen näille palveluille on erityisen ongelmallista, viitataan *kolmansilla osapuolilla* (engl. third parties) tässä tutkimuksessa lähinnä käyttäjien dataa keräämiin analytiikkapalveluihin.

[4] https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data

Taulukko 1. Analytiikkapalveluille usein lähteviä henkilötietoja.

Kategoria	Henkilötieto
Tekniset tunnistetiedot	IP-osoite
	Käyttäjätunniste
	Laite-/selaintunniste
	Muu tuntematon tunniste
	Selain
	Käyttöjärjestelmä
	Näytön resoluutio
	Ikkunan koko
	Värisyvyys
	Aikaleima
	Kieli
	Internet-yhteyden tyyppi
Käyttäjän toiminta	Ostetut tuotteet/palvelut
	Tarkasteltu sisältö / Suoritettu toiminto

2.3.2 Kolmansille osapuolille lähetettäviä henkilötietoja

GDPR:n artiklassa 4 määritellään, että *henkilötiedoilla* (engl. personal data) tarkoitetaan ”kaikkia tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön liittyviä tietoja”[5]. Tunnistettavissa oleva on henkilö, joka voidaan suoraan tai epäsuorasti tunnistaa erilaisten tunnistetietojen, kuten esimerkiksi nimen, henkilötunnuksen, sijaintitiedon tai verkkotunnistetietojen perusteella. Yleisesti termiä tulkitaan siis varsin laajasti [31, s. 113] ja monenlaisia analytiikkapalvelujen keräämiä teknisiä tunnistetietoja voidaan pitää henkilötietoina.

Taulukkoon 1 on kerätty henkilötietoja, joita verkkopalveluista usein lähtee kolmansille osapuolille sivuilla olevien analytiikkapalvelujen kerääminä [7]. Varsin oleellinen tekninen tieto on asiakaslaitteen IP-osoite, joka lähetetään kolmannelle osapuolelle jokaisen HTTP-pyynnön yhteydessä, kun kolmannen osapuolen palvelimelle siirtyy dataa verkkosivustolta[6].

IP-osoitteet (engl. IP addresses) ovat numerosarjoja, joiden avulla voidaan yleensä yksilöidä verkkoon yhdistetty laite ja sen käyttäjä. Käyttäjien IP-osoitteet ovat yleensä dynaamisia, mutta esimerkiksi kiinteissä laajakaistaliittymissä on kuitenkin tavallista, että sama IP-osoite säilyy tietyllä laitteella yli kuukaudenkin ajan [32]. Erityisesti suurilla globaaleilla analytiikkapalvelujen tarjoajilla voi myös olettaa olevan

[5] https://gdpr-info.eu/art-4-gdpr/

[6] Tämä on tarpeellista, koska IP-osoitteita käyttäen yksilöidään verkossa dataa vastaanottava laite.

tehokkaita keinoja yhdistää dynaamisetkin IP-osoitteet tiettyyn käyttäjään. Vaihtuvakin IP-osoite on siis mahdollinen keino seurata henkilön tekemisiä internetissä.

On syytä huomata, että esimerkiksi Google Analytics mahdollistaa sitä käyttävän sivuston ylläpitäjälle vierailijoiden IP-osoitteiden anonymisoinnin. Google Analyticsin versiossa 3 anonymisointi tapahtui poistamalla IP-osoitteesta viimeinen oktetti[7]. Vaikka anonymisoitua IP-osoitetta ei enää laillisessa mielessä yksinään tulkittaisi henkilötiedoksi[8], IP-osoitteen viimeisen oktetin poistaminen ei mitä todennäköisimmin useimmissa tapauksissa estä analytiikkapalvelua seuraamasta käyttäjän tekemisiä, koska tämä anonymisointitapa ei merkittävästi vielä laajenna mahdollisten IP-osoitteiden joukkoa[9] ja koska yksittäinen käyttäjä on lisäksi edelleen tunnistettavissa muiden teknisten tietojen ja yksilöivien tunnisteiden avulla. Google Analyticsin versio 4 menee anonymisoinnissa selvästi pidemmälle poistamalla koko IP-osoitteen[10]. Tämä on askel oikeaan suuntaan, mutta *koska anonymisointi tehdään yhä Googlen omalla palvelimella*, käyttäjä joutuu luottamaan siihen, että analytiikkapalvelun tarjoaja hoitaa toimenpiteen asianmukaisesti. Lisäksi on mahdollista, että Yhdysvaltain viranomaiset voivat edelleenkin ottaa IP-osoitteen ja muut kerätyt tiedot talteen ennen anonymisointia – myös silloin, kun palvelin on EU-alueella.

Usein kolmansille osapuolille lähtevä verkkoliikenne sisältää IP-osoitteen ohella myös erilaisia laite- ja käyttäjäkohtaisia tunnisteita, jotka säilyvät evästeitä hyödyntäen useiden käyttäjäistuntojen ajan [33]. On syytä huomata myös, että Googlen ja Metan kaltaiset suuret teknologiayritykset voivat helposti seurata käyttäjiä myös silloin, kun nämä ovat kirjautuneina sisään niiden omiin palveluihin (kuten esimerkiksi Gmail ja Facebook) [34]. Tällöin IP-osoitteet tai muut laitetunnisteet voidaan usein helposti liittää suoraan käyttäjän oikeaan nimeen. Analytiikkapalvelujen verkkoliikenteessä esiintyy säännöllisesti myös tuntemattomia ja dokumentoimattomia tunnisteita [7]. Nämä tunnisteet ovat tavallisesti pitkiä merkkijonoja, jotka saattavat olla esimerkiksi käyttäjään, laitteeseen tai istuntoon liittyviä tunnistetietoja. Tällaisilla yksilöllisillä tunnisteilla käyttäjiä voidaan seurata selvästi tehokkaammin kuin yksittäisillä järjestelmään liittyvillä teknisillä tiedonpalasilla, joita käsitellään seuraavaksi.

Jokaisen HTTP-pyynnön otsikkotiedoissa lähetetään palvelimelle User-Agent-otsikkokenttä. Tämä otsikkokenttä paljastaa kolmannelle osapuolelle käyttäjän selaimen ja käyttöjärjestelmän, useissa tapauksissa myös näiden versionumerot. User-Agent-merkkijono näyttää seuraavan kaltaiselta:

```
Mozilla/5.0 (Windows NT 10.0; Win64; x64)
```

[7] https://support.google.com/analytics/answer/2763052

[8] https://tietosuoja.fi/pseudonymisointi-anonymisointi

[9] Yhden oktetin anonymisoimisen jälkeen mahdollisia IP-osoitteita on IPv4:ssä 256 kappaletta.

[10] https://support.google.com/analytics/answer/12017362?hl=en

```
AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/123.0.0.0 Safari/537.36
```

Monet käyttäjät eivät välttämättä tiedä, että tällaisia tietoja, joita voidaan hyödyntää yksittäisen vierailijoiden tunnistamiseen ja profilointiin, lähetetään myös muille osapuolille kuin käytettävälle verkkopalvelulle. Koska laitteiden ja käyttöjärjestelmien valikoima on jatkuvassa kasvussa, ja esimerkiksi suositusta Chrome-selaimesta tulee jatkuvasti uusia versioita, sisältää User-Agent-merkkijono yhä enemmän informaatiota [35]. Jo vuonna 2010 on laskettu User-Agent-merkkijonon sisältävän keskimäärin 10,5 bittiä informaatiota[11], ja on syytä olettaa, että tämä luku on ajan myötä vielä huomattavasti kasvanut. Käyttäjät voidaan 10 bitin avulla jakaa jo 2^{10} eli 1024 erilliseen ryhmään. Muihin teknisiin tietoihin yhdistettynä User-Agent-merkkijono auttaa siis merkittävästi käyttäjän yksilöinnissä jopa ilman IP-osoitetta.

Kolmansille osapuolille ja analytiikkaa kerääville tahoille lähtee monissa tapauksissa myös muita teknisiä tietoja, kuten esimerkiksi näytön ja selainikkunan koko. Usein lähetettäviä teknisiä tietoja ovat niin ikään näytön värisyvyys ja järjestelmän kieli. Kolmannet osapuolet seuraavat toisinaan myös käytettävän internet-yhteyden tyyppiä (esimerkiksi CONN_CELLULAR_5G). Laajaa henkilötiedon määritelmää mukaillen tietosuojavaltuutetun toimisto määrittelee henkilötiedoiksi myös ”tiedot, joiden perusteella henkilö voidaan tunnistaa suoraan tai välillisesti esimerkiksi yhdistämällä yksittäinen tieto johonkin toiseen tietoon, joka mahdollistaa tunnistamisen”[12]. Edellä mainitut tekniset yksityiskohdat siis voidaan myös tulkita henkilötiedoiksi. Yhdessä nämä tekniset tiedonpalaset rajaavat jo huomattavasti käyttäjien joukkoa ja auttavat yksilöimään ja profiloimaan tietyn sivuston vierailijan.

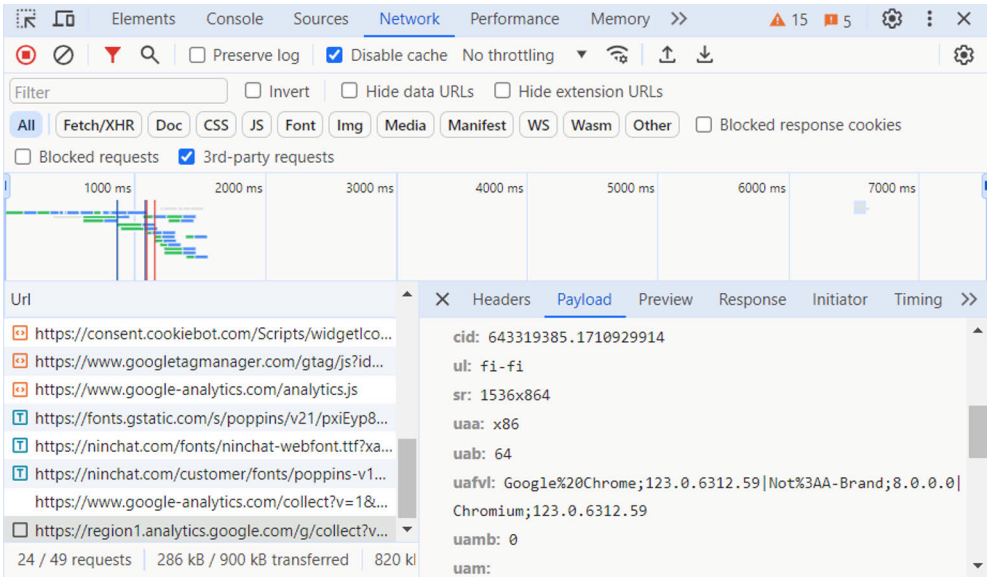
Monet analytiikkapalvelut kykenevät varsin tarkasti rakentamaan tällaisen *digitaalisen sormenjäljen* (engl. digital fingerprint) eli kompaktin yhteenvedon laitteen ja verkkoselaimen teknisistä tiedoista [36]. Digitaalinen sormenjälki on siis henkilötieto, joka koostuu pienistä teknisistä yksityiskohdista, ja luodaan monissa tapauksissa käyttäjän tietämättä ja ilman hänen suostumustaan. Monet käyttäjät eivät ymmärrä teknisten tietojen keräämisen merkitystä [37], eivätkä osaa ajatella, mitä seurauksia näiden henkilötietojen luovuttamisesta kolmansille osapuolille voi olla [38].

Kuvassa 3 on esimerkki Google Analytics -palvelulle lähtevästä HTTP-pyynnöstä ja eräistä hyötykuorman sisältämistä henkilötiedoista. Kuvassa näkyy Google Analyticsin CID-tunniste, joka yksilöi tietyn selain-laite-parin ja auttaa Google Analyticsiä yksilöimään käyttäjät sivustolla. Tämä tunniste voi säilyä Googlen evästeessä kaksikin vuotta[13]. Lisäksi kuvassa näkyy käyttäjän kieliasetus, näytön koko, prosessoriarkkitehtuuri, käyttöjärjestelmän bittisyys ja selaimen versiotieto.

[11] https://www.eff.org/deeplinks/2010/01/tracking-by-user-agent

[12] https://tietosuoja.fi/mika-on-henkilotieto

[13] https://policies.google.com/technologies/cookies?hl=en-US

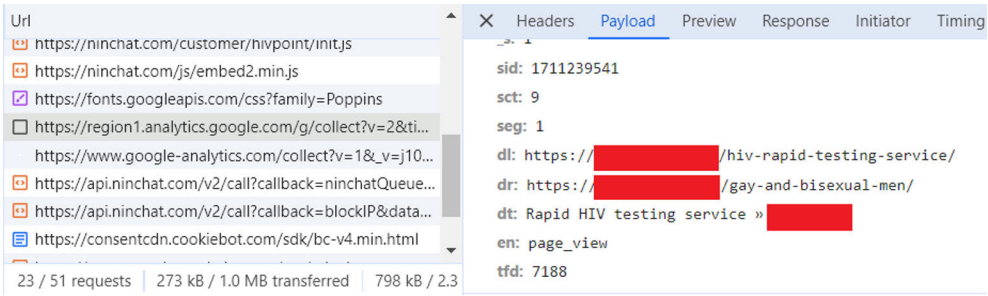


Kuva 3. Esimerkki Googlelle lähtevästä HTTP-pyynnöstä ja eräistä hyötykuorman sisältämistä henkilötiedoista.

Edellä käsitellyt tiedot liittyvät tietyn käyttäjän tunnistamiseen. Niitä siis voidaan pitää *yksilöivinä* eli *identifioivina* tietoina. Kolmansille osapuolille lähtee kuitenkin myös *asiayhteyteen liittyvää* eli *kontekstuaalista* tietoa siitä, mitä käyttäjä verkkosivulla tekee. Vaikkapa sisällöstä, jota käyttäjä sivulla tarkastelee (kuten tietyn reseptilääkkeen tuotesivu) tai toiminnosta, jonka käyttäjä suorittaa verkkosivulla (kuten ajanvaraus lääkärille) voi välittyä tietoa. Juuri käyttäjän toimista kertovan asiayhteyteen liittyvän arkaluonteisen tiedon ja käyttäjän yksilöivän tiedon yhdistelmä tekee tietovuodoista vaarallisia – kummankaan tiedon vuotaminen ei yksinään vielä olisi erityisen ongelmallista ainakaan yksittäiselle käyttäjälle. Monissa tapauksissa tieto käyttäjän tarkastelemasta sisällöstä tai suoritetusta toiminnosta välittyy kolmannelle osapuolelle yksinkertaisesti vieraillun sivun URL-osoitteen muodossa. Tämä osoite voi joko suoraan sisältää vihjeen sisällöstä tai toiminnosta, tai nämä saa selville yleensä ainakin kyseisessä osoitteessa vierailemalla ja selvittämällä sen merkityksen verkkosivun sisällön perusteella.

Kuvassa 4 on esimerkki hyvin arkaluonteisesta asiayhteyteen liittyvästä tiedosta HTTP-pyynnön hyötykuormassa. Tässä tapauksessa Googlelle vuotaa käyttäjän avaaman sivun URL-osoite, josta selviää, että käyttäjä on kiinnostunut HIV-pikatesteistä (parametri `dl`). Edelleen hyötykuormasta selviää, että käyttäjä on tullut nykyiselle sivulle seksuaalivähemmistöille suunnatun sivun kautta (parametri `dr`). Näin kolmas osapuoli voi päätellä, että käyttäjä on mahdollisesti seksuaalivähemmistön edustaja, joka epäilee HIV-tartuntaa. Näistä tiedoista kumpikin on GDPR:n valossa erityisen arkaluonteinen, vaikka sinänsä sivulla vieraileminen ei niitä vielä vahvis-

takaan todeksi. Tämänluonteisilla verkkosivuilla kolmannen osapuolen analytiikan käytön tarpeellisuutta olisi kuitenkin syytä miettiä hyvin tarkkaan.



Kuva 4. Esimerkki arkaluonteisesta asiayhteyteen liittyvästä tiedosta HTTP-pyynnön hyötykuormassa.

On huomionarvoista, että URL-osoitteen merkityksen päättely voidaan nykyään monissa tapauksissa antaa myös tekoälyn tehtäväksi – ottaen huomioon erityisesti viimeaikaisen asiayhteydet ymmärtävän tekoälyn nopean kehityksen. Silti URL-osoitteen tai muiden vuotaneiden tietojen merkityksen selvittäminen ei toki kaikissa tapauksissa välttämättä onnistu ilman syvempää tietämystä sivuston rakenteesta ja toiminnasta. Näihin perehtyminen voi vaatia manuaalista, ihmisen tekemää selvitystyötä.

On myös hyvin tärkeää ymmärtää aikaulottuvuuden merkitys asiayhteyteen liittyvän tiedon keräämisessä. Kun käyttäjän lukuisat käynnit samassa verkkopalvelussa voidaan liittää toisiinsa yksilöivien tunnisteiden avulla, voi kolmas osapuoli rakentaa hänestä hänen toimintansa ja asiayhteyksien avulla yhä tarkemman profiilin. Vaikkapa verkkoapteekkiesimerkin tapauksessa tämä voi tarkoittaa yhä suurempaa tietomäärää käyttäjän ostamista lääkkeistä ja sitä kautta hänen terveydentilastaan ja potilashistoriastaan. Vaikka hyvin harvoin voidaan varmasti sanoa, että kolmannet osapuolet varastoisivat ja käyttäisivät kerättyjä tietoja ainakaan tietyn käyttäjän yksilöivässä muodossa, on tällaisen toiminnan mahdollisuuskin ongelmallinen – onhan yksityisyyden ja tietoturvan varmistaminen pitkälti myös riskianalyysiä. Jos tietoja tallennetaan, on aina olemassa riski, että ne päätyvät luovuttamisen, myynnin tai tietomurron kautta niin sanotulle neljännelle osapuolelle (ks. kuva 2).

2.3.3 Tiedonkeruu verkkosivulla

Usein analytiikkatoiminnallisuuden asentamisen jälkeen verkkosivustolle on tarpeen lisätä sivuston eri osiin vierailijan toimien ja toteutuneiden konversioiden seurantaa. Seurannan hallinta manuaalisesti on työlästä, koska jokaisen tapahtuman seurantaa vastaava koodinpätkä pitäisi erikseen upottaa sivustolle. Esimerkiksi Google tarjoaakin Google Tag Manager -työkalun, jolla esimerkiksi Google Analyticsin pieniä koo-

dinpätkiä eli tageja voi hallita helpommin. Google Tag Managerissa voi lisätä, muokata ja poistaa mittauksia[14]. Google Tag Managerilla onnistuu myös käyttäjän suostumusten hallinta, sillä erilaiset analytiikkaan ja mainontaan liittyvät koodinpätkät voi asettaa toimimaan vain, jos käyttäjä on antanut niitä vastaavan suostumuksen.

Vaikkapa lääkärikeskus voi näin Google Tag Manageria käyttämällä lisätä esimerkiksi ajanvarauslomakkeeseen niin sanotun käynnistystriggerin[15], joka aktivoi analytiikkakoodin, kun lomakkeen tiedot lähetetään (tai jo sen täyttämisen aikana). Lääkärikeskus saa tärkeitä tietoja ajanvarauksesta, joka myös lienee yksi olennainen konversio markkinoinnin ja liiketoimintatavoitteiden seuraamisen kannalta. Samalla kuitenkin käyttäjän täyttämät tiedot, kuten vaikkapa lääkärin nimi ja ajanvarauksen aika ja paikka, saatetaan lähettää Googlen palvelimelle.

Lääkärikeskukselle analytiikka voi näyttäytyä hyödyllisenä käyttäjän toimien ja konversioiden mittaamisena ja tarkkailuna, mutta käyttäjän yksityisyyden näkökulmasta tilanteessa on hyvin paljon samaa kuin selainympäristössä tapahtuvassa väliintulohyökkäyksessä eli man-in-the-browser-hyökkäyksessä [39, 40]. Hyökkäyksessä pahantahtoinen taho onnistuu lisäämään verkkosivulle haittakoodia, joka kerää sivulle syötettyjä tietoja ja lähettää niitä eteenpäin hyökkääjän palvelimelle. Tämä koodi tarkkailee siis sivun käyttäjän (asiakaspuoli) ja sivua ylläpitävän tahon (palvelinpuoli) välistä viestintää ilman, että nämä osapuolet ovat tietoisia asiasta. Tilanteessa, jossa yritys ei ole kunnolla tietoinen sivuillaan olevasta analytiikasta (analytiikkaratkaisu voi tulla esimerkiksi oletuksena alihankkijan toimittaman alustan mukana) ja käyttäjää ei informoida kunnolla tästä käyttäjäseurannasta, ero analytiikan ja väliintulohyökkäyksen välillä on hyvin pieni. Asetelma, jossa analytiikkaratkaisu kerää vaivihkaa tietoa, on esitetty kuvassa 5.

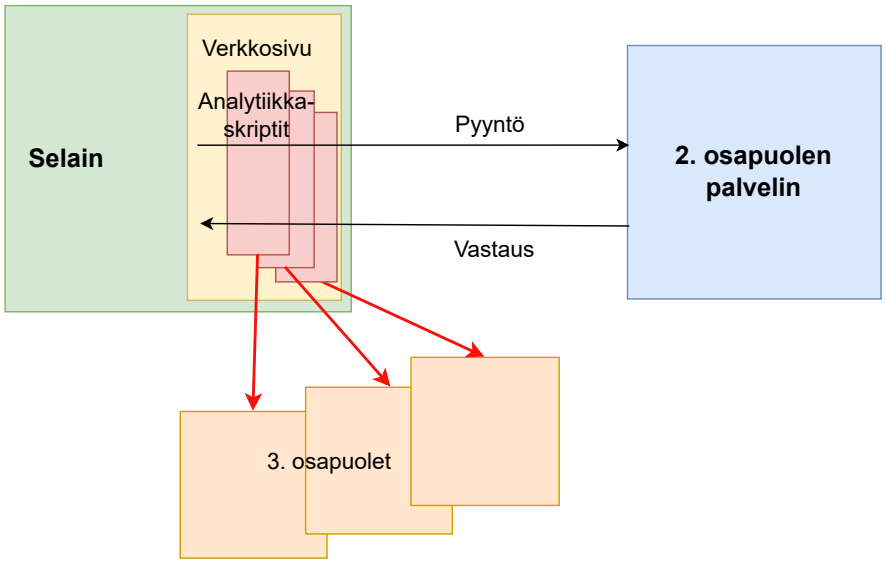
Kuten edellä kuvatun kaltaiset datan salakuunteluhyökkäykset, myös web-analytiikan toiminta on käyttäjälle hyvin usein täysin näkymätöntä [41]. Tämän takia tarvitaan selkeitä tietosuojaselosteita, joissa kerrotaan läpinäkyvästi, mitä tietoja käyttäjästä lähetetään kolmansille osapuolille ja mitä tahoja nämä kolmannet osapuolet ovat [7, 42]. Tämä taas vaatii väkisinkin, että ohjelmistokehittäjät ja tietosuojavastaava ovat tarkasti perillä yrityksen tai organisaation verkkosivuilla olevan analytiikan toiminnasta ja siihen liittyvästä henkilötietojen käsittelystä.

2.3.4 Analytiikkaan liittyvä aiempi tutkimus

Web-analytiikan aiheuttamista yksityisyysriskeistä on runsaasti aiempaa tutkimusta [43, 34, 44]. Vaikka useimmat analytiikkapalvelut ovat erinomaisia digitaalisen sormenjäljen rakentamisessa ja käyttäjän profiloinnissa [36], suurimman hyödyn tästä

[14] https://support.google.com/tagmanager/answer/6102821?hl=en

[15] Teknisesti kyse on DOM-rajapinnan ja JavaScriptin avulla lomakkeen lähetystapahtumaan kytketystä tapahtumankuuntelijasta.



Kuva 5. Kolmannen osapuolen tiedonkeruu verkkosivulla.

tiedonkeruusta saavat yleensä analytiikkapalvelujen tarjoajat. Esimerkiksi Alby [45] toteaa, että useimmat ylläpitäjät eivät hyödynnä web-analytiikkaa dominoivan Google Analyticsin kehittyneitä seurantatoimintoja eivätkä saa palvelun käytöstä läheskään täyttä hyötyä. Samalla tavallisten käyttäjien on vaikeaa ymmärtää analytiikkapalveluiden keräämien henkilötietojen merkitystä [37], vaikka kerätyt tiedot voivat monissa tapauksissa olla äärimmäisen arkaluonteisia, kun käyttäjän yksilöivät ja asiayhteyteen liittyvät tiedot yhdistyvät [46].

Samarasinghe ym. [47] tutkivat kolmannen osapuolen analytiikkaa 150244 julkisen sektorin verkkosivulla (206 eri maasta) ja 1166 Android-sovelluksessa (71 maasta). Sivuilta löytyi runsaasti seurantateknologiaa, esimerkiksi 17 % julkisen sektorin verkkosivustoista ja 37 % Android-sovelluksista sisälsi Googlen analytiikkaa. Julkisen sektorin Android-sovelluksista 27 % vuoti arkaluonteista tietoa (kuten käyttäjä- ja laitetunnisteita ja salasanoja) kolmansille osapuolille ja monissa tapauksissa tämä data myös välittyi suojaamattoman HTTP-yhteyden yli. Evästehallintatyökaluja kehittävä Cookiebot [48] puolestaan totesi raportissaan 2019, että 89 % EU-jäsenmaiden hallinnollisista ja viranomaisten ylläpitämistä verkkosivuista sisälsi kolmannen osapuolen analytiikkaa.

Maris ym. [49] tutkivat pornosivuilla käytettyjä seurantateknologioita ja niiden keräämiä tietoja. Tutkimuksessa selvisi, että 93 % sivuista vuoti henkilötietoja kolmannelle osapuolelle. Sivut, jotka vuotivat henkilötietoja, lähettivät niitä keskimäärin seitsemään eri verkko-osoitteeseen. Vastaavasti Samarasinghe [50] ym. tarkastelivat 62373 uskonnollisen verkkosivuston ja 1454 Android-sovelluksen yksityisyyttä. Tutkijoiden mukaan 32 % uskonnollisista verkkosivustoista ja 78 % Android-sovelluk-

sista käytti Googlen seurantatyökaluja. Useat kolmannen osapuolen palvelut keräsivät tarkastelluilla sivustoilla arkaluonteisia henkilötietoja.

Nämä esimerkit osoittavat varsin selkeästi, että kolmannen osapuolen web-analytiikka vaikuttaa muodostavan vakavan yksityisyysriskin hyvin erilaisilla web-sivustoilla riippumatta siitä, selaileeko käyttäjä vaikkapa julkisen sektorin verkkosivuja, pornosivuja vai uskonnollisten yhdyskuntien sivuja. On syytä huomata, että tällaiset sivustot ja palvelut voivat paljastaa kolmansille osapuolille hyvin arkaluonteisia tietoja monenlaisiin vähemmistöihin ja haavoittuviin ryhmiin kuuluvista käyttäjistä [49]. Esimerkiksi uskonnollinen vakaumus tai seksuaalinen suuntautuminen tai käyttäytyminen ovat GDPR:n mukaan tarkasti suojeltavia erityisiä henkilötietoryhmiä. Seuraavaksi tarkastellaan analytiikkaa ja yksityisyysriskejä verkkoapteekkien ja webpohjaisten terveyspalvelujen yhteydessä.

2.4 Verkkoapteekit ja yksityisyys

Lukuisat suomalaiset apteekit tarjoavat nykyään verkkopalveluita, joiden kautta voi tilata sekä itsehoito- että reseptilääkkeitä. Lääkealan turvallisuus- ja kehittämiskeskus Fimean laillisten verkkoapteekkien listalla on jo yli 200 verkkopalvelua[16]. Vaikkei verkkoapteekeissa asioinnin suhteen ole Suomessa vielä tapahtunut suurta läpimurtoa, ovat koronaepidemian kiihdyttämä digitalisaatio ja muutos kuluttajakäyttäytymisessä tukeneet myös apteekkipalvelujen parempaa saavutettavuutta ajasta ja paikasta riippumatta [51].

Apteekkiala on voimakkaasti säädelty, mikä näkyy myös verkkoapteekkien toiminnassa. Reseptilääkkeiden tilaaminen verkkoapteekista vaatii asiakkaalta vahvan tunnistautumisen, ja asiakkaan tulee saada neuvontaa lääkkeen oikeaan ja turvalliseen käyttöön esimerkiksi chat-toiminnallisuuden välityksellä. Juuri reseptilääkkeiden tilaukseen liittyvä toiminnallisuus on verkkoapteekkien osalta kriittinen myös tietoturvan ja käyttäjän yksityisyyden kannalta. Asiakkaan reseptitiedot ovat arkaluonteisia terveystietoja, joten yksityisyydensuojan säilymiseen ja reseptitietojen asianmukaiseen käyttöön tulisi voida verkkoapteekissa luottaa samalla tavalla kuin kivijalka-apteekissa.

Verkkoapteekkien riskejä käsittelevä kirjallisuus näyttää lähes täysin keskittyvän siihen, ovatko apteekit laillisia ja onko niistä ostettuja lääkkeitä turvallista käyttää [52, 53, 54]. Silloinkin, kun puhutaan yksityisyydestä, tarkastelu näyttää koskevan lähinnä laittomia apteekkeja, joissa käyttäjän henkilötiedot voivat joutua rikollisten käsiin [55]. Verkkoapteekkien tietoturvan suhteen on tutkittu tavanomaisia websovellusten haavoittuvuuksia ja painotettu esimerkiksi tietoliikenteen salaamisen merkitystä [56], mutta kolmansille osapuolille lähtevä verkkoliikenne on jäänyt lähes

[16] https://www.fimea.fi/apteekit/verkkopalvelutoiminta/lailliset_apteekin_verkkopalvelut

huomiotta. Laillisten apteekkien käyttämän analytiikan ongelmaa ei siis näytetä laajalti tiedostavan tieteellisessä kirjallisuudessa.

Ainoa verkkoapteekkien analytiikkaa käsittelevä julkaisu vaikuttaa olevan Zheutlinin ym. [57] lyhyehkö artikkeli Yhdysvaltalaisten verkkoapteekkien sivuilla olevasta seurantateknologiasta. Tutkimus osoittaa, että analytiikkaa on laillisten verkkoapteekkien sivuilla runsaasti, yli kaksi kolmasosaa tarkastelluista verkkoapteekeista sisälsi ainakin kaksi dataa keräävää kolmannen osapuolen palvelua. Yleisimmät analytiikkapalvelut olivat Google Analytics ja Facebook Pixel (nykyään Meta Pixel). Kirjoittajat ilmaisevat löydösten perusteella huolensa siitä, että terveystietoja, kuten tietoja reseptilääkkeistä, saatetaan jakaa eteenpäin. Tutkimuksessa ei kuitenkaan tarkastella lainkaan verkkoliikennettä tai kolmansille osapuolille lähetettäviä tietoja. Siksi tämä väitöskirja voidaan verkkoapteekkeja käsittelevän osuutensa puolesta nähdä pitkälti jatkona Zheutlinin ym. aloittamalle työlle.

Apteekkialan yksityisyyshaasteet on huomioitu myös mobiilialustojen puolella. Tietoturvayhtiö Sophos raportoi vuonna 2020, että GoodRx-mobiilisovellus, jolla yhdysvaltalaiset kuluttajat voivat vertailla lääkkeiden hintoja eri apteekeissa, vuoti etsittyjen reseptilääkkeiden tietoja Googlelle, Facebookille ja useille muille kolmansille osapuolille[17]. Kuzma ym. puolestaan tutkivat tietosuojaselosteita 90 eurooppalaisen, pohjois-amerikkalaisen ja aasialaisen verkkoapteekin sivuilla [58]. Vain noin 70 % tutkituista verkkoapteekkien sivuista sisälsi tietosuojaselosteen ja nämäkin selosteet olivat usein hyvin heikkolaatuisia. Selkeät ja läpinäkyvät tietosuojaselosteet verkkoapteekkien sivuilla ovat asiakkaan informoiduksi tulemisen kannalta hyvin tärkeitä, vaikka eivät voikaan korvata teknisiä toimia yksityisyyden parantamiseksi.

Verkkoapteekkien yksityisyyttä käsittelevän tutkimuksen vähäisyys on yllättävää, sillä yksityisyys tuodaan kyllä verkkoapteekkeja käsittelevässä tutkimuksessa esiin mahdollisena riskinä [59]. Kyselyissä myös merkittävä osa asiakkaista näkee verkkoapteekissa asioinnin riskinä yksityisyydelle. Esimerkiksi Rattay ym. [60] selvittivät vuonna 2018 mielipiteitä sähköisen reseptin tietosuojasta Suomessa, ja melkein joka neljäs (23 %) vastaaja oli huolissaan reseptitietojen joutumisesta asiattomiin käsiin. Verkkoapteekit voi nähdä eräänlaisena rajapintana, jossa terveystiedot joutuvat kosketuksiin julkisen internetin ja kolmannen osapuolen analytiikkapalvelujen kanssa, joten aihe kaipaa enemmänkin tutkimusta. On ilmeistä, että verkkoapteekit ovat jääneet yksityisyysriskien kannalta sokeaksi pisteeksi, eikä aihetta ole tarkasteltu syvällisesti ohjelmistotekniseltä kannalta. Osa tutkimuksista jopa esittää vahvan yksityisyyden verkkoapteekkien etuna [61, 62] perinteisiin kivijalka-apteekkeihin nähden, ja lähtökohtaisesti näin onkin hyvin helppo ajatella.

Lopuksi on vielä todettava, etteivät aivan kaikki tutkimukset näe terveystietojen yksityisyyttä itseisarvona. Kes ym. [63] tuovat esiin valtavirrasta poikkeavan

[17] https://nakedsecurity.sophos.com/2020/03/03/goodrx-stops-sharing-personal-medical-data-with-google-facebook/

näkemyksen siitä, että arkaluonteistenkin henkilötietojen kerääminen terveydenhuollon verkkosivustoilla on käyttäjän kannalta hyödyllistä. Tutkijat toteavat, että esimerkiksi verkkoapteekit voivat näin mukautua paljon paremmin käyttäjiensä tarpeisiin ja osaa esimerkiksi suositella tälle sopivia lääketarjouksia. Kes ym. vertaavat tätä tilannetta asiakassuhteeseen, joka voisi kivijalka-apteekissa kehittyä kanta-asiakkaan ja apteekkarin välille, ja jossa asiakas hyötyy saamastaan paremmasta ja yksilöllisemmästä palvelusta ja avusta. On kuitenkin hyvin kyseenalaista, onko tällainen kohdennettu markkinointi verkkopalveluissa asiakkaalle menetetyn yksityisyyden arvoista, kun otetaan huomioon terveystietojen vuotamiseen liittyvät mahdolliset riskit. Terveystietojen huoleton jakaminen kolmansille osapuolille ei sovi ainakaan eurooppalaiseen tietosuoja-ajatteluun. Asiakkaita voidaan kuitenkin hyvin seurata ja palvella yksilöidysti myös sellaisilla analytiikkaratkaisuilla, jotka eivät luovuta tietoja kolmannelle osapuolelle [3].

2.5 Terveydenhuollon verkkopalvelut ja yksityisyys

Toisin kuin verkkoapteekeista, muista terveydenhuollon verkkopalveluista ja niiden yksityisyydestä on olemassa paljon tutkimusta. Jo vuonna 2012, kun web-analytiikan käyttö ei vielä ollut niin laajalle levinnyttä kuin nykyään, Masters [64] tarkasteli kolmannen osapuolen analytiikkaratkaisujen esiintyvyyttä terveydenhuollon verkkosivustoilla ja osoitti, että nämä työkalut keräsivät monissa tapauksissa henkilötietoja ilman käyttäjän asianmukaista informointia. Samoihin aikoihin Burkell ja Fortier [65, 66] totesivat, että useat terveysaiheiset verkkosivustot eivät paljastaneet käyttäjilleen asianmukaisesti tietojenkeruukäytäntöjään, ja että käyttäjiä usein seurattiin analytiikkatyökaluilla yksityiskohtaisten profiilien luomiseksi. Huesch puolestaan julkaisi tutkimuksen, joka sisälsi havaintoja lääketieteellisten hakutermien joutumisesta vääriin käsiin terveysaiheisilla sivustoilla [67]. Brown ja Levy [68] puolestaan esittelivät mittausinstrumentin, jolla voidaan vertailla lääkealan verkkosivustojen todellisia ja dokumentoituja tietojenkeruukäytäntöjä.

Terveysalan verkkosivujen yksityisyyden tutkimus on jatkunut myös viime vuosina. Huo ym. [69] tutkivat vuonna 2022 analytiikkaa ja potentiaalisia terveystietovuotoja 459 verkkopohjaisen terveyspalvelun sivuilla. Sivustoista 14 % käytti Google Analytics -palvelua ja 9 sivuston todettiin siirtävän arkaluonteisia terveystietoja (esimerkiksi lääketietoja ja laboratoriotuloksia) kolmansille osapuolille. Tutkijoiden mukaan on hyvin todennäköistä, ettei terveysalan toimijoilla aina ole tarvittavaa teknistä osaamista ongelman ymmärtämiseen ja korjaamiseen. Useissa lääketieteen alan tutkimuksissa kolmannen osapuolen analytiikkaa käytetään verkkopalvelun keskeisenä komponenttina käyttäjäseurannassa (ks. esimerkiksi [70] ja [71]), todennäköisesti ymmärtämättä täysin tästä seuraavia yksityisyysriskejä. Nämä havainnot korostavat selvästi sisäänrakennetun yksityisyyden puutetta ja tarvetta valistaa ohjelmisto-

kehittäjiä ja tietosuojavastaavia verkkopohjaisten terveyspalvelujen yksityisyyshaasteista.

Libert [34] tutki terveystietojen vuotamista kolmansille osapuolille URL-osoitteissa. Hän totesi, että 91 % tutkituista sivustoista tekee HTTP-pyyntöjä kolmansille osapuolille vaarantaen käyttäjän yksityisyyden. HTTP-pyynnöissä on usein nykyisen sivun URL-osoite ja Libertin mukaan noin 70 % tarkastelluista URL-osoitteista sisälsi esimerkiksi sairauksia tai oireita paljastavaa arkaluonteista tietoa. Zheutlin ym. [72] puolestaan tutkivat kolmansien osapuolien web-analytiikkaa eri sektorien (yksityinen, julkinen, kolmas sektori) terveysaiheisilla verkkosivustoilla. Tutkituilla sivustoilla oli keskimäärin 2,11 kolmannen osapuolen palvelua. Zheutlin ym. eivät kuitenkaan yksityiskohtaisesti selvittäneet, mitä henkilötietoja kolmansille osapuolille lähtee sivustoilta.

Friedman ym. [73] käsittelivät kolmannen osapuolen seurantateknologioiden riskejä sairaaloiden verkkosivustoilla. Vastaavasti Yu yms. [74] tekivät laajan automatisoidun analyysin sairaaloiden verkkosivustoilla ympäri maailmaa käytetystä analytiikasta. Tutkimus paljasti, että 53,5 % niistä käytti seurantatyökaluja, jotka keräsivät henkilötietoja. Friedman ym. [75] tarkastelivat kolmannen osapuolen seurantatyökalujen yleisyyttä aborttiklinikoiden verkkosivustoilla ja totesivat, että useimmilla sivustoilla (99,1 %) oli käytössä jonkinlaista web-analytiikkaa, joka vuoti henkilötietoja kolmansille osapuolille.

Yleisesti ottaen näyttää selvältä, että web-pohjaisten terveyspalvelujen yksityisyydessä ja tietosuojassa on selkeitä puutteita [76]. Vaikka verkkopohjaisten terveyspalvelujen yksityisyyttä, web-analytiikkaa ja tietovuotoja on siis analysoitu ennenkin, tilanne ei näytä korjaantuneen. On sekä valitettavaa että mielenkiintoista, että sama vakava yksityisyyden ongelma, joka on havaittu terveysaiheisilla verkkosivuilla jo yli 10 vuotta sitten, on edelleen vesittämässä käyttäjän yksityisyyttä sivustoilla asioidessa.

2.6 Motivaatioita tutkimukselle

Vaikka web-analytiikkaa ja verkkopohjaisten palvelujen tietovuotoja kolmansille osapuolille onkin aiemmassa tutkimuksessa käsitelty runsaasti, on tutkimuksessa yhä useita aukkoja ja lisätutkimukselle on nähtävissä selvää tarvetta. Seuraavaksi käsitellään näitä aiemman tutkimuksen pohjalta esiin nousevia motivaatioita tutkimukselle, jota tässä väitöskirjassa on tehty.

Tietovuotojen yksityiskohtainen tarkastelu ja niiden vertailu eri sovellusaloilla ja sektoreilla

Yhtenä ilmeisenä motivaatiota työssä on selvittää kolmannen osapuolen tietovuotojen luonnetta yleisesti. Vaikka esimerkiksi URL-vuotojen [34] ongelma terveys-

palveluissa on tunnettu, ei tarkkaa vuotavien henkilötietojen erittelyä koko palvelun osalta useinkaan ole tehty. Tässä tutkimuksessa onkin tarkoituksena erilaisten yleisten ja arkaluonteisten käyttöskenaarioiden (esimerkiksi lääkäriajan varaaminen, lääkkeen ostaminen, tiedon hakeminen sairaudesta) pohjalta kattavammin eritellä vuotavia henkilötietoja. Usein esimerkiksi automaattisesti tai puoliautomaattisesti tehdyissä tutkimuksissa tietyt vuotavat henkilötiedot voivat jäädä huomiotta eivätkä tarkastellut palveluiden käyttöskenaariot välttämättä ole kovinkaan kattavia ja tarkkaan kuvattuja. Tämän väitöskirjan lähestymistapa on siksi pitkälti verkkoliikenteen manuaalinen ja yksityiskohtainen tarkastelu kattavien tulosten saavuttamiseksi.

Tämä kuvauksen kattavuus tuo tutkimukseen uutuusarvoa myös esimerkiksi lääkäriasemien tarkastelun osalta. Vaikka sairaaloita ja lääkäriasemia on aiemmin tarkasteltu, vuotavia henkilötietoja on harvemmin eritelty sen tarkemmin (esim. [69]), todennäköisesti suurien otosmäärien ja analyysimenetelmien rajoitusten vuoksi. Aihe siis kaipaa lisää huomiota. Väitöskirja päivittää ilmiön tarkastelua lähivuosiin ja analysoi sitä teknisestä näkökulmasta syvällisesti ja monitahoisesti.

Verkkopohjaisten mielenterveyspalveluiden osalta tutkimuksessa todettiin myös selvä aukko: näitä palveluja ei ole tarkasteltu kuin tietosuojaselosteiden ja mobiilisovellusten osalta. Näin ollen verkkopohjaisten mielenterveyspalvelujen tietovuodot ovat tärkeä tutkimuksen kohde. Samoin tarkoituksena on selvittää verkkoapteekkien tilanne kolmannen osapuolen tietovuotojen osalta. Kuten aiemmin todettiin, tätä aihetta ei ole vuotavien henkilötietojen osalta lainkaan tutkittu. Kuitenkin verkkoapteekki on potentiaalisesti äärimmäisen arkaluonteisia henkilötietoja käsittelevä verkkopalvelu; kun lääkettä tilataan, on diagnoosi jo tehty, ja käyttäjän terveydentilasta voi hänen lääkeostostensa (tai käyttäjän tarkastelemien lääkkeiden) perusteella tehdä päätelmiä.

Kolmannen osapuolen tietovuotojen tutkimuksessa ei myöskään ole tehty paljonkaan vertailua sen osalta, eroaako tietovuotojen määrä ja laatu eri sovellusalojen välillä (vaikkapa terveydenhuollon sivustot, kuntien sivut ja web-pohjaiset vaalikoneet) tai vastaavasti yksityisen ja julkisen sektorin toimijoiden välillä. Esimerkiksi julkisen sektorin voisi olettaa olevan tarkempi käytettyjen kolmannen osapuolen analytiikkapalveluiden osalta ja toimivan myös yksityisyyskysymysten osalta esimerkkinä muille tahoille. Vastaavasti julkisen sektorin verkkosivuilla ei ole samanlaista tarvetta vaikkapa liiketoimintatavoitteiden seuraamiselle tai kohdennetulle mainonnalle.

Web-kehittäjän näkökulma: tietovuotojen syyt ja keinot niiden korjaamiseen

Tarkoituksena on myös tutkia kolmannen osapuolen tietovuotojen syitä. Vaikka vuodoista itsestään on aikaisempaa tutkimusta, niiden syihin ei useinkaan ole perehdytty. Tarkoituksena on selvittää syitä ja perusteluja kolmannen osapuolen analytiikkapal-

velujen käytölle ohjelmistokehittäjien näkökulmasta. On erityisen mielenkiintoista, miksi yksityisyydelle ongelmallisia kolmannen osapuolen palveluja käytetään myös hyvin arkaluonteisia tietoja käsittelevillä verkkosivuilla ja miksi näitä toiminnallisuuksiltaan usein päällekkäisiäkin palveluja näyttää usein olevan käytössä samalla sivustolla monia.

Tutkimuksen motivaationa on lisäksi etsiä tapoja ennaltaehkäistä ja korjata tietovuotoja. Kolmannen osapuolen palveluiden kohdalla on tietysti helppoa ajatella, että niiden käyttöä vain tulisi välttää tai ne pitäisi kaikki poistaa, kun ymmärretään, että ne ovat riski yksityisyydelle. Näiden itsestäänselvien keinojen ohella voidaan kuitenkin pohtia keinoja turhien palvelujen käytön ja tietovuotojen välttämiseen ohjelmistokehityksen suunnittelu-, toteutus-, testaus- ja ylläpitovaiheissa. IDA-projektin tavoitteiden mukaisesti motivaationa on siis myös tutkia reiluja tapoja henkilökohtaisen datan käsittelemiseen, hallitsemiseen ja suojaamiseen. Tähän esitetään ohjelmistoteknisiä keinoja ja ratkaisuja omien empiiristen havaintojemme, kirjallisuuden ja asiantuntijahaastatteluiden pohjalta.

Käyttäjän näkökulma: mahdollisuus vaikuttaa tietojen käsittelyyn ja informoiduksi tuleminen

Työssä tarkastellaan myös, mitkä ovat normaalin käyttäjän mahdollisuudet vaikuttaa omien tietojensa käsittelyyn verkkopalveluissa suostumuksen tai tietojen käsittelyn kieltämisen kautta. Tällaista vertailua on terveyspalvelujen osalta tehty vain vähän. Siksi on mielenkiintoista selvittää, miten evästeiden käytön tai henkilötietojen keräämisen kieltäminen vaikuttaa todelliseen, verkkoliikenneanalyysin keinoin havaittuun henkilötietojen välittämiseen kolmansille osapuolille. Tarkoituksena on siis tarkastella tilannetta normaalin käyttäjän näkökulmasta. Vaikka myös regulaatio kulkee työssä tämän tarkastelun rinnalla, tarkastellaan yksityisyyttä ja datan käsittelyn reiluutta lähinnä käyttäjän ja hänen vaikutusmahdollisuuksiensa kautta.

Normaalin käyttäjän näkökulmaan liittyy myös ymmärrys siitä, miten hänen henkilötietojaan käsitellään. Tähän liittyen väitöskirjassa tarkastellaan myös tietosuojaselosteiden läpinäkyvyyttä sen suhteen, mitä henkilötietoja luovutetaan kolmansille osapuolille. Tämä on käyttäjän kannalta usein ainoa keino saada yksityiskohtaista tietoa siitä, miten hänen henkilötietojaan käsitellään. Ellei läpinäkyvyys toteudu, käyttäjän mahdollisuudet selvittää omien henkilötietojensa käsittelyä ja eteenpäin jakamista ovat yleensä hyvin heikot. Näin ollen on tärkeää selvittää, miten käyttäjän informointi toteutuu suomalaisissa verkkopohjaisissa terveyspalveluissa. Myös pimeät käytännöt (engl. dark patterns) ovat asia, joka voi vaikuttaa vahvasti siihen, että käyttäjä antaa suostumuksensa henkilötietojensa keräämiseen ja jakamiseen [77]. Pimeillä käytännöillä tarkoitetaan tavallisimmin verkkosivujen yhteydessä käyttöliittymien suunnittelussa käytettyjä keinoja, joilla harhautetaan käyttäjä tekemään jotain sellaista, mitä hän ei alun perin aikonut tehdä. Tässä väitöskirjassa tarkastelluilla

pimeillä käytännöillä käyttäjä saadaan antamaan suostumus evästeiden käyttöön ja henkilötietojen keruuseen. Vaikka pimeitä käytäntöjä on jo tutkittu runsaasti, on niiden esiintyvyyttä suomalaisissa terveyspalveluissa syytä kartoittaa.

Suomen konteksti ja yhteiskunnallinen vaikuttavuus

Kansalaisille suunnattujen digitaalisten sosiaali- ja terveydenhuollon palvelujen määrä on Suomessa ollut jatkuvassa kasvussa [78]. Kun välttämättömien terveydenhuollon palveluiden digitalisoituminen edelleen etenee ja tavoitteena on kehittää paitsi saavutettavia ja helppokäyttöisiä, myös tietoturvallisia, luotettavia ja käyttäjien yksityisyyttä kunnioittavia palveluita, on verkkopalvelujen tietovuotojen tarkastelu ensiarvoisen tärkeää. Siksi tämän työn yhtenä motivaationa on välttämättömien terveydenhuollon verkkopalvelujen yksityisyystilanteen laajamittainen kartoittaminen nimen omaan Suomen kontekstissa. Kyseessä on myös tiettävästi ensimmäinen tällainen analyysi suomalaisten terveydenhuollon verkkopalvelujen tietovuodoista.

Tähän kytkeytyy myös työn yhteiskunnallinen vaikuttavuus; tavoitteena on saada tarpeettomia kolmansia osapuolia ja tietovuotoja karsittua tutkituilta suomalaisilta verkkosivulta ja informoida sivustojen ylläpitäjiä (tai joissakin tapauksissa tietosuojavaltuutetun toimistoa) löydöksistä. Tämä antaa myös mahdollisuuden tarkastella tutkimuksessa sitä, millaisia positiivisia muutoksia saadaan aikaan informoimalla sivustojen ylläpitäjiä ja tietosuojaviranomaista tietovuodoista. Esimerkiksi medioiden kiinnostuessa tietovuodoista ja tietosuojaviranomaisen toimien kautta yksityisyysongelmat myös voivat saada julkisuutta, mikä toivottavasti edelleen kannustaa useampia arkaluonteisia henkilötietoja käsittelevien verkkopalvelujen ylläpitäjiä harkitsemaan käyttämiään analytiikkapalveluja uudelleen.

3 Tutkimustyön kuvaus

3.1 Tutkimuskysymykset

Edellisen luvun lopussa esitetyt tutkimusmotivaatiot liittyvät seuraaviin väitöskirjan tutkimuskysymyksiin:

TK1: *Mitä kolmannen osapuolen palveluita verkkoapteekkien ja terveysalan verkkopalveluiden sivuilla on ja mitä apteekkien ja terveyspalvelujen asiakkaiden henkilö- ja terveystietoja näille kolmansille osapuolille mahdollisesti vuotaa?*

Ensimmäinen tutkimuskysymys keskittyy selvittämään, millaisia kolmannen osapuolen palveluja verkkoapteekkien ja terveysalan verkkopalveluiden sivuilla on ja mitä asiakkaiden terveystietoja näille mahdollisesti vuotaa. Kuten edellisessä luvussa nähtiin, erityisesti verkkoapteekkien sivuilla olevan analytiikan ja potentiaalisten tietovuotojen osalta kysymys on tärkeä, sillä aiheesta ei ole olemassa lähes lainkaan tutkimusta huolimatta sen lisääntyvästä merkityksestä digitalisoituvassa yhteiskunnassa. Samoin muiden terveysalan sivustojen kuten lääkäriasemien ja mielenterveyden verkkopalvelujen osalta kaivataan lisää tietoa kolmansille osapuolille mahdollisesti vuotavien arkaluonteisten henkilötietojen luonteesta, erityisesti myös suomalaisissa verkkopalveluissa.

Tutkimuskysymyksen TK1 tiimoilta selvitetään tarkasti, millaista arkaluonteista tietoa kullekin kolmannen osapuolen palvelulle mahdollisesti vuotaa erilaisten yleisten käyttöskenaarioiden (tai analytiikkapalveluiden näkökulmasta usein konversioiden) tapauksissa. Tämän takia useimmissa väitöskirjan tutkimuksissa keskitytään suhteellisen pieniin otoksiin ja niiden olennaisimpien käyttöskenaarioiden tarkkaan ja manuaaliseen analyysiin sen sijaan, että analysoitaisiin suuri määrä verkkopalveluja automaattisin keinoin.

Ensimmäinen tutkimuskysymys on tämän väitöstutkimuksen tutkimuskysymyksistä tärkein ja siihen myös vastataan selvästi laajimmin. Tutkimuksen ensisijainen tarkoitus onkin saada käsitys terveysalan verkkosivustojen yksityisyyden tasosta Suomessa. Tähän tutkimuskysymykseen on mahdollista vastata teknisin keinoin tarkasti ja kattavasti, ja siihen liittyvät tulokset muodostavat siksi hyvän faktapohjan, jonka varaan pohdintaa havaittujen tietovuotojen vaikutuksista voidaan rakentaa. Muut tutkimuskysymykset, TK2–TK4, rakentuvat osittain näiden havaintojen varaan.

TK2: *Ovatko terveysalan verkkopalvelujen tietosuojapuutteet yhtä vakavia yksityisellä ja julkisella sektorilla, ja mikä on tilanne verrattuna muihin toimialoihin?*

Voidaan väittää, että julkisen sektorin verkkopalvelujen tulisi toimia esimerkkeinä tietosuoja-asioissa [6]. Niillä ei myöskään ole yksityiseen sektoriin verrattavaa kaupallista motiivia analytiikkapalvelujen käyttöön vaikkapa konversioiden ja liiketoimintatavoitteiden toteutumisen seuraamisen muodossa. Näistä syistä voidaan hypoteesina esittää, että kolmansien osapuolten ja vastaavasti niille vuotavien henkilötietojen määrä on julkisen sektorin terveyspalveluissa pienempi kuin yksityisellä sektorilla.

Vastaavasti terveysalaa on kiinnostavaa verrata muiden toimialojen verkkosivustoihin. Näin nähdään, onko analytiikkapalvelujen käyttö verkkopohjaisissa terveyspalveluissa yhtä runsasta ja huoletonta kuin muiden alojen ja toimijoiden verkkosivuilla. Tähän tutkimuskysymykseen liittyen väitöskirjassa tarkastellaan esimerkiksi vaalikoneiden ja kuntasektorin web-analytiikkaa.

TK3: *Onko verkkopalvelujen käyttäjillä mahdollisuus vaikuttaa yksityisyytensä tasoon evästesuostumusten kautta ja tulevatko he riittävissä määrin informoiduiksi henkilötietojensa käsittelystä?*

Kolmas tutkimuskysymys liittyy asiakkaan kykyyn vaikuttaa itsestään kolmansille osapuolille lähteviin tietoihin ja toisaalta siihen, tuleeko hän esimerkiksi tietosuojaselosteen kautta informoiduksi tietojensa käsittelystä. Tämän kysymyksen ensimmäistä osaa testataan analysoimalla terveysalan verkkopalveluista lähtevää dataa tilanteessa, jossa kaikki evästeet ja analytiikka on hyväksytty, ja toisaalta minimisuostumuksella. Näissä tilanteissa kolmansille osapuolille lähteviä tietoja verrataan toisiinsa.

Riittävän informoiduksi tulemisen tarkastelu taas ei ole aivan yksiselitteistä, mutta tässä tutkimuksessa asiaa ajatellaan tavallisen käyttäjän näkökulmasta: saako käyttäjä tiedot siitä, mille kolmansille osapuolille ja millaista tietoa hänestä annetaan? Mikäli näin ei ole, ei voida puhua informoidusta tai tietoisesta päätöksenteosta suostumusta annettaessa. Kysymystä riittävästä informoinnista tarkastellaan lähinnä terveystietojen osalta. Kerrotaanko terveystietojen luovutuksesta tietosuojaselosteessa ja mainitaanko ne kolmannet osapuolet, joille tietoja lähtee? Sen lisäksi, että välitetyt henkilötiedot ja niiden vastaanottajat ovat käyttäjälle olennaisia tietoja, niiden ilmoittamiselle on myös laillisia perusteita. Vallitsevassa oikeuskäytännössä on tehty tulkinta, että tiedonsiirtoja koskevien tietojen yhteydessä tulisi mainita myös henkilötietotyypit[1]. Lisäksi GDPR:n artiklassa 13 edellytetään, että palvelun käyttäjälle pitää selkeästi kertoa ”henkilötietojen vastaanottajat tai vastaanottajaryhmät.”

[1] Irlannin tietosuojaviranomaisen ratkaisu 20.8.2021 koskien Whatsappin käyttäjien informointia henkilötietojen käsittelystä, https://www.edpb.europa.eu/our-work-tools/consistency-findings/register-decisions/2021/decision-matter-whatsapp-ireland_en

TK4: *Mistä apteekkien ja terveyspalvelujen verkkosivujen tietosuojapuutteet johtuvat ja miten niiden yksityisyyttä voitaisiin parantaa ohjelmistoteknisestä näkökulmasta?*

Lopulta on syytä myös miettiä, mikä ohjelmistonkehitysprosessissa on mennyt vikaan, mikäli luottamuksellisia terveystietoja vuotaa verkkopalvelusta kolmansille osapuolille. On melko ilmeistä, että tietojen vuotaminen kolmansille osapuolille voidaan välttää tai havaita useissakin ohjelmistokehitysprosessin vaiheissa (kuten arkkitehtuuria suunnitellessa tai valmiin sivuston tietoturvaa testatessa), mutta jostakin syystä mahdollisia tietovuotoja ei vain useinkaan ole osattu ottaa huomioon. Tämän takia on syytä miettiä, millainen vallitseva tietosuojakulttuuri verkkopalvelujen kehittäjien keskuudessa on ja mistä sivustoilla havaitut puutteet johtuvat.

Tämän ilmiön tarkastelun pohjalta voidaan myös antaa suosituksia sellaisista teknisistä toimista ja menetelmistä, joiden avulla terveysalan verkkopalvelujen sisäänrakennettua tietoturvaa ja yksityisyyttä voidaan parantaa kehityksen ja elinkaaren eri vaiheissa. Tutkimuksen tuloksena ei ole tarkoitus synnyttää täysin valmista, uutta yksityisyyskehystä, vaan ennemminkin listata asioita, jotka olemassaolevia tekniikoita ja menetelmiä sovellettaessa tulisi ottaa paremmin huomioon erityisesti kriittisten verkkopalvelujen ja niihin mahdollisesti sisältyvien kolmannen osapuolen teknologioiden osalta. Yhtenä hypoteesina tässä voidaan pitää sitä, että modernissa webkehityksessä analytiikkapalveluista on pitkälti muodostunut de facto -standardi, joka verkkosivujen kehittäjien mielestä kuuluu vakiokomponenttina toimitettavaan pakettiin. Verkkosivustoilla käytetyt valmiit alustaratkaisut, joihin analytiikkapalvelut ovat hyvin helposti liitettävissä esimerkiksi lisäosina, saattavat myös olla yksi potentiaalinen selitys tietovuodoille.

3.2 Tutkimusasetelma, aineistot ja menetelmät

Empiirinen tutkimus on kokemusperäistä, eli tutkimuskohteita havainnoidaan tai mitataan tarkan suunnitelman pohjalta [79, s. 3]. Näin se eroaa teoreettisesta tutkimuksesta, jossa tutkimuskohteita ei havainnoida suoraan, vaan niitä tarkastellaan ja hahmotetaan käsitteellisten mallien ja olemassaolevan tutkimuskirjallisuuden avulla. Tämä väitöstutkimus sijoittuu pitkälti empirian puolelle hyödyntäen muun muassa nauhoitettua verkkoliikennettä ja asiantuntijahaastatteluilla kerättyjä aineistoja. Näitä aineistoja tarkastellaan ensinnäkin laadullisen analyysin keinoin, esimerkiksi pyrkien saamaan verkkoliikennettä analysoimalla kokonaisvaltainen ja syvällinen käsitys verkkopalveluissa esiintyvistä tietovuodoista. Kolmansien osapuolien ja tietovuotojen tapauksessa käytetään myös määrällistä analyysiä tilastoimalla niiden lukumääriä eri sovellusalojen sivustoilla.

3.2.1 Aineistot

Taulukossa 2 näkyvät kussakin julkaisussa käytetyt aineistot. Verkkoliikenteen osalta kerättiin ja tallennettiin raakadatana kaiken tiettyä sivustoa selattaessa syntyneen verkkoliikenteen sisältäviä HAR-tiedostoja[2]. Näitä tiedostoja jalostettiin edelleen kirjaamalla kunkin tarkastellun sivuston kohdalla ylös kaikki havaitut kolmannet osapuolet ja niille lähteneet henkilötiedot. Tästä jalostetusta dokumentista poimittiin edelleen julkaisuissa raportoidut olennaisimmat tutkimustulokset. Verkkoliikenneanalyysiä hyödynnettiin julkaisuissa I-VIII.

Useimmissa tutkimuksissa kerättiin talteen myös sivustojen tietosuojaselosteet, joissa olevaa tietoa voitiin sitten verrata varsinaiseen verkkoliikenteeseen. Tietosuojaselosteiden osalta tarkasteltiin esimerkiksi, informoivatko ne käyttäjää siitä, että IP-osoitteen kaltaisia käyttäjän yksilöiviä tietoja päätyy kolmansien osapuolien haltuun. Edelleen tarkasteltiin myös, kerrottiinko tietosuojaselosteissa asiayhteyteen liittyvien tietojen, kuten URL-osoitteiden ja hakutermien vuotamisesta sekä siitä, että ne voivat sisältää arkaluonteista tietoa. Samaten kolmansien osapuolien nimeäminen on tietosuojaselosteessa käyttäjän informoinnin kannalta tärkeä tarkasteltava seikka.

Taulukko 2. Julkaisuissa käytetyt aineistot. Suuremmat X-kirjaimet osoittavat julkaisussa käytetyt aineistot ja pienet x-kirjaimet tarkoittavat, että aineistoa on käytetty vähäisemmässä määrin.

Aineisto	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**
Verkkoliikenne	X	X	X	X	X	X	X	X	
Tietosuojaselosteet	X	X	X	X	X	X		X	
Evästebannerit / pimeät käytännöt		x	x		X				
Haastattelut							x		X

Julkaisun VI osalta tallennettiin myös evästebannerien kuvankaappaukset, joista näkyy, miten käyttäjiltä pyydetään suostumusta evästeisiin ja tiedonkeruuseen. Näistä kuvankaappauksista paljastuvat käyttäjän harhaanjohtamiseen mahdollisesti hyödynnetyt pimeät käytännöt. Tarkasteltuja pimeitä käytäntöjä evästebannereissa olivat 1) hyväksymispainikkeen korostaminen värin tai kontrastin avulla; 2) hylkäämispainikkeen kätkeminen niin, ettei se ole heti näkyvillä hyväksymispainikkeen tavoin evästebannerin ensimmäisellä tasolla, vaan vaatii useamman napsautuksen; sekä 3) valmiiksi rastitetut valintalaatikot[3]. Kaikki nämä houkuttelevat käyttäjää an-

[2] HAR tai HTTP Archive on tiedostomuoto, joka sisältää lokin verkkoliikenteestä verkkoselaimen ja palvelimen välillä tietyllä sivustolla.

[3] Nämä tutkitut pimeät käytännöt on määritelty Euroopan tietosuojaneuvoston ”Cookie Banner Taskforcen”laatimassa dokumentissa, https://www.edpb.europa.eu/our-work-tools/our-documents/other/report-work-undertaken-cookie-banner-taskforce_en

tamaan suostumuksen siten, ettei se enää ole GDPR:n tarkoittamalla tavalla ”vapaaehtoinen, yksilöity, tietoinen ja yksiselitteinen tahdonilmaisu”[4]. Myös julkaisuissa II ja III tarkasteltiin evästebannereja vähäisessä määrin (pienet x-kirjaimet taulukossa 2).

Julkaisussa IX aineistona toimivat asiantuntijahaastattelujen nauhoitteet. Nauhoitteiden puhe muutettiin edelleen tekstimuotoon automaattisella, tekoälyä hyödyntävällä työkalulla. Myös julkaisu IV sisältää kommunikaatiota web-kehittäjien ja tietosuojavastaavien kanssa, vaikkei se hyödynnäkään menetelmänä systemaattisia haastatteluja (pieni x-kirjain taulukossa 2).

3.2.2 Verkkoliikenteen nauhoittaminen ja analyysi

Verkkoliikennettä nauhoitettiin ja analysoitiin julkaisuissa I-VIII. Menetelmänä verkkoliikenneanalyysi kytkeytyy pitkään perinteeseen, jossa päämääränä on tavallisesti ollut haittaohjelmien tunnistaminen tai tunkeutumisen estäminen [80, 81]. Tässä väitöstutkimuksessa tutkittujen palvelujen verkkosivuja testattiin ensin navigoimalla tarkasteltavalle sivulle, tyhjentämällä selaimen välimuisti ja evästeet, ja lataamalla sitten sivu uudestaan. Sivun uudelleenlataamisesta lähtien kaikki verkkoliikenne nauhoitettiin käyttäen Google Chrome -selaimeen sisäänrakennettuja kehittäjätyökaluja (Chrome DevTools). Ensimmäisenä toimenpiteenä sivun uudelleenlataamisen jälkeen halutusta koeasetelmasta riippuen joko hyväksyttiin tai hylättiin kaikki evästeet.

Kehittäjätyökalujen avulla suodatettiin muun liikenteen joukosta kolmansille osapuolille menevä liikenne, jota vastaanottavat tahot (esimerkiksi Google, Meta) listattiin. Nauhoitettu verkkoliikenne analysoitiin sitten huolellisesti yksi HTTP-pyyntö kerrallaan, ja hyötykuormasta sekä otsikkotiedoista etsittiin henkilötietoja. Vuotaneet henkilötiedot kirjattiin ylös erikseen kunkin niitä vastaanottaneen tahon osalta. Näihin kuuluvat erilaiset tunnistetiedot, kuten käyttäjä- ja laitekohtaiset tunnisteet, IP-osoitteet sekä useita teknisiä tietoja, kuten käyttöjärjestelmä ja sen versio, selain ja näytön koko. Toiseksi etsittiin asiayhteyteen liittyviä tietoja, erityisesti mahdollisia terveystietoja. Esimerkiksi apteekkien tapauksessa näitä voivat olla avattujen lääkesivujen URL-osoitteet sekä lääkkeen tilaamisaikomusta kuvaava painikkeen napsautus.

Tarkasteltavien sivustojen sovellusalasta (kuten apteekit, lääkärikeskukset ja mielenterveyspalvelut) ja sivuston yksilöllisestä rakenteesta riippuen sivustoilla nauhoitetut navigointipolut vaihtelevat jonkin verran. Näitä on kuvattu tarkemmin kussakin väitöskirjan julkaisussa. Sivustoilta pyrittiin nauhoittamaan yleisimmät arkaluonteista tietoa sisältävät käyttöskenaariot, jotka on listattu taulukossa 3.

Verkkoapteekkien sivuilla tarkasteltu skenaario oli reseptilääkkeen tilaus. Hakutoiminnon avulla päädytään tietylle tuotesivulle, jonka kautta tieto tietyn resep-

[4] https://www.privacy-regulation.eu/fi/r32.htm

Taulukko 3. Terveyspalvelujen verkkosivuilla nauhoitettuja käyttöskenaarioita.

Käyttöskenaario	**Terveyspalvelu**
Reseptilääkkeen tilaus	Verkkoapteekit
Ajanvaraus Haku Tietosivut Omien tietojen/tulosten tarkastelu	Terveysasemat ja mielenterveyspalvelut

tilääkkeen tarkastelusta saattaa päätyä kolmannelle osapuolelle. Vastaavasti tältä tuotesivulta voidaan jatkaa tilauspainiketta napsauttamalla seuraavaan skenaarioon eli reseptilääkkeen tilaukseen, jolloin tieto aiotusta tilauksesta ja mahdollisesti tilattavasta reseptilääkkeestä voi lähteä eteenpäin analytiikkapalvelulle. Joidenkin apteekkien kohdalla tilausprosessi oli hieman edellä kuvatusta eroava: aina reseptilääkkeillä ei esimerkiksi ollut varsinaisia tuotesivuja, vaan niitä tilatakseen piti aluksi suorittaa vahva tunnistautuminen.

Terveysasemien ja mielenterveyteen liittyvien verkkopalvelujen kohdalla testattuja käyttöskenaarioita oli enemmän. Skenaarioista tärkein ja kriittisin on ajanvaraus. Tästä tarkasteltiin terveysaseman tarjoamista palveluista riippuen monenlaisia eri variaatioita: ajanvarausta lääkärin tai psykiatrin vastaanotolle, ajanvarausta influenssarokotukseen ja ajanvarausta koronatestiin. Kaikissa näissä tapauksissa on vaara ajanvarauksen yksityiskohtien (kuten lääkärin nimi, ajanvarauksen ajankohta, varatun palvelun tyyppi) vuotamisesta kolmansille osapuolille.

Ajanvarausta edeltää usein hakutoiminnon käyttö. Hakuterminä saattaa olla vaikkapa tietty erikoisala (esim. sukupuolitaudit), sairaus (esim. masennus) tai lääkärin nimi. Hakujen yhteydessä on aina olemassa riski tiedon vuotamisesta – usein tämä voi tapahtua puolivahingossa, kun analytiikkapalvelulle raportoidaan haun tulossivulla kyseisen sivun URL-osoite, jonka perään on liitetty parametrina myös hakutermi.

Terveysasemien sivuilla ja mielenterveyspalveluissa on myös monenlaisia sairauksia koskevia tietosivuja, joiden URL-osoitteiden päätyminen kolmansien osapuolien haltuun voi paljastaa arkaluonteisia tietoja. Viimeisenä skenaariona testattiin myös omien tietojen ja laboratoriotulosten tarkastelua vahvan tunnistautumisen takana ja tutkittiin, ilmenikö sivustojen näissä osioissa tietovuotoja kolmansille osapuolille.

3.2.3 Tapaustutkimus

Tapaustutkimus on lähestymistapa tai tutkimusstrategia, jossa tarkastellaan yhtä tai useampaa tapausta [82]. Tässä väitöskirjassa useimpia julkaisuja voi pitää monia verkkosivustoja analysoivina monitapaustutkimuksina [83], joissa sivustoja (tapauksia) vertaillaan toisiinsa ja niiden yksityisyyden tilaa ja tietovuotoja tarkastellaan ilmiönä etsien näistä eroja ja yhteisiä piirteitä. Tapaustutkimukseen voi sisältyä laadullisia ja määrällisiä osia, kuten tässä väitöskirjassa. Tapaustutkimukseen liittyy toisinaan myös asetelma, jossa tutkittavaa ilmiötä tarkastellaan kahtena (tai useampana) erillisenä ajankohtana, joiden välissä tapahtuu olennainen muutos [84]. Näin on julkaisujen III, IV ja VII tapauksissa, joissa verkkoapteekkien, terveysalan verkkopalvelujen ja kuntien verkkosivujen ylläpitäjillä on ollut mahdollisuus korjata sivustojensa yksityisyyttä tarkastelujen välissä.

Monitapaustutkimukset antavat mahdollisuuden myös eri aihealueiden verkkosivujen keskinäiseen vertailuun ja esimerkiksi julkisen sektorin tilanteen vertaamisen yksityisiin ja kaupallisiin palveluihin. Esimerkiksi julkaisussa VI on tapauksina tutkittu erilaisia terveysalan palveluja (verkkoapteekit, lääkärikeskukset, mielenterveyssivustot), jolloin eri aihealueita voidaan vertailla keskenään laadullisesti ja määrällisesti. Tapaustutkimukselle ja useimmille tämän väitöskirjan julkaisuille on ominaista, ettei niissä pyritä tekemään laajoja yleistyksiä. Sen sijaan tavoitteena on syvällisesti ja havainnollisesti kuvata kohteena olevien verkkosivustojen tietovuotoja ja saada ilmiöstä monipuolista ja kattavaa tietoa.

3.2.4 Tilastollinen analyysi

Vaikkei tämän väitöskirjan otoskokoja useissa julkaisuissa voi pitää tarpeeksi suurina siihen, että niiden perusteella voisi tehdä tilastollisesti päteviä päätelmiä, on tiettyihin aihealueisiin kuuluvien sivustojen tietovuotojen määriä kuitenkin tilastoitu. Yksittäiset sivustot antavat myös määrällisiä esimerkkejä siitä, millainen tilanne tietyn sovellusalan verkkopalveluissa voi tietovuotojen osalta olla. Joissakin tapauksissa (terveydenhuollon palvelut, mielenterveyssivustot) valittujen sivustojen joukon voi myös perustellusti väittää olevan siinä mielessä suhteellisen kattava, ettei kyseisen sovellusalan merkittäviä verkkopalveluja Suomessa ole paljoakaan enempää. Laajan verkkoapteekkitutkimuksen (julkaisu III) ja kuntien (julkaisu VI) osalta mukana ovat kaikki mahdolliset tapaukset, joten näiden alojen osalta tilastollinen analyysi on kattavaa ja luotettavaa.

Tilastoinnin kohteina tässä väitöskirjatutkimuksessa ovat lähinnä verkkosivustoilta löytyneiden kolmansien osapuolien määrät, sekä verkkoliikenneanalyysissä tietovuotojen määrä ja vuotaneiden henkilötietotyyppien määrät. Näitä lukuja tarkastellaan osassa tutkimuksista (julkaisut III ja VI) myös toistomittausaineiston keinoin, kun tietovuotojen määrää mitataan toistuvasti. Suuntaa-antavaa vertailua tehdään

myös eri sovellusalojen välillä ja käyttäjän antaman suostumuksen vaikutuksen osalta.

Tässä väitöstutkimuksessa käytetään deskriptiivistä eli kuvailevaa tilastotiedettä. Kuvaileva tilastotiede tarkoittaa numeerisia menetelmiä ja graafisia kuvaustapoja, joilla havainnollistetaan, tiivistetään ja jäsennetään tietyn aineiston ominaisuuksia [85]. Vaikkapa tietyssä verkkosivustojoukossa havaituista tietovuodoista saatetaan esittää esimerkiksi tilastollinen keskiarvo, mediaani, vaihteluväli ja kolmansien osapuolien osalta esimerkiksi frekvenssi. Tietovuotojen ja kolmansien osapuolien määrien kuvaamiseen käytetään taulukoita, mutta usein myös tilastografiikkaa, kuten pylväs-, ympyrä- ja alluviaalidiagrammeja.

3.2.5 Asiantuntijahaastattelut

Web-kehittäjiä haastattelemalla pyrittiin selvittämään, mistä verkkopalveluiden tietovuodot kolmansille osapuolille kehittäjien mielestä johtuvat ja miten niitä voisi heidän näkemyksensä mukaan välttää. Haastattelun alussa selvitettiin myös ylipäätään, tuleeko kolmansien osapuolien käytön aiheuttama riski mainituksi yksityisyysuhkista puhuttaessa. Väitöstutkimusta varten haastateltiin 12:ta asiantuntijaa julkaisussa IX käyttäen puolistrukturoituja asiantuntijahaastatteluja.

Asiantuntijahaastattelu on tutkimusmenetelmä, jossa haastatellaan alan asiantuntijoita tai henkilöitä, joilla on syvällistä tietämystä halutusta aihepiiristä [86]. Asiantuntijahaastattelu sopii erityisesti tilanteisiin, joissa tarvitaan syvällistä ja yksityiskohtaista tietoa, jota ei välttämättä löydy kirjallisista lähteistä. Haastateltavat valitaan heidän asiantuntemuksensa perusteella. Esimerkiksi tämän väitöskirjan haastattelututkimukseen valittiin henkilöitä, joilla on vuosien kokemus web-kehityksestä ja tätä kautta vahvaa käytännön osaamista.

Asiantuntijahaastattelun suoritustapa vaihtelee. Strukturoidussa haastattelussa on ennalta määrätty lista kysymyksiä ja jokaiselta haastateltavalta kysytään samat kysymykset. Avoin haastattelu taas on vapaamuotoinen keskustelu, jossa keskustelun kulku määrittää haastattelun suunnan. Puolistrukturoitu haastattelu sijoittuu edellisten välille; haastattelussa on valmis kysymysrunko, mutta haastattelija voi myös esittää lisäkysymyksiä ja mukauttaa haastattelua tilanteen mukaan [87, s. 47]. Tässä väitöstutkimuksessa asiantuntijahaastattelut tehtiin puolistrukturoituina haastatteluina, joissa kysymykset kyllä kysyttiin tietyn listan mukaisesti, mutta valmiita vastausvaihtoehtoja ei ollut, ja lisäkysymyksille ja haastateltavien itse esiin nostamille teemoille annettiin tilaa.

3.2.6 Gioia-menetelmä

Haastatteluilla kerätyn aineiston analysointiin käytettiin julkaisussa IX Gioia-menetelmää. Gioia-menetelmä [88] on laadullisen tutkimuksen lähestymistapa, jossa joh-

detaan käsitteitä kerätystä aineistosta systemaattisesti ja induktiivisesti. Menetelmä pohjautuu grounded theory -tutkimusmenetelmään [89], joka kehitettiin laadullisen tutkimuksen systemaattisuuden parantamiseksi. Menetelmässä luodaan teoriaa johtamalla aineistosta käsitteitä ja kategorisoidaan niitä. Gioia-menetelmän ensimmäisenä vaiheena tarkastellaan avoimen koodauksen avulla, mitä käsitteitä (concepts) aineistosta ja haastateltavien käyttämästä kielestä nousee esiin. Avoimessa koodauksessa siis laadullisesta aineistosta tunnistetaan ja nimetään merkityksellisiä käsitteitä ilman ennakko-olettamuksia [89]. Gioia-menetelmän toisessa vaiheessa ryhmitellään käsitteet teemoiksi (themes) ja noustaan abstraktimmalle tasolle haastateltavien omien termien ja näkökulmien yläpuolelle. Kolmannessa vaiheessa teemoista yhdistellään vielä kokoavia, teoreettisia ulottuvuuksia (aggregate dimensions). Näistä ulottuvuuksista voidaan muodostaa teoreettinen viitekehys tai malli.

3.3 Eettiset kysymykset

Väitöskirjan tutkimusten tekotapa noudattaa tutkimuseettisen neuvottelukunnan laatimaa ohjeistusta hyvästä tieteellisestä käytännöstä[5]. Verkkoliikenneanalyysi on asiakaspuolella tehtynä täysin luvallista ja eettistä eikä aiheuta vahinkoa testatuille apteekeille ja verkkosivustoille. Myös muuta haittaa vältettiin aiheuttamasta tehdyissä kokeissa. Esimerkiksi apteekkien tapauksessa lääketilaukset keskeytettiin aina ennen kuin apteekkihenkilökuntaan otettiin yhteyttä chatin kautta, eikä rekisteröitymisprosessin aikana verkkosivustoille rekisteröintiä vaadittaessa luotu tilejä väärillä tiedoilla (yksi tutkimuksen tekijöistä rekisteröityi sivustoille omalla nimellään).

Lisäksi kaikissa tutkimuksissa päätettiin olla nimeämättä tutkittuja tahoja, kuten apteekkeja tai lääkärikeskuksia, ja olla paljastamatta niiden verkkosivustojen osoitteita. Tämä on tehty asianomaisten tahojen maineen suojelemiseksi ja niiden negatiivisen leimautumisen välttämiseksi. Väitöskirjan tekijän ja tutkimusryhmän näkemyksen mukaan tämä päätös on parhaiten linjassa eettisten tutkimuskäytäntöjen kanssa. Lisäksi tämä väitöstutkimus keskittyy muutenkin yksittäisten toimijoiden sijaan itse ilmiöön, eli tunnistettuihin vakaviin tietosuojaongelmiin ja niiden mahdollisiin ratkaisuihin. Lisäksi tutkittujen tahojen nimettömänä pitäminen mahdollistaa osaltaan sen, että tietosuojaviranomainen voi suorittaa tutkimuksia rauhassa ja asianomaiset tahot saavat tilaa käsitellä ja korjata verkkosivustoillaan havaitut tietosuojaongelmat.

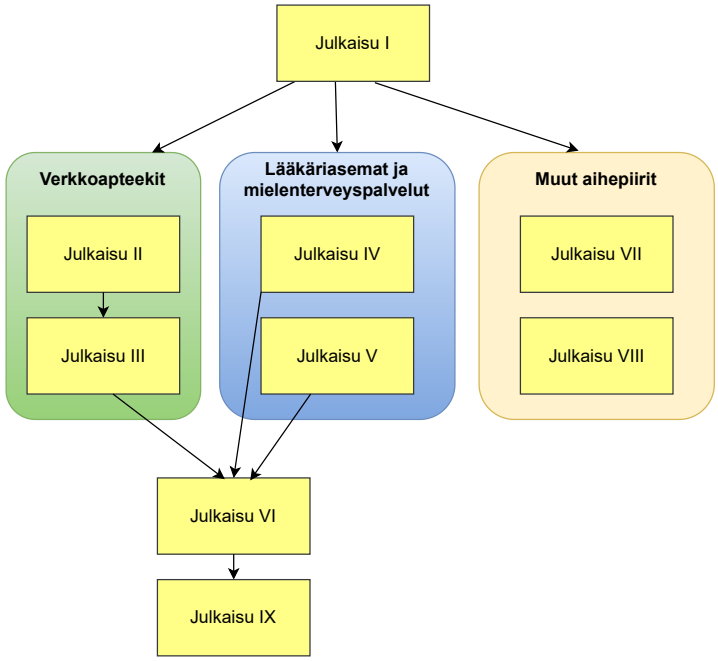
Myös asiantuntijahaastatteluissa noudatettiin eettisiä tutkimusperiaatteita. Haastattelun suunnittelussa otettiin huomioon, että tietyt asiat (kuten kolmannet osapuolet) mainittiin vasta haastattelun edetessä, ettei vaikutettaisi vääristävästi aiempien kysymysten vastauksiin. Haastatteluissa kerättiin nauhoitettu suostumus haastattelun tekoon ja vastausten käyttöön tutkimustarkoituksiin. Henkilöt ja yritykset pidet-

[5] https://tenk.fi/fi/tiedevilppi/hyva-tieteellinen-kaytanto-htk

tiin lopullisessa julkaisussa anonyymeinä ja esitetyistä vastauksista poistettiin kaikki tieto, josta voisi tunnistaa henkilön tai yrityksen. Haastateltavia informoitiin haastattelun luottamuksellisuudesta, aiheesta ja tarkoituksesta, sekä siitä, miten saatuja tietoja käytetään. Haastattelujen nauhoitteet ja niiden tekstiversiot tallennettiin tietosuoja huomioiden.

3.4 Tutkimusprosessi ja julkaisut

Väitöskirjan tutkimustyö on tehty osana IDA – Intiimiys datavetoisessa kulttuurissa -projektia. Alun perin tämän tutkimusprojektin ohjelmistoteknisessä osuudessa tutkittiin julkisen sektorin web- ja mobiilipalveluja kolmannen osapuolen analytiikan ja siitä aiheutuvien datavuotojen osalta. Näiden palvelujen ohella tarkasteltiin myös erilaisten haavoittuvien ryhmien usein käyttämiä verkkopalveluja, ja tässä yhteydessä esille nousivat myös verkkoapteekit.



Kuva 6. Tutkimusprosessin kulku ja julkaisut.

Kuvassa 6 näkyvät väitöskirjaan liittyvän tutkimusprosessin eteneminen ja siihen liittyvät julkaisut. Tutkimustyö lähti käyntiin edellä mainitusta julkisen sektorin verkkopalvelujen tarkastelusta (julkaisu I), jonka yhteydessä selvitettiin tutkittavien tietovuotojen luonnetta ja testattiin systemaattista verkkoliikenneanalyy-

sia. Tätä ensimmäistä julkaisua voi osaltaan pitää myöhempää verkkoapteekkien ja terveysalan verkkopalvelujen tarkastelua tukevana esikartoituksena. Samoin ennen verkkoapteekkien ja terveyspalvelujen toteutettu pienimuotoinen kirjallisuuden läpikäynti osoitti, että terveysalan verkkopalveluiden ja erityisesti verkkoapteekkien sivuilla esiintyvästä analytiikasta kaivataan lisää tutkimusta.

Verkkoapteekkien tarkastelu aloitettiin julkaisussa II, jossa tarkastellaan 20 suomalaisen verkkoapteekin sivuilta löytyvää analytiikkaa. Huolestuttavien, arkaluonteisten terveystietojen vuotamiseen liittyvien tulosten takia tätä tarkastelua oli luontevaa laajentaa kattamaan kaikki suomalaiset verkkoapteekit, mistä seurasi julkaisu III.

Verkkoapteekeista tehdyt löydökset herättivät luonnollisesti huolta laajemminkin koko sosiaali- ja terveysalan verkkopalveluiden yksityisyydestä ja niissä mahdollisesti käytetystä analytiikasta. Tämän takia tarkastelua laajennettiin julkaisussa IV kattamaan terveysasemien (lääkärikeskukset, terapiatalot, laboratoriot) verkkosivustoja sekä julkaisussa V erilaisia mielenterveyteen liittyviä sivustoja. Kaikki viisi ensimmäistä julkaisua vastaavat laajasti ensimmäiseen tutkimuskysymykseen (TK1) terveysalan verkkopalvelujen sivuilla esiintyvästä analytiikasta ja sille vuotavista terveystiedoista.

Tarkasteltujen verkkopalvelujen osalta kiinnostus kohdistui myös siihen, voiko käyttäjä itse kunnolla vaikuttaa vuotavien terveystietojen määrään suostumukseen liittyvillä valinnoilla ja tuleeko käyttäjä riittävissä määrin informoiduksi kolmansille osapuolille vuotavista tiedoista ja niitä käsittelevistä tahoista (TK3). Ensimmäiseen kysymykseen liittyen julkaisussa VI tarkasteltiin muutamaa verkkopalvelua kustakin aiempien terveyspalveluja käsittelevien julkaisujen kategoriasta – verkkoapteekkeja, lääkärikeskuksia ja mielenterveyspalveluja – antamatta suostumusta tietojenkeruuseen. Toista kysymystä käyttäjän informoinnista tarkastellaan useissa julkaisussa (I, II, III, V, VII, VIII).

IDA-projektissa on tarkasteltu tietovuotoja hyvin moneen erilaiseen aihepiiriin liittyvillä verkkosivuilla, ja tähänkin väitöskirjaan nähtiin tarpeelliseksi sisällyttää vertailun vuoksi muutamia muita aihepiirejä terveysalan lisäksi. Julkisen sektorin verkkopalveluja tarkastelevan julkaisun I lisäksi julkaisussa VII tarkastellaan suomalaisten kuntien verkkosivujen tietovuotoja ja julkaisussa VIII vaalikoneiden tietovuotoja.

Lopulta julkaisussa IX etsittiin vielä kirjallisuuden ja asiantuntijahaastatteluiden avulla syitä siihen, miksi tarkastelluissa verkkopalveluissa esiintyi tutkimuksessa havaittuja yksityisyyspuutteita. Lisäksi julkaisussa pohditaan, mitä ohjelmistokehityksen osalta olisi voinut tehdä toisin paremman yksityisyyden tason saavuttamiseksi. Tältä pohjalta laadittiin suosituksia siitä, millaisilla toimilla kriittisissä verkkopalveluissa voisi välttää arkaluonteisten tietojen vuotamisen kolmansille osapuolille (TK4). Julkaisujen ja tutkimuskysymysten kaikki yhteydet – joita on enemmänkin kuin edellä mainitut – näkyvät taulukossa 4.

Taulukko 4. Julkaisujen ja tutkimuskysymysten väliset yhteydet.

Tutkimuskysymys	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**
TK1: Kolmannet osapuolet ja niille vuotavat tiedot verkkopalveluissa?	X	X	X	X	X	X			
TK2: Julkisen sektorin tilanne? Tilanne muilla toimialoilla?	X				X	X	X	X	
TK3: Käyttäjän informointi ja mahdollisuudet vaikuttaa vuotoihin suostumuksen kautta?	X	X	X		X	X	X	X	
TK4: Tietovuotojen syyt ja niiden korjaaminen?	X	X	X	X	X	X	X	X	X

3.5 Yhteenvedot alkuperäisjulkaisujen sisällöistä

Seuraavassa esitetään yhteenveto jokaisen alkuperäisjulkaisun sisällöstä. Varsinaisen sisällön kuvaamisen lisäksi käsitellään jokaisen julkaisun merkitystä väitöskirjan ja sen tutkimuskysymysten kannalta. Samaten kuvataan väitöskirjan tekijän kontribuutio kunkin julkaisun osalta.

3.5.1 Julkaisu I: Assessing discrepancies between network traffic and privacy policies of public sector web services

Yhteenveto sisällöstä

Julkaisussa I [7] tarkastellaan 34 julkisen sektorin toimijan ylläpitämän, yleisesti kansalaisten käyttämän verkkopalvelun yksityisyyttä. Palvelujen kolmansille osapuolille lähettämää verkkoliikennettä analysoidaan ja selvitetään, mitä henkilötietoja tutkitut palvelut mahdollisesti välittävät kolmansille osapuolille ja analytiikkapalveluille. Tutkimuksessa tarkastellaan myös verkkopalveluiden tietosuojaselosteita ja arvioidaan niiden läpinäkyvyyttä ja selkeyttä vertaamalla niiden sisältöä palveluiden lähettämän verkkoliikenteen sisältöön.

Tutkimus keskittyy monia muita aikaisempia tutkimuksia tarkemmin erittelemään, millaisia yksilöiviä, tietyn käyttäjän identifioivia henkilötietoja kolmansille osapuolille vuotaa. Tutkimus jatkaa ja laajentaa näin esimerkiksi Libertin [34] ja Zheutlinin yms. [57] aikaisempaa tutkimusta kriittisillä verkkosivustoilla olevista

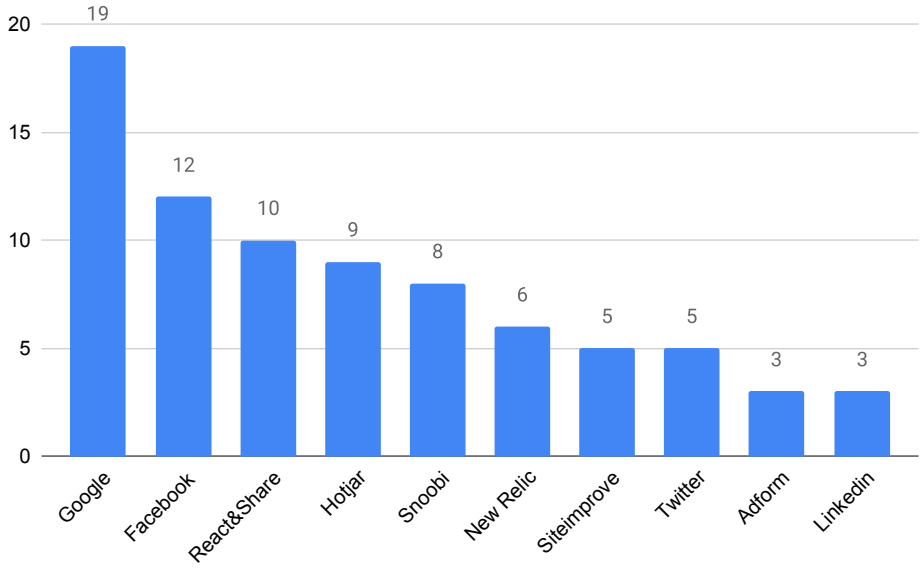
kolmansista osapuolista ja osin myös päivittää niitä. Kuten taulukossa 1 esitettiin, IP-osoitteen lisäksi tällaisia ovat monet muutkin tekniset laitteeseen, käyttöjärjestelmään ja selaimeen liittyvät tiedot, sekä evästeissä säilytettävät yksilölliset tunnisteet. Useiden tarkasteltujen sivustojen osalta vaikuttaa siltä, etteivät kehittäjät ole loppuun asti miettineet, millaisiin asiayhteyteen liittyviin tietoihin yksilöivät tiedot voivat yhdistyä.

Löydöksissä paljastuu lukuisia epäjohdonmukaisuuksia tietosuojaselosteiden ja tutkittujen verkkopalveluiden todellisen liikenteen välillä. Käyttäjille ei monissa tapauksissa kerrottu selkeästi edes sitä, että heidät voidaan yksilöidä kolmansien osapuolien keräämien tietojen perusteella. Toinen tärkeä havainto on tutkittujen verkkosivustojen käyttämien analytiikkapalvelujen korkea määrä. Sivustot käyttivät keskimäärin 3,1 analytiikkapalvelua, mitä voidaan pitää suurena lukuna, kun kyse on julkisen sektorin sivustoista. Tutkitusta 34 verkkopalvelusta 31 (91,2 %) lähetti vähintään yhdelle analytiikkapalvelulle sekä käyttäjän yksilöiviä tunnisteita (mm. IP-osoite) että asiayhteyden paljastavan URL-osoitteen. Erityisesti julkisen sektorin tarjoamissa välttämättömissä verkkopalveluissa tämä voi olla hyvin ongelmallista. Esimerkiksi tilanteessa, jossa käyttäjä vierailee sivulla, jolla annetaan ohjeita turvapaikan hakemiseen Suomesta, sivuvierailun paljastavan tiedon vuotaminen yhdistettynä yksilöivään henkilötietoon on tietovuotona hyvin arkaluonteinen. Analytiikkapalveluista Google Analytics oli käytössä 19 sivustolla, Metan analytiikkaa taas 12 sivustolla. Nämä ovatkin johdonmukaisesti kaksi suurinta henkilötietoja vastaanottavaa kolmatta osapuolta kaikissa tämän väitöskirjan tutkimuksissa. Kuvassa 7 näkyvät yleisimmät tutkimuksessa havaitut kolmannet osapuolet.

Johtopäätöksenä voidaan todeta, että web-kehittäjien ja julkisen sektorin toimijoiden on edelleen parannettava tietoisuuttaan verkkopalveluidensa kolmansille osapuolille välittämästä henkilötiedoista ja olemassa olevista tietosuojaregulaatiosta. Paljon työtä on myös tehtävä yksityisyyteen liittyvien asioiden selkeässä ja läpinäkyvässä ilmaisemisessa käyttäjille. Tämä on erityisen tärkeää, koska julkaisussa tarkasteltujen julkisen sektorin toimijoiden ja viranomaistahojen voisi olettaa toimivan tässä asiassa muille esimerkkeinä.

Julkaisun merkitys

Julkaisu I on IDA-projektin ja tämän väitöskirjan ensimmäinen kolmannen osapuolen tietovuotoja web-ympäristössä tarkasteleva tutkimus. Tarkoituksena on selvittää tällaisten tietovuotojen luonnetta ja yleisyyttä käyttäjien jokapäiväisessä web-asioinnissa. Julkaisussa osoitetaan tiettävästi ensimmäisen kerran, että käyttäjän henkilötietojen päätyminen kolmansille osapuolille käyttäjän tätä huomaamatta on ongelma myös suomalaisilla julkisen sektorin verkkosivuilla ja tärkeissä kansalaispalveluissa. Tämä antaa aihetta tarkastella samaa ilmiötä laajemminkin.



Kuva 7. Yleisimmät (vähintään 3 esiintymää) tutkimuksessa havaitut analytiikkapalvelut.

Vaikka julkaisussa kuvaillut URL-vuodot kolmansille osapuolille eivät sinänsä ole yksityisyystutkimuksessa täysin uusi ilmiö [34], ne ovat edelleen selvä sokea piste monien kriittisten verkkopalvelujen yksityisyydessä. Tämän takia ilmiön tarkka kuvaus ja sen tarkastelu erilaisten aihepiirien verkkosivustoilla ja yksityisyysuhan esiintuominen sekä tiedeyhteisölle että web-kehittäjille on arvokasta. On tärkeää, että ongelma saadaan korjattua erityisesti julkisella sektorille ja muissa kriittisissä web-pohjaisissa palveluissa.

Tässä tutkimuksessa käytettiin myös ensimmäistä kertaa IDA-projektin aikana verkkoliikenneanalyysiä aineistonkeruumenetelmänä ja kehitettiin tähän liittyvä, aiemmin tässä luvussa kuvailtu prosessi. Verkkoliikenteen analysointi tähän tapaan ei ole pelkästään toimiva aineistonkeruumenetelmä tieteelliseen tutkimukseen, vaan myös yrityksille ja organisaatioille hyvin käyttökelpoinen keino tarkastella, mitä henkilötietoja modernit ja monimutkaisia riippuvuuksia sisältävät verkkosivustot lähettävät kolmansille osapuolille.

Tekijän kontribuutio

Väitöskirjan laatija on ideoinut tutkimuksen, ollut mukana kerätyn aineiston analysoinnissa, kirjoittanut suurimman osan julkaisusta, ja johtanut tutkimuksen tekemistä.

3.5.2 Julkaisu II: Several Online Pharmacies Leak Sensitive Health Data to Third Parties

Yhteenveto sisällöstä

Julkaisussa II [90] keskitytään verkkoapteekkeihin, jotka ovat hyvin tärkeä esimerkki välttämättömistä digitaalisista palveluista. Näissä verkkopalveluissa käsitellään arkaluonteisia tietoja, kuten reseptilääkkeiden tilauksia, minkä vuoksi on tärkeää asettaa verkkoapteekeille tiukkoja yksityisyysvaatimuksia. Perinteisten kivijalka-apteekkien on pidettävä asiakkaan terveydentilaa koskevat tiedot luottamuksellisina, ellei niiden luovuttamiselle kolmansille osapuolille ole laillista perustetta, kuten asiakkaan nimenomainen suostumus tai laillinen velvoite. Olisi tärkeää säilyttää digitaalisissa palveluissa samat luottamuksellisuuden ja yksityisyyden periaatteet.

Kuitenkin useissa välttämättömissä verkkopalveluissa arkaluonteisia henkilötietoja voidaan tahattomasti toimittaa kolmansille osapuolille, mikäli kehittäjät ja tietosuojavastaavat eivät ole suunnitelleet verkkosivustoa huolellisesti yksityisyyskysymykset huomioiden. Tässä tutkimuksessa analysoidaan 20 suomalaisen verkkoapteekin verkkoliikennettä ja tarkastellaan, vuotaako arkaluonteisia henkilötietoja (mahdollisesti jopa terveystiedoiksi luokiteltavia tietoja) kolmansille osapuolille. Erityisesti mielenkiinnon kohteena ovat käyttäjän tarkastelemat ja tilaamat reseptilääkkeet.

Tutkimus on Zheutlinin yms. [57] tekemän analyysin ohella ainoa, jossa tarkastellaan kolmannen osapuolen palveluja apteekkien verkkosivuilla. Tästä aiemmasta tutkimuksesta poiketen tähän väitöskirjaan kuuluva tutkimus ei kuitenkaan tarkastele vain apteekkien sivuilla olevia kolmansia osapuolia, vaan erityisesti myös niiden keräämiä henkilötietoja. Lisäksi tarkastelemme ilmiötä nimen omaan suomalaisissa verkkoapteekeissa.

Apteekkien verkkoliikenteen analysoimiseen käytettiin Google Chromen kehittäjätyökaluja, joilla nauhoitettiin apteekkien verkkosivuilla navigoitaessa syntynyt verkkoliikenne. Testeissä käytettiin apteekin etusivulla tavallisesti olevaa hakutoimintoa, jonka kautta päästiin halutun lääkkeen tuotesivulle, josta edelleen edettiin tilaamaan lääkettä. Varsinaista tilausta ei suoritettu loppuun asti. Verkkoliikenteestä etsittiin reseptilääkkeen nimeä tai reseptilääkkeen paljastavan tuotesivun URL-osoitetta. Koeasetelmassa annettiin suostumus evästeisiin ja tiedonkeruuseen, mutta on syytä huomata, ettei terveystietojen luovuttaminen silti ole hyväksyttävää sitä ja tietojen vastaanottajia erikseen mainitsematta. Verkkoliikenteen lisäksi tarkasteltiin myös apteekkien tietosuojaselosteita, joista etsittiin mahdollisia mainintoja reseptilääketietojen lähettämisestä kolmansille osapuolille.

Tarkastelluista 20 verkkoapteekista 7 vuoti tiedon tietyn nimetyn reseptilääkkeen tarkastelusta (tuotesivun avaaminen) ja vastaavasti samat 7 apteekkia vuotivat tiedon aikomuksesta tilata tietty lääke. Kaikkiaan 14 apteekin sivuilla oli analytiikkaa, ja nämä kaikki vuotivat tiedon siitä, että käyttäjä aikoi tilata jonkin lääkkeen (tieto aikomuksesta tehdä tilaus, mutta lääkettä ei välttämättä nimetty). Kolmansille osapuo-

lille välitetyt lääkkeiden katsomiseen ja tilaamiseen liittyvät tiedot näkyvät taulukossa 5. Niistä 14 apteekista, jotka vuotivat tietoja, vain kolmen tietosuojaselosteita saattoi tulkita niin, että tietoja voidaan mahdollisesti luovuttaa kolmansille osapuolille. Selosteista ei kuitenkaan suoraan voinut tehdä sellaista johtopäätöstä, että nimen omaan terveystietoja tai reseptilääketietoja luovutettaisiin.

Taulukko 5. Kolmansille osapuolille välitetyt lääkkeiden katsomiseen ja tilaamiseen liittyvät tiedot.
© Springer Nature 2024

	Sivustolla analytiikkaa	Tietyn lääkkeen katselusta lähtee tieto	Aikomus tilata lääke paljastuu	Aikomus tilata lääke voidaan yhdistää tiettyyn lääkkeeseen
Apteekki 1	X	X	X	X
Apteekki 2	X	X	X	X
Apteekki 3	X	X	X	X
Apteekki 4	X	X	X	X
Apteekki 5				
Apteekki 6				
Apteekki 7				
Apteekki 8				
Apteekki 9	X	X	X	X
Apteekki 10				
Apteekki 11	X		X	
Apteekki 12	X		X	
Apteekki 13	X		X	
Apteekki 14	X		X	
Apteekki 15	X		X	
Apteekki 16	X		X	
Apteekki 17	X		X	
Apteekki 18	X	X	X	X
Apteekki 19	X	X	X	X
Apteekki 20				

Vaikkei otos ollut tässä tutkimuksessa verkkoapteekien kokonaismäärään nähden suuri, oli sen perusteella hyvin selvää, että 1) joissakin suomalaisissa verkkoapteekeissa on vakava tietosuojaongelma ja 2) on erittäin todennäköistä, ettei ongelma rajoitu vain tässä tarkasteltuihin verkkoapteekkeihin. Vaikka yksityisyyden suojassa on sivustoilla vakavia puutteita, eikä niitä selvästikään ole testattu ja auditoitu kattavasti ja asianmukaisesti, ovat löydetyt yksityisyys- ja tietosuojaongelmat kuitenkin

ilmeisen tahattomasti syntyneitä. Tietovuodot hyödyttävät vain arkaluonteisten tietojen vastaanottajia, joten apteekeilla ei ole mitään syytä tietoisesti luovuttaa näitä tietoja kolmansille osapuolille.

On selvää, että reseptilääkkeet, joita käyttäjä aikoo tilata, liittyvät yleensä vahvasti hänen terveydentilaansa ja mahdollistavat oireisiin ja sairauksiin liittyvät päätelmät. On myös tärkeää ymmärtää aikaulottuvuuden merkitys: kun kolmannen osapuolen analytiikkapalvelulla on mahdollisuus havaita lukuisia tilauksia pidemmällä aikavälillä, paljastuu yksityiskohtia henkilön potilashistoriasta. Tällaisten tietojen siirtäminen kolmansille osapuolille on epäeettistä ja tarpeetonta. Näin on siitä huolimatta, ettei tutkimuksemme perusteella luonnollisestikaan voi sanoa, säilyttävätkö reseptilääketietoja vastaanottavat tahot niitä ja käytetäänkö niitä mitenkään hyväksi. Riski on joka tapauksessa olemassa, ja tietojen lähettäminen kolmansien osapuolien palvelimille on jo itsessään suuri ongelma.

Ohjelmistokehittäjillä ei välttämättä aina ole selvää ymmärrystä siitä, että verkkoapteekki ei yksityisyysvaatimusten osalta vastaa tavallista verkkokauppaa. Siksi käytetyt ratkaisut alustoineen ja analytiikkapalveluineen saattavat olla sopimattomia verkkoapteekeille. Ohjelmistokehittäjien ja tietosuojavastaavien tulisi kiinnittää enemmän huomiota siihen, millaisia datavirtoja heidän verkkosivustoiltaan lähtee kolmansille osapuolille. Tämä on helppo toteuttaa samanlaisella koeasetelmalla ja verkkoliikenneanalyysillä, joita tässä tutkimuksessa on käytetty.

Julkaisun merkitys

Julkaisu II soveltaa julkaisun I yhteydessä kehitettyä tiedonkeruu- ja analyysimenetelmää verkkoapteekkien tietovuotojen analysoimiseen. Aihealueena tätä voidaan pitää vuotavien tietojen arkaluonteisuuden vuoksi tämän väitöskirjatutkimuksen kaikkein kriittisimpänä. Vaikka julkisen sektorin ylläpitämien kansalaisille suunnattujen verkkopalvelujen yksityisyys on hyvin tärkeää, ovat verkkoapteekkien yksityisasiat vielä selvästi arkaluonteisempia. Näin ollen tutkimus osoittaa, etteivät kaikki hyvin kriittisetkään verkkopalvelut ole yksityisyydeltään riittävällä tasolla, kun tarkastellaan kolmannen osapuolen tietovuotojen esiintymistä. Tämä tutkimus antoi myös selkeää aihetta laajempaan jatkotutkimukseen verkkoapteekkien suhteen, sillä Suomessa oli tutkimuksen tekohetkellä 163 reseptilääkkeitä myyvää verkkoapteekkia.

Tekijän kontribuutio

Väitöskirjan laatija on ideoinut tutkimuksen, ollut mukana kerätyn aineiston analysoinnissa, kirjoittanut suurimman osan julkaisusta, ja johtanut tutkimuksen tekemistä.

3.5.3 Julkaisu III: Analyzing third-party data leaks on online pharmacy websites

Yhteenveto sisällöstä

Koska julkaisun II tulokset olivat hyvin huolestuttavia, päätettiin julkaisussa III [91] tehdä kaikki Suomen repseptilääkkeitä myyvät verkkoapteekit kattava laajempi analyysi, jossa selvitettiin niiden sivustoilla olevia kolmansia osapuolia ja tietovuotoja.

Toimivia reseptilääkkeitä myyviä verkkoapteekkeja oli Suomessa tutkimuksen tekohetkellä 163; 48 apteekin verkkoliikenne analysoitiin manuaalisesti ja loput 125 automaattisella työkalulla. Manuaaliseen analysointiin käytettiin Google Chromen kehittäjätyökaluja aiemmin julkaisun II yhteydessä kuvatulla tavalla. Automaattinen analysointi tehtiin hyödyntämällä Website Evidence Collector -työkalua [6]. Kummallakin menetelmällä kerätystä datasta etsittiin käyttäjiä yksilöiviä tietoja ja katseltavan tai tilattavan reseptilääkkeen nimeä.

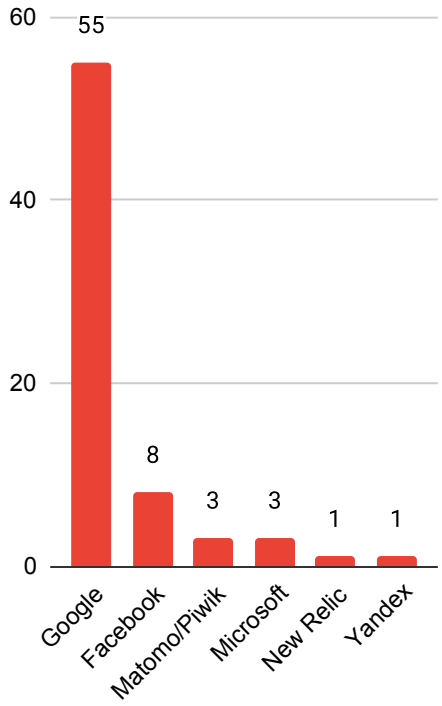
Kaikkia apteekkeja tarkastellessa tutkimus paljastaa, että 163 verkkoapteekista 57 (35,0 %) vuoti tietyn reseptilääkkeen nimen kolmansille osapuolille joko lääkkeen tuotesivun tai tilausta aloitettaessa, ja reseptilääkkeen nimi oli edelleen liitettävissä tietyn käyttäjän yksilöiviin tunnistetietoihin (kuten IP-osoite). Tämä tarkoittaa siis, että tutkimuksen tekohetkellä noin joka kolmannessa verkkoapteekissa oli riski tarkasteltujen tai tilattujen lääkkeiden paljastumiselle!

Tutkimuksessa tarkasteltiin myös, minne kerätyt arkaluonteiset lääketiedot päätyivät. Huomattavia löydöksiä olivat esimerkiksi tiedonsiirrot Bingille (Microsoft) ja Yandexille, joiden palvelimet hyvin todennäköisesti sijaitsivat Pohjois-Amerikassa ja Venäjällä. Näin käyttäjän terveyteen liittyviä tietoja saatetaan apteekkien ymmärtämättä siirtää EU:n ja Euroopan talousalueen ulkopuolelle, mikä on myös GDPR:n valossa ongelmallista.

Esimerkiksi usein EU-alueella ongelmallisena pidetty Google Analytics [4] oli käytössä 134 (81,6 %) apteekin verkkosivuilla. Lääkkeen nimi vuoti Googlelle peräti 55 (33,7 %) verkkoapteekista. Tämä antaa Googlelle hyvin kattavan näkymän suomalaisten verkkoapteekkien käyttöön. Tulos on huolestuttava, kun otetaan huomioon Googlen ja Metan kaltaisten suurten teknologiajättien kyky yhdistää käyttäjän toiminta verkossa hänen oikeaan nimeensä. Kuvassa 8 näkyy, kuinka moni verkkoapteekki vuoti lääkkeen nimen kullekin kolmannelle osapuolelle.

Tutkimuksessa tarkasteltiin myös, miten löydösten raportointi viranomaisille vaikutti tietovuotojen määrään yhden vuoden aikana. Kuvassa 9 näkyy lääkkeen nimen sisältäneiden tietovuotojen määrän lasku vuoden aikana 48 apteekin otoksessa. Määrä on laskenut selvästi 33 tietovuodosta 4 tietovuotoon, ja asian tutkinta on edelleen (alkuvuodesta 2025) vireillä tietosuojavaltuutetun toimistossa. Vaikka tutkimuk-

[6] https://joinup.ec.europa.eu/collection/free-and-open-source-software/solution/website-evidence-collector

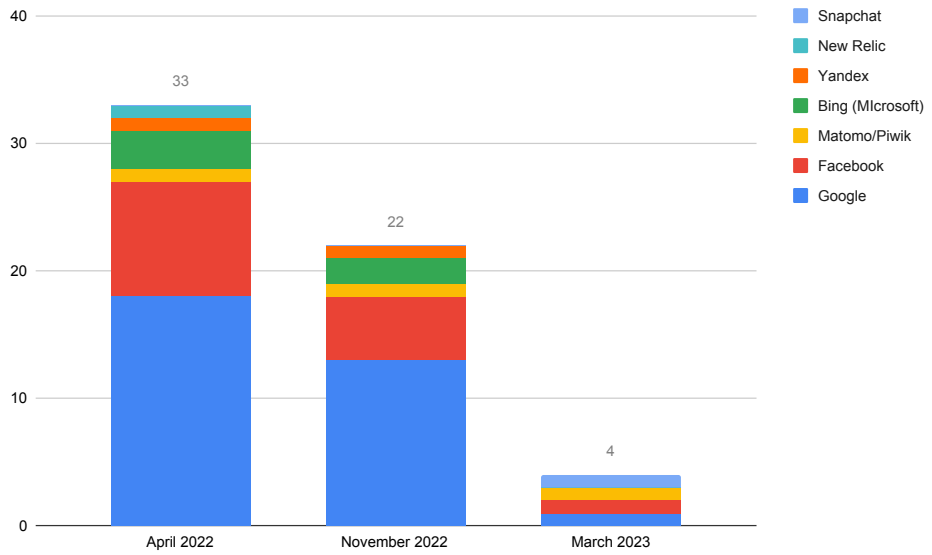


Kuva 8. Kullekin kolmannelle osapuolelle lääkkeiden nimiä vuotaneiden verkkoapteekkien määrä.

sen havainnot siis alun perin ovatkin huolestuttavia, on hyvin positiivista, että suurin osa vuodoista oli vuoden päästä tutkimuksen tekemisestä (keväällä 2023) korjattu.

Tutkimus osoittaa myös, että verkkoapteekkien käyttämät valmiit verkkokauppa-alustat ovat usein selvässä yhteydessä käytettyihin kolmannen osapuolen analytiikkapalveluihin ja siihen, mitä arkaluonteisia henkilötietoja apteekin sivuilla kerätään. Vaikka monissa tapauksissa kyse on siitä, että käytetään yleisiä verkkokauppa-alustoja, jotka eivät sovellu apteekeille, esiintyi myös varta vasten apteekkeja varten tehdyissä alustoissa samanlaisia analytiikan laajamittaiseen käyttöön liittyviä ongelmia.

Oikeustieteelliseltä kannalta tutkimuksessa todetaan, että koska tietoja yksilöitävissä olevan henkilön lääkehankinnoista voidaan käyttää henkilön terveystilanteen selvittämiseen (ks. esim. [92, s. 39–41]) ja suurilla teknologiayrityksillä on resursseja analysoida tietoja edistyneillä tekniikoilla (ks. [93, s. 239–241]), ”terveystiedon”käsitettä tulisi tulkita laajasti siten, että verkkoapteekkien vuotamat tiedot katsotaan GDPR:n tarkoittamiksi terveystiedoiksi. Koska tämän arkaluonteisen tiedon ei-toivotulla käytöllä voi olla apteekkien asiakkaille merkittäviä vaikutuksia ja tiedon paljastaminen kolmansille osapuolille kasvattaa henkilötietojen käsittelyn riskitasoa, muu tulkinta vähentäisi asiakkaiden oikeutta tietosuojaan ja yksityisyyteen.



Kuva 9. Muutos käyttäjän yksilöivien ja lääkkeen nimen sisältävien tietovuotojen määrässä vuoden aikana (N=48). Yhdeksi tietovuodoksi lasketaan lääkkeen nimen vuotaminen yhdelle kolmannelle osapuolelle, eli tietovuotoinstansseja voi olla yhdessä verkkoapteekissa useita.

Vaikka tämän tutkimuksen tulokset eivät olleet määrällisesti aivan niin vakavia kuin julkaisun II tulosten perusteella olisi voinut odottaa, on lääkkeiden nimien vuotaminen yhdistettynä tiettyyn käyttäjään yli kolmasosassa verkkoapteekeista erittäin huolestuttava tulos. Löydösten myötä jo julkaisussa II web-kehittäjille annetut suositukset mm. kolmannen osapuolen palvelujen tarkasta läpikäynnistä ja verkkoliikenteen analysoinnista ovat edelleen olennaisen tärkeitä ja hyvin perusteltuja. Tutkimuksen tulokset korostavat tarvetta tarkalle apteekkien sivustojen tietoturvatestaukselle ja auditoinnille.

Julkaisun merkitys

Julkaisu III on tämän väitöskirjan tärkein yksittäinen julkaisu siinä mielessä, että se antaa laajassa mittakaavassa esimerkin siitä, mihin verkkopalvelujen yksityisyyden huolimaton toteuttaminen voi johtaa. Yli kolmasosa verkkoapteekeista vuosi reseptilääkkeiden nimiä niitä tarkastellessa tai tilattaessa, mikä on tuloksena hyvin hälyttävä. Tutkimus osoittaa myös, mitä voi tapahtua, kun tietty hyvin kriittinen toiminto, tässä tapauksessa lääkkeiden myynti verkossa on hajautettu suurelle määrälle pienten toimijoiden ylläpitämiä verkkosivustoja. Halu seurata liiketoimintatavoitteiden toteutumista analytiikan avulla, vähäiset resurssit yksityisyyden ylläpitämiseen,

tietämättömyys ja auditoinnin puute ovat mahdollisia syitä ei-toivotuille tietovuodoille.

Tutkimus on ensimmäinen laajamittainen katsaus verkkoapteekkien kolmannen osapuolen tietovuotoihin. Lisäksi se yhdistää poikkitieteellisesti uudella tavalla kolmannen osapuolen tietovuotojen teknisen analyysin oikeustieteelliseen analyysiin. Tieteellisen merkityksen ohella sillä on suuri yhteiskunnallinen merkitys; tutkimuksen tulokset on raportoitu tietosuojavaltuutetun toimistolle, joka selvittää edelleen tätä kirjoitettaessa tietovuotoja lukuisien verkkoapteekkien osalta. On myös hyvin tärkeää, että tutkimuksen löydösten perusteella on jo saavutettu huomattava lasku verkkoaptekkien palveluissa esiintyvien tietovuotojen määrässä.

Tutkimuksen tulokset antavat aihetta tarkastella tilannetta apteekkien lisäksi myös muiden mahdollisesti käyttäjän terveydentilaan liittyviä tietoja vuotavien palvelujen osalta. Tällaisia sovellusaloja ovat selvästi ainakin lääkäriasemien, laboratorioiden ja terapiatalojen verkkosivut (julkaisu IV) sekä erilaiset mielenterveyteen liittyvät sivustot ja verkkopalvelut (julkaisu V).

Tekijän kontribuutio

Väitöskirjan laatija on ideoinut tutkimuksen, ollut mukana kerätyn aineiston analysoinnissa, kirjoittanut suurimman osan julkaisusta, raportoinut löydökset tietosuojavaltuutetun toimistolle ja johtanut tutkimuksen tekemistä.

3.5.4 Julkaisu IV: Health data leaks to third parties in web-based health services

Yhteenveto sisällöstä

Julkaisujen II ja III verkkoapteekkeihin liittyvien tutkimustulosten perusteella päätettiin julkaisussa IV [94] laajentaa tarkastelua myös muihin terveydenhuollon palveluihin – lääkäriasemiin, laboratorioihin ja terapiataloihin. Näiden toimijoiden osalta selvitettiin niiden verkkosivuilla olevia kolmansia osapuolia ja niille mahdollisesti vuotavia henkilötietoja useissa eri skenaarioissa.

Määrällisesti laajan tarkastelun sijaan tässä julkaisussa keskityttiin kymmeneen yksityiseen terveydenhuollon palveluntarjoajaan, joiden sivustoilta nauhoitettiin verkkoliikennettä 1) sairauksiin ja oireisiin liittyvien tietosivujen katselun, 2) hakutoiminnon käytön, 3) ajanvarauksen tekemisen ja 4) omien terveystietojen tarkastelun osalta. Nämä neljä käyttöskenaariota kattavat suuren osan tarkasteltujen sivustojen arkaluonteisten tietojen käsittelystä ja sisältävät laajan joukon mahdollisesti vuotavia henkilötietoja ja terveystietoja.

Lääkärikeskusten, laboratorioiden ja terapiatalojen verkkosivujen liikenteen analysointiin käytettiin Google Chromen kehittäjätyökaluja, ja eri käyttöskenaarioissa

syntynyt verkkoliikenne nauhoitettiin. Verkkoliikenteestä etsittiin henkilö- ja terveystietoja. Esimerkiksi eri sairauksiin liittyvien tietosivujen selaamisen kannalta oli mielenkiintoista, vuotaako tietosivun aiheen paljastava URL-osoite kolmannelle osapuolelle. Vastaavasti hakutoiminnon yhteydessä tarkasteltiin, vuotaako hakutermi. Ajanvarauksen yhteydessä on suuri määrä potentiaalisia vuotavia tietoja – kriittisimpinä esimerkiksi lääkärin nimi ja toimenpide, johon aika varataan (esimerkiksi rokotus). Omien tietojen tarkastelun osalta analysoitiin, vuotavatko esimerkiksi käyttäjän laboratoriotulokset kolmansille osapuolille. Tutkimuksen asetelmaan kuului, että siinä annettiin suostumus evästeisiin ja tiedonkeruuseen – mikä ei toki silti tarkoita, että käyttäjän voisi ajatella ymmärtävän terveystietojensa mahdollisesti vuotavan.

Yleisimmät kolmannet osapuolet tutkituissa 10 terveyspalvelussa olivat Google Analytics ja Meta Pixel. Google oli läsnä jokaisella 10 verkkosivulla ja Meta 8 palvelussa. Kolmansien osapuolten keskimääräinen määrä terveyspalvelua kohden oli 5,2, mitä voi pitää hyvin suurena määränä verkkosivustoilla, jotka käsittelevät arkaluonteisia terveyteen liittyviä tietoja. Esimerkiksi verkkosivulla WS1 oli käytössä peräti 9 kolmannen osapuolen palvelua. On yllättävää, että ainakin osittain samoja toimintoja sisältäviä kolmannen osapuolen palveluja on käytössä näin runsaasti!

Kuvassa 10 näkyvät tiedot, jotka vuotivat kolmansille osapuolille tutkituilla verkkosivustoilla (joulukuussa 2022). Jokainen solu kuvaa tietyntyyppisen tiedon vuotamisen tietyssä terveyspalvelussa. Numerot osoittavat, kuinka monelle kolmannelle osapuolelle tiedot vuotivat. Esimerkiksi tieto ajanvarauksen aloittamisesta vuoti viidelle eri kolmannelle osapuolelle palvelussa WS1.

Vuotava tieto	WS1	WS2	WS3	WS4	WS5	WS6	WS7	WS8	WS9	WS10
Ajanvarauksen aloitus	5	3	3	4	2	3		4	1	1
Ajanvarauksen vaihe	1					3			1	
Valittu toimenpide			2			3			1	
Toimipiste	4		1	1						1
Ajanvarauksen päivä	4		2	1						
Ajanvarauksen aika			1							
Lääkärin nimi		1	1	1						
Lääkärin erikoisala	4	1								
Yksityinen vai työterveysasiakas			3	1						
Hakutermi		2	1	4		2	1	4	2	
Tietosivun aihe	5	6	3	4	2	5	1	4	3	3

Kuva 10. Tietovuodot tarkastelluissa terveydenhuollon verkkopalveluissa (joulukuu 2022). Yhdeksi tietovuoksi lasketaan yhden arkaluonteisen tiedot vuotaminen yhdelle kolmannelle osapuolelle. Tietovuotoinstansseja on siis yhdellä terveydenhuollon sivustolla tässä tapauksessa aina useita, myös saman henkilötiedon kohdalla (esimerkiksi tieto valitusta toimenpiteestä vuotaa usealle taholle). Solujen numerot kuvaavat, kuinka monelle kolmannelle osapuolelle tietty tieto vuoti tietystä palvelusta.

Yleinen tietovuoto koski ajanvarausominaisuuden käyttöä. Vaikka ajanvarausta ei viety loppuun tässä tutkimuksessa, tieto sen aloittamisesta osoittaa käyttäjän aikomuksen tehdä varaus. Kaikissa palveluissa yhtä (WS7) lukuun ottamatta tieto ajanvarausprosessin aloittamisesta vuoti vähintään yhdelle kolmannelle osapuolelle. Myös ajanvarausprossesin tarkempi vaihe vuoti kolmessa palvelussa. Ajanvarausprosessiin liittyvän tiedon vuotaminen on ongelma, koska se osoittaa suhteen potilaan ja terveydenhuollon palvelun tarjoajan välillä. Apulaisoikeusasiamies on todennut, että tällainen suhde on salassapidettävää tietoa[7].

Seitsemän tutkituista verkkopalveluista vuoti myös tarkempia tietoja ajanvarauksista kolmansille osapuolille. Näitä olivat toimipisteen sijainti (3 verkkopalvelua), ajanvarauksen päivämäärä (3), ajanvarauksen aika (1), asiantuntijan (esim. lääkärin) nimi (3), asiantuntijan erikoisala (2) ja se, oliko ajanvaraus tehty yksityisenä vai työterveysasiakkaana (2). Valittu toimenpide (esim. influenssarokotus tai sukupuolitautitesti) vuoti myös kolmella (3) tutkitulla verkkosivustolla. Tietosuojavaltuutetun toimisto on linjannut, että rokotuksen saamista koskeva tieto on terveystieto[8].

Tiedot valitusta toimenpiteestä tai lääkärin nimestä paljastavat potilaan ja terveyspalvelun tarjoajan välisestä suhteesta vielä paljon pelkkää ajanvarauksen aloittamista enemmän. On myös tärkeää huomioida, että kolmas osapuoli voi usein seurata tietyn henkilön toimintaa verkkosivuilla pitkälläkin aikavälillä. Kun käyttäjä tekee useita ajanvarauksia, alkaa muodostua selkeä kuva potilaan hoitotoimenpiteistä ja terveydentilasta.

Kuva 10 näyttää myös, kuinka käyttäjien tekemiä hakuja seurattiin. Huomionarvoista on, että kaikissa seitsemässä tapauksessa, joissa terveyspalvelun verkkosivustolla ylipäätään oli hakutoiminto, mahdollisesti arkaluonteiset hakutermit välitettiin vähintään yhdelle kolmannelle osapuolelle – ja pahimmissa tapauksissa (WS4 ja WS8) jopa neljälle eri analytiikkapalvelulle. Vuotava hakutermi voi paljastaa käyttäjän oireita tai sairauksia.

Kaikissa 10 tutkitussa terveyspalvelussa käyttäjän avaamien tietosivujen URL-osoitteet vuotivat vähintään yhdelle kolmannelle osapuolelle. Yhdessä palvelussa (WS2) URL lähetettiin kuudelle (6) kolmannelle osapuolelle. Tiettyyn sairauteen liittyvän tietosivun katselu ei toki välttämättä tarkoita, että käyttäjällä on kyseinen sairaus tai edes epäilys siitä. Kuitenkaan arkaluonteisten selailtujen sivujen vuotaminen kolmannen osapuolen analytiikkapalveluille ei ole suotavaa.

Tehdyssä verkkoliikenneanalyysissä ei löytynyt viitteitä siitä, että käyttäjän sisäänkirjautumisen jälkeen nähtävillä olevia terveystietoja, kuten laboratoriotuloksia, olisi vuotanut kolmansille osapuolille. Näin ollen vaikuttaa siltä, että nämä sivustojen kaikkein arkaluonteisimmat osiot on toteutettu käyttäjän yksityisyyden suoja huomioiden.

[7] https://yle.fi/a/3-11213545

[8] https://tietosuoja.fi/koronavirus

Vuotava tieto	WS1	WS2	WS3	WS4	WS5	WS6	WS7	WS8	WS9	WS10
Ajanvarauksen aloitus	3	1	1	3		2		2	2	1
Ajanvarauksen vaihe	1					2			2	
Valittu toimenpide			1			1				
Toimipiste	2		1							1
Ajanvarauksen päivä	2		1							
Ajanvarauksen aika			1							
Lääkärin nimi			1							
Lääkärin erikoisala	2									
Yksityinen vai työterveysasiakas			1	2						
Hakutermi		1	1	4		1	1	2	1	1
Tietosivun aihe	3	2	1	4		5	1	2	1	3

Kuva 11. Tietovuotojen päivitetty tilanne tarkastelluissa terveydenhuollon verkkopalveluissa (helmikuu 2024).

Kun terveydenhuollon palveluntarjoajia oli informoitu niiden verkkosivuilla havaituista yksityisyysongelmista ja niille oli annettu runsaasti aikaa korjata tilanne, sivustoja analysoitiin uudestaan helmikuussa 2024. Kuvassa 11 näkyy päivitetty tilanne. Tietovuotojen määrä on vähentynyt; esimerkiksi kaikkien tietovuotojen yhteismäärä kuvassa 10 on 116 mutta kuvassa 11 enää 70. Silti tätäkin lukua voi vielä pitää varsin suurena. Ajanvarausprosessin aloittamisen paljastaminen ja selattujen sivujen ja hakutermien vuotaminen kolmansille osapuolille näyttävät yhä päivitetyssäkin kuvassa olevan merkittävä ongelma suurimmassa osassa terveyspalveluja. On myös yllättävää, että joissakin tapauksissa hyvin arkaluonteisia tietoja, kuten potilaan varaama toimenpide tai asiantuntijan nimi, vuoti yhä. Vain yksi palvelu, WS5, oli poistanut sivuiltaan kaiken kolmannen osapuolen web-analytiikan ja välttänyt näin tietovuodot. Yksi syy esimerkiksi verkkoapteekkeja vähäisempään positiiviseen muutokseen lienee se, ettei lääkärikeskuksen verkkosivujen yksityisyysongelmista ilmoitettu tietosuojaviranomaiselle, vaan suoraan lääkärikeskusten tietosuojavastaaville.

Tutkimuksen tulokset osoittavat, ettei kolmannen osapuolen palvelujen oikea paikka ole terveystietoja käsittelevissä verkkopalveluissa. Löydettyjen kaltaiset tietovuodot lisäävät useiden haavoittuvien käyttäjäryhmien haavoittuvuutta entisestään.

Julkaisun merkitys

Julkaisu IV laajentaa julkaisun II ajatusta kriittisten terveystietoja sisältävien vuotojen analyysistä terveydenhuollon verkkopalveluissa ja keskittyy lääkärikeskusten ja laboratorioiden verkkosivuihin. Tämä on luontevaa jatkoa verkkoapteekkien tarkastelulle. Kun apteekkien yhteydessä tarkasteltiin käytännössä vain lääkkeiden katseluun ja tilaamiseen liittyviä skenaarioita, julkaisussa IV käyttöskenaarioiden joukko

on laajempi sisältäen tietosivujen katselun, hakutoiminnon käytön, ajanvarauksen ja omien tietojen katselun sisäänkirjautumisen takana.

Tutkimus osoittaa, etteivät useiden suomalaisten terveydenhuollon verkkopalvelujenkaan sivustot ole yksityisyydeltään riittävällä tasolla, sillä selvästi käyttäjän terveydentilaan liittyviä tietojakin vuoti kolmansille osapuolille. Tämä on siinä mielessä yllättävää, että useilla lääkäriasemilla ja laboratorioilla voi olettaa olevan enemmän resursseja ja tietoutta vahvan yksityisyyden ylläpitämiseksi kuin esimerkiksi pienillä apteekeilla. Toisaalta tutkimus näyttää, että yksityisillä lääkäriasemilla on tarvetta seurata liiketoimintatavoitteidensa toteutumista ja sivustojensa käytettävyyttä analytiikan avulla, monessa tapauksessa yksityisyyden kustannuksella.

Aiempaan tutkimukseen verrattuna julkaisu IV tarkastelee myös laajempaa joukkoa erilaisia skenaarioita ja potentiaalisia vuotavia henkilötietoja. Luultavasti juuri tämän ja kattavan manuaalisen verkkoliikenneanalyysin vuoksi vuotoja myös löytyi selvästi enemmän kuin monessa muussa aiemmassa terveydenhuollon palveluiden yksityisyyteen liittyvässä tutkimuksessa (esim. [69]). Aiemmassa tutkimuksessa on usein tarkasteltu sivustoja vähemmän kattavasti automaattisilla työkaluilla, joilla ei välttämättä voisi havaita esimerkiksi julkaisussa IV löydettyjä runsaita ajanvaraustapahtumaan liittyviä vuotoja.

Tekijän kontribuutio

Väitöskirjan laatija on ideoinut tutkimuksen, ollut mukana aineiston keräämisessä ja analysoinnissa, kirjoittanut suurimman osan julkaisusta, raportoinut löydökset tarkastelluille terveydenhuollon palveluntarjoajille ja johtanut tutkimuksen tekemistä.

3.5.5 Julkaisu V: Analysis of Third-Party Data Leaks on Finnish Mental Health Websites

Yhteenveto sisällöstä

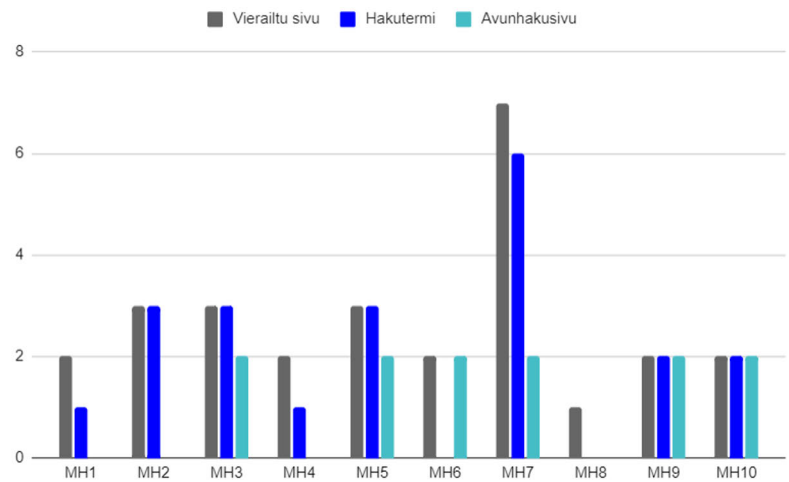
Julkaisussa V [95] tutkittiin kymmenen suomalaisen mielenterveyssivuston yksityisyyttä verkkoliikenneanalyysin keinoin. Koska mielenterveyden verkkopalveluista ei ole olemassa varsinaista listaa, sopivia palveluita etsittiin Google-hakukoneella käyttäen avainsanaa ”mielenterveyspalvelu”. Mukana oli niin yksityisen, julkisen kuin yhdistyskentänkin toimijoita.

Aiempien julkaisujen tapaan tarkasteltiin, vuotavatko nämä verkkopalvelut tahattomasti arkaluonteisia tietoja kolmansille osapuolille. Tutkimuksessa tarkasteltiin, millaisia henkilötietoja löydetyt tietovuodot sisältävät ja annetaan suosituksia yksityisyyden suojan parantamiseksi verkkopohjaisissa mielenterveyspalveluissa. Lisäksi tutkittiin pimeitä käytäntöjä, joilla huomaamattomasti suostutellaan käyttäjä hyväksymään tiedonkeruu, sekä sivustojen tietosuojaselosteita. Sen sijaan, että tutkimuk-

seen olisi valittu tarkastelun kohteeksi laaja joukko verkkopalveluja, tarkoitus oli pyrkiä syvällisempään analyysiin pienemmällä otoksella.

Mielenterveyspalveluja tarkastellessa aloitettiin verkkosivuston etusivulta, minkä jälkeen hyväksyttiin kaikki evästeet ja tietojen keruu. Tämän jälkeen suoritettiin haku ja valittiin ensimmäinen tulossivulla näkyvä linkki. Käytetyt hakusanat olivat kaikki mielenterveyteen liittyviä suomenkielisiä sanoja, kuten ahdistus ja masennus. Sanavalinnat vaihtelivat eri sivustojen välillä, koska kaikki sivustot eivät antaneet hakutuloksia samoilla hakusanoilla. Jos kyseisellä verkkosivustolla ei ollut hakutoimintoa, tutkimus suoritettiin siirtymällä tarjottuja mielenterveyspalveluja kuvaavalle sivulle. Tämän lisäksi lisäksi käytiin vielä ajanvaraukseen tai avun hakemiseen liittyvällä sivulla. Kuvattu testisarja antoi mahdollisuuden tarkastella, vuotiko seuraavia tietoja kolmansille osapuolille: 1) vierailtujen sivujen osoitteet, 2) tieto vierailusta ajanvaraus- tai avunhakusivulla, ja 3) käyttäjien syöttämät arkaluonteiset hakutermit.

Kuvassa 12 näkyvät tutkituilla mielenterveyssivustoilla (M1–M10) havaitut tietovuodot kolmeen edellämainittuun kategoriaan jaettuina. Kaikki 10 tutkittua sivustoa vuotivat vierailtujen sivujen URL-osoitteet vähintään yhdelle kolmannelle osapuolelle. Hakusanat, joissa voi mahdollisesti olla mukana arkaluonteista tietoa käyttäjän mielenterveydestä, vuotivat kahdeksalla sivustolla. Jäljelle jäävillä kahdella sivustolla (MH6 ja MH8) ei kuitenkaan ollut hakukenttää, mikä tarkoittaa, ettei yksikään tutkituista sivustoista tarjonnut käyttäjilleen luottamuksellista hakutoimintoa! Kuusi tutkituista sivustoista paljasti yhdelle tai useammalle kolmannelle osapuolelle, että käyttäjä on käynyt ajanvaraus- tai avunhakusivulla. Kaikissa vuototyypeissä Google ja Meta olivat yleisimpiä henkilötietojen vastaanottajia.



Kuva 12. Mielenterveyspalveluissa havaitut vuototyypit kullakin tarkastellulla verkkosivulla.
© IEEE 2024

Vaikka tutkimuksen otos on pieni, on selvää, että nämä luvut maalaavat synkän kuvan suomalaisten mielenterveyssivustojen yksityisyydestä. Vaikka valitussa koeasetelmassa suostumus evästeiden käyttöön ja tietojen keräämiseen olikin annettu, mahdollisesti käyttäjän mielenterveyteen liittyvien arkaluonteisten tietojen jakaminen on silti eettisesti kyseenalaista. Ongelma on sama kuin lääkärikeskustenkin sivuilla: vaikka käyttäjät hyväksyisivät henkilötietojen keräämisen sivustolla, he tuskin ymmärtävät tai odottavat, että heidän mielenterveyteensä mahdollisesti liittyvät tiedot voivat vuotaa URL-osoitteiden mukana yhdistettynä käyttäjät yksilöiviin tunnisteisiin.

On myös huomionarvoista, että kaikissa mielenterveyspalveluissa yhtä (MH8) lukuun ottamatta oli vähintään kaksi erillistä kolmatta osapuolta. Yksityinen mielenterveysklinikka MH7 käytti seitsemää erillistä kolmannen osapuolen palvelua seuraamaan vierailtuja sivuja ja kuutta palvelua seuraamaan syötettyjä hakusanoja! Näin laajamittainen arkaluonteisten henkilötietojen jakaminen lisää merkittävästi tietojen väärinkäytön riskiä. Muutoin eri sektoreiden mielenterveyspalvelut näyttivät olevan keskenään melko tasoissa kolmansien osapuolten tietovuotojen suhteen. Tosin julkisen sektorin sivustoilla (MH1, MH2, MH4) ei havaittu tietovuotoja niiden avunhakusivuilla.

Suostumuksen ja pimeiden käytäntöjen osalta havaittiin, että 3 tarkastelluista sivustoista (MH8–MH10) ei edes kysynyt käyttäjän suostumusta evästeisiin ja henkilötietojen keruuseen. Loput 7 sivustoa pyysivät suostumuksen, mutta 4 niistä (MH3, MH5, MH6, MH7) korosti hyväksymispainiketta näkyvällä värillä, jolloin käyttäjä usein saattaa antaa suostumuksen rutiininomaisesti. Harhaanjohtavat värit ovat moderneilla verkkosivuilla valitettavan tavanomainen pimeä käytäntö. Vastaavasti tietosuojaselosteita tarkastellessa käyttäjän informointi henkilötietotojen keräämisestä oli hyvin puutteellista. Vain 2 palvelua mainitsi tietosuojaselosteessaan, että käyttäjä voidaan yksilöllisesti tunnistaa ja vain 2 palvelua nimesi kolmannet osapuolet, joille henkilötietoja luovutetaan. Edelleen 3 sivustoa mainitsi vierailtujen sivujen seurannan, mutta yksikään ei maininnut hakutermien vuotamisesta. Suostumuksen pyytämisessä ja tietosuojaselosteissa olisi siis paljon parannettavaa.

Julkaisun merkitys

Julkaisu V laajentaa aiemmissa julkaissa tarkasteltujen verkkopohjaisten terveyspalvelujen joukkoa edelleen mielenterveyspalveluihin jatkaen pitkälti siitä, mihin lääkärikeskusten verkkosivujen tarkastelussa jäätiin. Aiempiin julkaisuihin verrattuna tässä julkaisussa tutkimusasetelmaan on lisätty myös pimeiden käytäntöjen tutkiminen. Lisäksi tässä tutkimuksessa analysoitiin verkkopalveluja niin julkiselta, yksityiseltä kuin kolmanneltakin sektorilta.

Verkkopohjaisten mielenterveyspalvelujen tarkastelun tekee mielenkiintoiseksi myös se, että niiden yksityisyyttä on tieteellisessä kirjallisuudessa käsitelty hyvin

rajoitetusti. Aikaisempi aiheen tutkimus näyttää kohdistuvan lähinnä tietosuojaselosteisiin [96, 76] ja mobiilisovelluksiin [97]. Näin ollen julkaisu V laajentaa aiempaa tietovuototutkimusta verkkopohjaisiin mielenterveyspalveluihin ja erityisesti Suomen tilanteeseen.

Aivan kuten lääkäriasemien tapauksessa, myös mielenterveyspalvelujen yksityisyydessä on runsaasti korjattavaa. Mielenterveyden häiriöiden tai oireiden ei koskaan tulisi päätyä kolmansien osapuolten tietoon. Näyttää siltä, ettei tälläkään alalla ole täysin herätty tietosuoja-asioista huolehtimiseen, vaikka tarvittavia resursseja monilla mielenterveysalankin toimijoilla luulisi olevan. Analytiikan hyödyntäminen ja liiketoimintatavoitteiden seuraaminen näitä toimijoita kuitenkin vaikuttaa kiinnostavan yhtä paljon kuin lääkäriasemia aiemmassa tutkimuksessa. Vaikka esimerkiksi ajanvaraustietojen ja valittujen asiantuntijoiden nimien vuotamista ei tässä tutkimuksessa havaittu, tulisi verkkopohjaisten mielenterveyspalvelujenkin ryhtyä pikaisiin toimiin käyttäjän yksityisyyden takaamiseksi.

Mielenterveyspalvelujen tietovuodoilla voi olla vakavia seurauksia. Ne voivat heikentää käyttäjien luottamusta mielenterveyspalveluihin ja johtaa siihen, ettei verkkopohjaisia palveluja haluta enää käyttää samoissa määrin kuin ennen. Pahimmillaan seurauksena käyttäjille voi olla syrjintä tai sosiaalinen leimautuminen, jos heidän mielenterveystietonsa paljastuvat myös ulkopuolisille. Tietovuodot voivat aiheuttaa stressiä ja ahdistusta pahentaen olemassa olevia mielenterveysongelmia. Äärimmäinen esimerkki tästä on jo nähty Suomessakin psykoterapiakeskus Vastaamon tapauksen yhteydessä, jossa Vastaamon potilaita on päätynyt jopa itsemurhaan tietojensa vuotamisen takia[9]. Tietovuodoista vastuussa olevat organisaatiot voivat myös kärsiä maineen menetyksestä.

Tekijän kontribuutio

Väitöskirjan laatija on ideoinut tutkimuksen, ollut mukana aineiston keräämisessä ja analysoinnissa, kirjoittanut suurimman osan julkaisusta ja johtanut tutkimuksen tekemistä.

3.5.6 Julkaisu VI: Data Leaks to Third-Party Services on Medical Websites

Yhteenveto sisällöstä

Julkaisussa VI [98] tarkasteltiin 20 suomalaisella terveysalan verkkosivulla esiintyviä tietovuotoja. Tällä kertaa otettiin mukaan kaikki kolme aiempaa kategoriaa: verkkoapteekkeja, terveysasemia ja mielenterveyssivustoja. Aiempaan verrattuna koeasetelmaa muutetiin myös siten, ettei evästeisiin ja tiedonkeruuseen annettu lain-

[9] https://yle.fi/a/74-20077270

kaan suostumusta. Tämä antoi mahdollisuuden tarkastella käyttäjän vaikutusmahdollisuuksia – kerätäänkö arkaluonteisia henkilötietoja silloinkin, kun käyttäjä haluaa nimenomaisesti estää kaiken mahdollisen tiedonkeruun? Julkaisussa tarkastellaan myös minimisuostumuksella 8 sellaista terveydenhuollon verkkopalvelua, joita julkaisussa IV aiemmin tarkasteltiin täydellä suostumuksella. Näin voidaan tutkia, vaikuttaako käyttäjän suostumus näissä palveluissa esiintyviin kolmansiin osapuoliin ja tietovuotoihin.

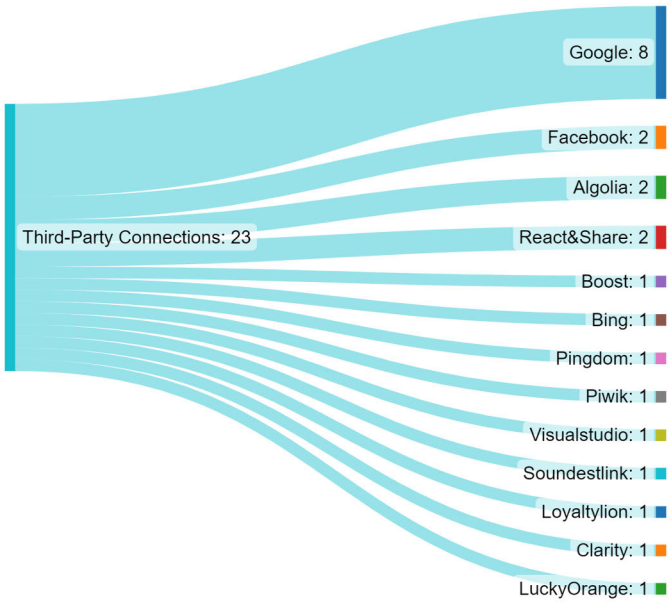
Testiasetelmassa oli minimisuostumusta lukuun ottamatta samanlainen kuin verkkoapteekkien, lääkärikeskusten ja mielenterveyssivustojen osalta aiemminkin. Lääkärikeskusten sivuilla tehtiin haku (esimerkiksi jonkin sairauden nimi) ja hakutulossivun kautta siirryttiin tietosivulle. Tämän jälkeen testattiin vielä ajanvarausprosessia ja tarkasteltiin, vuotaako sen yhteydessä terveystietoja. Ajanvaraus keskeytettiin ennen sen vahvistamista. Vastaavasti mielenterveyssivuilla tehtiin haku, jonka jälkeen tulossivulta valittiin linkki ja navigoitiin esimerkiksi tiettyyn mielenterveyden häiriöön liittyvälle tietosivulle. Tämän jälkeen siirryttiin sivulle, joka vihjaa käyttäjän olevan kiinnostunut avun hakemisesta – joissakin tapauksissa tämä oli ajanvarauslomake, toisissa vain sivu, jolla oli mielenterveyspalvelun puhelinnumero. Lopulta verkkoapteekkien tapauksessa haettiin tiettyä reseptilääkettä hakutoiminnon avulla, siirryttiin tämän lääkkeen tuotesivulle, ja edettiin sitten tilaamaan lääkettä. Prosessi keskeytettiin ennen varsinaisen tilauksen tekemistä.

Verkkoliikenteen analysoinnin lisäksi tarkasteltiin myös verkkosivustojen tietosuojaselosteita ja evästebannereita. Tässä yhteydessä tutkittiin, oliko henkilötietojen jakamista kolmansille osapuolille kuvattu läpinäkyvästi eli vastasiko kuvaus todellisesta verkkoliikenteestä tehtyjä havaintoja. Erityisen kiinnostuksen kohteena tutkimuksen tässä osuudessa oli, oliko mahdollinen terveystietojen siirtäminen kolmansien osapuolien haltuun ilmoitettu käyttäjille asianmukaisesti.

Kuvassa 13 näkyvät 20 sivustolla olleet kolmannet osapuolet, joille vuoti arkaluonteisia henkilötietoja tarkastelluilla verkkosivustoilla. Ilman käyttäjän antamaa suostumustakin Google Analytics esiintyi 8 sivulla, mitä voi terveydenhuollon palveluissa pitää hyvin korkeana lukemana. Samoin Facebook, React&Share (käyttäjäpalautetta keräävä palvelu) sekä Algolia (hakutoiminnon ulkoistava ja hakutermien indeksointia tarjoava kolmannen osapuolen palvelu) olivat edustettuina 2 kertaa. Lisäksi 9 muuta kolmannen osapuolen palvelua esiintyi tarkastelluilla sivuilla yhden kerran.

Mielenterveyssivustojen osalta 5/7 sivustoa vuoti potentiaalisesti arkaluonteisia henkilötietoja. Kaikki 5 mainittua sivustoa vuotivat vieraillut sivut, ja 4 näistä vuoti myös käyttäjän syöttämät hakutermit. Hyvin vastaavia lukuja esiintyi terveysasemien osalta, joista 6/8 vuoti mahdollisesti arkaluonteista tietoa. Vieraillut sivut vuotivat 5 sivustolla, hakutermit taas 3 tapauksessa. Vierailu ajanvaraussivulla vuoti 3 tapauksessa.

Tarkastelluista 5 verkkoapteekista vain yhdessä oli kolmansia osapuolia, mutta tämä apteekki vuoti henkilötietoja ilman suostumusta 6 kolmannelle osapuolelle!



Kuva 13. Verkkopohjaisissa terveyspalveluissa havaittujen tietovuotojen vastaanottajat.
© IEEE 2024

Tarkempi analyysi osoitti, että tämä apteekki vuoti lääketilausta tehdessä kolmannelle osapuolelle tilatun lääkkeen nimen, asiakkaan sähköpostiosoitteen, sekä vahvan tunnistautumisen kautta saadun asiakkaan koko nimen! Kyseinen löydös on raportoitu tietosuojavaltuutetun toimistolle.

Kaikkiaan 12 tutkitusta 20 terveydenhuollon verkkopalvelusta vuoti arkaluonteisia tietoja kolmansille osapuolille. Vierailtu sivu vuoti 11/20 tapauksessa ja hakutermi 8/20 tapauksessa. Kun otetaan huomioon se, että tiedonsiirrot tehtiin ilman käyttäjän suostumusta ja kyseessä ovat arkaluonteiset terveysalan sivustot, nämä luvut ovat hyvin korkeita. On kuitenkin huomionarvoista, että tutkimuksessa mukana olleet kaksi julkisen sektorin mielenterveyspalvelua eivät sisältäneet lainkaan kolmannen osapuolen analytiikkaa. Yksityiset ja kolmannen sektorin sivustot taas menestyvät vertailussa selvästi huonommin.

Tässä julkaisussa tarkasteltujen terveydenhuollon verkkopalvelujen joukossa oli myös 8 sellaisen lääkäriaseman verkkosivut, joita tarkasteltiin aiemmin julkaisussa IV täydellä suostumuksella. Kaikista palveluista täydellä suostumuksella löytyi tietovuotoja, ja minimisuostumuksellakin 6 näistä sivustoista vuoti edelleen henkilötietoja. Vain 2 tarkastelluista 8 sivustosta ei aktivoinut mitään kolmannen osapuolen palvelua, kun käyttäjä ei antanut suostumustaan henkilötietojen käsittelyyn.

Tietosuojaselosteita tarkasteltiin 12 henkilötietoja kolmansille osapuolille vuotaneen sivuston osalta. Näistä selosteista 7:stä saattoi päätellä, että käyttäjä voidaan tunnistaa seurannassa yksilöllisesti. Vastaavasti vain 5 sivuston selosteissa mainittiin URL-osoitteiden kerääminen ja ainoastaan 3 selostetta mainitsi kaikki kolmannet osapuolet nimeltä. Terveystietojen vuotamisesta ei yhdessäkään selosteessa mainittu mitään, puhumattakaan esimerkiksi reseptilääkkeiden vuotamisen mainitsemisesta. Selosteet eivät myöskään antaneet käyttäjälle käsitystä siitä, että tällaista tiedonkeruuta voi tapahtua myös ilman suostumusta. Kaikkiaan käyttäjän informointi henkilötietojen käsittelystä oli tarkastelluissa tietosuojaselosteissa hyvin puutteellista.

Julkaisun merkitys

Tämä tutkimus antoi aiemmissa julkaisuissa I-IV käsitellyistä sovellusaloista eräänlaisen läpileikkauksen, nyt kuitenkin ilman käyttäjän suostumusta tiedonkeruuseen ja evästeisiin. Tulosten perusteella on selvää, että tilanne on hyvin ongelmallinen silloinkin, kun käyttäjä kieltäytyy kaikista mahdollisista evästeistä eikä anna lupaa henkilötietojen keräämiseen. Analytiikkapalveluiden käyttö näyttää siis menevän käyttäjän yksityisyyden edelle silloinkin, kun käyttäjä ei niin missään tapauksessa halua. Tämä näyttää olevan ongelma erityisesti yksityisten lääkärikeskusten ja mielenterveyden verkkopalvelujen osalta.

Vaikka otos oli tässä tutkimuksessa oli yksittäisten sovellusalojen osalta pieni, on muistettava, ettei Suomessa ole vaikkapa mielenterveyteen liittyviä toimijoita ja lääkäriasemia lopulta kovinkaan suurta määrää. Henkilötietoja vastaanottavat kolmannet osapuolet useiden toimijoiden sivuilla ovat näin ollen suuri yksityisyysriski. On kestämätöntä, että vaikkapa sairauksiin tai oireisiin liittyvät hakutermit tai tilattavan lääkkeen nimi voivat vuotaa kolmannelle osapuolelle silloinkin, kun käyttäjä on varta vasten yrittänyt kieltää henkilötietojen keruun.

Tämä tutkimus myös osoitti aikaisemmista tutkimuksista poiketen, että IP-osoitteen, laitetunnisteen tai vaikkapa Google-tiliin liittyvän identiteetin vuotamisen sijaan on myös täysin mahdollista, että käyttäjän vahvan tunnistautumisen kautta saatu koko nimi lääketilaukseen liitettynä voi päätyä kolmannen osapuolen haltuun ilman suostumusta. Kaikkiaan tutkimus korostaa entisestään vahvaa tarvetta karsia kolmannen osapuolen palvelut pois kriittisiltä terveysalan sivustoilta, sillä ne voivat huomaamatta ja ilman minkäänlaista käyttäjän suostumusta kerätä hyvin arkaluonteisia terveystietoja.

Tekijän kontribuutio

Väitöskirjan laatija on ideoinut tutkimuksen, ollut mukana aineiston keräämisessä ja analysoinnissa, kirjoittanut suurimman osan julkaisusta ja johtanut tutkimuksen tekemistä.

3.5.7 Julkaisu VII: Third-party data leaks on municipal websites

Yhteenveto sisällöstä

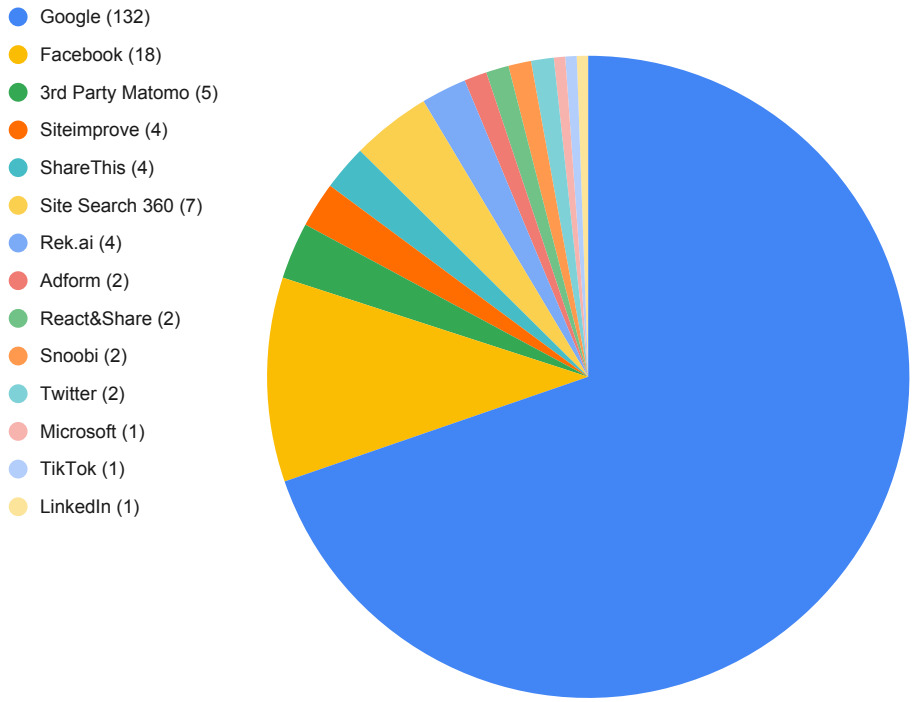
Julkaisussa VII [99] tarkastellaan kolmannen osapuolen palveluiden käyttöä suomalaisten kuntien verkkosivuilla. Tutkimuksessa analysoidaan verkkoliikennettä 309 verkkosivustolla ja selvitetään, mitä henkilötietoja siirtyy kolmansille osapuolille. Verkkoliikenneanalyysi tehtiin suurimmaksi osaksi itse kehitetyllä automaattisella työkalulla, joka analysoi, vuotivatko vieraillut URL-osoitteet ja käytetyt hakutermit kolmansille osapuolille.

Tutkimus paljastaa, että 51,8 % Suomen kuntien verkkosivuista vuotaa käyttäjän tarkastelemat sivut kolmansille osapuolille, useimmiten Google Analyticsille, vaikka käyttäjä ei olisi antanut suostumustaan henkilötietojen käsittelyyn. Lisäksi 48,6 % niistä kuntien sivuista, joilla oli hakutoiminto, vuoti hakutermit ilman suostumusta kolmansille osapuolille. Tämä tarkoittaa, että suomalaisen kunnan sivustolla tehty haku vuotaa lähes 50-prosenttisen varmasti ilman käyttäjän suostumusta kolmannelle osapuolelle, kun kunta on valittu satunnaisesti! Kuvassa 14 näkyy, kuinka moni kunta vuoti hakutermin kullekin kolmannelle osapuolelle. Huolestuttavaa on myös, että 30,7 % kunnista ei tarjonnut vaihtoehtoa kieltäytyä henkilötietojen keruusta, vaikka samaan aikaan lähetti näitä tietoja kolmansille osapuolille.

On syytä huomioida, että kuntien sivuilla käytettävät hakutermit voivat olla hyvinkin arkaluonteisia. Esimerkkinä voi miettiä vaikkapa kunnan alueelta päihdekuntoutuspalveluja etsivää käyttäjää. Vaikka sosiaali- ja terveyspalvelujen järjestämisvastuu siirtyikin kunnilta hyvinvointialueille vuoden 2023 alussa, kuntien sivuilla saatetaan edelleen tehdä arkaluonteisia terveyspalveluihin liittyviä hakuja. Apulaistietosuojavaltuutettu on todennut, että viranomaistahojen, jollaisia kunnatkin ovat, tulisi tarkkaan harkita, millaisia kolmannen osapuolen palveluja ne sivuillaan käyttävät[10]. Lisäksi apulaistietosuojavaltuutettu suosittelee, ettei viranomaisten sivuilla tulisi ilman suostumusta käyttää sellaisia palveluja, joiden tarjoajat saattavat käyttää henkilötietoja kaupallisiin tarkoituksiin.

On kuitenkin rohkaisevaa, että 43,7 % kunnista, joihin otettiin yhteyttä, vastasi, tunnusti tehdyt havainnot ja sitoutui parantamaan sivustojensa yksityisyyttä. Vastaavasti kunnista, joihin otettiin yhteyttä, 51,3 % oli poistanut kolmannen osapuolen analytiikkaa verkkosivuiltaan kolmen kuukauden kuluttua yhteydenotosta ja useat olivat myös siirtymässä Matomon kaltaisiin lokaaleihin analytiikkaratkaisuihin, joissa kunta itse kontrolloi säilyttämäänsä dataa. Tästä huolimatta kuntien vastauksista kävi ajoittain ilmi myös resurssien ja teknisen asiantuntemuksen puute Suomen kunnissa. Useat kunnat eivät esimerkiksi ensimmäisellä yrityskerralla saaneet Google Analyticsiä poistettua sivuiltaan onnistuneesti.

[10] https://tietosuoja.fi/-/apulaistietosuojavaltuutettu-antoi-huomautuksen-kirjastojen-aineistohakutietojen-valittymisesta-yhdysvaltalaisyritys-googlelle



Kuva 14. Kolmannet osapuolet, joille käyttäjän syöttämät hakutermit vuotivat kuntien sivuilla.
© Springer Nature 2024

Tutkimus osoittaa kolmannen osapuolen analytiikkapalvelujen olevan merkittävä ongelma myös kuntasektorilla, vaikka tilanne onkin nyt parantunut. Se korostaa tarvetta lisätä kuntien virkamiesten ja web-kehittäjien tietoisuutta mahdollisista tietovuodoista ja yksityisyyttä ja tietosuojaa koskevasta regulaatiosta. On myös huomattavissa, että tietosuojaviranomaisen tulisi antaa vielä selkeämpi ohjeistus liittyen kolmannen osapuolen palvelujen käyttämiseen julkisen sektorin verkkosivustoilla, jotta käyttäjien yksityisyyteen liittyvät riskit saataisiin minimoitua.

Julkaisun merkitys

Kuntasektorin sivustoja analysoivan tutkimuksen rooli tässä väitöskirjassa on tarkastella kolmannen osapuolen palvelujen esiintymistä ja henkilötietojen keräämistä muilla kuin suoranaisilla terveysalan sivustoilla. Lisäksi sen tarkoitus on tarkastella myös, kerätäänkö henkilötietoja ilman suostumusta, kun useimmissa muissa väitöskirjan julkaisuissa (I, II, III, IV, V, VIII) oletetaan käyttäjän antaneen suostu-

muksen tiedonkeruuseen. Tutkimuksen tulokset osoittavat selvästi, että henkilötietoja lähetetään laajamittaisesti kolmansille osapuolille ilman suostumusta jopa kuntasektorilla. Yksityisen sektorin tavoin Googlen ja Metan kaltaiset teknologiajätit ovat tiedonkerääjinä vahvasti läsnä.

Kuntasektorilla on perinteisesti ollut vahva yhteys myös sosiaali- ja terveyspalveluihin. Kuten aiemmin todettiin, esimerkiksi kuntien sivustojen hakutoimintoja käytetään varmastikin edelleen arkaluonteisten sosiaali- ja terveyspalvelujen etsimiseen tietyltä seudulta. Tässä mielessä kuntien sivustot ovat käyttäjille yksi tärkeä portti terveyspalveluihin ja liittyvät niihin kiinteästi.

Koska kuntiin ja niiden sivustojen ylläpitäjiin oltiin myös suoraan yhteydessä, tutkimus valaisee myös omalta osaltaan sitä, miten kunnat itse näkevät kolmansien osapuolien tiedonkeruun. Vastausten voi nähdä osoittavan ongelman tunnistamista ja vahvaa halua parantaa sivustojen yksityisyyttä, vaikkeivät resurssit ja tekninen tietämys aina olekaan täysin riittävät.

Tekijän kontribuutio

Väitöskirjan laatija on ideoinut tutkimuksen, ollut mukana kerätyn aineiston analysoinnissa, kirjoittanut suurimman osan julkaisusta, hoitanut yhteydenotot kuntiin ja johtanut tutkimuksen tekemistä.

3.5.8 Julkaisu VIII: Leaky Democracy: Third Parties in Voting Advice Applications

Yhteenveto sisällöstä

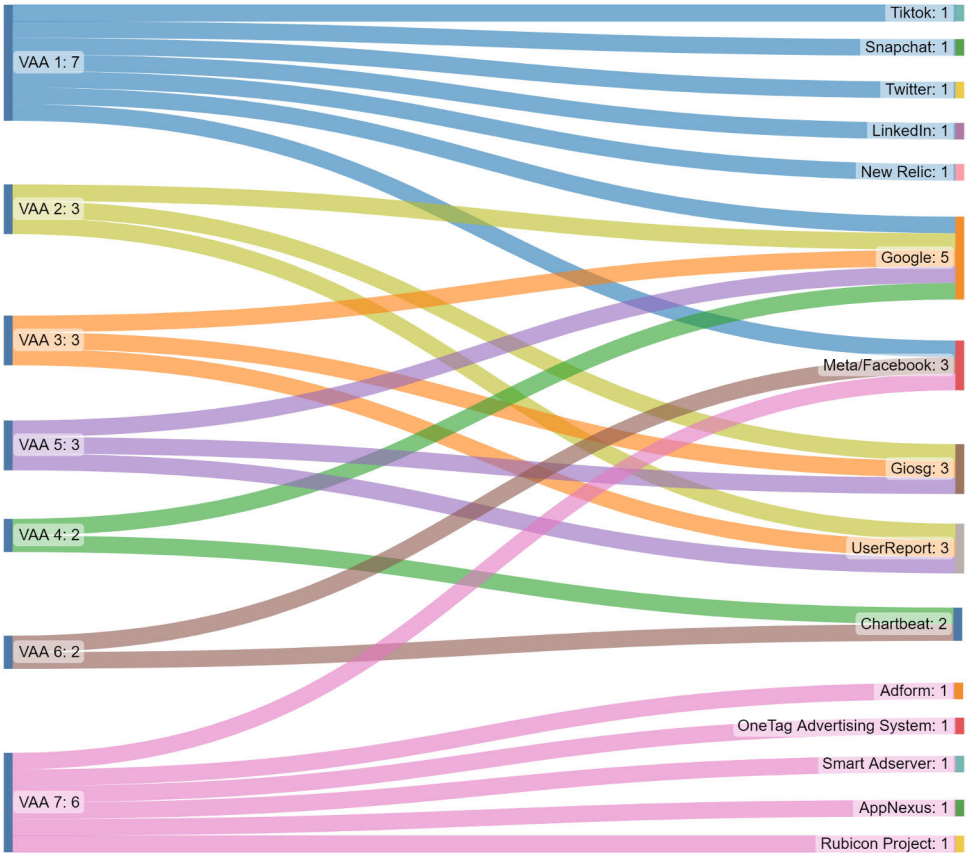
Julkaisu VIII [100] käsittelee vaalikoneiden tietovuotoja. Vaalikoneet ovat sovelluksia, jotka auttavat äänestäjiä tekemään informoituja äänestyspäätöksiä. Suomessa erityisesti useat mediat tarjoavat vaalikoneita osana verkkosivustojaan. Sovellukset esittävät käyttäjälle joukon ajankohtaisia poliittisia kysymyksiä, joihin myös ehdokkaat ovat aiemmin vastanneet. Käyttäjän vastausten perusteella hänelle voidaan sitten esittää luettelo ehdokkaista ja poliittisista puolueista, joiden näkemykset ovat yhteensopivia käyttäjän esittämien mielipiteiden kanssa. Suomessa vaalikoneet ovat suosittuja – Tilastokeskuksen mukaan ennen vuoden 2019 eduskuntavaaleja 49 % äänioikeutetuista käytti niitä[11].

Tutkimuksessa tarkasteltiin 11 suosittua suomalaista vaalikonetta. Koeasetelmassa hyväksyttiin kaikki evästeet ja tiedonkeruu. Vaalikoneen kysymyksiin vastattiin ja tämän jälkeen avattiin kärkiehdokkaiden sivuja, joilla omia vastauksia saattoi vertailla ehdokkaan vastauksiin. Verkkoliikenne nauhoitettiin Chromen kehittäjätyökaluilla

[11] https://www.stat.fi/til/sutivi/2019/sutivi_2019_2019-11-07_kat_002_fi.html

ja tarkasteltiin, vuotavatko kysymysten vastaukset ja tarkastellut kärkiehdokkaat vaalikoneista kolmansille osapuolille.

Kuvassa 15 näkyy alluviaalidiagrammi arkaluonteisen tiedon vuotamisesta kolmansille osapuolille. Diagrammissa näkyvät vain sellaiset datavirrat, joissa käyttäjän poliittisia kantoja koskevat tiedot potentiaalisesti vuotivat. Voidaan nähdä, että suurin osa tutkituista sovelluksista, yhteensä 7, vuoti näitä tietoja. Kaikissa seitsemässä tapauksessa tämä tapahtui tietovuotoina käyttäjän avaamilta ehdokassivuilta, kun käyttäjä avasi hänelle suositeltujen ehdokkaiden sivuja vaalikonesovelluksessa. Kun käyttäjä vierailee näillä ehdokkaiden sivuilla, kolmannet osapuolet saavat sivun URL-osoitteen. URL sisältää usein suoraan ehdokkaan nimen, mutta vaikka näin ei olisi, tieto ehdokkaasta voidaan katsoa vuotaneeksi, koska kolmas osapuoli voi yksinkertaisesti vierailla ehdokkaan sivulla tutkiakseen sitä tarkemmin. Tällainen analyysi voidaan myös helposti automatisoida tekoälyn avulla.



Kuva 15. Alluviaalidiagrammi arkaluonteisen poliittisen datan vuotamisesta kolmansille osapuolille. © Springer Nature 2024

Kuvasta 15 näkyy, että vaalikone 1, jota ylläpitää suomalainen työnhakusivusto, vuoti tiedot käyttäjän tarkastelemista ehdokkaista seitsemälle (7) eri kolmannelle osapuolelle, mikä on lähes puolet kaikista tässä tutkimuksessa löydetyistä yksittäisistä kolmansista osapuolista! Suomalaisen mediatalon tarjoama vaalikone 7 ei menesty tarkastelussa paljoakaan paremmin: käyttäjän katselemat ehdokkaat vuotavat kuudelle (6) kolmannelle osapuolelle. Vaalikoneet 2, 3 ja 4 vuotavat kaikki katsotut ehdokkaat kolmelle (3) kolmannelle osapuolelle. Nämä vaalikoneet ovat kaikki saman suomalaisen mediatalon sivustoilla, mikä lienee syynä identtisille kolmansille osapuolille niiden sovelluksissa. Vaalikoneet 5 ja 6 vuotavat katsotut ehdokkaat kahdelle (2) kolmannelle osapuolelle. Tutkimuksessa oli neljä vaalikonetta (8, 9, 10 ja 11), joissa ei ollut poliittisten mielipiteiden vuotoja. Vaikka poliittisten mielipiteiden voi nähdä vuotavan avattujen ehdokassivujen kautta, käyttäjän vastaukset varsinaisiin vaalikoneen kysymyksiin eivät vuotaneet yhdessäkään tapauksessa.

Kuva 15 näyttää myös sivustoilla olevat kolmannet osapuolet, joille vuotaa potentiaalisia poliittisia mielipiteitä. Google on yleisin kolmas osapuoli (5 esiintymää), mikä ei ole yllättävää. Muita useasti esiintyviä kolmansia osapuolia ovat Meta, UserReport ja Giosg. Vaikka kaikki havaitut kolmannet osapuolet tuskin tietoisesti keräävät ainakaan poliittisia mielipiteitä, on selvää, ettei tällaisten tietojen tulisi päätyä niiden palvelimille ja mahdollisesti Euroopan ulkopuolelle.

Tutkimuksessa tarkasteltiin myös vaalikoneiden tietosuojaselosteita. Vain kolmella vaalikoneella oli erikseen vaalikoneen mainitseva seloste ja muissa tapauksissa saatavilla oli vain laajemman sivuston yleinen tietosuojaseloste. Yksikään seloste ei maininnut mitään siitä, että tieto katselluista ehdokkaista välitetään kolmansille osapuolille. Vaikka URL-osoitteista oli joissakin selosteissa mainintoja, on kohtuutonta olettaa, että käyttäjä ymmärtää, että hänen katselemistaan ehdokkaista voi vuotaa tieto URL-osoitteen mukana.

Voidaan luonnollisesti esittää väite, etteivät käyttäjän tarkastelemat ehdokkaat paljasta hänen poliittista kantaansa, ja osin näin varmasti onkin. Monet käyttäjät kuitenkin varmasti tarkastelevat lähinnä omia kärkiehdokkaitaan, jotka yleensä paljastavat ainakin käyttäjän jonkinasteisen sijoittumisen arvokartalla (oikeisto–vasemmisto, konservatiivi–liberaali), vaikkei tarkka puoluekanta tai suosikkiehdokas olisikaan varma tieto. Silti vuotanut poliittinen kanta ei tässä tilanteessa ole läheskään yhtä eksakti ja käyttökelpoinen tieto kuin vaikkapa käyttäjän tilaama lääke tai ajanvaraukseen liittyvä toimenpide.

Julkaisun merkitys

Web-pohjaisten vaalikoneiden yksityisyys ei suoraan liity terveysalaan tai terveystietojen vuotamiseen, mutta mahdollistaa mielenkiintoisen vertailun kahden aihealueen

välillä. Sekä terveystiedot että poliittiset mielipiteet ovat GDPR:n mukaan erityisiä henkilötietoryhmiä, joiden käsittely on lähtökohtaisesti kiellettyä [12].

Aiempi esimerkki siitä, miten tietovuodot voivat uhata demokratiaa, oli Cambridge Analytica -skandaali. Vuonna 2018 paljastui, että Facebook oli luovuttanut yli 87 miljoonan käyttäjän psykologiset profiilit brittiläiselle poliittiselle konsultointiyritykselle, Cambridge Analyticalle [101]. Kerätyn datan avulla luotiin äänestäjäprofiileja, joiden perusteella voitiin kohdentaa poliittista mainontaa ja pyrkiä vaikuttamaan tietyn äänestäjäryhmän mielipiteisiin [102]. Vaalikoneiden kautta kerätyn datan avulla voitaisiin pyrkiä samankaltaiseen mielipidevaikuttamiseen.

Julkaisu VIII osoittaa, vaalikoneissa on tapahtunut runsaasti aivan samanlaisia erehdyksiä arkaluonteisten URL-vuotojen osalta kuin esimerkiksi apteekkien ja lääkärikeskusten tapauksissa. Tällaiset tietovuodot eivät siis ole ainoastaan terveysalan verkkosivujen ongelma. Vaalikoneiden yhteydessä riskejä on ehkä vaikeampikin tiedostaa, sillä ne ovat usein osa medioiden sivustoja, joissa ei yleisesti ottaen käsitellä kovinkaan arkaluonteisia tietoja. Näin mediasivustoilla yleisesti käytetty kolmannen osapuolen analytiikka päätyy helposti myös osaksi vaalikoneita. Tarvetta sisäänrakennetulle yksityisyydelle ei niissä siksi liene yhtä helppoa tunnistaa kuin terveyspalveluissa.

Löydetyt tietovuodot saivat myös julkisuutta Suomen Kuvalehdessä julkaistun artikkelin[13] ja useiden medioiden lainauutisten kautta[14]. Alkukesästä 2024, noin vuoden artikkelin julkaisun jälkeen, suomalaisista vaalikoneista on pitkälti tämän julkisuuden seurauksena korjattu valtaosa tietovuodoista. Tätä kautta julkaisulla on ollut merkittävää yhteiskunnallista vaikuttavuutta.

Tekijän kontribuutio

Väitöskirjan laatija on ideoinut tutkimuksen, ollut mukana kerätyn aineiston analysoinnissa, kirjoittanut suurimman osan julkaisusta, hoitanut yhteydenpidon mediaan ja vaalikoneiden ylläpitäjiin sekä johtanut tutkimuksen tekemistä.

[12] https://tietosuoja.fi/erityisten-henkilotietoryhmien-kasittely

[13] https://suomenkuvalehti.fi/politiikka-ja-talous/useiden-vaalikoneiden-tietosuojassa-aukkoja-kayttajan-tarkastelemia-tietoja-vuotanut-ulkopuolisille-tahoille-kuten-googlelle-metalle-ja-tiktokille/

[14] Ks. esim. https://www.tivi.fi/uutiset/suositut-suomalaiset-vaalikoneet-vuotaneet-tietoja-ulkopuolisille-sk-huolestuttavia-aukkoja/90ea5750-228b-45be-afc6-c22edce6f2a1 ja https://www.tekniikkatalous.fi/uutiset/vaalikoneet-vuotaneet-suomessa-tietoja-ulkopuolisille-sk-huolestuttavia-aukkoja/6ad3905b-45bd-4565-8f35-ba9c2dbfb473

3.5.9 Julkaisu IX: Insights from Developers: Understanding Third-Party Data Leaks in Web Environment

Yhteenveto sisällöstä

Väitöstutkimuksen viimeinen julkaisu [103] selvittää kyselytutkimuksella web-kehittäjien näkemyksiä kolmannen osapuolen tietovuotojen syistä ja tavoista korjata niitä. Tutkimuksen aineiston keräämiseen käytettiin puolistrukturoituja asiantuntijahaastatteluja. Tutkimusta varten haastateltiin 12:ta web-kehittäjää. Mukaan valittiin asiantuntijoita, jotka työskentelivät IT-alalla ja joilla oli kokemusta web-kehityksestä. Haastattelussa pyrittiin ensin selvittämään, millaisia yksityisyysuhkia kehittäjät näkevät web-sovelluksissa yleisesti ja millaisia ratkaisuja näihin uhkiin nähdään. Tässä haastattelun ensimmäisessä osiossa haastateltaville ei vielä mainittu kolmansia osapuolia, jolloin oli mahdollista seurata, ottivatko haastateltavat niiden aiheuttaman riskin itse puheeksi.

Haastattelun toisessa osiossa haastateltavilta kysyttiin, miten he arvioivat web-sovelluksiin lisäämiään kolmansia osapuolia yksityisyyden kannalta. Heitä pyydettiin myös arvioimaan, miksi analytiikkapalvelujen kaltaisia kolmansia osapuolia, joilla on päällekkäisiäkin toiminnallisuuksia, käytetään runsaasti samalla verkkosivulla. Vastaavasti kysyttiin myös syitä siihen, että kolmansia osapuolia käytetään kriittisilläkin verkkosivuilla kuten terveysalan palveluissa. Kyselyn viimeisessä ja keskeisimmässä kysymyksessä tiedusteltiin keinoja torjua kolmansien osapuolien aiheuttamia yksityisyysuhkia. Kysymykset on esitetty taulukossa 6.

	Kysymykset
Osa 1	Millaisia yksityisyysriskejä näet niissä web-sovelluksissa, joita kehität? Millaisia yksityisyyden kannalta olennaisia tietoturvatoimia kriittisissä web-sovellusten kehityksessä ja ylläpidossa tulisi ottaa huomioon?
Osa 2	Miten tavallisesti arvioit kolmannen osapuolen palveluiden integroimiseen liittyviä riskejä toteuttamissasi web-sovelluksissa? Miksi verkkosivustoilla on käytössä niin monia päällekkäisiä analytiikkapalveluja? Miksi analytiikkaa on ylipäätään käytössä kriittisillä sivuilla? Millaisia suunnittelun, toteutuksen, testauksen ja ylläpidon aikaisia toimia voitaisiin toteuttaa kolmannen osapuolen tietovuotojen välttämiseksi?

Taulukko 6. Haastattelun kysymysrunko

Haastattelut nauhoitettiin, muutettiin tekoälyavusteisesti tekstimuotoon ja koodattiin NVivo-ohjelmiston avulla. Gioia-menetelmän [88] ja avoimen koodauksen periaatteiden mukaisesti koodatuista tekstinpätkistä luotiin käsitteitä. Käsitteet yhdisteltiin edelleen teemoiksi ja nämä vielä yläkategorioiksi eli teoreettisiksi ulottu-

vuuksiksi. Esiin nousevat ulottuvuudet kertovat yhteenvedonomaisesti siitä, millaisia asioita web-kehittäjät painottavat yksityisyysriskien ja niiden torjumisen osalta. Näiden pohjalta esitettiin myös teoria siitä, millaisia toimia kolmannen osapuolen tietovuotojen estämiseksi tarvitaan ja mitkä ovat näiden toimenpiteiden väliset yhteydet.

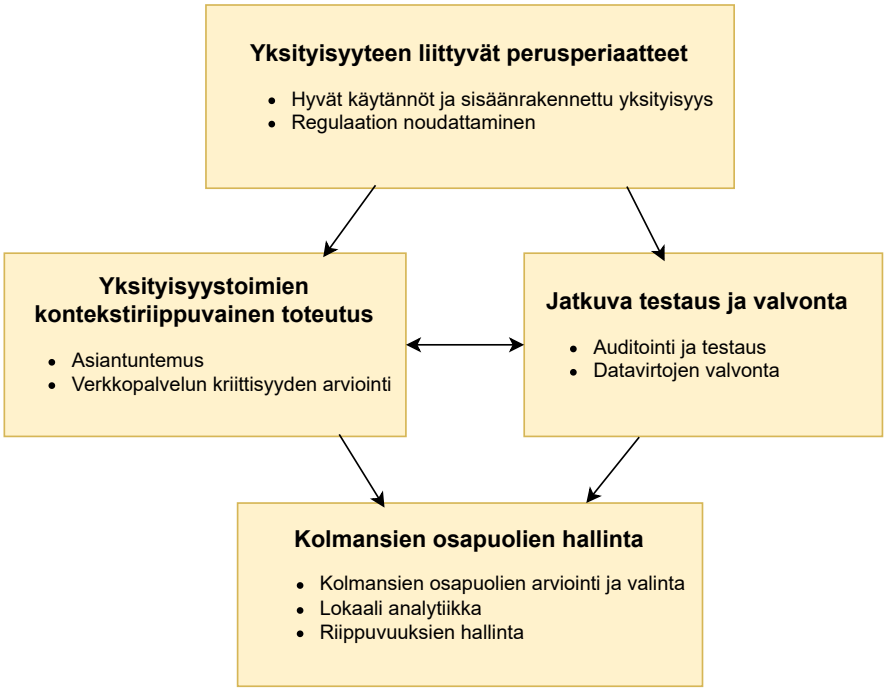
Yleisesti verkkosovellusten yksityisyyteen liittyvien kysymysten vastaukset osoittivat, että vaikka kolmannet osapuolet mainittiin yksityisyysriskien ja ratkaisujen yhteydessä muutamia kertoja, ei tätä yksityisyysuhkaa näytetä tiedostavan samalla tasolla kuin vaikkapa yleisesti hyvin tunnettuja ja OWASP Top 10 -listaukseenkin[15] sisältyviä tunnistautumisen ja pääsynhallinnan toteutuksen ongelmia. Yksityisyysriskien ehkäisemiseen käytettävistä keinoista kysyttäessä kehittäjät kuitenkin toivat esiin useita kolmannen osapuolen tietovuotojenkin estämiseen sopivia keinoja, kuten regulaation noudattamisen, auditoinnin ja tietoturvatestauksen.

Kolmansien osapuolien valinnassa korostuivat niiden toiminnan, luotettavuuden ja regulaationmukaisuuden arviointi. Esimerkiksi ymmärrystä siitä, missä kolmas osapuoli lopulta säilyttää kerättyä dataa, pidettiin oleellisena tietona. Mielenkiintoisena löydöksenä kehittäjät näkivät myös kolmannen osapuolen suuren suosion lisäävän sen luotettavuutta – näin totesi peräti 7 kaikista 12:sta kehittäjästä. Suurien teknologiajättien palveluista puhuttaessa tästä näkemyksestä voi kuitenkin olla toistakin mieltä; vaikka tiedot säilöttäisiin tietoturvallisesti, voi kolmas osapuoli itse hyödyntää niitä kyseenalaisesti. Sovellusalan erityisvaatimusten huomiointia ja kolmansien osapuolien käyttämättä jättämistä oletuksena pidettiin myös tärkeinä seikkoina.

Syyt sekä useiden päällekkäisten kolmansien osapuolien käytölle sekä kolmansien osapuolien integroinnille osaksi kriittisiäkin sivustoja olivat melko samankaltaisia. Usein syynä saattaa kehittäjien mukaan olla puhdas vahinko; kehittäjät eivät ole tietoisia kolmansista osapuolista, eivät ymmärrä niiden toimintaa, tai ne tulevat valmiin alustan mukana. Dataa myös usein kerätään varmuuden vuoksi tulevan päätöksenteon tueksi miettimättä sen suuremmin yksityisyysriskejä. Ilmeisiä syitä ovat myös liiketoiminnan edistäminen ja käyttökokemuksen parantaminen, joita edistetään usein käyttäjän yksityisyyden kustannuksella.

Kuvassa 16 on korkealla tasolla esitetty kehittäjien ehdottamia toimia kolmannen osapuolen tietovuotojen ehkäisemiseksi ja näiden välisiä suhteita. Hyvät käytännöt, sisäänrakennettu yksityisyys ja regulaation noudattaminen toimivat pohjana, jolle kolmansiin osapuoliin liittyvä riskienhallintaa ja tietovuotojen estäminen teknisin ja organisatorisin toimenpitein voidaan rakentaa. Näiden toimien voidaan ajatella muodostavan eettisen ja oikeudellisen viitekehyksen yksityisyyden huomioimiseksi ohjelmistokehityksessä. Organisaation tulisikin varmistaa, että nämä perusperiaatteet

[15] https://owasp.org/www-project-top-ten/



Kuva 16. Kolmannen osapuolen tietovuotojen ehkäiseminen.

tunnetaan ja otetaan aina huomioon sen toiminnassa ja ohjelmistokehitysprosessissa; kaikki muut käytännön toimet tietovuotojen estämiseksi nojaavat tähän perustaan.

Edelleen kehittäjien vastauksissa tuli selkeästi esiin, että organisaatioiden tulisi mukauttaa yksityisyystoimiaan käsittelemiensä henkilötietojen arkaluonteisuuden ja verkkopalvelun sovellusalan mukaisesti. Kriittisten ja arkaluonteisten verkkosivustojen, esimerkiksi verkkopohjaisten terveyspalveluiden, kohdalla tiukkojen yksityisyysarviointien, vahvan sovellusalatietämyksen ja asiantuntijuuden tarve on erityisen suuri. Palvelun käsittelemän datan arkaluonteisuutta arvioimalla ja asiantuntijalähtöisellä päätöksenteolla organisaatiot voivat varmistaa, että verkkopalvelussa on käytössä asianmukaiset toimenpiteet parhaan mahdollisen yksityisyyden tason varmistamiseksi. Näin varmistetaan, että regulaatio ja parhaat käytännöt tulevat erityisesti huomioiduiksi silloin, kun verkkopalvelu käsittelee erityisen luottamuksellisia henkilötietoja. Vaikkapa GDPR:n erityisten henkilötietokategorioiden, kuten terveystietojen, osalta tämä on erityisen tärkeää. Sovellusalatuntemus ja yksityisyyteen liittyvä asiantuntemus toimivat myös yhdessä jatkuvan yksityisyystestauksen ja -auditoinnin kanssa lopullisten kolmansiin osapuoliin liittyvien arkkitehtuuripäätösten perustana.

Sovellusalatietämystä ja yksityisyyteen liittyvää asiantuntemusta voidaan sitten hyödyntää jatkuvassa yksityisyystestauksessa ja -auditoinnissa. Esimerkiksi organisaation sisäiset ja ulkoiset auditoinnit, kriittisten tietovirtojen reaaliaikainen seuranta ja haavoittuvuuksien analysointi ovat tärkeitä toimenpiteitä vahvan yksityisyyden ja luottamuksellisuuden takaamiseksi. Tällainen verkkopalvelun yksityisyyden jatkuva verifiointi paljastaa mahdolliset heikkoudet luottamuksellisuudessa ja auttaa tekemään päätöksiä tietovuotojen ehkäisemiseksi.

Tehokas riskienhallinta myös edellyttää, että kehittäjillä on syvällinen ymmärrys kolmannen osapuolen palveluista. Yritysten on valittava verkkosivustollaan käytettävät kolmansien osapuolien komponentit huolellisesti ja perehdyttävä kattavasti kolmansien osapuolten käytön mahdollisiin riskeihin. Yksityisyysriskien vähentämiseksi on tärkeää olla tietoinen yrityksen verkkosivuston riippuvuuksista ja varmistaa, että kolmansien osapuolten palvelut ovat läpinäkyviä ja regulaation mukaisia. Kolmansien osapuolten aiheuttaman verkkoliikenteen analysoiminen, testaaminen ja auditointi antavat hyödyllistä tietoa niiden arviointiin, sillä ne paljastavat usein yksityisyyteen liittyviä ongelmia. Näin voidaan tehdä päätös tietyn kolmannen osapuolen valitsemisesta, poisjättämisestä tai korvaamisesta lokaalilla analytiikkaratkaisulla. Tässä päätöksentekoprosessissa hyödynnetään edelleen sovellusalatietämystä ottaen huomioon verkkopalvelun kriittisyys ja vaadittu yksityisyyden taso. Kolmansien osapuolien huolellinen arviointi ja valinta seuraa edelleen myös regulaation noudattamisesta ja parhaiden käytäntöjen huomioon ottamisesta: jokaisen arkkitehtuuripäätöksen, johon liittyy kolmas osapuoli, on oltava regulaation mukainen ja käyttäjän yksityisyyttä tukeva.

Julkaisun merkitys

Tutkimus lienee ensimmäinen nimen omaan web-analytiikkaan liittyviin yksityisyyskysymyksiin pureutuva yksityiskohtainen haastattelututkimus. Julkaisu IX rikastaa muutoin lähinnä teknistä verkkoliikenneanalyysiä hyödyntävää väitöstutkimusta menetelmällisesti ja tuo siihen empiirisen tutkimuksen ja alan aiemman kirjallisuuden tueksi mukaan myös web-kehittäjien käytännön tietämystä ja asiantuntijuutta. Näin pyritään saamaan monipuolisempaa ja kattavampaa ymmärrystä kehittäjien organisaatioiden asenteista ja olennaisiksi näkemistä tavoista parantaa yksityisyyttä. Teknisten riskien tunnistamisen ja tietovuotojen havaitsemisen lisäksi voidaan siten ymmärtämää myös niiden taustalla olevia prosesseja ja inhimillisiä tekijöitä.

Tutkimuksen avulla saatiin vahvistusta jo aiemmissa julkaisuissa estetylle arveluille siitä, että kolmansien osapuolien läsnäolo sivustolla on monessa tapauksessa vahinko, jonka merkitystä käyttäjän yksityisyyden kannalta ei aina täysin ymmärretä. Kolmannen osapuolen palvelut saattavat olla osa valmista alustaa ja dataa kerätään usein varmuuden vuoksi tulevan päätöksenteon tueksi. Toki käytetyillä palveluilla on

usein myös aivan oikeaa liiketoimintaan ja verkkopalvelun laatuun liittyvää merkitystä.

Tutkimuksella on luotu pohja kehykselle, jota voidaan käyttää kolmansien osapuolien karsimiseen ja niiden aiheuttamien tietovuotojen ehkäisyyn. Tämän kehyksen olennaisia osia ovat hyvien käytäntöjen seuraaminen ja yksityisyysregulaation noudattaminen, palvelun sovellusalueen ja kriittisyyden huomiointi yksityisyystoimien toteutuksessa, jatkuva testaus ja valvonta, sekä kolmannen osapuolen palvelujen huolellinen hallinta. Tällaisen kehyksen käytännön sovellukseen liittyvien yksityiskohtien määrittäminen vaatii kuitenkin vielä lisätutkimusta.

Tekijän kontribuutio

Väitöskirjan laatija on ideoinut tutkimuksen, tehnyt useimmat haastattelut, ollut mukana kerätyn aineiston koodaamisessa ja analysoinnissa, kirjoittanut suurimman osan julkaisusta sekä johtanut tutkimuksen tekemistä.

4 Johtopäätökset

4.1 Keskeisimmät löydökset

Väitöstutkimuksen keskeisimmät löydökset voidaan tiivistää seuraavasti:

- Suomalaisissa terveydenhuollon verkkopalveluissa on laajamittainen tietovuotoongelma, vaikka tilanne onkin löydösten tekemisen jälkeen parantanut.
- Kolmansille osapuolille on vuotanut jopa reseptilääkkeiden ja käyttäjän valitsemien lääkärien nimiä yhdistettyinä käyttäjän identiteettiin, mikä asettaa käyttäjän yksityisyyden vakavaan vaaraan.
- Käyttäjän mahdollisuudet vaikuttaa arkaluonteisten henkilötietojen keräämiseen omilla valinnoillaan ovat hyvin rajalliset.
- Kehittäjät eivät näytä kovinkaan hyvin tiedostavan henkilötietojen kolmansille osapuolille vuotamisesta aiheutuvaa yksityisyysongelmaa, vaikka suoraan ratkaisuja kysyttäessä osaavatkin ehdottaa tehokkaita toimenpiteitä yksityisyyden parantamiseksi verkkopalveluissa.
- Kolmannen osapuolen tietovuotojen korjaaminen vaatii terveydenhuollon palveluiden yksityisyys- ja auditointikäytäntöjen sekä henkilötietojen käsittelyn läpinäkyvyyden selkeää parantamista.

Seuraavaksi käydään läpi väitöskirjan tärkeimmät kontribuutiot, linkitetään ne tutkimuskysymyksiin ja tarkennetaan siten tutkielman pääväitteitä ja keskeisiä löydöksiä.

4.2 Kontribuutiot

Tässä luvussa vastataan väitöstutkimuksen tutkimuskysymyksiin ja esitetään niihin liittyvät kontribuutiot. Kuhunkin tutkimuskysymykseen liittyvät kontribuutiot sekä työn muut kontribuutiot on listattu taulukossa 7. Näitä käsitellään tarkemmin seuraavaksi.

Taulukko 7. Väitöskirjan tutkimuskysymyksiin liittyvät kontribuutiot

TK	Kontribuutio
TK1	K1: Terveysalan verkkosivujen kolmansien osapuolten kartoittaminen
	K2: Tietovuotojen ja vuotomekanismin tarkka tekninen kuvaus
	K3: Vuotavien henkilötietojen erittely ja vuotoihin liittyvät käyttöskenaariot
TK2	K4: Julkisen ja yksityisen sektorin palvelujen vertailu
	K5: Eri sovellusalojen vertailu
TK3	K6: Käyttäjän vaikutusmahdollisuuksien tarkastelu
	K7: Käyttäjän informoinnin riittävyyden tarkastelu
TK4	K8: Kolmannen osapuolen tietovuotojen syiden selittäminen
	K9: Tavat ehkäistä ja korjata kolmannen osapuolen tietovuotoja
Muut	K10: Yhteiskunnallinen vaikuttavuus
	K11: Verkkoliikenneanalyysi työkaluna
	K12: Tietovuotojen seuraukset eri sovellusaloilla

4.2.1 Tutkimuskysymys TK1

Ensimmäiseen tutkimuskysymykseen *Mitä kolmansien osapuolien palveluita verkkoapteekkien ja terveysalan verkkopalveluiden sivuilla on ja mitä apteekkien ja terveyspalvelujen asiakkaiden henkilö- ja terveystietoja näille kolmansille osapuolille mahdollisesti vuotaa?* haettiin vastausta tarkastelemalla verkkosivuilta niitä käytettäessä lähtevää verkkoliikennettä sekä erittelemällä vuotavia henkilötietoja ja niiden vastaanottajia erilaisissa käyttöskenaarioissa.

K1: Terveysalan verkkosivujen kolmansien osapuolten kartoittaminen

Suomalaisilla terveysalan verkkosivustoilla käytettyjä kolmansia osapuolia tarkastellessa on selvää, että vaikka Google Analyticsin tiedonkeruusta on esiintynyt EU-alueella paljon huolta viime vuosina [4], käytetään palvelua vielä hyvin laajalti kriittisilläkin terveystietoja käsittelevillä sivustoilla ja se dominoi vahvasti tässä tutkimuksessa tarkasteltua suomalaisilla verkkosivuilla käytettyjen kolmansien osapuolien joukkoa. Esimerkiksi 163 suomalaista verkkoapteekkia tarkasteltaessa (julkaisu III) Googlen palvelut (lähinnä Google Analytics) olivat käytössä 134 (81,6 %) apteekin verkkosivustoilla, mikä tarjoaa teknologiajätille varsin kattavan näkymän suomalaisten verkkoapteekkien käyttöön. Muiden kolmansien osapuolten määrät olivat tähän verrattuna hyvin pieniä, esimerkiksi toiseksi sijoittuneella Metalla oli vain 12 esiintymää. Käyttäjän tarkasteleman tai tilaaman lääkkeen nimi vuoti Googlelle 55 apteekin verkkosivustolla, yli kolmasosassa (33,7 %) kaikista suomalaisista verkkoapteekeista. Kaikista 57 apteekista, jotka vuotivat arkaluonteisia terveystietoja, 96,4 % lähetti näitä tietoja Googlelle. Verrattuna Googleen Metan osuus arka-

luonteisten tietojen vastaanottajana ei ollut läheskään yhtä suuri (8 tapausta, 4,9 % kaikista apteekeista).

Tilanne oli hyvin vastaavankaltainen muuallakin kuin verkkoapteekkien tapauksessa: Google oli suurin vuotojen vastaanottaja ja Meta toisella sijalla myös esimerkiksi lääkärikeskusten, mielenterveyspalvelujen ja kuntien tapauksessa. Muiden kolmansien osapuolten osuus on kokonaisuutta tarkastellessa marginaalinen. Silti mukana oli joitakin huolta herättäviä yksittäistapauksia, kuten vaikkapa reseptilääketietojen vuotaminen venäläiselle Yandexille tai Microsoftin Bing-palvelulle Pohjois-Amerikkaan. Vaalikoneiden osalta samanlainen kriittinen tietovuoto oli katseltujen ehdokkaiden vuotaminen kiinalaiselle TikTokille.

Positiivisina asioina on mainittava lokaalin Matomo-analytiikkapalvelun käyttö useilla tarkastelluilla verkkosivuilla sekä lukuisien verkkosivujen ylläpitäjien suorittama siirtyminen käyttämään Matomoa, kun heille oli ilmoitettu Googlen ja Metan palvelujen ongelmallisuudesta. Kuten aiemmin on mainittu, esimerkiksi apteekki- ja kuntasektoreilla on väitöskirjan löydösten seurauksena menossa laajamittainen muutos kohti käyttäjien yksityisyyttä kunnioittavan analytiikan käyttöä.

Tässä väitöskirjassa on kolmannen osapuolen palvelujen osalta keskitytty lähinnä analytiikkapalveluihin, mutta havaittujen kolmansien osapuolien joukossa on muitakin kuin varsinaisia analytiikkapalveluja. Usein kuitenkin esimerkiksi sosiaalisen median palveluihin (Facebook, TikTok) tai mainospalveluihinkin (Adform) liittyy tietynlaista analytiikkaa. Mukana on toki myös joitakin kolmansia osapuolia, kuten chat-palveluita, jotka melko selvästi eroavat perinteisestä web-analytiikasta. Näiden määriä ei ole useimmissa julkaisuissa tarkasti eritelty, mutta vaikkapa verkkoapteekkien tapauksessa tällaisten palvelujen osuus tietovuotojen vastaanottajina oli selvästi alle 5 % kaikista kolmansista osapuolista.

K2: Vuotavien henkilötietojen erittely ja vuotoihin liittyvät käyttöskenaariot

Väitöstutkimuksessa keskityttiin analysoimaan ja kuvaamaan tarkasti verkkopohjaisissa terveyspalveluissa vuotavia henkilötietoja. Moniin aiempiin terveydenhuollon verkkopalveluiden tietovuotoja käsitteleviin tutkimukseen verrattuna olemme tarkastelleet laajempaa joukkoa erilaisia käyttöskenaarioita ja mahdollisesti vuotavia henkilötietoja. Henkilötietoja siis saatiin manuaalisen verkkoliikenneanalyysin keinoin tarkasteltua syvällisemmin ja kategorisoitua tarkemmin.

Tästä syystä ja kattavan manuaalisen verkkoliikenneanalyysin ansiosta tietovuotoja löytyi enemmän kuin monessa aiemmassa terveydenhuollon palveluiden yksityisyyteen liittyvässä tutkimuksessa. Tätä eroa korostaa esimerkiksi tutkimus, jossa Huo ym. [69] tarkastelivat kolmansia osapuolia ja terveystietovuotoja 459 verkkopohjaisen terveyspalvelun sivuilla. Vain 9 sivuston todettiin vuotavan käyttäjän terveydentilaan mahdollisesti liittyviä tietoja (kuten lääketietoja ja laboratoriotuloksia) kolmansille osapuolille. Sitä vastoin tämän väitöskirjan terveydenhuollon palvelu-

ja (lääkäriasemat, laboratoriot ja terapiatalot) tarkastelevassa tutkimuksessa kaikista 10 sivustosta löydettiin ainakin jonkintasoisia tietovuotoja. Aiemmassa tutkimuksessa käytettyjen automaattisten menetelmien haasteena onkin tähän väitöskirjaan verrattuna erilaisten tietovuototyyppien havaitsematta jääminen sekä liian suppea henkilötietojen tarkastelu. Laaja erilaisten vuotojen tarkastelu antaa verkkopalveluiden kehityksessä osallisina oleville henkilöille paremman kuvan siitä, millaisten henkilötietojen vuotamiseen pitäisi varautua.

Väitöstutkimus on kuvannut arkaluonteisten henkilötietojen vuotamista erilaisten käyttöskenaarioiden pohjalta. Verkkosivujen laatijoiden on tärkeää ymmärtää, mitkä ovat potentiaalisesti kriittisiä tietovuodoille alttiita kohtia ja toiminnallisuuksia heidän sivustollaan. Terveystietojen vuotamisen osalta tällaisiksi on tunnistettu mm. lääkkeiden tilaukseen ja ajanvaraukseen liittyvät toiminnallisuudet (julkaisut I, III ja IV). Näistä kolmansilla osapuolille vuotaviin reseptilääkkeiden nimiin, potilaan varaamiin hoitotoimenpiteisiin tai lääkärien nimiin on kiinnitettävä erityistä huomiota terveysalan verkkopohjaisten sovellusten kehityksessä. Samaten hyvin kriittinen toiminnallisuus lähes millä tahansa verkkosivustolla on hakutoiminto. Koska käyttäjä saa tässä vapaasti päättää antamansa syötteen ja terveysalan palveluissa nämä hakutermit voivat useinkin olla sairauksia ja niiden oireita, on ensiarvoisen tärkeää, ettei näitä hakuja vuodeta kolmansilla osapuolille.

K3: Tietovuotojen ja vuotomekanismin tarkka tekninen kuvaus

Havaittujen tietovuotojen tutkimisen kannalta on myös äärimmäisen tärkeää, että vuotojen mekanismi kuvataan tarkasti. Muussa tapauksessa vuotojen merkitys ja vakavuus ei käy selväksi. Väitöstutkimuksessa on pyritty vahvasti korostamaan sitä, että tietovuodot koostuvat aina kahdesta eri komponentista, jotka yhdessä tekevät niistä vakavia: yksilöivästä ja asiayhteyteen liittyvästä tiedosta. Kun voidaan esimerkiksi sanoa, että tietty tunnistettava käyttäjä on aikeissa ostaa tiettyä reseptilääkettä, ja tämä tieto vuotaa kolmannelle osapuolelle, on käyttäjän yksityisyys selvästi uhattuna.

Paljon huomiota saaneen ja usein tutkimuksissa tarkastellun IP-osoitteen ohella tässä tutkimuksessa on pyritty painottamaan käyttäjän tunnistamiseen useita muitakin mekanismeja. Käyttäjä voidaan yksilöidä kolmannen osapuolen evästeissä olevan tunnisteen avulla, UserAgent-merkkijonoa ja lukuisia muita teknisiä laitetietoja hyväksi käyttäen, tai hänet voidaan helposti tunnistaa, jos hän on jo valmiiksi kirjautunut esimerkiksi Googlen tai Metan palveluihin samalla laitteella.

Näihin yksilöiviin tietoihin yhdistettävien asiayhteyteen liittyvien tietojen vuodot tapahtuvat lähes aina URL-osoitteiden kautta, sillä monet URL-osoitteet joko itsessään sisältävät valmiiksi arkaluonteisia tietoja tai joka tapauksessa paljastavat sivun, jolla näitä tietoja on. URL-osoitteiden lisäksi monet hyötykuormassa olevat tiedot voivat olla hyvin ongelmallisia, näiden joukossa voi olla vaikkapa lääkärikeskuk-

sen ajanvarauslomakkeeseen täytettyjä tietoja tai esimerkiksi tieto reseptilääkkeiden tilauspainikkeen painalluksesta (julkaisut III ja IV). Tällaisten asiayhteyteen liittyvien tietojen vuotamisen merkityksen ymmärtäminen ja vuotomekanismien huomioiminen on kriittisten verkkosivustojen kehittäjille ensiarvoisen tärkeää.

Vastauksena tutkimuskysymykseen TK1 voidaan siis kootusti todeta, että merkittävimpiä henkilötietoja kerääviä ja vastaanottavia kolmansia osapuolia ovat Google ja Meta. Vuotavista arkaluonteisista henkilötiedoista erityisen ongelmallisia ovat käyttäjän identiteettiin yhdistettyinä reseptilääkkeiden nimet ja ajanvarauksen yhteydessä vuotavat varatut hoitotoimenpiteet ja lääkärien nimet.

4.2.2 Tutkimuskysymys TK2

Toiseen tutkimuskysymykseen *Ovatko terveysalan verkkopalvelujen tietosuojapuutteet yhtä vakavia yksityisellä ja julkisella sektorilla, ja mikä on tilanne verrattuna muihin toimialoihin?* haettiin vastausta sisällyttämällä väitöstutkimukseen verkkopalveluita sekä julkiselta että yksityiseltä sektorilta ja vastaavasti myös muilta sovellusaloilta kuin terveysalalta (kunnat, vaalikoneet).

K4: Julkisen ja yksityisen sektorin palvelujen vertailu

Väitöstutkimus sisältää myös jonkin verran julkisen ja yksityisen sektorin verkkopalvelujen tietovuototilanteen vertailua. Julkaisun VI löydökset antavat viitteitä siitä, että julkisen sektorin mielenterveyspalveluissa yksityisyys on huomioita yksityisen sektorin palveluita paremmin. Todennäköisesti tällä tuloksella on myös yhteyttä yksityisen sektorin tarpeeseen seurata liiketoimintatavoitteiden toteutumista ja konversioita (esimerkiksi ajanvaraukset). Otos on tässä kuitenkin vain 7 mielenterveyssivustoa – eikä Suomen kokoisessa maassa olekaan merkittäviä alan toimijoita kovin suurta määrää. Kuntatutkimus puolestaan osoittaa 309 verkkosivun aineistolla, että kuntien verkkosivujen yksityisyydessä on runsaasti parannettavaa, kun yli puolet verkkosivustoista vuotaa apulaistietosuojavaltuutetun suosituksen vastaisesti vieraillut sivut ilman käyttäjän suostumusta tahoille, jotka hyvin todennäköisesti hyödyntävät tietoja kaupallisiin tarkoituksiin. Voidaan siis turvallisesti sanoa, etteivät läheskään kaikki suomalaiset julkisen sektorinkaan verkkosivut ole mallikelpoisessa kunnossa terveysalalla tai sen ulkopuolella. Olisikin toivottavaa, että julkinen sektori voisi toimia tässä asiassa esimerkkinä yksityiselle ja kolmannelle sektorille.

K5: Eri sovellusalojen vertailu

Verkkoapteekkien, terveydenhuollon palvelujen ja mielenterveyssivustojen lisäksi väitöstutkimus tarkasteli myös kuntien verkkosivuja (julkaisu VII) sekä vaalikonei-

ta (julkaisu VIII), jotta terveysalan löydöksiä voitaisiin verrata muihin aihealueisiin. Tulokset eivät anna viitteitä siitä, että erilaisilla arkaluonteisia tietoja käsittelevillä sovellusaloilla olisi tietovuotojen tai kolmannen osapuolen palvelujen osalta merkittäviä eroja.

Esimerkiksi vaalikoneissa tietovuotoja oli runsaasti ja merkittävimmät kolmannet osapuolet olivat pitkälti samoja kuin terveysalankin sivustoilla – kummallakin sovellusalalla vuotomekanismit ovat samat ja erityisesti URL-osoitteiden vuotaminen näyttää olevan ylläpitäjille ja kehittäjille pitkälti sokea piste. Kuntienkin sivustoista 51,8 % sisälsi kolmannen osapuolen palveluja, ja sama osuus sivustoista myös vuoti vierailtujen sivujen osoitteet eteenpäin. Samaten hakutoiminnallisuuden käyttäminen kuntien sivuilla johti lähes 50 % todennäköisyydellä hakutermin vuotamiseen kolmannelle osapuolelle. On syytä huomata, että kuntien sivuilla ei edes annettu suostumusta evästeisiin tai tietojen keräämiseen. Nämä monitapaustutkimukset osoittavat, että sama tietovuoto-ongelma on läsnä monellakin sovellusalalla, eivätkä ne merkittävästi eroa toisistaan yksityisyytensä tason suhteen.

Vastauksena tutkimuskysymykseen TK2 voidaan todeta, että vaikka julkisen sektorin palvelut erityisesti terveysalalla sisältävät yksityistä sektoria vähemmän kolmansia osapuolia ja tietovuotoja, on myös julkisen sektorin verkkopalvelujen yksityisyydessä runsaasti parannettavaa. Terveysalan ja muiden sovellusalojen verkkopalvelujen välillä ei ollut merkittäviä eroja tietovuotojen esiintyvyydessä.

4.2.3 Tutkimuskysymys TK3

Kolmanteen tutkimuskysymykseen, *Onko verkkopalvelujen käyttäjillä mahdollisuus vaikuttaa yksityisyytensä tasoon evästesuostumusten kautta ja tulevatko he riittävissä määrin informoiduiksi henkilötietojensa käsittelystä?*, haettiin vastausta tarkastelemalla julkaisussa VI useita terveysalan verkkopalveluja antamatta suostumusta tietojenkeruuseen ja vertaamalla tulosta tilanteeseen, jossa tietojenkeruuseen oli annettu lupa. Vastaavasti kuntia tarkastellessa (Julkaisu VII) saadaan kattava kuva tilanteesta, jossa käyttäjä ei ole antanut suostumustaan evästeisiin ja tietojenkeruuseen. Käyttäjän informointia tarkasteltiin vertailemalla tietosuojaselosteiden sisältöä oikeaan verkkoliikenteeseen.

K6: Käyttäjän vaikutusmahdollisuuksien tarkastelu

Tutkimukset, joissa ei annettu verkkosivun käyttäjänä suostumusta evästeisiin ja tiedonkeruuseen (julkaisut VI ja VII) osoittavat, että kolmannet osapuolet ovat yhä laajamittaisesti läsnä silloinkin, kun käyttäjä ei ole antanut niiden käytölle hyväksyntäänsä. Terveyspalveluja minimisuostumusasetelmalla tarkasteleva julkaisu VI osoitti, että 12 verkkosivustoa 20 tutkitusta vuoti henkilötietoja kolmansille osapuolille il-

man suostumustakin. Vastaavasti kuntien verkkosivuja tarkasteleva tutkimus osoitti, että 51,8 % tarkastelluista sivustoista vuoti henkilötietoja kolmansille osapuolille ilman suostumusta. Kummassakin tutkimuksessa Google pysyi suurimpana vuotojen vastaanottajana ja Meta sijoittui toiseksi.

Julkaisussa VI tarkastellaan minimisuostumusasetelmalla myös 8 sellaista lääkärikeskuksen verkkosivua, joita tarkasteltiin julkaisussa IV täydellä suostumuksella. Täydellä suostumuksella kaikista verkkopalveluista löytyi vuotoja. Vaikka kolmannen osapuolen palvelujen ja tietovuotojen määrä vähenee, vain 2 tarkastelluista sivustoista ei minimisuostumuksella enää aktivoinut mitään kolmannen osapuolen palvelua eikä sisältänyt tietovuotoja.

Näiden löydösten perusteella voidaan todeta, ettei käyttäjä edes kriittisillä verkkosivuilla useinkaan pysty täysin estämään henkilötietojensa keruuta ainakaan sivustojen evästeasetuksia muuttamalla, eikä hänellä myöskään ole realistisia mahdollisuuksia havaita, milloin henkilötietojen vuotamista tapahtuu. Edelleen käyttäjän vaikutusmahdollisuuksia pienentävät osaltaan mahdollinen vakiintunut rutiini hyväksyä kaikki tiedonkeruu, evästebannereissa usein ilmenevät pimeät käytännöt (kuten hyväksymispainikkeen korostaminen niin, että käyttäjä houkutellaan painamaan sitä) sekä se, ettei käyttäjä voi edes tarkkaan saada selville, mitä tietoja sivusto hänestä kerää, eikä näin mitenkään voi tehdä täysin tietoista ja informoitua päätöstä sivuston käytöstä tai tiedonkeruun hyväksymisestä.

Lopuksi on syytä huomata, että vaikka käyttäjällä esimerkiksi selainasetuksia muuttamalla tai selainlaajennoksia käyttämällä onkin mahdollisuuksia rajoittaa henkilötietojen keräämistä, ei tällaista tulisi käyttäjältä vaatia. Tämä myös sotii GDPR-asetusta vastaan siinä mielessä, ettei käyttäjän pitäisi joutua tekemään aktiivisia toimenpiteitä oletuksena päällä olevan tiedonkeruun pois päältä kytkemiseksi (opt-out), vaan päinvastoin käyttäjän olisi aktiivisesti hyväksyttävä tiedonkeruu (opt-in).

K7: Käyttäjän informoinnin riittävyyden tarkastelu

Väitöstutkimuksessa tarkasteltiin myös käyttäjän informointia kolmansille osapuolille lähetettävistä tiedoista normaalin käyttäjän näkökulmasta (esim. julkaisut I, II, III ja V). Tutkimus osoittaa, että käyttäjän informointi tietosuojaselosteissa on lähes aina puutteellista esimerkiksi lähetettävien henkilötietojen erittelyn ja kolmansien osapuolien nimeämisen osalta. Käyttäjille ei myöskään läheskään aina tehdä selväksi sitä, että heidät voidaan usein yksilöidä vaikkapa IP-osoitteiden tai laitetunnisteiden avulla. Samoin esimerkiksi tietosuojaselosteissa on kyllä saatettu mainita URL-osoitteiden päätyminen kolmannen osapuolen tietoon, mutta tämän tiedonsiirron seurauksia ei ole millään tavalla selitetty.

Jos tavallinen käyttäjä ei saa selkeästi tietää vaikkapa sitä, että hänen tarkastelemansa tai tilaamansa reseptilääkkeet päätyvät kolmannen osapuolen tietoon, ei informointi ole läpinäkyvää eikä suostumus voi olla täysin tietoinen. Toisaalta taas on

selvää, että koska tässä väitöskirjatutkimuksessa löytyneet tietovuodot hyvin usein ovat mitä ilmeisimmin kehittäjien ja palveluiden omistajienkin osalta vahingossa syntyneitä, ei niistä tietenkään anneta realistista kuvaa myöskään tietosuojaselosteissa. Joka tapauksessa tietosuojaselosteiden läpinäkyvyys ja käyttäjän informointi todettiin riittämättömiksi ja puutteellisiksi kaikissa väitöskirjan tutkimuksissa, joissa tätä osa-aluetta tarkasteltiin.

Vastauksena tutkimuskysymykseen TK3 voidaan siis todeta, ettei käyttäjä useinkaan omilla aktiivisilla toimillaan voi estää tietovuotoja, ja käyttäjän informointi henkilötietojen siirroista kolmansille osapuolille on yleensä riittämätöntä.

4.2.4 Tutkimuskysymys TK4

Neljättä tutkimuskysymystä *Mistä apteekkien ja terveyspalvelujen verkkosivujen tietosuojapuutteet johtuvat ja miten niiden yksityisyyttä voitaisiin parantaa ohjelmistoteknisestä näkökulmasta?* haastateltiin verkkopalvelujen kehitystehtävissä toimivia henkilöitä. Tällä tavoin selvitettyjä kolmannen osapuolen tietovuotojen syitä ja tapoja ehkäistä niitä vertailtiin sekä kirjallisuuteen että väitöskirjan kirjoittajan omiin näkemyksiin. Näin voidaan esittää perusteltuja suosituksia verkkosivustojen kehittäjille kolmannen osapuolen tietovuotojen välttämiseksi.

K8: Kolmannen osapuolen tietovuotojen syiden selittäminen

Kuten tämän väitöskirjan tutkimuksissa on jo aiemmin arveltu, merkittävinä syinä analytiikkapalvelujen käytölle kolmannen osapuolen tietovuotoihin nähtiin myös haastattelututkimuksessa tietämättömyys ja ajattelemattomuus. Valmiit alustat, joita usein käytetään modernien verkkosivustojen kehittämisessä, tarjoavat usein helppoja vaihtoehtoja lisätä kolmannen osapuolen analytiikkapalveluita tai joskus ottavat nämä palvelut käyttöön oletuksena. Kehittäjät eivät välttämättä ole aina tietoisia niiden olemassaolosta ja vaikutuksista yksityisyyteen. Haastatteluissa kehittäjät myös totesivat dataa kerättävän päätöksenteon tueksi ilman selkeitä tavoitteita ja päämääriä, ”varmuuden vuoksi”. Usein datan keräämisellä ajatellaan tuettavan liiketoimintatavoitteita ja saavutettavan kilpailuetua siitäkin huolimatta, että datan käyttötarkoitus ei ole selvä, eikä analytiikkatyökalusta osata ottaa täyttä hyötyä irti (ks. [45]).

Tärkeitä ja yleisiä syitä kolmansien osapuolen käytölle ja tietojen vuotamiselle ovat kehittäjien mukaan myös liiketoimintatavoitteet ja käyttäjäkokemuksen parantaminen. Liiketoimintatavoitteiden saavuttaminen verkossa vaatii käyttäjien seurantaa, navigointipolkujen tarkastelua ja markkinoinnin tuloksellisuuden mittaamista konversioiden tarkastelun avulla. Verkkopalvelun käytettävyyden ja suorituskyvyn arviointi ja parantaminen vaatii myös käytön mittaamista. Sen sijaan voidaan perus-

tellusti kysyä, miksi nämä toiminnot pitäisi tehdä kolmannen osapuolen tarjoamilla työkaluilla, jotka kierrättävät kerätyn datan sen palvelimien kautta [104].

K9: Tavat ehkäistä ja korjata kolmannen osapuolen tietovuotoja

Paras ja ilmeinen ratkaisu olisi poistaa kolmannen osapuolen palvelut kokonaan, sillä niiden ei yleisesti ottaen tulisi olla käytössä verkkosivustoilla, jotka käsittelevät terveystietojen kaltaisia arkaluonteisia henkilötietoja. Jos esimerkiksi analytiikkaa tarvitaan, voidaan käyttää lokaalia ratkaisua [104, 3]. Näin yritys tai organisaatio säilyttää itsellään kerätyn datan täyden hallinnan. Kaikkien kolmannen osapuolen palveluiden käyttö tulisi perusteellisesti arvioida käyttäjän yksityisyyden kannalta ja niiden käyttö tulisi aina perustella tarkasti. Terveyspalvelujen yhteydessä esimerkiksi luotettavien chat-palveluiden tai ajanvarausjärjestelmien käyttö voi olla perusteltua.

Sisäänrakennettu yksityisyys ja regulaation noudattaminen muodostavat haastateltujen kehittäjien mukaan perustan, johon kolmansien osapuolen tietovuotojen estäminen teknisin ja organisatorisin toimenpitein voi pohjautua. Näiden perusperiaatteiden tunteminen ja huomioiminen ohjelmistokehitysprosessissa on ensiarvoisen tärkeää. Yksityisyystoimien toteutus on myös vahvasti kontekstiriippuvainen asia, jossa käsiteltyjen henkilötietojen arkaluonteisuus ja verkkopalvelun sovellusala ja kriittisyys on aina huomioitava. Verkkopohjaisten terveyspalvelujen yksityisyyden varmistaminen vaatii vahvaa sovellusalatietämystä ja asiantuntemusta. Tätä osaamista voidaan edelleen hyödyntää jatkuvaan yksityisyystestauksessa ja -auditoinnissa. Tietosuoja-auditoinnit, testaus ja järjestelmän ulkopuolelle kulkevien tietovirtojen reaaliaikainen seuranta ovat tärkeitä toimenpiteitä vahvan yksityisyyden ja luottamuksellisuuden takaamiseksi. Tällainen teknisiin toimiin perustuva verkkopalvelun yksityisyyden jatkuva varmistaminen paljastaa kolmannen osapuolen palveluiden mahdolliset ongelmat ja auttaa kolmansien osapuolien hallinnassa.

Näiden perustavanlaatuisten ja yleisten toimien lisäksi kehittäjien tulisi tarkasti perehtyä verkkopalveluun integroitaviin kolmansiin osapuoliin ja tehdä sen perusteella huolellisia ja perusteltuja valintoja. Kolmansiin osapuoliin syventyminen voi sisältää esimerkiksi niiden tietosuojaselosteisiin, palveluntarjoajan taustaan, kehitykseen liittyviin seikkoihin tai jopa lähdekoodiin tutustumista. Näin voidaan päättää tietyn kolmannen osapuolen valitsemisesta, poisjättämisestä tai korvaamisesta lokaalilla analytiikkaratkaisulla. Edelleen myös sovellusalatietämys, verkkopalvelun vaadittu yksityisyyden taso ja mahdollisten auditointien ja testausten tulokset otetaan huomioon päätöksenteossa. Lopulta kolmansien osapuolien arvioinnin ja valinnan taustalla ovat myös regulaatio ja hyvät yksityisyyskäytännöt.

Tässä väitöstutkimuksessa esitetään, että soveltamalla näitä periaatteita verkkopalvelun koko kehitys- ja elinkaaren ajan siitä saadaan käyttäjän yksityisyyttä tukeva. Sisäänrakennetun yksityisyyden periaatteiden mukaisesti yksityisyys on huomioitava terveysteknologisissa ratkaisuissa jo suunnitteluvaiheessa [105].

Vastauksena tutkimuskysymykseen TK4 voidaan siis todeta, että kolmannen osapuolen tietovuotojen taustalla näyttävät olevan tietämättömyys ja huolimattomuus, datan kerääminen ilman selkeää päämäärää, sekä liiketoimintatavoitteiden edistäminen ja käyttäjäkokemuksen parantaminen käyttäjän yksityisyyden kustannuksella. Kolmannen osapuolen tietovuotoja voidaan ehkäistä seuraamalla regulaatiota ja hyviä käytäntöjä, ottamalla huomioon verkkopalvelun sovellusalakohtaiset vaatimukset, jatkuvalla testauksella ja valvonnalla, sekä kolmansien osapuolien huolellisella hallinnalla, eli yleisesti koko palvelun kehitys- ja elinkaaren kestävällä sisäänrakennetulla yksityisyydellä.

Muut kontribuutiot

Varsinaisiin tutkimuskysymyksiin vastaamisen ohella väitöstutkimuksella on myös muita merkittäviä kontribuutioita, joista merkittävin on luonnollisesti työn yhteiskunnallinen vaikuttavuus (K10) kolmansien osapuolien ja tietovuotojen vähenemisen muodossa useilla eri toimialoilla. Tähän liittyy myös mahdollisuus informoida verkkosivujen kehittäjiä ja verkkopalveluiden käyttäjiä tietovuodoista tämän väitöskirjan ja sen julkaisujen jo saaman julkisuuden kautta. Tämän lisäksi teoreettisempana mutta myös käytäntöön vahvasti soveltuvana kontribuutiona tässä väitöskirjassa on esitelty systemaattinen tapa analysoida verkkosivujen verkkoliikennettä ja havaita tätä kautta tietovuotoja kolmansille osapuolille (K11). Viimeisenä teoreettisena lisäkontribuutiona väitöskirjassa pohditaan myös tietovuotojen mahdollisia seurauksia eri sovellusaloilla (K12).

K10: Yhteiskunnallinen vaikuttavuus

Merkittävin väitöskirjan yhteiskunnallinen vaikutus on tietovuotojen väheneminen useiden eri toimialojen verkkosivustoilla. Esimerkiksi apteekkien osalta tietovuotojen määrä oli 48 verkkoapteekin otoksessa pudonnut noin 88 % vuoden aikana, kun tietosuojavaltuutetun toimisto oli aloittanut toimenpiteet näiden tietosuojapuutteiden korjaamiseksi. Vastaavasti lääkärikeskusten, laboratorioiden ja terapiatalojen tietovuodot vähenivät, vaikkakaan eivät yhtä merkittävästi.

Myös käytettyjen kolmansien osapuolien määrä on pudonnut useissa tapauksissa (millä tavallisesti on myös vaikutus tietovuotojen määrän vähenemiseen). Esimerkiksi kuntasektorilla 60 kaikkiaan 117 (51,3 %) kunnasta, joihin otettiin yhteyttä, oli poistanut Google Analyticsin tai Meta Pixelin kolme kuukautta yhteydenoton jälkeen. Jotkut kunnat totesivat myös, että kolmansien osapuolien tarve ollaan arvioimassa uudelleen verkkosivuston seuraavan suuren päivityksen yhteydessä, joten tulevaisuudessa todennäköisesti tapahtuu lisää positiivisia muutoksia kolmansien osapuolien ja tietovuotojen suhteen.

Tärkeää on myös kehittäjien informointi tietovuodoista. Tämä ei tapahdu ainoas-

taan julkaisujen ja tämän väitöskirjan kautta, vaan henkilötietoja vuotavien sivustojen kehittäjiin on myös laajamittaisesti oltu yhteydessä joko tietosuojaviranomaisen tai suoraan tämän väitöskirjan kirjoittajan toimesta. Tietosuojavaltuutetun toimisto on esimerkiksi ollut yhteydessä lukuisiin apteekkeihin tietovuotojen tiimoilta, ja vaikkapa osana kuntien verkkosivuja käsitellyttä tutkimusta väitöskirjan laatija oli sähköpostitse yhteydessä yli sataan Suomen kuntaan.

Edelleen on myös tärkeää informoida suurta yleisöä ja verkkopalveluiden käyttäjiä löydetyistä tietovuodoista. Näin on siitäkin huolimatta, ettei käyttäjillä ole kovinkaan hyviä mahdollisuuksia tietovuotojen välttämiseen. Esimerkiksi verkkoapteekkien (julkaisu III) ja vaalikoneiden (julkaisu VIII) osalta väitöskirjan tulokset ovat olleet esillä suomalaisessa mediassa.

K11: Verkkoliikenneanalyysi työkaluna

Osana tätä väitöskirjaa on tietovuotoja analysointia varten kehitetty Google Chrome -selaimen kehittäjätyökaluja hyödyntävä systemaattinen verkkoliikenteen analyysimenetelmä, jota kuvattiin luvussa 3.2.2. Menetelmä perustuu verkkoliikenteen hyötykuormien systemaattiseen ja tarkkaan manuaaliseen tarkasteluun ja löytyvien kriittisten henkilötietojen listaamiseen. Kuvatun kaltainen verkkoliikenneanalyysi soveltuu työkaluksi sekä tutkimukseen että yrityksillä ja organisaatioille, jotka haluavat tarkastella sivustojensa yksityisyyttä ja sivustolta lähteviä datavirtoja. Sekä käytetyt kehittäjätyökalut että menetelmä itsessään ovat sen verran yksinkertaisia ja intuitiivisia, että verkkoliikennettä voinee oppia analysoimaan melko lyhyelläkin perehdytyksellä ilman teknistäkin taustaa.

K12: Tietovuotojen seuraukset eri sovellusaloilla

Tässä väitöskirjassa on alustavasti kunkin tarkastellun sovellusalan (verkkoapteekit, lääkärikeskukset, mielenterveyspalvelut, kunnat jne.) osalta pohdittu tietovuotojen mahdollisia seurauksia. Ihmiset pitävät terveystietoja erittäin arkaluonteisina ja henkilökohtaisina [106, 107], minkä vuoksi näiden tietojen vuotaminen on vakava yksityisyyden loukkaus – ei vain lain silmissä, vaan myös ihmisten mielissä. Hyvin harvat ihmiset haluavat sairauksiaan, hoitojaan ja lääkitystään koskevien tietojen päätyvän väärien tahojen käsiin. Terveystietojen vuotaminen kolmansille osapuolille loukkaa vakavasti yksilön oikeutta yksityisyyteen, mikä voi johtaa haavoittuvuuden tunteeseen ja siihen, että ihmiset menettävät luottamuksensa terveydenhuollon verkkopalveluihin ja terveydenhuoltojärjestelmiin yleisesti [108]. Tämä voi olla erityisen haitallista jo valmiiksi haavoittuville ihmisryhmille.

Jos terveystietoja päätyy useille kolmansille osapuolille, kuten monissa tutkituissa palveluissa tapahtui, on aina suurempi mahdollisuus, että ne päätyvät vääriin käsiin. Terveystietojen vuotaminen ja mahdollisesti julkiseksi tuleminen voi pahim-

millaan aiheuttaa sosiaalista syrjäytymistä ja vaikkapa uhrin ennakkoluuloista kohtelua hänen henkilökohtaisessa elämässään ja työpaikallaan. Myöskään vuotaneiden terveystietojen hyödyntäminen taloudellisen hyödyn saavuttamiseksi ei ole mahdotonta. Niitä voidaan käyttää esimerkiksi käyttäjien profilointiin terveydentilan perusteella. Terveystiedot ovat erittäin arvokkaita ja voivat sisältää sellaista informaatiota, joka pysyy voimassa henkilön elämän loppuun asti.

Kun tarkastellaan konkreettisella tasolla tutkittuja suomalaisia terveyspalveluja, voidaan sanoa, että niillä on näiden potentiaalisten seurausten valossa paljon parannettavaa. Esimerkiksi Puumalainen ja Huttunen [51, s. 19] totesivat verkkoapteekkien kehittämistä käsittelevässä tutkimuksessaan 2022, että tietoturvaan liittyvät riskit huolestuttavat monia apteekkien asiakkaita. Vastaavasti Kyytsönen ym. [109] toteavat tarkastellessaan väestön kokemuksia sosiaali- ja terveydenhuollon sähköisestä asioinnista, että suurta osaa vastaajista (37 %) huolestutti henkilökohtaisten tietojen turvallisuus. Valitettavasti tämän väitöstutkimuksen löydökset osoittavat, etteivät nämä huolet ole olleet lainkaan perusteettomia. Mitä julkisen sektorin palveluihin tulee, läheskään kaikki terveys- ja sosiaalialan verkkopalvelut eivät myöskään näytä täysin vastaavan Valtionvarainministeriön digitalisoinnin periaatteita, joiden mukaan palveluja kehitetään asiakaslähtöisesti rakentamalla ”turvallisia palveluja, joista koituu hyötyä asiakkaille”.[1]

On silti hyvin tärkeää ymmärtää, että tämä väitöskirja voi vasta aloittaa keskustelua tällaisten tietovuotojen seurauksista. Vaikkapa verkkoapteekkien, lääkärikeskusten tai mielenterveyspalvelujen tietovuotojen seurauksien arviointi ei kuulu ainakaan pelkästään IT-alan tai tietosuoja-alan ammattilaisille vaan vaatii sovellusalan tuntemusta. Tämän takia aihe ei ole väitöstutkimuksen keskiössä. Silti maallikollekin on varmasti täysin selvää, että kun vaikkapa kaikista suomalaisista reseptilääkkeitä myyvistä verkkoapteekeista yli kolmasosa vuoti tietyn reseptilääkkeen nimen yhdistettynä asiakkaan yksilöiviin tietoihin, on tämä löydös, jolla voi olla vakavia seurauksia käyttäjien yksityisyyden kannalta.

4.3 Tutkimuksen rajoitukset

Tässä väitöskirjassa tehdyllä tutkimuksella on useita rajoituksia. Suoritettu verkkoliikenneanalyysi kattaa vain selvätekstillä lähetetyt henkilötiedot, joita tutkittavat verkkopalvelut välittävät kolmansille osapuolille. Joitakin tietoja siis saattaa jäädä havaitsematta tapauksissa joissa lähetettävää dataa on salattu tai koodattu. Tämä ei luonnollisestikaan vaikuta havaittuihin tietovuotoihin sinänsä, niitä vain saattaa olla enemmänkin.

Vastaavasti, koska monet tutkituista verkkopalveluista ovat laajoja, verkkoliikenneanalyysi ei kata kaikkia tutkittavien palveluiden tarjoamia toimintoja. Tutkimuk-

[1] https://vm.fi/digitalisoinnin-periaatteet

sessa keskityttiin tiettyihin, erityisen arkaluonteisiksi miellettyihin käyttöskenaarioihin pyrkimättä koko sivuston täysin kattavaan analyysiin. Tämä skenaariopohjainen lähestymistapa auttaa paljastamaan sivustojen kriittisimpien toiminnallisuuksien tyypillisiä tietovuotoja ja antamaan hyvän yleiskuvan ilmiöstä.

On myös tärkeää ymmärtää, että tehty verkkoliikenneanalyysi rajoittuu asiakaspuolelle. Vaikka voidaan sanoa varmasti, että tietyt henkilötiedot on lähetetty tietylle kolmannelle osapuolelle, ei tässä tutkimuksessa tehdyn analyysin perusteella voida tietää, mitä kolmas osapuoli tekee näillä tiedoilla omilla palvelimillaan. Ne voidaan poistaa välittömästi niiden vastaanottamisen jälkeen tai tallentaa myöhempää käyttöä varten. Ei voida myöskään tietää, vuotaako jokin verkkopalvelu palvelinpuolen kautta vielä lisää tietoja kolmansille osapuolille.

Vaikka olisi myös mielenkiintoista tutkia, siirretäänkö henkilötietoja Euroopan ulkopuolelle, kolmansien osapuolien palvelinten tarkkojen sijaintien määrittäminen on pilviteknologioiden laajan käytön vuoksi haastavaa. Siksi tutkimuksessa ei ole keskitytty maantieteellisiin sijainteihin kovinkaan paljon. Tästä huolimatta on hyvä muistaa, että Suomessa tai Euroopassa sijaitsevilta palvelimilta voidaan helposti siirtää tietoa eteenpäin.

Lisäksi otosten koot eivät tämän väitöskirjan sisältämissä tutkimuksissa ole kovinkaan suuria, jolloin selkeitä tilastollisia johtopäätöksiä ei voida tehdä. Tutkimukset kuitenkin auttavat monitapaustutkimukselle ominaiseen tapaan havaitsemaan tietyn sovellusalan verkkopalvelujen tietovuodoille ominaisia piirteitä ja pohtimaan näiden seurauksia. On myös huomionarvoista, että kaikilla sovellusaloilla (esim. lääkärikeskukset tai vaalikoneet) otoskoot eivät voisikaan olla kovin suuria, sillä merkittäviä aihealueeseen liittyviä sivustoja tai palveluja ei ole suurta joukkoa. Kaikkia Suomen verkkoapteekkeja ja kaikkia Suomen kuntien verkkosivuja tarkastelevat tutkimukset antavat aiheeseen myös suppeampia monitapaustutkimuksia selvästi laajempaa perspektiiviä.

Lopuksi on vielä syytä huomata, että tässä väitöskirjassa on analysoitu toisaalta verkkoliikenteen sisältämiä henkilötietoja teknisestä näkökulmasta ja toisaalta tietosuojaselosteita tavallisen käyttäjän näkökulmasta (selkeys ja läpinäkyvyys). Kuitenkaan yksityiskohtainen oikeustieteellinen analyysi tietosuojaselosteista ja kolmansille osapuolille tehtävien tiedonsiirtojen ongelmallisuudesta ei kuulu tämän tutkimuksen aihepiiriin.

4.4 Jatkotutkimusaiheet

Tässä väitöskirjassa ja IDA-projektissa yleensä on tarkasteltu moniin aihealueisiin liittyviä verkkosivustoja, mutta kiinnostavia teemoja on silti vielä lisääkin. Erityisen mielenkiintoisia tarkasteltavia voisivat olla GDPR:n mukaan erityisen arkaluonteiset aihealueet, kuten sivustot, joilla voi paljastua esimerkiksi henkilön uskonnolli-

nen vakaumus tai ammattiliiton jäsenyys. Samoin erilaisille haavoittuville ryhmille suunnattujen sivustojen yksityisyys olisi tarkastelemisen arvoinen asia.

Myös pelkästään terveydenhuollon verkkopalvelujen tietovuotojen tarkastelua voisi laajentaa paitsi suomessa, myös muiden maiden osalta. Erityisesti verkkoapteekkien ja lääkärikeskusten sivustojen huolestuttavat löydökset herättävät kysymyksen, onko tilanne muualla EU:ssa ja muualla maailmassa samanlainen kuin Suomessa.

Kolmannen osapuolen tietovuotojen laajempaan tarkasteluun liittyy läheisesti sen automatisointi. Tämän väitöskirjan verkkoapteekkeihin (julkaisu III) ja kuntien verkkosivuihin liittyvissä (julkaisu VII) tutkimuksissa käytettiin osin automaattisia työkaluja, ja verkkoliikenneanalyysin ja tietovuotojen tarkastelun automatisointia tullaan jatkamaan IDA-projektin puitteissa.

Edelleen tällaisen automaattisen työkalun ei tarvitse olla ainoastaan verkkoliikenneanalyysin väline, vaan olisi kiinnostavaa tarkastella tässä yhteydessä myös esimerkiksi pimeiden käytäntöjen automaattista tunnistamista ja tietosuojaselosteiden totuudenmukaisuuden ja läpinäkyvyyden arviointia esimerkiksi tekoälyn avulla. Työkalu voisi näin olla paitsi tutkimustyökalu, myös automaattinen ja helppokäyttöinen väline yrityksille ja organisaatioille verkkosivujen yksityisyyden ja siihen liittyvien hyvien käytäntöjen arviointiin. On myös selvää, että tällainen verkkosivustojen arviointi on liitettävissä laajempaan keskusteluun vastuullisuudesta ja datan reilusta käsittelystä [110]. Näin kyseessä itse asiassa olisi osa yrityksen tai organisaation vastuullisuuden ja sosiaalisen kestävyyden arviointia.

Lopuksi tässä väitöskirjassa myös selvitettiin web-kehittäjien näkemyksiä kolmansien osapuolien käytön ja tietovuotojen syntymisen syistä. Tätä selvitystä voisi vielä laajentaa esimerkiksi toteuttamalla aiheesta kyselyn, jossa kolmannen osapuolen tietovuotoihin liittyen kerättäisiin käsityksiä suuremmalta joukolta web-kehittäjiä ja otettaisiin mukaan myös verkkosivustot omistavia tahoja.

Lähteet

[1] Sekhar Somenahalli and Matthew Shipton. Examining the distribution of the elderly and accessibility to essential services. *Procedia-social and behavioral sciences*, 104:942–951, 2013.

[2] Fernando Almeida, José Duarte Santos, and José Augusto Monteiro. The challenges and opportunities in the digitalization of companies in a post-COVID-19 World. *IEEE Engineering Management Review*, 48(3):97–103, 2020.

[3] Jonas Gamalielsson, Björn Lundell, Simon Butler, Christoffer Brax, Tomas Persson, Anders Mattsson, Tomas Gustavsson, Jonas Feist, and Erik Lönroth. Towards open government through open source software for web analytics: The case of Matomo. *JeDEM-eJournal of eDemocracy and Open Government*, 13(2):133–153, 2021.

[4] Denise Quintel and Robert Wilson. Analytics and Privacy: Using Matomo in EBSCO's Discovery Service. *Information Technology and Libraries*, 39(3), 2020.

[5] Tim Wambach and Katharina Bräunlich. The evolution of third-party web tracking. In Olivier Camp, Steven Furnell, and Paolo Mori, editors, *Information Systems Security and Privacy*, pages 130–147. Springer International Publishing, 2017.

[6] Nik Thompson, Ravi Ravindran, and Salvatore Nicosia. Government data does not mean data governance: Lessons learned from a public sector application audit. *Government information quarterly*, 32(3):316–322, 2015.

[7] Timi Heino, Robin Carlsson, Sampsa Rauti, and Ville Leppänen. Assessing discrepancies between network traffic and privacy policies of public sector web services. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*, pages 1–6, 2022.

[8] Richard Spinello. *Cyberethics: Morality and Law in Cyberspace*. Jones and Bartlett Publishers, Sudbury, 2003.

[9] William Stallings, Lawrie Brown, Michael D Bauer, and Michael Howard. *Computer security: principles and practice*, volume 2. Pearson Upper Saddle River, 2012.

[10] Sandra Petronio. *Boundaries of privacy: Dialectics of disclosure*. Suny Press, 2002.

[11] Giuseppe Pino De Francesco. The general data protection regulation's practical impact on software architecture. *Computer*, 52(4):32–39, 2019.

[12] Stephen Breen, Karim Ouazzane, and Preeti Patel. GDPR: Is your consent valid? *Business Information Review*, 37(1):19–24, 2020.

[13] Norita Ahmad. Do data almost always eventually leak? *Computer*, 54(2):70–74, 2021.

[14] Prasad Jadhav and P Chawan. Data leak prevention system: A survey. *Virus*, 6(10):197–199, 2019.

[15] Nader Sohrabi Safa and Rossouw Von Solms. An information security knowledge sharing model in organizations. *Computers in Human Behavior*, 57:442–451, 2016.

[16] Omer Bin Tauqeer, Sadeeq Jan, Alaa Omar Khadidos, Adil Omar Khadidos, Fazal Qudus Khan, and Sana Khattak. Analysis of security testing techniques. *Intelligent Automation & Soft Computing*, 29(1):291–306, 2021.

[17] Kenza Meridji, Khalid T Al-Sarayreh, Alain Abran, and Sylvie Trudel. System security requirements: A framework for early identification, specification and measurement of related software requirements. *Computer Standards & Interfaces*, 66:103346, 2019.

[18] Dimitrios Zissis and Dimitrios Lekkas. Addressing cloud computing security issues. *Future Generation computer systems*, 28(3):583–592, 2012.

[19] James BD Joshi, Walid G Aref, Arif Ghafoor, and Eugene H Spafford. Security models for web-based applications. *Communications of the ACM*, 44(2):38–44, 2001.

[20] J Efrim Boritz. Is practitioners' views on core concepts of information integrity. *International Journal of Accounting Information Systems*, 6(4):260–279, 2005.

[21] William Stallings. *Information privacy engineering and privacy by design: Understanding privacy threats, technology, and regulations based on standards and best practices*. Addison-Wesley Professional, 2019.

[22] Suhail Qadir and SMK Quadri. Information availability: An insight into the most important attribute of information security. *Journal of Information Security*, 7(3):185–194, 2016.

[23] Mingwei Zhang and R Sekar. Control flow and code integrity for cots binaries: An effective defense against real-world rop attacks. In *Proceedings of the 31st Annual Computer Security Applications Conference*, pages 91–100, 2015.

[24] WAA Standards Committee et al. Web analytics definitions. *Washington DC: Web Analytics Association*, 2008.

[25] Jan H Schumann, Florian Von Wangenheim, and Nicole Groene. Targeted online advertising: Using reciprocity appeals to increase acceptance among users of free web services. *Journal of Marketing*, 78(1):59–75, 2014.

[26] Dave Chaffey and Mark Patron. From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. *Journal of Direct, Data and Digital Marketing Practice*, 14(1):30–45, 2012.

[27] Guangzhi Zheng and Svetlana Peltsverger. Web analytics overview. In *Encyclopedia of Information Science and Technology, Third Edition*, pages 7674–7683. IGI Global, 2015.

[28] Layla Hasan, Anne Morris, and Steve Probets. Using google analytics to evaluate the usability of e-commerce sites. In *International conference on human centered design*, pages 697–706. Springer, 2009.

[29] Janne Nielsen. Using mixed methods to study the historical use of web beacons in web tracking. *International Journal of Digital Humanities*, 2(1):65–88, 2021.

[30] Tietosuojavaltuutetun toimisto. Apulaistietosuojavaltuutetun päätös käsittelyn lainmukaisuutta, käsittelyn turvallisuutta, sisäänrakennettua ja oletusarvoista tietosuojaa, rekisteröityjen informointia ja henkilötietojen siirtoa kolmansiin maihin koskevassa asiassa. 13.12.2022.

[31] L. Bygrave and L. Tosoni. Article 4(1). personal data. In C. Kuner, L. Bygrave, C. Docksey, and L Drechsler, editors, *The EU General Data Protection Regulation: A Commentary*. Oxford University Press, Oxford, United Kingdom, 2020.

[32] Vikas Mishra, Pierre Laperdrix, Antoine Vastel, Walter Rudametkin, Romain Rouvoy, and Martin Lopatka. Don't count me out: On the relevance of ip address in the tracking ecosystem. In *Proceedings of The Web Conference 2020*, pages 808–815, 2020.

[33] Oleksii Starov, Yuchen Zhou, Xiao Zhang, Najmeh Miramirkhani, and Nick Nikiforakis. Betrayed by your dashboard: Discovering malicious campaigns via web analytics. In *Proceedings of the 2018 World Wide Web Conference*, pages 227–236, 2018.

[34] Timothy Libert. Privacy implications of health information seeking on the web. *Communications of the ACM*, 58(3):68–77, 2015.

[35] Jeff Kline, Paul Barford, Aaron Cahn, and Joel Sommers. On the structure and characteristics of user agent string. In *Proceedings of the 2017 Internet Measurement Conference*, pages 184–190, 2017.

[36] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. Mitigating browser fingerprint tracking: multi-level reconfiguration and diversification. In *2015 IEEE/ACM 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, pages 98–108. IEEE, 2015.

[37] Yabing Liu, Han Hee Song, Ignacio Bermudez, Alan Mislove, Mario Baldi, and Alok Tongaonkar. Identifying personal information in internet traffic. In *Proceedings of the 2015 ACM on Conference on Online Social Networks*, pages 59–70, 2015.

[38] Stephanie Winkler and Sherali Zeadally. Privacy policy analysis of popular web platforms. *IEEE technology and society magazine*, 35(2):75–85, 2016.

[39] Sampsa Rauti. A survey on countermeasures against man-in-the-browser attacks. In *International Conference on Hybrid Intelligent Systems*, pages 409–418. Springer, 2019.

[40] Timothy Dougan and Kevin Curran. Man in the browser attacks. *International Journal of Ambient Computing and Intelligence (IJACI)*, 4(1):29–39, 2012.

[41] Robert Epstein and Vanessa R Zankich. The surprising power of a click requirement: How click requirements and warnings affect users' willingness to disclose personal information. *Plos one*, 17(2):e0263097, 2022.

[42] Jacob Klemovitch, Lauren Sciabbarrasi, and Alan Peslak. Current privacy policy attitudes and fair information practice principles: A macro and micro analysis. *Issues In Information Systems*, 22(3), 2021.

[43] Jonathan R. Mayer and John C. Mitchell. Third-party web tracking: Policy and technology. In *2012 IEEE Symposium on Security and Privacy*, pages 413–427, 2012. doi: 10.1109/SP.2012.47.

[44] Jukka Ruohonen and Ville Leppänen. Invisible pixels are dead, long live invisible pixels! In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pages 28–32, 2018.

[45] Tom Alby. Popular, but hardly used: Has google analytics been to the detriment of web analytics? In *Proceedings of the 15th ACM Web Science Conference 2023*, pages 304–311, 2023.

[46] Esko Vuorinen, Panu Puhtila, Sampsa Rauti, and Ville Leppänen. From whistle to echo: Data leaks in web-based whistleblowing channels. In *Nordic Conference on Secure IT Systems*, pages 37–53. Springer, 2023.

[47] Nayanamana Samarasinghe, Aashish Adhikari, Mohammad Mannan, and Amr Youssef. Et tu, brute? privacy analysis of government websites and mobile apps. In *Proceedings of the ACM Web Conference 2022*, pages 564–575, 2022.

[48] Cookiebot. Ad tech surveillance on the public sector web. https://www.cookiebot.com/media/1121/cookiebot-report-2019-medium-size.pdf, 2019.

[49] Elena Maris, Timothy Libert, and Jennifer R Henrichsen. Tracking sex: The implications of widespread sexual data leakage and tracking on porn websites. *new media & society*, 22(11): 2018–2038, 2020.

[50] Nayanamana Samarasinghe, Pranay Kapoor, Mohammad Mannan, and Amr Youssef. No salvation from trackers: Privacy analysis of religious websites and mobile apps. In *International Workshop on Data Privacy Management*, pages 151–166. Springer, 2022.

[51] Inka Puumalainen and Janne Huttunen. Lääkkeet läheltä vai kaukaa – Verkkoapteekkien kehittäminen osana lähiapteekkien palvelua. Apteekkariliitto, 2022.

[52] Chiau Soon Long, Harshily Kumaran, Khang Wen Goh, Faizah Safina Bakrin, Long Chiau Ming, Inayat Ur Rehman, Jagjit Singh Dhaliwal, Muhammad Abdul Hadi, Yee Wai Sim, and Ching Siang Tan. Online pharmacies selling prescription drugs: Systematic review. *Pharmacy*, 10(2):42, 2022.

[53] Bassam M Alwon, Gennifer Solomon, Faseeha Hussain, and David J Wright. A detailed analysis of online pharmacy characteristics to inform safe usage by patients. *International journal of clinical pharmacy*, 37(1):148–158, 2015.

[54] Jack E Fincham. Negative consequences of the widespread and inappropriate easy access to purchasing prescription medications on the internet. *American Health & Drug Benefits*, 14(1): 22, 2021.

[55] Grazia Orizio, Anna Merla, Peter J Schulz, Umberto Gelatti, et al. Quality of online pharmacies and websites selling prescription drugs: a systematic review. *Journal of medical Internet research*, 13(3):e1795, 2011.

[56] Joanne Kuzma. Web vulnerability study of online pharmacy sites. *Informatics for Health and Social Care*, 36(1):20–34, 2011.

[57] Alexander R Zheutlin, Joshua D Niforatos, and Jeremy B Sussman. Data-tracking among digital pharmacies. *Annals of Pharmacotherapy*, pages 958–962, 2022.

[58] Joanne Kuzma, Kate Dobson, and Andrew Robinson. An examination of privacy policies of global on-line e-pharmacies. *European Journal of Research and Reflection in Management Sciences*, 4(6):23–28, 2016.

[59] Long Ma. Understanding non-adopters' intention to use internet pharmacy: Revisiting the roles of trustworthiness, perceived risk and consumer traits. *Journal of Engineering and Technology Management*, 59:101613, 2021.

[60] Kim Rattay, Hanna Kauppinen, Riitta Ahonen, and Johanna Timonen. Mielipiteitä sähköisen reseptin tietosuojasta – kyselytutkimus apteekkien asiakkaille. *Dosis: Farmaseuttinen aikakauskirja*, (4):264–275, 2018.

[61] András Fittler, Róbert György Vida, Mátyás Káplár, Lajos Botz, et al. Consumers turning to the internet pharmacy market: cross-sectional study on the frequency and attitudes of hungarian patients purchasing medications online. *Journal of medical Internet research*, 20(8):e11115, 2018.

[62] Isaac D Montoya and Elda Jano. Online pharmacies: safety and regulatory considerations. *International Journal of Health Services*, 37(2):279–289, 2007.

[63] Isabelle Kes, Daniel Heinrich, and David M. Woisetschläger. Behavioral Targeting in Health Care Marketing: Uncovering the Sunny Side of Tracking Consumers Online. *Developments in Marketing Science: Proceedings of the Academy of Marketing Science*, page 297, 2016.

[64] Ken Masters. The gathering of user data by national medical association websites. *The Internet Journal of Medical Informatics*, 6(2), 2012.

[65] Jacquelyn Burkell and Alexandre Fortier. Privacy policy disclosures of behavioural tracking on consumer health websites. In *Proceedings of the American Society for Information Science and Technology*, volume 50, pages 1–9. Wiley Online Library, 2013.

[66] Jacquelyn Burkell and Alexandre Fortier. Consumer health websites and behavioural tracking. In *Proceedings of the Annual Conference of CAIS/Actes du congrès annuel de l'ACSI*, 2012.

[67] Marco D Huesch. Privacy threats when seeking online health information. *JAMA Internal Medicine*, 173(19):1838–1840, 2013.

[68] Shonda D Brown and Yair Levy. Towards a development of an index to measure pharmaceutical companies' online privacy practices. *Online Journal of Applied Knowledge Management (OJAKM)*, 1(1):93–108, 2013.

[69] Mingjia Huo, Maxwell Bland, and Kirill Levchenko. All eyes on me: Inside third party trackers' exfiltration of phi from healthcare providers' online systems. In *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, pages 197–211. ACM, 2022.

[70] Olinda Santin, Theresa McShane, Peter Hudson, and Gillian Prue. Using a six-step co-design model to develop and test a peer-led web-based resource (plwr) to support informal carers of cancer patients. *Psycho-oncology*, 28(3):518–524, 2019.

[71] Jake Linardon, John Rosato, and Mariel Messer. Break binge eating: Reach, engagement, and user profile of an internet-based psychoeducational and self-help platform for eating disorders. *International Journal of Eating Disorders*, 53(10):1719–1728, 2020.

[72] Alexander R Zheutlin, Joshua D Niforatos, and Jeremy B Sussman. Data-tracking on government, non-profit, and commercial health-related websites. *Journal of general internal medicine*, pages 1–3, 2021.

[73] Ari B Friedman, Raina M Merchant, Amey Maley, Karim Farhat, Kristen Smith, Jackson Felkins, Rachel E Gonzales, Lujo Bauer, and Matthew S McCoy. Widespread third-party tracking on hospital websites poses privacy risks for patients and legal liability for hospitals. *Health Affairs*, 42(4):508–515, 2023.

[74] Xiufen Yu, Nayanamana Samarasinghe, Mohammad Mannan, and Amr Youssef. Got sick and tracked: Privacy analysis of hospital websites. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 278–286. IEEE, 2022.

[75] Ari B Friedman, Lujo Bauer, Rachel Gonzales, and Matthew S McCoy. Prevalence of third-party tracking on abortion clinic web pages. *JAMA Internal Medicine*, 182(11):1221–1222, 2022.

[76] Aishwarya Surani, Amani Bawaked, Matthew Wheeler, Braden Kelsey, Nikki Roberts, David Vincent, and Sanchari Das. Security and privacy of digital mental health: An analysis of web services and mobile apps. In *Conference on Data and Applications Security and Privacy*, 2023.

[77] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–18, 2021.

[78] Eeva Pyörälä. Potilaan näkökulmia terveydenhuollon digitaalisiin palveluihin. *Suomen lääkärilehti*, 76(46):2713–2716, 2021.

[79] Mildred Patten. *Proposing empirical research: A guide to the fundamentals*. Routledge, 2016.

[80] Christian Rossow, Christian J Dietrich, Herbert Bos, Lorenzo Cavallaro, Maarten Van Steen, Felix C Freiling, and Norbert Pohlmann. Sandnet: Network traffic analysis of malicious software. In *Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, pages 78–88, 2011.

[81] Mohammed Abdul Qadeer, Arshad Iqbal, Mohammad Zahid, and Misbahur Rahman Siddiqui. Network traffic analysis and intrusion detection using packet sniffer. In *2010 Second International Conference on Communication Software and Networks*, pages 313–317. IEEE, 2010.

[82] Markus Laine, Jarkko Bamberg, and Pekka Jokinen. Tapaustutkimuksen käytäntö ja teoria. In *Tapaustutkimuksen taito*, pages 9–38. 2007.

[83] Robert E Stake. *Multiple case study analysis*. Guilford press, 2013.

[84] Päivi Eriksson and Katri Koistinen. Monenlainen tapaustutkimus. kuluttajatutkimuskeskuksen julkaisuja 4. *Kerava: Kuluttajatutkimuskeskus*, 2005.

[85] Murray J Fisher and Andrea P Marshall. Understanding descriptive statistics. *Australian critical care*, 22(2):93–97, 2009.

[86] Michael Meuser and Ulrike Nagel. The expert interview and changes in knowledge production. In *Interviewing experts*, pages 17–42. Springer, 2009.

[87] Sirkka Hirsjärvi and Helena Hurme. Teemahaastattelu: teemahaastattelun teoria ja käytäntö. *Helsinki: Yliopistopaino*, 2001.

[88] Dennis A Gioia, Kevin G Corley, and Aimee L Hamilton. Seeking qualitative rigor in inductive research: Notes on the gioia methodology. *Organizational research methods*, 16(1):15–31, 2013.

[89] Anselm Strauss, Juliet Corbin, et al. *Basics of qualitative research*, volume 15. SAGE, Newbury Park, CA, 1990.

[90] Robin Carlsson, Sampsa Rauti, Sini Mickelsson, Tuomas Mäkilä, Timi Heino, Elina Pirjatanniemi, and Ville Leppänen. Several online pharmacies leak sensitive health data to third parties. In *World Conference on Information Systems and Technologies*, pages 164–175. Springer, 2023.

[91] Sampsa Rauti, Robin Carlsson, Sini Mickelsson, Tuomas Mäkilä, Timi Heino, Elina Pirjatanniemi, and Ville Leppänen. Analyzing third-party data leaks on online pharmacy websites. *Health and Technology*, 14(2):375–392, 2024.

[92] Werner Schäfke-Zell. Revisiting the definition of health data in the age of digitalized health care. *International Data Privacy Law*, 12(1):33–43, 2022.

[93] Gianclaudio Malgieri and Giovanni Comandé. Sensitive-by-distance: quasi-health data in the algorithmic era. *Information & Communications Technology Law*, 26(3):229–249, 2017.

[94] Sampsa Rauti, Robin Carlsson, Samuli Laato, Timi Heino, Panu Puhtila, and Ville Leppänen. Health data leaks to third parties in web-based health services. Accepted to TKTP2024., 2024.

[95] Sampsa Rauti, Esko Vuorinen, Panu Puhtila, and Robin Carlsson. Analysis of third-party data leaks on finnish mental health websites. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*, pages 1543–1548. IEEE, 2024.

[96] Kewa Jiang. Mental health mobile apps and the need to update federal regulations to protect users. *Mich. Tech. L. Rev.*, 28:421, 2021.

[97] Lisa Parker, Vanessa Halter, Tanya Karliychuk, and Quinn Grundy. How private is your mental health app data? an empirical study of mental health app privacy policies and practices. *International Journal of Law and Psychiatry*, 64:198–204, 2019. ISSN 0160-2527.

[98] Sampsa Rauti, Esko Vuorinen, Robin Carlsson, and Panu Puhtila. Data leaks to third-party services on medical websites. In *2023 16th International Conference on Security of Information and Networks (SIN)*, pages 1–7. IEEE, 2023.

[99] Sampsa Rauti, Robin Carlsson, Panu Puhtila, and Ville Leppänen. Third-party data leaks on municipal websites. In *International Congress on Information and Communication Technology*, pages 599–610. Springer, 2024.

[100] Timi Heino, Sampsa Rauti, Samuli Laato, Robin Carlsson, and Ville Leppänen. Leaky democracy: Third parties in voting advice applications. In *International Conference on Smart Computing and Communication*, pages 351–360. Springer, 2024.

[101] Jim Isaak and Mina J Hanna. User data privacy: Facebook, cambridge analytica, and privacy protection. *Computer*, 51(8):56–59, 2018.

[102] Vian Bakir. Psychological operations in digital political campaigns: Assessing cambridge analytica's psychographic profiling and targeting. *Frontiers in Communication*, 5:67, 2020.

[103] Sampsa Rauti, Panu Puhtila, Samuli Laato, and Tuomas Mäkilä. Insights from developers: Understanding the reasons for third-party data leaks in web environment. Submitted to..., 2024.

[104] Adam Chandler and Melissa Wallace. Using piwik instead of google analytics at the cornell university library. *The Serials Librarian*, 71(3-4):173–179, 2016.

[105] Outi Linnaranta and Ilkka Pelkonen. Teknologia-avusteiset mielenterveyden hoitomenetelmät suomessa. Terveyden ja hyvinvoinnin laitos THL, 2023.

[106] Filiz İşleyen, Mustafa Mahir Ülgü, and Kemal Hakan Gülkesen. A national minimum health log standard; samilog. *Journal of International Health Sciences and Management*, 10(19):20–24, 2024.

[107] Andrew J Rohm and George R Milne. Just what the doctor ordered: The role of information sensitivity and trust in reducing medical information privacy concern. *Journal of Business Research*, 57(9):1000–1011, 2004.

[108] Vahid Rafe and Maryam Monfaredzadeh. A qualitative framework to assess hospital/medical websites. *Journal of medical systems*, 36:2927–2939, 2012.

[109] Maiju Kyytsönen, Anna-Mari Aalto, and Tuulikki Vehko. Sosiaali-ja terveydenhuollon sähköinen asiointi 2020-2021: Väestön kokemukset. THL, 2021.

[110] Annamari Vänskä, Sampsa Rauti, Timi Heino, Robin Carlsson, Sini Mickelsson, and Natalia Särmäkari. Fair data is the new black: Online shopping, data leaks, and broadening the understanding of sustainable fashion. *Fashion Theory*, 28(3):305–333, 2024.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2025

ISBN 978-952-02-0120-3 (Painettu)
ISBN 978-952-02-0121-0 (PDF)
ISSN 2736-9390 (Painettu)
ISSN 2736 9684 (Sähköinen)