Radboud University Nijmegen

Faculty of Science

Negotiating Privacy and Utility: A Study on Google Analytics Implementation

Thesis Research

Master's thesis for Information Sciences

Author:
Tu Ngoc Huy

Supervisor:
dr. I. (Ilona) Wilmont
mr. drs. M.S.L. (Marvin) van Bekkum
prof. mr. F.J. (Frederik) Zuiderveen Borgesius

July 2023

Acknowledgement

This thesis is the culmination of many long days and nights spent writing over nearly 5 months, and I could not have completed this journey without the unending support from Ilona Wilmont and Marvin van Bekkum (my daily supervisors). They reviewed this thesis more than 14 times, spending countless hours providing feedback and offline meetings. They also gently nudged me in the right direction and encouraged me to persevere whenever I face my fear of imposter syndrome. Their unwavering motivation and unquestioning belief in my capabilities served as an invaluable source of hope during this journey, and I consider it a true privilege to have been under their esteemed guidance.

To Frederik Zuiderveen Borgesius, my second reader, your enlightening discourses on the course Law & Technology; the passion and advocacy you have for privacy issues fueled my curiosity and interest in this intricate domain. I am deeply appreciative of the time you took to be my second reader and to review my work, as well as to attend my final presentation.

I extend my sincere thanks to the eight participants whose participation enriched this study. Their generous contributions, in terms of time and expertise, provided invaluable insights, making this research a realistic and credible endeavor.

My journey of crafting this research was positively influenced by Jorrit Geels, Marene Dimmendaal, and Marieke de Vries. Their pivotal role during the proposal stage, coupled with their instrumental assistance in connecting me with Marvin and Ilona, set the wheels of this research journey in motion. The invaluable lessons I absorbed during their Judgment and Decision-Making for Information Sciences classes not only spurred me to choose this topic but also inspired the path I should follow.

Finally, my profound thanks go to Radboud University Nijmegen, for providing me with the opportunity to study here. While my journey may have been short, the interaction and lessons I got from all the lecturers have been profoundly impactful, marking one of the most rewarding years of my life. The knowledge and experiences I gained here will continue to resonate and inspire me as I advance in my academic and professional endeavors.

I cannot close this chapter without expressing my gratitude to all my friends studying in The Netherlands this academic year – Ale, Sam, Truc, Linh, Phuoc, Nguyen, Tin, Vu, Chi, Anh, Lam, An, Quynh, Chi, Michael, Karolis, Laura, Danny, Manos, Veselina, Joachim, Jokki, and the many others. Their companionship, mutual support, and collective wisdom made this journey memorable. They were my safe haven, my confidantes, and my pillars of support, adding a warm touch to my academic endeavors.

This thesis, an embodiment of perseverance, passion, and hard work, is also a testament to the unwavering support and encouragement I received from everyone around me. To all who were a part of this journey, I remain eternally indebted.

Nijmegen, July 1st, 2023

Tu Ngoc Huy

Contents

Abstract

The growing utilization of the data tracking tools, such as Google Analytics, has spurred concerns about user privacy and data protection. Legal authorities have declared the use of such tools illegal in certain jurisdictions and issued recommended measures for compliant usage. However, adhering to these measures could restrict the effectiveness of these tools, leading to a difficult task for developers who need to balance Google Analytics' utility with privacy protection. It is essential to understand how developers and analytics specialists navigate these challenges to prepare for potential legal changes. The current literature lacks a comprehensive understanding of the matter; this research aims to identify methods to reconcile utility and privacy legality in an ambiguous legal environment, offering insights that could influence future regulations and decision-making in organizations using Google Analytics. By deriving best practices from the literature and confirming them via semi-structured interviews with eight analytics specialists, this study explores the practicality and the extent of implementation in this intersection of privacy and utility. The findings highlight that recommended practices such as user consent, IP anonymization, and data minimization are still prevalent. At the same time, methods involving proxy servers and rigorous pseudonymization of referrers and URL parameters are less widespread due to constraints such as cost, technical complexity, and operational necessities. Nonetheless, emerging practices like the adoption of Google Analytics 4 and server-side tracking could potentially bridge this divide. Despite these advances, formulating a one-size-fits-all solution remains an unsolved challenge, underscoring the need for a concerted effort among Google, legal entities, and developers to formulate clearer guidelines and ensure the long-term viability of data privacy measures.

1 Introduction

1.1 Uncertain legal climate surrounding Google Analytics

In recent times, numerous court cases have brought to light potential infringements of the General Data Protection Regulation (GDPR)'s data transfer principle by Google Analytics. Under the GDPR, transferring personal data outside the EU is prohibited unless the recipient can guarantee adequate data protection (Article 44 of GDPR, 2018); this act is also known as illegitimate data transfer. Google Analytics, a prevalent analytics service, has faced intense scrutiny for possible violations of this principle.

Schrems II (a data privacy verdict) was issued by the Court of Justice of the European Union in July 2020 (Court of Justice of the European Union, 2020). It invalidated the EU-US Privacy Shield, which made it more complicated to share data between the EU and the US. As a result, the decision raised profound concerns among researchers and professionals in the field of data privacy and protection due to the extensive access of US intelligence agencies to European citizens' data.

At the core of the invalidation lies the Foreign Intelligence Surveillance Act (FISA), a US federal law that allows for collecting foreign intelligence information through electronic surveillance and physical searches. The Court of Justice of the European Union (CJEU) argued that FISA fails to provide adequate protection for EU citizens' data against indiscriminate surveillance by US intelligence agencies, thus undermining the essence of the EU's fundamental right to respect private life (CJEU, 2020). The court raised concerns about the potential for indiscriminate access to and processing of EU citizens' data by US authorities without effective remedies for individuals to challenge such practices. The CJEU's argument reflects its interpretation that the Privacy Shield Framework, as built upon the principles of adequacy and equivalent protection, cannot overcome the deficiencies in the US legal framework governing surveillance activities.

The CJEU's decision highlights the significance the court places on upholding the fundamental rights of EU citizens, particularly the right to privacy. It emphasizes that any transfer of personal data from the EU to a third country must ensure high protection in line with the EU's legal standards. In the context of the Privacy Shield Framework, the CJEU found that the deficiencies in US law, specifically regarding surveillance practices, rendered the framework invalid and unable to guarantee the protection of EU citizens' data.

Following the Schrems II decision, several European data protection authorities have declared that Google Analytics is not compliant with the GDPR's data transfer principle. A 2022 decision by the Austrian Data Protection Authority (Austria DSB) determined that the use of Google Analytics contravened the GDPR's data transfer principle by transmitting user data to the United States without ensuring sufficient protection (European Center for Digital Rights - NOYB, 2022, January). The conclusion accentuated the importance of performing risk assessments and establishing safeguards when employing third-party services like Google Analytics to process EU residents' personal data. Similarly, France's CNIL claimed that Google Analytics' usage also breached the GDPR's data transfer principle (CNIL, 2022). The Italian Data Protection Authority (GPDP) also banned the use of Google Analytics (European Center for Digital Rights - NOYB, 2022, July). Norway's Data Protection Authority also issues a preliminary decision on the Google Analytics case, recommending that websites look into alternatives to Google Analytics in this uncertain landscape (Datatilsynet, 2023). These decisions emphasize the need for organizations using Google Analytics to adhere to GDPR's data transfer requirements. Non-compliance can result in severe consequences, such as substantial fines and reputational harm. For example, the Spanish Data Protection Authority (AEPD) imposed an €8.15 million fine on Vodafone Spain for violating GDPR rules on data transfer (European Data Protection Board - EDPB, 2021).

While the use of Google Analytics is not declared illegal in the Netherlands yet, Autoriteit Persoonsgegevens (2022) announced that they investigated two websites of Dutch providers, and the report is expected to draw up soon, which will be followed by the investigation. European Commission (2022, March), on the other hand, also announced that a new Trans-Atlantic Data Privacy Framework was in the making, which can facilitate the data follows between the EU and America again and address the related concern surrounding the matter by the CJEU, after the Schrems II decision in July 2020. In this uncertain climate, sites have to make a choice regarding the tracking solution on their tools to prepare for the upcoming changes in the legal landscape.

1.2 The prevalence of third-party tracking plugins and Google Analytics

Despite the situation, third-party web tracking continues to be widespread. Degeling et al. (2019) conducted a study to evaluate third-party tracking prevalence and service providers' compliance with GDPR requirements. Their findings indicate that third-party web tracking is pervasive, and service providers' compliance remains limited. Many popular websites adopted privacy policies and cookie notices in line with GDPR, but their actual privacy practices were often insufficient (Degeling et al., 2019). These inadequacies in privacy practices imply that service providers may prioritize user experience and advertising revenue over GDPR compliance.

Google Analytics, an eminent web analytics service proffered by Google, equips organizations with the capacity to monitor and evaluate website traffic, generating invaluable insights for data-driven decision-making processes – Google's "Analytics Tools & Solutions for your business" (n.d.). As a testament to its ubiquity and efficacy, Google Analytics has emerged as the industry standard in web analytics, supported by a study

conducted by Jansen et al. (2022), which compared two prevalent analytics approaches using data from 86 websites. Schelter & Kunegis (2018) found that Google dominated the tracking industry, with ownership of the three tracking websites that had the largest share (namely google-analytics.com, google.com, and googleapis.com). The study's findings revealed that as a widely used tracking service, Google Analytics had a substantial presence in the web tracking ecosystem.

In 2020, Google unveiled Google Analytics 4 (GA4), a new rendition of the analytics platform that utilizes event-based data from both websites and apps, paving the way for the future of measurement. GA4 incorporates enhanced features that foster a more in-depth comprehension of the customer journey, utilizing event-based data rather than session-based data. Furthermore, it encompasses privacy controls such as cookieless measurement, behavioral and conversion modeling, and predictive capabilities without necessitating intricate models. GA4 also streamlines direct integrations with media platforms in Google's ecosystem, facilitating actions on websites or apps. Google Analytics 4 will be the only version of Analytics operated from July 1, 2023, as Universal Analytics (the predecessor of GA4) will cease processing data.

1.3 CNIL issues recommended measures to enhance user's privacy and support the compliance of use of third-party plugins

The French Data Protection Authority (CNIL) followed the complaints regarding the use of Google Analytics by French companies, leading to orders for compliance as the tool resulted in insufficiently regulated transfers to the US (CNIL 2022a). CNIL emphasizes that simply implementing standard contractual clauses or changing Google Analytics' settings is insufficient for GDPR compliance. Instead, they propose using a Proxy Server as a potential solution that avoids direct contact between users' devices and the analytics tool's servers. To be effective, the Proxy Server must implement a set of measures to limit data transfers and ensure data pseudonymization, such as not transferring the IP address, replacing User Identifiers, removing external Referrer information, reprocessing Browser Fingerprinting information, and deleting any data that could lead to re-identification.

Moreover, the Proxy Server must be hosted in conditions that prevent data transfers outside the EU to countries lacking adequate protection levels. The recommended measures stress that data controllers must conduct a thorough analysis to verify the Proxy Server's compliance with GDPR rules and maintain these measures over time, adjusting to product evolutions as needed.

1.4 The competing interests in implementing privacy measures, and the responsibility of developers and analytics specialists to implement them

Spiekermann (2019) highlights the difficulties of implementing principles in technology systems to balance citizens' privacy rights with the data needs of businesses and governments. Similarly, Tahaei & Vaniea (2021) indicate that based on the languages of the advertising network, third-party plugins imply that developers and IT Professionals who set up the plugins take responsibility for implementing privacy measures and ensuring compliance with data protection regulations in the software development process. However, their experiences and the influences they face can be highly diverse. Studies found that developers are usually aware of their responsibilities in maintaining user privacy but are under various influences from work, in which businesses, clients, and

other stakeholders can impact their ability to prioritize privacy (Alhazmi & Arachchilage, 2021; Bednar, 2019; Stöver et al., 2023). Spiekermann et al. (2019) pointed out that engineers frequently encounter organizational barriers, including time limitations and a lack of autonomy, challenging the development of ethical systems. Besides, organizational privacy and security norms were found to be weak or even contradictory to the value of privacy by design principles, resulting in conflicts between engineers and their organizations.

Various strategies can be employed to enhance user privacy. However, Google notes that settings regarding privacy measures can also compromise the tool's efficiency as disabling those functions will affect how certain features behave and, as a result, affect the business goals; this dilemma has led to conflicts in decision-making. As the European Commission is developing a new Trans-Atlantic Data Privacy Framework, Google Analytics in its standard configuration may not be allowed in the near future. Therefore, it is essential to study how analytics practitioners balance these competing interests in this landscape to understand the challenges better and provide recommendations for using the tool.

1.5 Scope of the research

1.5.1 Problem Statement

The increasing reliance on tools such as Google Analytics underscores the need for businesses to address conflicting goals of data utility and user privacy[1]. The recent legal challenges surrounding Google Analytics' adherence to GDPR principles, notably the data transfer principle, have raised significant concerns for developers and specialists involved in the implementation of such tools. The prevalence of third-party tracking plugins and the persistent and prevalent use of Google Analytics further complicate the decision-making processes for these stakeholders. Measures suggested by the legal authority, such as those proposed by CNIL (2022a), call for a range of measures to limit data transfers and ensure against possible re-identification of data subjects, placing the responsibility on developers and specialists to adopt strategies that would satisfy these requirements while preserving the functionality of the analytics tool. However, the practical implementation of these measures presents a riddle: How can the functionality of Google Analytics be maintained without infringing on privacy rights and regulations? Furthermore, advertising networks and third-party plugins imply that developers and IT Professionals bear the responsibility for implementing privacy measures and ensuring compliance with data protection regulations in the software development process.

1.5.2 Research question

In implementing Google Analytics, how do developers and analytics specialists strike a balance between user privacy and utility?

In this research

- By developers and analytics specialists, the researcher means the stakeholders involved with setting up and utilizing the tools to achieve certain business goals,

[1] According to Quach et al. (2022) in "Digital technologies: Tensions in privacy and data - journal of the Academy of Marketing Science", privacy is the right of individuals to control their personal information and interactions, and it can be divided into three types: information, communication, and individual privacy. The utility here can be defined as the benefit firms can gain from using digital technologies to collect, process, share, and monetize data. There is a trade-off between privacy and utility because firms' data strategies can create value for themselves and their customers but also pose risks and tensions for consumers' privacy. To balance these conflicting goals, firms, consumers, and regulators need to interact and cooperate within a system of rules and resources that can protect privacy while enabling data-driven innovation.

such as web engineers, site owners, product analysts, data analysts, and marketing specialists.

- The research focuses on the Netherlands.
- By "striking a balance", the researcher means choosing technical measures that align with the recommendation for privacy from the legal authorities and, on the other hand, allow Google Analytics to remain useful for tracking user behaviors.
- By "implementing" Google Analytics, the researcher means technical measures that analytics specialists can use to modify the behaviors of Google Analytics, for example, changing the data settings in Google Analytics' Admin Dashboard or using Google Tag Manager to fire events that change the behavior of collecting a certain type of data.
- To research the strategies and methods, the researcher formulates specific practices for Google Analytics based on the recommendations by the document directly related to privacy implementation, such as Google Analytics Help, CNIL measures, and Handleiding privacyvriendelijk instellen van Google Analytics.
- After formulating the set of practices, the researcher asks the developers/ analytics specialists about these practices and how they think about the impact of those recommended measures on the utility of the tools.

1.5.3 Aim of the research

The primary aim of this research is to investigate the strategies and technical methodologies employed by developers and specialists to balance user privacy and utility when implementing Google Analytics into their applications. This study seeks to understand the factors that influence developers and specialists when making decisions about the technical measures that protect user privacy while maintaining the usefulness of Google Analytics for achieving business goals. The insight would elicit the complex relationship between utility and privacy regarding each privacy setting and how certain adopted strategies would affect the utility or the functions of the tool to uncover the best practices when implementing Google Analytics.

On a societal level, the research aims to provide personnel and firms working with Google Analytics with an understanding of the set of tools and considerations when making an informed decision of what practices to be considered and their potential trade-offs between the two aspects. As Google Analytics dominates the third-party tracking plugin market, the practice suggested here can be generalized and applied to the market-wide level. This can prompt a broader shift in societal expectations and norms around privacy, and constructive dialogues between businesses, governments, and users.

By exploring the trade-offs between privacy and utility in the context of Google Analytics implementation, this research contributes to the broader understanding of the challenges stakeholders face in managing the tension between these two objectives. The insights derived from this investigation may inform best practices and guidelines that can help organizations optimize the implementation of Google Analytics, effectively addressing user privacy concerns without compromising the tool's utility. Furthermore, this study aims to examine the perceptions of developers and specialists on the impact of various privacy-preserving techniques on the utility of Google Analytics, shedding light on the practical challenges they face when attempting to strike a balance between user privacy and the tracking of user behaviors for business goals. By focusing on the experiences of a diverse range of stakeholders, this research offers a comprehensive

perspective on the factors that influence the delicate balance between privacy and utility in the implementation of Google Analytics.

1.6 Related work

Prior work has studied the balance between privacy and utility in Software Engineering in general, and the decision-making process when choosing to adopt certain third-party plugins, also the technical configurations of third-party plugins in light of privacy compliance:

The balance between privacy and utility in Software Engineering The balance between privacy and utility in software engineering has become increasingly important due to the growing concerns surrounding data protection and legal regulations, such as the General Data Protection Regulation (GDPR) (European Union, 2016). Peters (2018) surveyed data privacy solutions for software engineering data and found that combining data minimization and obfuscation techniques produced high levels of privacy while maintaining the usefulness of the data. The study highlights the importance of balancing privacy and utility to ensure compliance with regulations and protect sensitive information without compromising the quality of software engineering processes. Further research is needed to explore various privacy solutions and their potential impacts on different sub-disciplines of software engineering.

Consideration when it comes to the adoption of third-party plugins: Utz et al. (2022) carried out a survey to find out whether privacy is taken into account in the decision-making process of IT professionals, to what extent they are aware of the privacy implication and the effort they put into it. The authors conclude that ease of integration and the popularity of the plugins play a notable role in how they choose the service. Legal requirements or guidelines are also the drives behind the consideration of user privacy, and the data collection awareness corresponds to how the plugin is used for.

Sneaky technologies adopted to track users without proper their consent: Papadogiannakis et al. (2022) in "User Tracking in the Post-cookie Era: How Websites Bypass GDPR Consent to Track Users" discovered that websites do use modern forms of tracking even before users had the opportunity to register their choice with respect to cookies, or when users chose to reject all cookies. The authors argue that websites do not respect user choices and use sophisticated forms of tracking to bypass consent mechanisms known as first-party ID leaking, ID synchronization, and Browser Fingerprinting. They also suggested some possible solutions, such as browser extensions, auditing tools, and regulatory actions, to enhance user privacy and enforce GDPR for those new technologies.

Technical implementation that protects user's privacy when using third-party plugins: Mayer & Mitchell (2012) explored the policy and technology aspects of third-party web tracking in early 2012. The authors delved into the ongoing policy discussion and offered explanations for the associated technology. They introduced the Fourth Party web measurement platform and shared findings from their research. The paper's primary objective was to equip the researcher with the necessary background and tools to contribute to the broader understanding and policy discourse on web tracking.

The relationship between technology and privacy Kröger (2022) noted that no technological solution could guarantee privacy through the test of time, it may be perceived as secure enough at one given time, but new techniques are introduced and render the solution outdated. The paper also mentioned that using technological methods usually leads to several implications, including the compromise of usability. When law and technology cannot present a reliable solution, the author suggests that transparency about practices and business can lead to a more comprehensive measure to

address the issue.

Developers' decision-making process toward privacy attitudes and practices: Ayalon et al. (2017) studied the effects on developers' professional privacy attitudes and practices, including organizational, professional, and personal factors. The organizational privacy climate, defined as participants' perceptions of how their organization refers to privacy, was found to have a greater impact on developers' privacy practices than the legal background. This finding highlights the importance of informal aspects of organizational privacy conduct and indicates that the climate mediates the legal and business environments in which the organization operates (Ayalon et al., 2017). Bednar et al. (2019) explore the attitudes of senior engineers towards privacy and their perceived control over and responsibility for privacy implementations and present three core findings: Engineers often expressed that privacy is a vague concept, and its value is uncertain, context-dependent, and not absolute, time-consuming and technically challenging. Second, the engineers have an inner conflict between recognizing the importance of privacy and feeling that they do not have the autonomy or resources to address privacy concerns adequately. Lastly, engineers often struggle with lawyers, finding it difficult to reach a shared level of understanding of privacy regulations. Tahaei et al. (2023) found that both developers and end-users shared concerns about unnecessary permissions undermining trust, harming the app's reputation, and potentially granting access to sensitive data. Developer participants sometimes requested multiple permissions due to confusion about the scope of certain permissions or requirements imposed by third-party libraries. Moreover, developers also believed the end-user was responsible for granting permission requests. Tahaei et al. (2023) emphasized that app functionality and features were the primary reasons for developers to include permissions and for end-users to grant them.

1.7 Contribution to prior work

Building upon the existing body of literature, this research extends our understanding of the balance between privacy and utility in software engineering, specifically in the context of Google Analytics implementation, which has not been previously explored. By focusing on the recommendations from legal authorities regarding the measures that can enhance the privacy of the user and their practical implications, this study provides a concrete example of the trade-offs developers and specialists face when attempting to maintain adherence to the principle of privacy and data protection while maximizing the utility of Google Analytics.

In contrast to prior work that investigated the considerations when it comes to the adoption of third-party plugins or general privacy practices in software engineering, this research focuses on the measures for balancing after Google Analytics is adopted and evaluating the impact of these technical privacy safeguards on the utility of the tracking plugin, this research contributes valuable insights into the concerns that should be kept in mind in relation to both their practical application and their functional limitations include as suggested by Kröger et al. (2022).

2 Literature research

This section reviews the practices suggested by the CNIL recommended measures in light of Google Analytics. It starts by introducing details of what data Google Analytics collects and detailing Google's advice on the best practices to avoid non-compliance with sending Personal Data. By presenting this information first, in the discussion of each CNIL recommended measure, we can see the competing interest of privacy and utility that presents the challenge for the implementation.

Additionally, it provides a list of measures that Google Analytics supports to control the data sent to Google Analytics. This information plays an important part in arguing and formulating the practice that bridges the balance gap between utility and privacy in each CNIL recommended measure. Next, the research aims to discuss the CNIL's recommended measures regarding data collection by explaining how this information is used in Google Analytics and the impact on the usage of Google Analytics if the collection of data is redacted or altered and using the information in this section combined with the information from Section 2.1 to 2.3, the researcher provides the solution that can address the gap between utility and privacy. Finally, In Chapter 2.4, the researcher will formulate the practices that developers and specialists can employ to meet the privacy requirements but still allow Google Analytics to be usable, which becomes the foundation for creating the interview questions.

2.1 What data does Google Analytics collect?

While Google started not collecting IP Addresses in Google Analytics 4, Google still collects other information that CNIL suggested removing or replacing to avoid non-compliance. According to "[ga4] data collection - analytics help", Google (2023, June) gathers the following data by default:

- Number of users
- Session statistics
- Approximate geolocation
- Browser and device information

The full list of user properties includes:

- **Age** (app, web): The age of the user by bracket: 18-24, 25-34, 35-44, 45-54, 55-64, and 65+.
- **App store** (app): The store from which the app was downloaded and installed.
- **App version** (app): The version Name (Android) or the Bundle version (iOS).
- **Browser** (web): The browser from which user activity originated.
- **City** (app, web): The city from which user activity originated.
- **Continent** (app, web): The continent from which user activity originated.
- **Country** (app, web): The country from which user activity originated.
- **Device brand** (app, web): The brand name of the mobile device (such as Motorola, LG, or Samsung).
- **Device category** (app, web): The category of the mobile device (such as mobile or tablet).
- **Device model** (app): The mobile device model name (such as iPhone 5s or SM-J500M).
- **Gender** (app, web): The gender of the user (male or female).
- **Interests** (app, web): The interests of the user (such as Arts & Entertainment, Games, Sports).

- **Language** (app, web): The language setting of the device OS (such as en-us or pt-br).
- **New/Established** (app):
  - New: First opened the app within the last 7 days.
  - Established: First opened the app more than 7 days ago.
- **Operating system** (app, web): The operating system used by visitors to your website or mobile app.
- **OS version** (app, web): The operating system version used by visitors to your website or mobile app (such as 9.3.2 or 5.1.1).
- **Platform** (app, web): The platform on which your website or mobile app ran (such as web, iOS, or Android).
- **Region** (app, web): The geographic region from which user activity originated.
- **Subcontinent** (app, web): The subcontinent from which user activity originated.

The list of default user properties collected by Google Analytics is taken from the documentation "[ga4] data collection - analytics help". For websites, Google Analytics stores a Client ID in a first-party cookie called _ga to differentiate unique users and their sessions. However, the Client ID is not stored when analytics storage is disabled through Consent Mode (which is explained in Chapter 2.4).

2.2 Google Analytics' measures that can assist the compliance with the CNIL requirements and address the concerns of user privacy

As concerns regarding user privacy and data protection continue to rise, both businesses and industry regulators are focusing on establishing higher standards for these practices. To address this need, Google Analytics and Google's tracking solutions offer various data controls that allow businesses to govern how data is collected, stored, and used in "[GA4] Google Analytics Data Controls Guide - analytics help" (Google, n.d.), which can help alter or remove the collection of data required by the CNIL. A key component of data control within Google Analytics is the use of Google Tag Manager (GTM). Google (n.d.) defines GTM in "Google's Tag Manager overview – tag manager help" as a versatile tool that facilitates the implementation of tags and the management of data collection and usage, allowing businesses to comply with data protection regulations. Google Tag Manager (GTM) is a comprehensive tag management system (TMS) that enables users to efficiently update measurement codes and related code fragments, known as tags, on their websites or mobile applications.

A key feature mentioned in the "Data controls in Google Analytics 4 - analytics help" is Consent Mode, which allows businesses to adjust how Google tags behave based on the consent status of their users. For instance, when consent for advertising storage or analytics storage is denied, Google Analytics adjusts its behavior accordingly, such as not reading or writing first-party analytics cookies when the parameter "analytics_storage" is denied. In "Consent Mode on websites and mobile apps – analytics help" by Google (n.d.), Consent Mode is noted to be an essential feature for website and app owners who utilize cookie consent banners or widgets to manage user consent. It allows communication of users' consent status to Google, enabling tags to adjust their behavior accordingly while respecting users' choices. This research examines the implementation

of Consent Mode in Google Analytics and explores its implications for data collection and user privacy.

Consent Mode interacts with Consent Management Platforms (CMPs) or custom implementations of cookie consent banners to obtain user consent choices. Google products like Google Analytics, Ads, and third-party tags will adapt their behaviors to the Consent Mode accordingly and ensure that the cookies are not stored when consent is denied. Instead, tags send cookieless pings to Google, providing minimal information about user activity. This mechanism allows Google Analytics 4 to fill data collection gaps through conversion modeling and behavioral modeling.

When Consent Mode is enabled, Google measurement products ensure the preservation of a visitor's consent state across pages. For denied consent, tags that fire do not store cookies but instead send cookieless pings to the Google server. These pings include practical information, such as timestamps, user agent details (web only), and Referrer data, along with aggregate/non-identifying information like the presence of ad-click information in the URL and boolean indicators of consent state. Additionally, random numbers generated on each page load and information about the consent platform used by the site owner are also included in the pings. The use of Consent Mode in Google Analytics provides website and app owners with a mechanism to respect user consent choices and adjust data collection practices accordingly. Google Analytics maintains a balance between data collection and user privacy by employing cookieless pings and preserving consent states across pages.

Other data controls

There are several other data types available within GA4 that users can control. Each data type has its own native analytics data controls and Server-side tracking controls, along with the potential impact if the data type is altered or redacted.

- Client ID: Users have control over the Client ID value used by Google Analytics, which can be modified or removed through Server-side Google Tag Manager (GTM) with a custom variable and sandboxed JavaScript. If the Client ID is different between the cookie value and the value used in Google Analytics, audience remarketing functionality may be affected.

- Advertising Identifiers: When Google signals are enabled, Google Analytics collects visitation information and associates it with Google information from accounts of signed-in users who have consented to this association. Controls are available at the property and regional level with Google Signal's admin settings on/off toggle and at the user level with gtag.js function GTM template option.

- User ID: A User ID is a unique, persistent, and non-personally identifiable ID string that represents a user. Users have control over the User ID value that Google Analytics will use, and it can be transmitted to Server-side GTM and then utilized by Server-side GA4 tags once the User ID is configured in the web container.

- Granular Location & Device: Users have control over the granular location and device data that is collected about their visitors. When the collection is disabled, city-level location data and certain device-level metadata are redacted prior to collection in Google Analytics servers.

- Referrer and URL Parameters: Users have control over the page Referrer and URL parameter values that Google Analytics uses. These values can be modified or redacted in Server-side GTM with a custom variable and sandboxed JavaScript.

Google also supports data deletion through Data Deletion Requests, where users can issue a request for the removal of data from the Analytics servers. Additionally, users

can delete a single user's data from Google Analytics by passing a single User Identifier to the Google Analytics User Deletion API or via the User Explorer report Google's "Tag manager overview - tag manager help" (2023, June).

2.3 Discussing the CNIL recommended measures on how to maintain the balance between Privacy and Utility when implementing Google Analytics

According to CNIL (2022a), when the Privacy Shield was invalidated in 2020, it meant that there were not enough guarantees of data protection to the citizens in the EU, which may be used by other parties such as the authorities and intelligence services. Google Analytics, published in the US, was ruled insufficient to regulate data transferred between the EU and the US. CNIL stated that the IP address while being modified, was still transferred to the US. Other changes regarding the identifiers of the users may not guarantee the chance the user was not re-identified due to the consistent processing of IP addresses. When a person was re-identified by Google Analytics, CNIL stated that it might lead to the reveal of their browsing history on the sites using Google Analytics. CNIL offered a Proxy Server as a possible measure to be a barrier between the direct contact between data subjects when maintaining the use of Google Analytics, which was also followed by other measures to make sure the user will not be re-identified. These measures will be discussed in this chapter.

2.3.1 The use of Proxy Sever

According to Kurose & Ross (2017, p. 110), "A Web cache—also called a Proxy Server—is a network entity that satisfies HTTP requests on behalf of an origin Web server. The Web cache has its own disk storage and keeps copies of recently requested objects in this storage. [...]. A user's browser can be configured so that all of the user's HTTP requests are first directed to the Web cache. Once a browser is configured, each browser request for an object is first directed to the Web cache". Hence, Proxy Servers act as an intermediary between the client and the destination server. When a client sends a request through a Proxy Server, the Proxy Server forwards the request to the destination server on behalf of the client. As a result, the destination server receives the request from the Proxy Server rather than directly from the client. This process, as a result, prevents Google Analytics from having direct contacts between users and Google Analytics' server as mandated by CNIL.

On the other hand, CNIL (2022a) stresses that as a proxy is also considered to be a data processor, it must be hosted in a condition that ensures the data it processes will not be transferred outside the European Union to a country with inadequate data protection measures compared to the European Economic Area. Furthermore, it is also mentioned the responsibility of data controllers is to conduct a comprehensive analysis of the hosting conditions and implement necessary measures in case they opt for Proxy Server solutions. This analysis should account for the maintenance and monitoring of these measures over time, considering the dynamic nature of digital products and services. CNIL also says it is necessary, in principle, to implement further extra measures (which are listed in Chapters 2.5.2 to 2.5.6 to protect the user's identity. However, CNIL also notes that these extra measures listed below can be costly and may compromise the operational use of the tracking feature.

Server-side tracking, an alternative to Proxy Server

In "An introduction to Server-side tagging — Google tag manager - Server-side — google developers" by Google (n.d.), it is noted that Server-side tracking represents a significant shift in how data is gathered and processed in comparison to the traditional

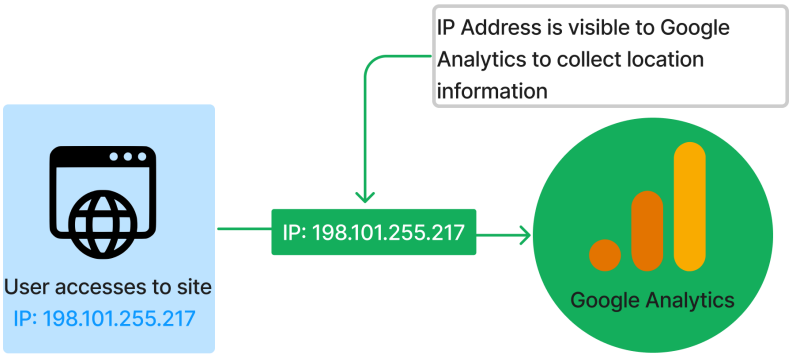

Figure 1: This is Google Analytics in the default Client-side mode. Typically, when a user accesses a website that integrates with Google Analytics, the user's IP is visible to the website and Google Analytics when making a direct request. Hence, Google Analytics can directly use the visible IP from the user to perform IP Address Lookup, which is to determine the location of the user (original figure)

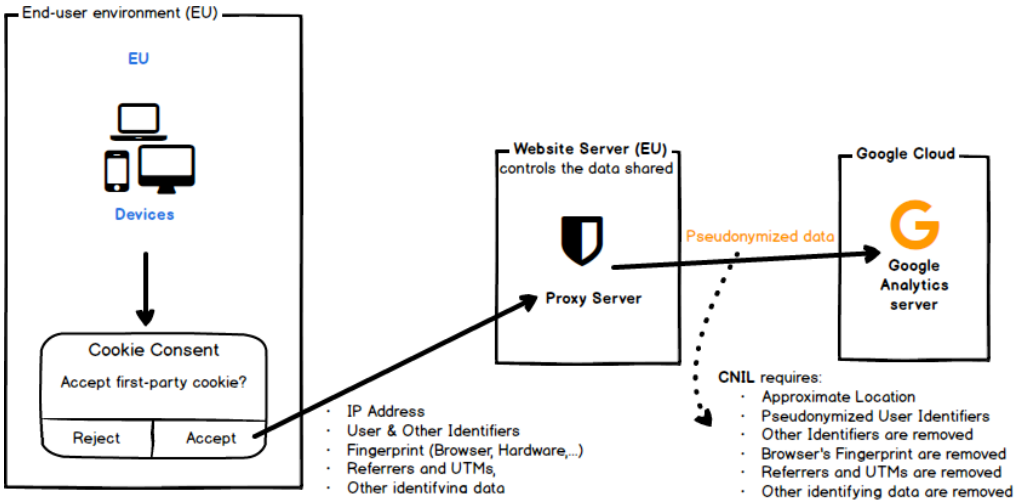

Figure 2: This is Google Analytics in the context of Proxy Server and Client-side tracking. The Proxy Server is used to send pseudonymized data to Google Analytics: If the user accepts cookie consent, the Proxy Server will process the tracking information and send it to Google Cloud instead; this prevents Google Analytics from having direct contact with the user's IP Address (original figure)

Client-side method, offering potential benefits in the areas of data privacy and utility. Unlike Client-side tracking, where data is sent directly from the user's browser to various collection servers, Server-side tracking moves this process to a server the organization controls.

According to "Why and when to use Server-side tagging? — Server-side tagging fundamentals — google developers" by Google (n.d.), Server-side tracking is a progressive method of data collection and processing that demonstrates significant benefits in areas of data privacy, website performance, and data quality. The advent of Server-side tracking has dramatically changed the landscape of data collection, introducing a new level of control and flexibility.

Unlike Client-side tracking, Server-side tracking inserts an additional layer of control between the user and the Google Cloud, thus offering improved privacy controls. This layer regulates the composition of data dispatched to vendors, thereby enhancing user privacy. Specifically, it allows for the removal of personal data, such as the user's IP address, from the data sent to the vendor. Additionally, cookies can be set on your domain, making them more secure and durable, unlike in Client-side tracking where they are set by JavaScript on the page.

Moreover, Server-side tracking has shown potential in optimizing website and app performance. It significantly reduces the amount of third-party code loaded in the user's browser, improving page speed. Further, Server-side tracking reduces the user's browser and device load as only a single stream of data needs to be sent to the server container, unlike in Client-side tracking where multiple almost identical requests are dispatched to different vendors. This can significantly reduce performance bottlenecks, thus enhancing the user experience.

Furthermore, Server-side tracking improves data quality. By moving data processing away from the client and into the server, mechanisms to enhance data quality can be leveraged. Server-side processes happen outside the user's browser, enabling the enrichment of data with information that should not be exposed to the browser, such as API secrets, business-sensitive data, and user data. Moreover, using custom templates in the Server-side tracking environment facilitates the normalization of data collected and processed by the server.

Potential practices employed to strike a balance

Server-side tracking offers a potential alternative to using a Proxy Server in the context of Google Analytics, providing a more privacy-compliant approach to data collection and handling. With Server-side tracking, data from a user's device is sent to a secure server before being transmitted to third-party platforms. This server acts as an intermediary, controlling which information is shared with specific platforms, thus providing the capability to decide on a case-by-case basis.

When using Google Analytics in conjunction with Server-side tracking, the analytics data is not automatically sent to Google's servers, as in traditional Client-side tracking. Instead, the data can be sent to a server located in a region that complies with data privacy regulations, such as the General Data Protection Regulation (GDPR) in Europe. This server can then modify the data, anonymizing or pseudonymizing it, before it is sent to Google Analytics. This ensures that no personal customer data is transferred to regions that may not provide adequate data privacy protections, such as the United States.

This same concept has been highlighted by data protection authorities, such as the CNIL, as a viable solution for using Google Analytics in a GDPR-compliant way. Although implementing Server-side tracking can be more time-consuming than traditional Google Analytics usage due to the need to maintain both the server and the data adjustments, Google provides support. For instance, Google allows users to set up and manage the server through their platform and decide on the server's location, providing

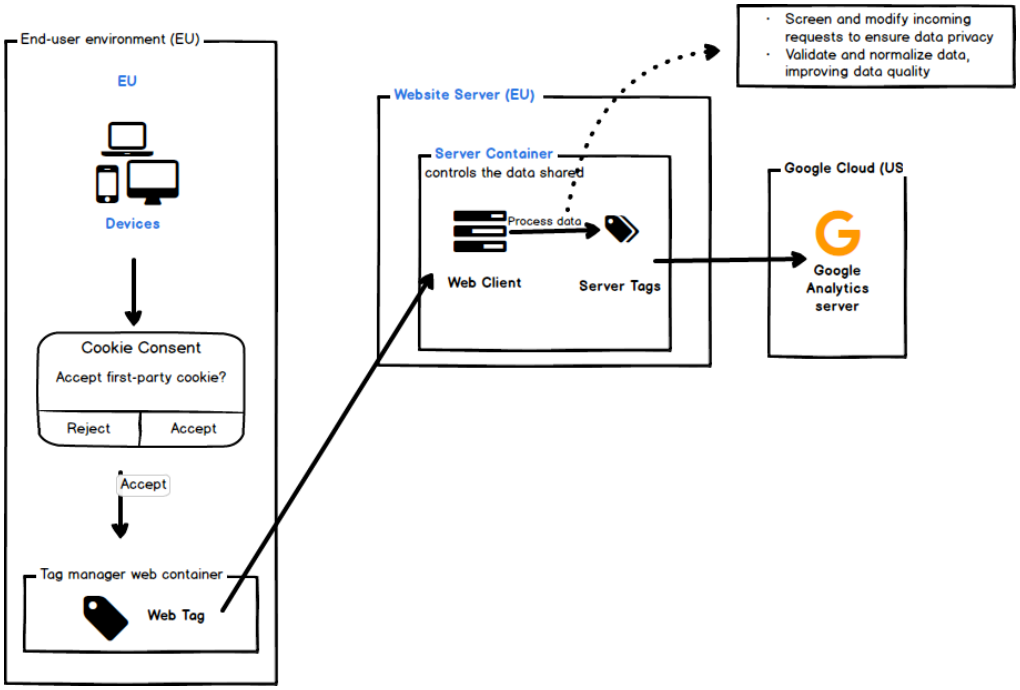

Figure 3: Google Analytics in the context of Server-side tracking. Server-side tracking can be an alternative to the Proxy Server that is easier to implement. By using Server-side tracking to send the data to Google through server side, Google Analytics is not directly in contact with the user and hence prevents personal data such as IP Address from being transferred out of the EU (original figure)

additional control over data privacy.

2.3.2 Stopping the transfer of IP addresses in Google Analytics

An IP address is a unique numerical identifier assigned to each device connected to a computer network using the Internet Protocol for communication. This hierarchical address, consisting of four bytes, serves as a means to route packets between source and destination end systems in the network. Much like postal addresses, IP addresses provide increasingly specific information about the host's location within the network as they are examined from left to right (Kurose & Ross, 2016)

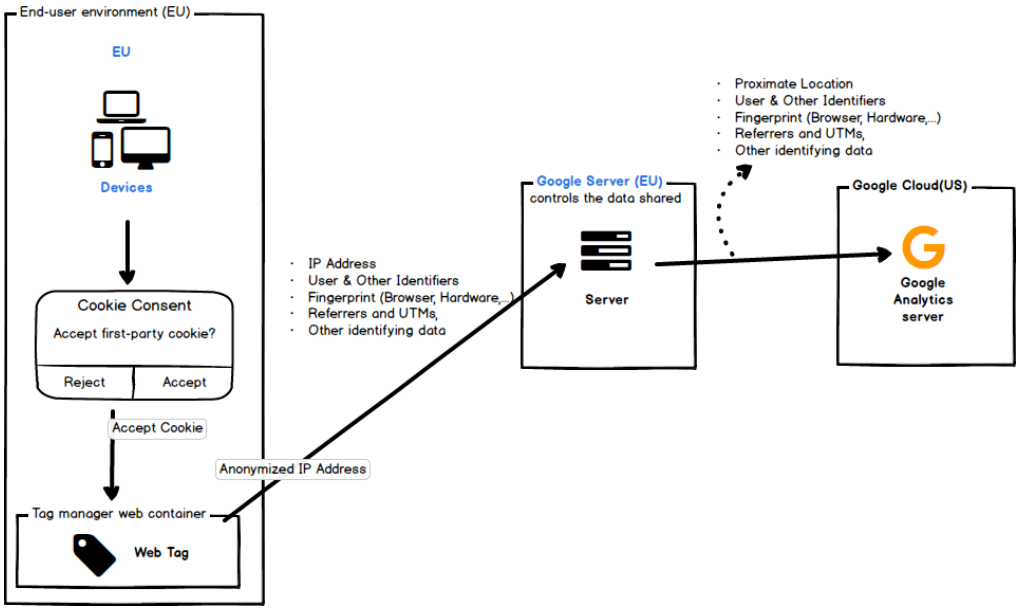

Figure 4: From Google Analytics 4, regarding the EU region, it no longer sends the IP Address to the GA server outside the EU. Instead, the IP Address is only processed (within the EU) to retrieve the proximate location data from the user, and only the location data is sent to the Google Analytics server outside Europe (original figure)

In the recommended measures by CNIL (2022a), it states that "If a location is transmitted to the servers of the measurement tool, it must be carried out by the Proxy Server and the level of precision must ensure that this information does not allow the person to be re-identified (for example, by using a geographical mesh ensuring a minimum number of Internet users per cell);". In the case of Google Analytics 4, IP addresses are utilized at the time of collection to ascertain location information, such as country, city, and geographical coordinates. Google (n.d.) claims in "About Geographical Data – Analytics Help" that the location data derived from the user IP is only approximate. Furthermore, according to the article "IP masking in Universal Analytics – Analytics Help" by Google (n.d.), in Google Analytics 4, IP addresses are only used to collect geographical data, and be discarded after that (IP data is not logged or stored in any data center or server) so the implementation of this strategy may be redundant.

Potential practices employed to strike a balance

In "EU-focused data and privacy – Analytics Help", Google (n.d.) offers various measures to help ensure compliance with the relevant EU regulation; Google states that the processing of IP addresses and collection of geolocation data will be different within the EU domains. Firstly, the IP address data is used exclusively to look up the EU-

based servers and derive approximate geolocation data (which includes the latitude/ longitude of the city, continent, country, and subcontinent, in this case, can "ensure a minimum number of Internet users per cell" as required by the CNIL). The IP address is discarded immediately after that. Google further stresses for the EU-based traffic that Google Analytics 4 (GA4) performs all the collection through domains and servers based in the EU based on the information, then the analytics data is forwarded to the Analytics server for processing. This means that in Google Analytics 4, IP addresses, even in pseudonymized form, will not be transferred outside the EU and will be solely used to derive other information. This information adheres to the legal requirement that IP must not be transferred outside the EU and ensure the level of geolocation precision of the user. However, according to The Danish Data Protection Agency (Denmark DPA): "In regard to Google Analytics 4, it is apparent from Google's documentation that IP addresses are used to determine the approximate location of the visitor, after which the address is discarded before the data is logged to a server. As with Universal Analytics, the same issue is also relevant for Google Analytics 4, as – depending on the location of the data subject – there can be a direct connection to, among others, American servers before the address is discarded" (Datatilsynet, 2022), which means Google Analytics approaches to IP address is not holistic enough to cover all cases. In this case, the best way to avoid sending personal data, in this case, is to either use Google Server-side tracking or implementation of a Proxy Server within the EU to send the location data instead, considering the cost of maintenance and difficulty in implementing the Proxy Server as noted by the CNIL recommended measures, Server-side tracking may be a more optimal solution.

2.3.3 The replacement of the User, Cross-site and Lasting identifiers in Google Analytics

In the "Measurement protocol parameter reference — analytics measurement protocol — google developers", Google (n.d.) distinguishes the two identifiers: Client ID and User ID, when a user is in a session. When the above identifiers are used to unify that user across different sites owned by the same web owner, they are known as Cross-site ID. The concept for each identifier is explained below:

Client ID

The Client ID plays an essential role in distinguishing individual users and tracking their behavior on a website, allowing web admins and marketers to gain valuable insights into user engagement, navigation patterns, and overall site performance (Weber, 2015). To achieve this, Google Analytics relies on a Client ID, which is stored in a first-party cookie named **_ga** - Google's " [GA4] Data collection – Analytics Help" (n.d.). Consent Mode is also introduced, which allows users to disable analytics storage, including the storage of the Client ID in the _ga cookie.

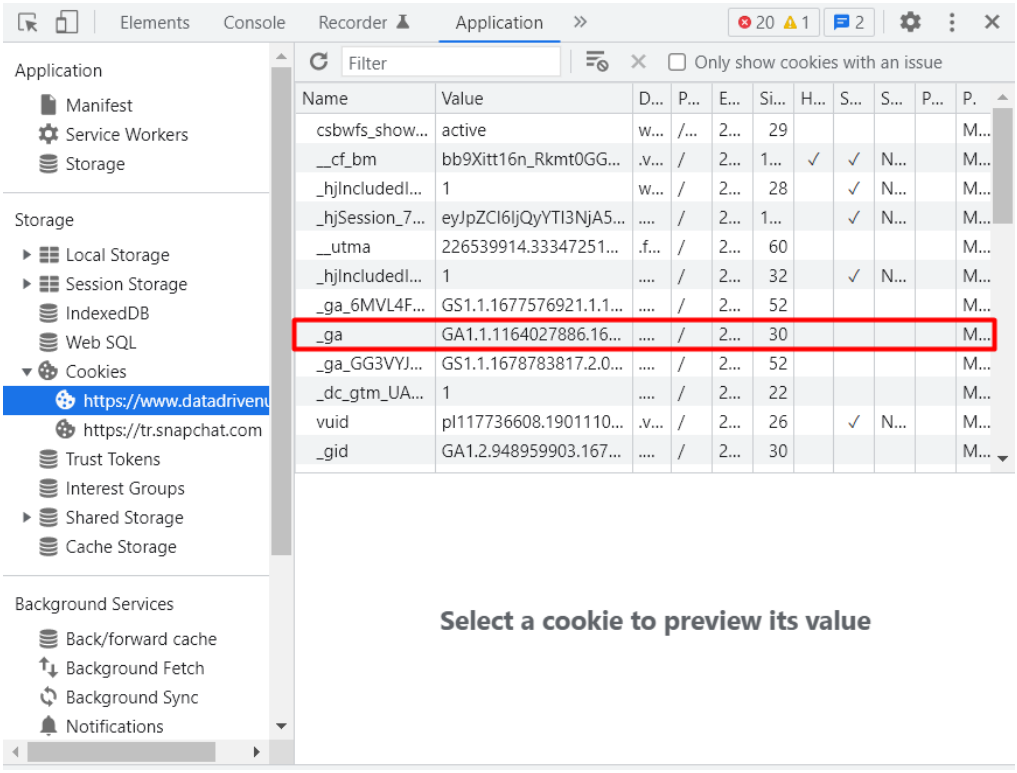

Figure 5: A Client ID was created when browsing the site datadrivenu.com

Thus, a visitor who lands receives a cookie containing the Client ID. Google Analytics checks to see if the cookie containing the Client ID is present in the browser when the same visitor navigates from the landing page to another page on the same website. If the answer is yes, Google Analytics recognizes that the visitor/device was on a previous page; as a result, two page views are combined and attributed to the same visitor.

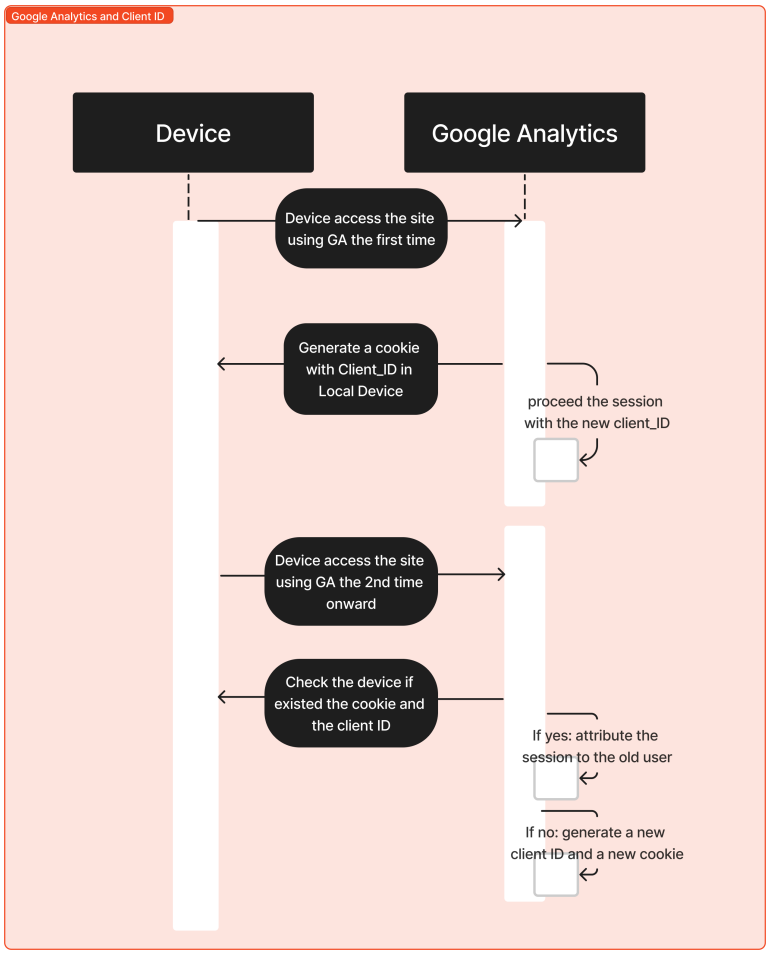

Figure 6: Client ID and Google Analytics' flow: Google Analytics uses Client ID inside the placed cookie in the user's browser to link users when they visit multiple times (original figure)

However, Weber (2015) noted that the Client ID is tied to a specific browser and device rather than an individual person. This means that users who browse a website using different devices or browsers will be counted as separate users in Google Analytics. Furthermore, users who delete their browser cookies or use private browsing modes may also impact the accuracy of the user metric and user behavior data, as Google will have to re-generate a new cookie with a new Client ID.

This type of behavior creates a problem because it is constrained by technical features of how browsers are made to function. Google Analytics will not be able to tell that a user is the same if they navigate the website or app using different browsers or devices. Every browser and device has a unique Client ID, and User ID is introduced to identify individuals rather than clients.

User ID

According to "[GA4] Measure activity across platforms with User ID - Analytics Help" by Google (n.d.), User ID feature is an intricate solution that helps businesses to thoroughly understand user behavior across diverse sessions, devices, and platforms.

By associating unique identifiers with individual users, companies can obtain highly accurate user counts and further uncover detailed insights into user interactions with their web service.

To implement the User ID feature, businesses must generate unique IDs for each user and consistently assign them, typically during the login process. This ensures that user behavior can be accurately tracked across multiple platforms and devices, creating a holistic view of user interactions. However, it is imperative to avoid using personal data or information that could be used by a third party to determine a user's identity. Utilizing the User ID provides many extra features for businesses:

- Session Unification: This User ID setting allows hits collected before a User ID is assigned to be associated with the ID. It helps associate user activities and devices, connect seemingly independent data points, and understand user interactions with the business more holistically.
- Comparing signed-in and non-signed-in users: The User ID feature enables businesses to compare the behavior of users signed in to those not signed in. By building a comparison that uses the "Signed in with User ID" dimension, businesses can evaluate user engagement, new users, engagement time, and revenue for both groups.
- User Exploration: This feature allows businesses to view detailed information about individual users, including acquisition, summary metrics, and a timeline of activities on the website or app. This information can be used to identify user preferences, patterns, and potential areas of improvement.
- Remarketing Audiences: User ID data can be used to create remarketing audiences. When Google Analytics and Ads accounts are linked, these audiences are available in the shared library in Google Ads, helping businesses develop more targeted advertising campaigns.
- When examining sessions with incomplete User ID collection, Google Analytics employs advanced techniques to associate Session IDs with User IDs. This ensures that user behavior is accurately represented, even when users trigger events before or after signing out. This level of detail and complexity demonstrates the robust nature of the User ID feature

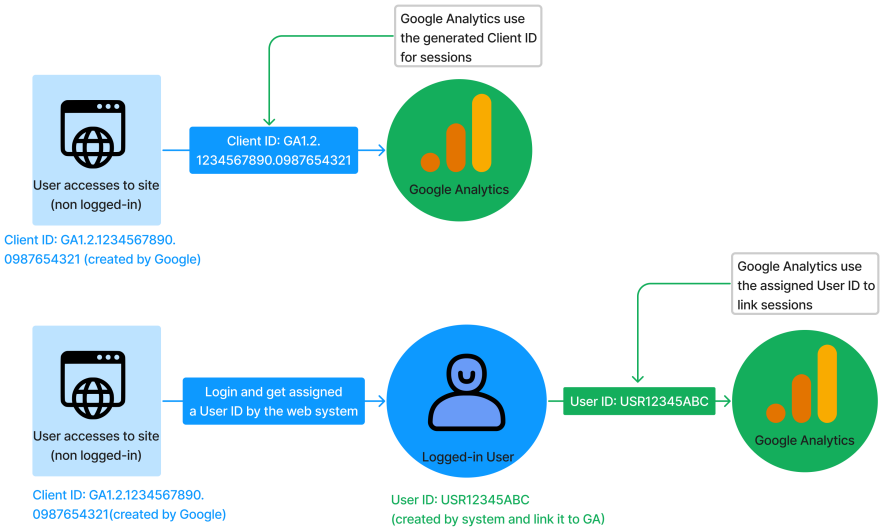

Figure 7: The use of User ID in Google Analytics 4: By using system-generated ID for users, it helps better understand behaviors across sessions, devices, and platforms. It also helps capture more accurate user counts and more detailed insights into user (original figure)

Cross-site or Lasting Identifiers

Cross-domain measurement is an essential aspect of Google Analytics 4 (GA4), enabling website owners to obtain a unified measurement across multiple domains, such as a custom website and a separate shopping cart domain - Google's "[ga4] set up cross-domain measurement - analytics help" (n.d.). This feature is particularly relevant for researchers and marketers who need to track user activity accurately as users navigate between different domains. Additionally, cross-domain measurement permits the precise attribution of user activity across domains. Outbound clicks that would typically generate an event through enhanced measurement are disregarded when the outbound link directs to a domain included in the cross-domain measurement configuration.

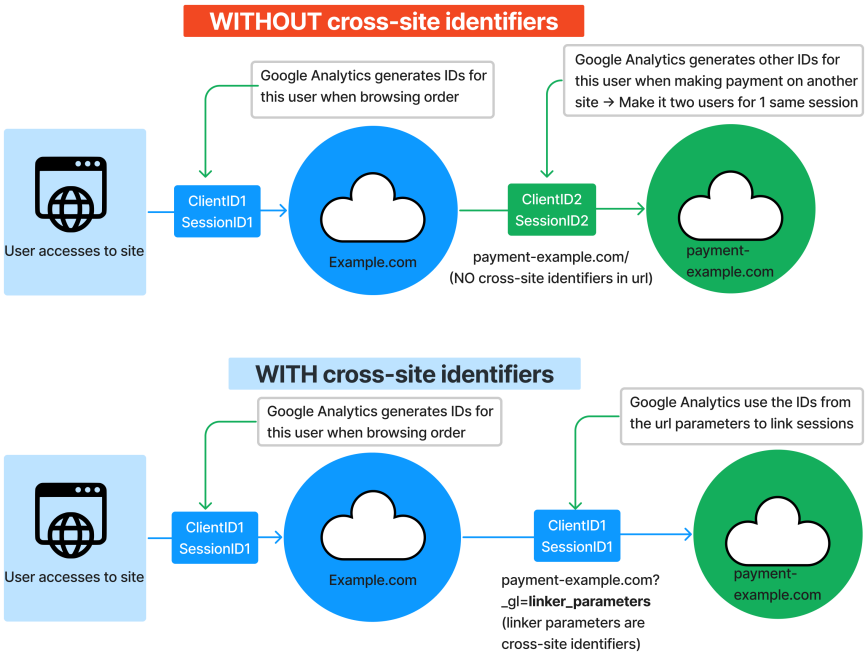

Figure 8: The difference in the use of Client ID and Session ID with and without the use of Cross-site Identifiers: With Cross-site Identifiers set up, the same user can be unified when they visit across multiple domains (original figure)

As required by the CNIL (2022a), the identifiers for users need to be replaced so that Google cannot recognize the returning user when they revisit the site. In the "Google Analytics Data Controls Guide - Analytics Help" Google (n.d.) stresses the implication of Analytics functionality when there is an alternation or disabling of the identifiers, including Client ID and User ID, which may reduce the accuracy and utility of the data collected. When the Client ID is absent, the audience marketing functionality will be rendered unusable, and the modification of such ID can lead to decreased accuracy in the user and visitor counts and ultimately cause potential misinterpretation of the data. Furthermore, removing the Client ID may cause Google Analytics to identify multiple users as a single user; this subsequently deflates the user counts. In contrast, changing the Client ID could lead to the recognition of a single user as multiple users, inflating user counts and, therefore, compromising the integrity of the data. In the case of User ID, Google notes that when this information is not collected, it results in a range of consequences. Firstly, User ID-based reporting when the Reporting Identity is set to Blended Views will be missing. Secondly, inflated user counts due to the unavailability of a unique identifier, as the same visitor on different browsers or devices would be recognized and reported as new unique users. Finally, an absence of cross-device or cross-platform advertising personalization.

Potential practices employed to strike a balance

Autoriteit Persoonsgegevens (2022) in a guideline "Handleiding privacyvriendelijk instellen van Google Analytics" advises that the use of User ID to comply with GDPR needs to obtain prior consent from users before implementing User ID and the use

of identifiers across sessions and devices. It is also noted that while the placing of Google Analytics cookies can happen without prior consent, the site must inform the visitors about this information, which includes: the use of Google Analytics cookies, the processing agreement that the site has made with Google, the use regarding IP; and the site has turned off "sharing data" that will not allow the combination of Google Analytics with other Google services. Additionally, Autoriteit Persoonsgegevens also advises that the site provide the visitor with an opt-out option for Google Analytics, which can facilitate the Consent Mode suggested in "Google Analytics Data Controls Guide – Analytics Help", that stops Google Analytics from placing the cookies and further collecting data from this traffic. Besides, following the advice of Google in "[GA4] Measure activity across platforms with User ID – Analytics Help", the site should not create a custom User ID field, and the naming convention for User ID must not include information that can lead to a user's identity being determined, this practice can minimize the risk of having the personal user data sent to Google Server and breach Article 44 of GDPR – General principle for transfers.

2.3.4 Removing Referrer information and URL parameters in Google Analytics

Referrers play a pivotal role in understanding user navigation patterns, analyzing website traffic sources, and enhancing overall user experience. Burby et al. (2007) define Referrer as the URL of the page responsible for generating the request for the current page view or object; it can also be used to identify the origin of a visitor and who referred them to the website. However, it is worth noting that the Referrer value might be empty or null in some instances. Additionally, Referrer URLs can hold valuable information, such as the content viewed or the searched keyword, further emphasizing the significance of Referrers in web analytics. MDN (n.d.) characterizes Referrers in "HTTP headers" as essential in gathering data for analytics, logging, and optimized caching, as it contains the address of the page from which a resource has been requested, allowing servers to identify the pages that users visit or the requested resources. For example, when a user clicks on a link, the Referrer header contains the address of the page containing that link that the user clicks to. Suppose the resource request is made to another domain. The Referrer header may contain an origin, path, and query string but typically excludes URL fragments or username and password information. The data included in the Referrer header is determined by the request's Referrer policy, which can be set to values such as "origin" or "origin-when-cross-origin."

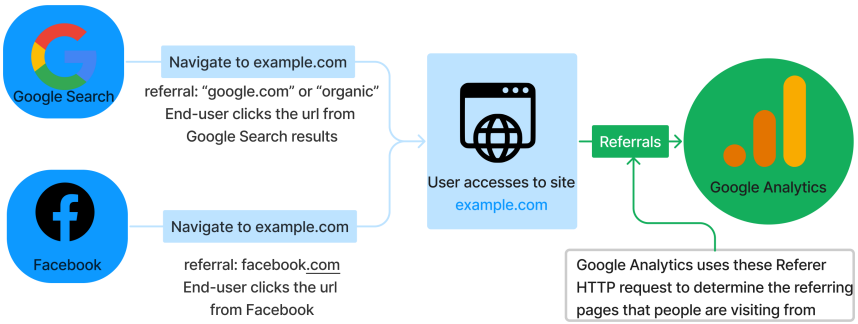

Figure 9: Referrers are used to determine the traffic source of the user when they land on the site, which is crucial for marketers to allocate better resources in their marketing strategies (original figure)

Clifton (2012) highlighted that understanding Referrer information can provide valuable insights into the origins of website traffic. For instance, an e-commerce site can analyze Referrer data using tools like Google Analytics to identify the primary sources of their customers, such as specific social networking sites or search engine results, which in turn enables businesses to allocate resources more effectively towards marketing channels that maximize website traffic and revenue. Additionally, Referrer data allows website administrators to study user journeys across multiple pages on their site, which can contribute to content and user experience improvements, ultimately resulting in increased user engagement and transactions.

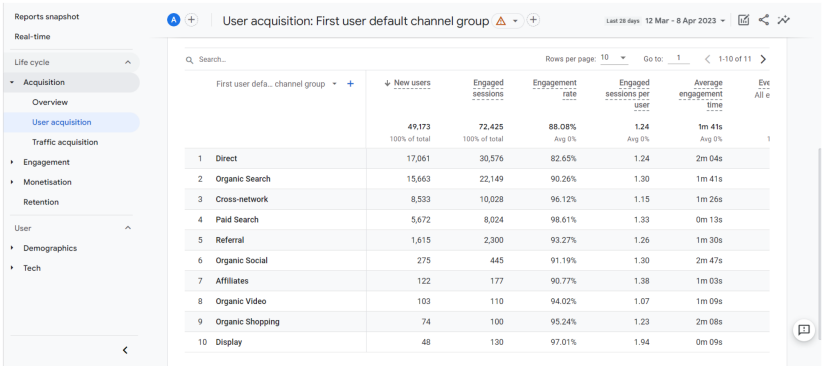

Figure 10: Referral traffic report in Google Analytics

Tonyan (2016) noted that UTM (Urchin Tracking Module) parameters are essential in tracking social media traffic and campaigns, enabling better data collection and analysis for marketing purposes. Google Analytics may sometimes misinterpret social media referral traffic as direct traffic due to missing referral information, leading to inaccurate reporting. Utilizing UTM parameters in Campaign URLs can help solve this issue and provide more accurate data regarding social media traffic. UTM parameters are added to the end of a link and do not affect the link's functionality.

For example, a Campaign URL might look like this: `http://libcal.uccs.edu/booking/groupstudy?utm_medium=social&utm_source=facebook&utm_campaign=studyroom-reservations-spring-2016`, taken the original example from the author's paper. In this example, the campaign medium is "social", the campaign source is "facebook", and the campaign name is "studyroom-reservations-spring-2016". Campaign URLs containing UTM parameters allow for better data analysis using Google Analytics Campaign reports and Social Users Flow report, providing insights into user behavior and campaign performance. These insights can guide future marketing efforts and improve the overall success of social media campaigns (Tonyan, 2016).

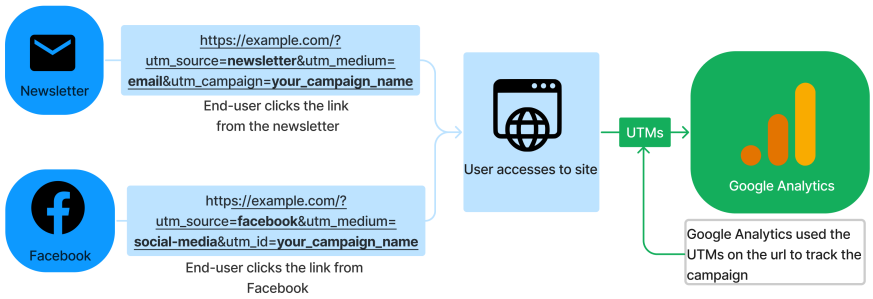

Figure 11: UTMs are used in Google Analytics to track the performance of the campaign every time a user clicks a URL from the campaign (original figure)

Potential practices employed to strike a balance

The concern regarding Referrer and URL parameters revolves around the fact that this information can reveal the browsing history of the user (where the user comes from and the user's journey), and CNIL recommends it be removed. However, the use of these parameters is mainly concerned with the study of user traffic and which channel users arrive at the site from, provided that there is no personal data attached to those parameters. Furthermore, Google states that "Removal of URL parameters or Referrer could impact the accuracy of conversions and traffic source attribution. For example, modification of UTM parameters could lead to misattributed traffic sources". In this case, it is possible that the site may change the Referrer policy within the site so it will not reveal the full URL, only the URL fragments that can be used for marketing and analysis purposes. Regarding the URL's parameters, we can follow the practice by Google in "Best practices to avoid sending Personally Identifiable Information (PII) - Analytics Help" to remove any personal data from the URL (that may be the case where users enter their personal information into the search box and end up in the URL) and only retains necessary parameters for marketing purpose before it is sent to Analytics.

2.3.5 Stopping Browser Fingerprinting in Google Analytics

Browser Fingerprinting is a sophisticated and covert tracking technique employed by websites and third-party entities to collect a unique combination of browser attributes and user-specific information. Fingerprinting amasses data related to HTTP headers, JavaScript properties, time zones, system fonts, screen resolutions, plugins, and hardware configurations, resulting in a distinctive user-specific profile. This distinctive collection of attributes enables tracking entities to identify users and reconstruct tracking cookies even after deletion, thereby circumventing conventional browser mechanisms designed to protect user privacy (Eckersley, 2010). The pervasive nature of remote resources, such as fonts, analytics scripts, and social media widgets, exacerbates the issue by allowing third-party tracking across multiple websites, significantly extending the reach and impact of Browser Fingerprinting on users' online privacy, these insights are combined from the three papers: Eckersley (2010), Krishnamurthy's & Wills (2006), and Szymielewicz & Bill Budington (2018).

Both Eckersley (2010) and Krishnamurthy & Will (2006) highlight the significant privacy risks of Browser Fingerprinting. Eckersley points out that fingerprints should be considered alongside cookies, IP addresses, and supercookies in discussions about web privacy and user trackability. Krishnamurthy's study also emphasizes that the size of privacy footprints should be a cause for concern across all types of websites. Szymielewicz & Budington's paper (2018) explores the implications of the GDPR for Browser Fingerprinting. They argue that Fingerprinting can be considered "personal data processing" under the GDPR and must adhere to its regulations. This means that companies using Browser Fingerprinting must obtain user consent or demonstrate a legitimate interest that does not infringe on users' rights and freedoms. Additionally, companies must provide users with detailed information about data processing and comply with user requests to stop processing their data.

Despite the GDPR's potential impact on Browser Fingerprinting practices, Szymielewicz & Bill Budington acknowledge that Fingerprinting is unlikely to disappear entirely. They expect some non-EU companies to continue Fingerprinting users believing they can avoid European law, while European companies may claim legitimate interests in tracking users. Browser Fingerprinting is a powerful and concerning tracking technique that has attracted attention from researchers and regulators alike. While the GDPR may help curb some Fingerprinting practices, browser companies, standards organizations, privacy advocates, and technologists need to continue working together to protect

user privacy and limit the effectiveness of Browser Fingerprinting.

Potential practices employed to strike a balance

While in default mode, Google collects an extensive range of data, including about users' device information, including browser minor version, Browser User-Agent string, Device brand, Device model, Device name, operating system minor version, and platform minor version. Google's "[GA4] Predefined user dimensions – Firebase Help" (n.d.), There is no explicit statement from Google indicating that Google Analytics employs Browser Fingerprinting techniques. Instead, Google Analytics primarily relies on cookies and User ID tracking to monitor user activities on a website. In "EU-focused data and privacy – Analytics Help", Google states that the collection of such information can be toggled (on/off) on a per-region basis for each data, which means that when adhering to this requirement by CNIL, it is recommended to minimize the data collected only to the necessary basis for their user base study, and further restrict more data on the region within Europe. For example, suppose it is the main goal to reduce the cost of development by optimizing the user interface. In that case, they can choose to collect mainly the screen resolution and the device name to focus on and disable the other data that may not serve that purpose in the EU region.

2.3.6 The deletion of any other data that could lead to re-identification

With any data that was collected previously by the site that can potentially lead to the re-identification of a natural person, Google offers the "Data Deletion" mechanism in "Google Analytics Data Controls Guide – Analytics Help" that allows users to issue a request to Google to remove the collected data from the server. Deleting this data can be dynamic as the user can select either to delete individual fields or all data as a whole (which can include URL, Event Category, Title,...). Sites can also perform the deletion of a particular user using the User Identifier, and the requested deletion from the report will be performed within 72 hours.

Besides the discussed information that Google Analytics collects above by default and its potential risk, in "[GA4] Custom dimensions and metrics – Analytics Help" Google (n.d.) states the ability to create "custom dimensions and metrics" from event parameters and user properties, enabling the analysis of data specific to cater to business needs. This feature allows the site to access and employ its custom event parameters or user properties in reports and explorations. For example, when doing A/B testing regarding the color palette of the site, the site can create a custom dimension for the custom color event parameter and deploy it in two colors; by doing that, the site can compare the performance of the two tests by the parameter value of "color".

However, using custom dimensions and metrics introduces potential compliance issues regarding handling personal data. When a site creates custom dimensions and metrics, it is essential to consider best practices to ensure data privacy and compliance. These practices include using default dimensions and metrics before creating custom ones and avoiding the creation of high-cardinality custom dimensions, which can negatively impact reports and cause data aggregation under the "other" row. For instance, Google advises that the site either disallows or should avoid using a custom dimension for unique User IDs or Session IDs and sending timestamps as custom dimensions. As an alternative, it is recommended to use the User ID feature and explore alternative methods to meet use cases without creating a custom dimension for User IDs. The same goes for other predefined dimensions as well; registering a custom dimension for a parameter that already exists as a predefined dimension is advised against, as it consumes part of the custom dimension quota without providing additional benefits. Finally, the site must not use the custom dimension for any personal information or any information that can increase the chance of the person being re-identifiable.

2.4 Formulation of expected practices

Developers and specialists are expected to adopt various strategies to balance data privacy and optimize the utility of Google Analytics. The following best practices represent a confluence of insights and understanding from the discussion in the above Chapter 2.5, aiming to harmonize adherence to privacy regulations while retaining the benefits of Google Analytics:

- GA4 Adoption: It is expected that companies may choose to continue using Google Analytics and migrate to GA4 if they have not already. This is because most of the complaints are toward Universal Analytics (the legacy version of Google Analytics 4); the prevalence and overwhelming popularity of Google Analytics are also part of the reasons, especially for those who also operate outside Europe. Businesses that transit to GA4 can benefit from this by having better privacy measures and controls. It is expected that businesses will embrace the adoption of GA4 by collecting proximate data that can only provide aggregated and generalized information rather than precise, individual-level data to increase privacy and compliance with the law. This change will result in a loss of granularity and detail in the analysis. Insights and patterns are specific to individual users, or segments may not be captured accurately, limiting the depth of understanding. Proximate data also limits the ability to deliver highly tailored content, and experiences to individual users may be compromised. Personalization strategies that rely on detailed user information may not be as effective, as well as hindering certain the evaluation of marketing efforts and optimization of campaigns.

- Proxy Server and the possibility of Server-side tracking: Given the adoption of GA4 and the IP collection only happens inside the EU, the deployment of a Proxy Server may be considered if the site also wants to benefit from other proxy features, but may be skipped if it is just solely to make sure that will not directly access the IP of the user as CNIL has noted that it is costly and challenging to maintain the infrastructure of the proxy. To further make sure that the IP Address, in any case, will not be transferred outside of the EU, sites should employ a Server-side tracking solution instead of a Client-side one and use Server-side tracking to send the user's analytics data to Google Analytics outside Europe. Regarding the utility, adopting Server-side tracking can improve page view and performance, unlock more precise user privacy controls, and improve data quality, which can lead to the betterment of privacy and utility.

- The use of Informed Consent and Consent Mode for identifiers: Developers should employ the use of Informed Consent with detailed information about obtaining the right from users to use identifiers (Client ID, User ID, Cross-site Identifiers) within Google Analytics; this will ensure that unique User IDs and Client IDs remain consistent for returning users is crucial for preserving GA4 functionality for user behavior analysis. If consent is not granted, the site will either not collect any information from the user to send it to Google (the use of Google Analytics Consent Mode). Besides, sites will follow the practice of not using any information that can determine the user's identity to create the identifier to link with Google.

- Notifying the use of Google Analytics users regarding the followings: In the consent banner, sites must inform users about the use of Google Analytics cookies, the processing agreement with Google, IP address usage, Data sharing restrictions with other Google services, offering an opt-out option to users about the use of Google Analytics to embrace users make their own choice regarding the use of third-party plugins on their data.

- Reducing the chance of Browser Fingerprinting: Sites should implement a selective approach to Browser Fingerprinting, considering the "principle of data minimization." But with fewer data, it may become more challenging for businesses to make informed decisions, personalize experiences, optimize marketing strategies, measure campaign success, and identify trends and areas for improvement.

- Deletion of Re-Identification Data: It is recommended that businesses check and delete information that may lead to identifying a natural person, which they have previously sent to Google Analytics. This proactive approach demonstrates their commitment to user privacy and regulatory compliance.

- URL Parameters and Referrers: It is suggested that for this type of data, sites should adjust their Referrer policy to avoid revealing full URLs but still maintain on the report which channels the traffic comes from. They also need to carefully perform the removal of any personal data from URL parameters before sending them to Analytics. Modification made to UTMs and Referrers can have a negative impact on the accuracy of conversions and traffics report in Google Analytics; when the traffics sources are misattributed, this can result in reduced insights, compromised user segmentation, and difficulties in understanding the effectiveness of marketing efforts and optimizing conversion tracking.

- Documentation: Companies should invest more resources in creating comprehensive documentation outlining their compliance efforts. Such documentation can prove their adherence to privacy regulations and help them navigate any potential legal challenges. While the documentation does not necessarily have a positive effect on the tool's utility, it serves as a valuable resource for navigating potential legal challenges and regulatory inquiries, helping mitigate risks and maintaining a favorable reputation. Additionally, comprehensive documentation enables clear communication and transparency with stakeholders, including customers, partners, and regulatory authorities, fostering trust and confidence in the company's privacy practices.

- Regional Tracking Plugins: For companies with smaller infrastructures or those that primarily operate within Europe and start considering adopting third-plugin tracking, it may be best to consider opting for tracking plugins that reside in Europe to avoid development costs for changes in the legal landscape in the future. This choice would ensure compliance with the principle of data transfer and potentially reduce the complexity of meeting data protection requirements. This transition, however, can lead to alterations within the businesses; for example, it also requires developers to spend more effort in implementing the solution and other departments (marketing, product) to learn to use the new product, and other tools, on the other hand, can also not entirely replace all features provided by Google Analytics, especially when it comes to integrating with other business intelligence tools that the use of Google Analytics can easily facilitate.

  For other companies that already implemented Google Analytics or operate on an international scale, deploying other third-party plugins may not be a good idea, as we are still awaiting further decisions from the legal authorities. Adopting new plugins requires development resources allocated to the change in the system and introduces new privacy and compliance issues if not implemented correctly. To fully avoid breaching privacy regulations in this sensitive time, companies may consider disabling the collection of granular location-and-device data on a per-region basis, especially in the EU region, while receiving further guidance.

3 Approach and framework

3.1 Method: Qualitative research

Qualitative research is appropriate for this investigation because it aims to gain insight into and comprehend developers' and specialists' perspectives on the complex interplay between privacy and utility in the use of Google Analytics. As noted by Bogdan & Biklen (2007), qualitative research can provide a detailed description of people, places, and conversations, which are not easily handled by statistical procedures, and is appropriate for this sensitive and multifaceted aspect of this topic. Additionally, when using the informant's frame of reference, the researcher prioritizes understanding meaning rather than external causes and allows deepening insight into the experience, perspectives, and decision-making processes of developers and analytics specialists in relation to user privacy and utility in Google Analytics.

The research utilizes qualitative research to dig into more profound insight into the thoughts of participants (Maso & Smaling, 2004). As this is a complex and dynamic topic, this self-reflexive stance can enable the researcher to engage actively with data and develop a nuanced understanding in an engaging way, which can reduce the bias in interpreting the result.

3.2 Framework: Semi-structured interview

Semi-structured interviews were used in this study to gather information from developers who work for European businesses that use Google Analytics. For semi-structured interviews, the researcher takes an active role in recruiting representative individuals to be interviewed, and this intentional selection ensures that participants have credible insights and expertise that can help answer the research question.

Additionally, the involvement of participants from different backgrounds in the interviews in this research pushes forward the representativeness that captures a wide range of experiences and insights from different sides of the industry. The use of 'semi-structured' helps the researcher lead the interview by addressing specific topics related to the research question, from which it can achieve the depth required with the respondents. This approach facilitates a comprehensive understanding of practitioners' factors, considerations, and strategies in balancing user privacy and utility.

Furthermore, the method (semi-structured interview) does not presume a predetermined answer, and it gives space for the respondents to provide their perspectives and insights based on their experiences and expertise instead. This openness enables the emergence of novel ideas and perspectives that align with the goals of the research, including depth and detail, nuance, liveliness, and richness of information. Depth and detail are achieved by probing for more information and exploring the specific meaning and implications of the respondents' statements. Nuance is pursued by recognizing that reality is often complex and situational, allowing for the exploration of different levels and circumstances. Liveliness is fostered by creating space for respondents to share stories, examples, and emotional or perceptual aspects related to the research topic. Richness is obtained by covering many ideas and themes through probes and follow-up questions.

3.3 Recruitment

Participants are recruited via a range of methods, including (1) referrals from people who work in the tech industry; (2) posters around campus; (3) advertisement on Facebook and Instagram that allows targeting a specific group of the demographic who are accustomed to the use of Google Analytics; (4) cold-call messages via a professional

network called LinkedIn. While there is no fixed number of participants needed for qualitative research, the researcher aims to achieve the saturation point.

Saturation in qualitative research refers to the point at which no new information is being found, and the researcher becomes confident that a category is saturated. It is a widely accepted methodological principle in qualitative research and is often used as a criterion for discontinuing data collection and/or analysis (Saunders et al., 2018). However, given the dynamic and multi-faceted nature of this topic, saturation can be a moving target. Influenced by rapidly evolving technology, regulatory environment, and user expectations, the digital landscape necessitates a continual reassessment of when saturation has been reached. To ensure the findings remain timely and relevant, the researcher employs an iterative approach, frequently revisiting the data collection and analysis methods, always prepared to incorporate new insights as they arise.

The participants' identities are also re-checked via LinkedIn to verify their professional experience. No prior relationships were established with the participants before the interview. Here are the recruitment steps for the participants:

Step 1:

The recruitment process via LinkedIn (which recruited 5 participants) involved identifying potential interviewees through LinkedIn, a professional networking platform, and engaging with them to schedule interviews. The first step in the recruitment process was to perform targeted searches on LinkedIn using specific keywords, such as "analytics", "Google Analytics," and "Campaign," in order to locate professionals with experience in implementing and utilizing analytics services. Particular emphasis was placed on identifying individuals who had Google Analytics keywords listed on their LinkedIn profiles. This approach helped ensure that the selected candidates had the expertise and experience related to the research topic. Once suitable candidates were identified, the researcher sent connection requests to these individuals on LinkedIn.

To target people via referrers, the researcher utilizes the connection in his network to ask for people with Google Analytics experience to join the interview. The referrers sent an introduction email to the potential candidates about the information of the study, which was followed by asking the participants about their interest in the study, their experience of working with Google Analytics, and their available timeslots for joining.

For social media recruitment, the researcher created a page on Instagram and Facebook dedicated to this research and used Meta Ads to boost the post to potential social media users. For the setting of the Ads, the researcher targets the demographic of the Facebook user **from 22 to 40 years old** with the **occupation of Analyst, Developer, Engineer** who live in **The Netherlands** and have interests in **Google Analytics and Privacy.**

Step 2:

After contacting the potential candidates, the researcher sent personalized invitations to the interviewees, inviting them to participate in the research study, especially explicitly asking them about their experience with Google Analytics to make sure the research could gain insight from them. These invitations included a brief introduction to the research topic, the purpose of the interview, and a proposed schedule for the interview session.

In order to encourage participation and provide a convenient experience for the interviewees, the researcher offered flexible scheduling options and assured participants of the confidentiality of their responses. Additionally, the researcher provided a clear outline of the interview process, addressing any concerns or questions the potential interviewees might have. By employing this meticulous and adaptive recruitment strategy, the researcher could secure interviews with knowledgeable and experienced professionals in the field of analytics engineering.

3.4 Participants

Selection will be made from European companies, predominantly The Netherlands, that employ Google Analytics. To ensure a representative sample, specialists from various industries and domains and those with differing levels of experience in Google Analytics will be included. Here is the description of the participants in the study.

- Participant 1:
  - Age: 30
  - Gender: Male
  - Experience:
    - Current role: Business Data Analyst
    - Past experience: Developer
- Interviewee 2:
  - Age: 40
  - Gender: Male
  - Experience:
    - Current role: Owner of the site
    - Past experience: Developer and application architecture
- Interviewee 3:
  - Age: 33
  - Gender: Male
  - Experience:
    - Current role: Technical Web Analyst
    - Past experience: Analytics Developer
- Interviewee 4:
  - Age: 34
  - Gender: Male
  - Experience:
    - Current role: Data Engineer
    - Past experience: Business Intelligence
- Interviewee 5:
  - Age: 67
  - Gender: Male
  - Experience:
    - Current role: Software Programmer and Data Analyst
    - Past experience: Solution Architect and Chief Technology Officer
- Interviewee 6:
  - Age: 36
  - Gender: Male

  - Experience:
    * Current role: Analytics Engineer
    * Past experience: Data Scientist and Business Development

- Interviewee 7:
  - Age: 28
  - Gender: Male
  - Experience:
    * Current role: Head of Marketing
    * Past experience: Digital Marketing Specialist

- Interviewee 8:
  - Age: 28
  - Gender: Female
  - Experience:
    * Current role: Conversion Specialist and Site Owner
    * Past experience: Marketing Specialist and SEO manager

3.5 Data Collection

The interviews were conducted in English. Except for Participant 2, who was interviewed in person, the others were interviewed online via Google Meet and Microsoft Team platform. The researcher re-introduced the purpose and process of the research interview for the participants and sent the consent form for the participants to agree to and sign. After that, with the permission of the interviewees, the researcher proceeded to audio-record the interview for later transcription. The audio files were then transcribed using the Audio Premiere transcription feature on the local device and then were uploaded to Atlas.ti for coding. The interview lasts 45 to 60 minutes, with the recorded contents ranging from 30 to 45 minutes.

3.6 Interview Guide Development

In this section, the researcher will outline the interview process for participants and explain the rationale behind each question. The interview aims to gather insights into the technical strategies employed by developers and specialists to strike a balance between data privacy and optimizing the utility of Google Analytics while also ensuring alignment with CNIL-recommended measures.

The first section focuses on the candidate's familiarity with legal issues surrounding third-party tracking plugins and GDPR requirements related to Google Analytics. The interview proceeds by asking the participants questions about the practices suggested in Chapter 2.4 and their thoughts on the impacts those practices bring about on privacy and utility. The following section serves to ask the candidate about the course of using cookies and consent. The consideration of an alternative EU-based third-party plugin is also mentioned. The interview ends with any other experience in balancing those two aspects that the participants adopt and wants to address so that other unknown practices are also covered.

I. Legal and Regulatory Issues:

1. **How familiar are you with the current legal issues surrounding the use of third-party tracking plugins and GDPR requirements when it comes to using Google Analytics?* This question assesses the candidate's familiarity with legal and regulatory aspects related to third-party tracking plugins, such as Google Analytics. It helps determine their overall understanding of the legal landscape and the potential implications for data privacy.

2. *The legal landscape continues to change, and some information that may not be considered Personal Data now can be identified as one in the future. In that case, how would you handle the situation, and what prompted this decision? And how do you think it will impact the data analysis?* This question evaluates the candidate's ability to adapt to changing legal requirements and their understanding of the potential impact on data analysis. It also assesses their critical thinking skills in considering the factors that prompt decision-making in the face of evolving data classification.

II. Google Analytics Implementation and Compliance Strategies:

This section focuses on the candidate's knowledge of Google Analytics implementation, compliance strategies, and their impact on data privacy and utility. The questions aim to assess the candidate's experience with GA4, their thoughts on default settings, and their approach to balancing privacy and utility.

3. **Have you migrated to Google Analytics 4 (GA4), and if so, what improvements or changes have you noticed in terms of privacy and utility?* This question explores the candidate's experience with GA4 and their observations regarding improvements or changes in privacy and utility. It helps assess their familiarity with the latest version of Google Analytics and its implications for data privacy.

4. **What do you think about Google Analytics in its default mode? Do you think we need to make any additional changes to the settings to make it work better? (and in which way, better for utility or privacy)* This question examines the candidate's perspective on the default settings of Google Analytics and their opinion on whether additional changes are required. It helps identify their awareness of privacy-related settings and their understanding of the trade-offs between privacy and utility.

5. **Could you discuss how Client-side vs. Server-side tracking impacted both the privacy of user data and the utility of the collected data? And which one would you recommend (for better privacy and utility)?* This question assesses the candidate's understanding of the differences between Client-side and Server-side tracking and their impact on privacy and utility. It helps determine their preferred approach and the factors influencing their recommendation.

6. **What measures have you taken to minimize Browser Fingerprinting while still collecting valuable data? Is there a difference between how you collect data in Europe and other regions?* This question explores the candidate's knowledge of Browser Fingerprinting mitigation techniques and their approach to collecting data while prioritizing privacy. Additionally, it investigates their understanding of regional differences and the specific measures employed in Europe vs. other parts of the world.

7. **Some measures have been recommended by legal regulators, such as the use of a Proxy to change the IP address, removing the Referrers and UTMs from the header, User ID anonymization, and reducing certain data collected. How do you*

think each of these strategies applies, and what impact do they have on the data you collect? This question assesses the candidate's familiarity with specific privacy-enhancing strategies recommended by legal authorities. It helps determine their understanding of the applicability and impact of each strategy on the data collected through Google Analytics.

8. **How do you remove personal data before sending data to Google Analytics? And do you have any thoughts on URL parameters and Referrers being modified to enhance privacy? Will changing the Referrer policy to restrict the full URL affect the insight derived from collected data?* This question explores the candidate's approach to removing personal data before sending it to Google Analytics. It also investigates their thoughts on modifying URL parameters and Referrers to enhance privacy. Additionally, it assesses their understanding of the potential impact of changing the Referrer policy on data insights.

9. **How have you documented your compliance efforts, and how does this documentation help navigate potential privacy and utility?* This question investigates the candidate's documentation practices regarding compliance efforts related to Google Analytics. It aims to assess their understanding of the importance of documentation for managing privacy and utility concerns effectively.

III. User Consent and Communication:

This section focuses on the candidate's perspective on user consent and communication-related to Google Analytics. The questions aim to evaluate their understanding of the role of cookies in obtaining consent and the potential impact of Consent Mode on data quality.

10. **Do you think the use of cookies to obtain consent and explicitly inform users about what data Google Analytics will collect affects the user's choice regarding the use of Google Analytics?* This question explores the candidate's perspective on using cookies to obtain consent and provide explicit information to users. It helps determine their understanding of how this process may influence the user's decision to allow or disallow Google Analytics.

11. **If the Consent Mode is implemented, meaning the user has the right to withdraw from data collection, do you think you can still get comprehensive quality data from the user?* This question investigates the candidate's opinion on the impact of implementing Consent Mode on the quality and comprehensiveness of the data collected through Google Analytics. It assesses their understanding of user consent mechanisms' potential challenges and limitations.

IV. Alternatives and Future Considerations:

This section explores the candidate's thoughts on alternatives to Google Analytics and their considerations for future decisions. The questions aim to assess their awareness of alternative tracking plugins and the factors influencing their decision-making process.

12. **Have you considered using a Proxy Server or alternative Europe-based tracking plugins for your website? If so, what factors influenced your decision?* This question investigates the candidate's consideration of Proxy Servers or alternative Europe-based tracking plugins as alternatives to Google Analytics. It explores the factors influencing their decision-making process, such as privacy requirements, data sovereignty concerns, or regional regulations.

V. Balancing Data Privacy and Utility:

This section focuses on the candidate's experiences and challenges in balancing data privacy and utility within Google Analytics. The questions aim to explore their strategies for achieving this balance and the impact on usability.

13. **What challenges have you faced in balancing data privacy and the utility of Google Analytics, and how have you addressed these challenges?* This question encourages the candidate to share their experiences navigating the challenges of balancing data privacy and utility. It provides insights into their problem-solving skills and their ability to find practical solutions.

14. **Is there any implementation that helps strike a balance between those two that you want to add? and How have the technical strategies you've implemented for privacy impacted the usability of GA4, and have any specific features or functionalities been affected?* This question allows the candidate to suggest additional implementations or strategies to help balance data privacy and utility better. It encourages creative thinking and highlights the candidate's knowledge of emerging practices in the field. It also explores the candidate's understanding of the relationship between technical privacy strategies and Google Analytics's usability. It assesses their ability to evaluate the impact of privacy measures on the platform's features and functionalities.

VI. Professional Experience and Best Practices:

This section delves into the candidate's professional experience and knowledge of best practices related to data privacy compliance and Google Analytics. The questions evaluate their expertise and approach to staying up-to-date with industry changes.

15. *Can you share examples of successful data privacy compliance practices implemented in organizations you have worked for?* This question allows the candidate to showcase their experience implementing data privacy compliance practices in real-world scenarios. It provides insights into their ability to navigate privacy challenges within organizational settings effectively.

16. **How do you stay up-to-date with changes in data privacy regulations and best practices related to Google Analytics?* This question investigates the candidate's approach to continuous learning and professional development in the context of data privacy regulations and best practices. It assesses their commitment to staying informed and adapting to evolving industry standards.

17. **Can you describe any instances where data privacy regulations have conflicted with your data analysis goals or business objectives, and how did you resolve this conflict?* This question explores the candidate's ability to manage conflicts between data privacy regulations and data analysis goals or business objectives. It assesses their problem-solving skills and their capacity to find mutually beneficial solutions.

Because the research aims for 30–45 minutes, the question with the asterisk will be prioritized in case the interview runs short of time.

3.7 Analysis

3.7.1 Content Analysis

This study will use content analysis to scrutinize interview transcripts, identifying and analyzing patterns within the data. Content analysis is a suitable method for this research due to its ability to analyze text-based data, such as interviews, written responses, and observations. Unlike other qualitative research methods associated with specific disciplines, content analysis is not linked to any particular science, reducing the risk of confusion in philosophical concepts and discussions. The content analysis method offers quantitative and qualitative methodologies, making it adaptable to the research's goals and objectives. This flexibility allows the researcher to choose between manifest analysis, which describes the surface structure of the text, or latent analysis, which seeks the underlying meaning (Bengtsson, 2016). In the context of this research, content analysis enables an in-depth exploration of the nuances surrounding balancing utility and privacy in Google Analytics.

3.7.2 Process

This methodology was structured in five distinct stages.

Following the inductive approach, the research process comprised five stages, primarily focused on allowing the data to guide the formation of codes and categories.

In Stage 1, the initial task was transcribing the collected text, laying the groundwork for the upcoming coding process. Instead of relying on preconceived codes derived from existing literature or theories, the researcher approached the data without predefined codes, ensuring an open-minded perspective that would allow the data to guide the formation of the codebook.

Stage 2 involved closely examining the data, allowing new codes to emerge naturally. This inductive approach revealed unanticipated codes, particularly those centered around developer narratives about privacy norms and strategies for optimizing utility. The codes were directly extracted from the data, so they authentically represented participants' experiences and perceptions.

During Stage 3, these emergent codes were thoroughly reviewed and consolidated, merging similar codes to ensure clarity and avoid redundancy. Clear definitions and descriptions for each code were developed, facilitating transparency in the coding process and setting a foundation for future research replication.

In Stage 4, the researcher engaged in a cyclical process of refining the codes. Revisiting the transcribed text repeatedly, the researcher adjusted the codes as needed until no further modifications were necessary. This iterative process, inherent to the inductive approach, ensured that the coding framework was fluid, adaptable, and accurately reflective of the complexities within the data. Continuous revisiting and revising of the codes bolstered the reliability and validity of the coding process. The complete code book can be found in Table 2 of the appendix.

In the final stage, Stage 5, the researcher identified patterns and emerging categories within the revised codes. Code clusters that appeared together frequently were grouped to form distinct categories. Each group presented a cohesive theme that shed light on the interplay between user privacy and utility in the context of Google Analytics implementation, thereby revealing critical facets of the balance developers and analytics specialists attempt to strike.

3.7.3 Trustworthiness in qualitative research

Trustworthiness is paramount in qualitative research, including content analysis, and it is critical to ensure the research's credibility, transferability, dependability, confirmability,

and authenticity (Shenton, 2004).

Credibility: Credibility refers to confidence in the truth of the findings. In this research, credibility is enhanced by collecting data from various perspectives, including data analysts, data engineers, developers, site owners, and marketing specialists. This practice ensures a diverse range of viewpoints on the issue at hand. Furthermore, an iterative approach to coding is adopted. The text is repeatedly revisited, and the codes are refined, enhancing the depth and credibility of the analysis. Triangulation is also employed to corroborate the research findings further, bolstering the credibility of the results.

Transferability: Transferability concerns the applicability of the research findings to similar contexts, circumstances, or situations. This study facilitates transferability by providing a clear, detailed description of each code. This level of granularity allows for the potential replication of the results in similar scenarios, thus enhancing the transferability of the findings.

Dependability: Dependability refers to the consistency and reliability of the findings, established regardless of any changes within the research setting or participants during data collection. To ensure dependability, a comprehensive codebook is provided, complete with descriptions. This allows for the replication of results by other researchers, thus fostering dependability.

Confirmability: Confirmability refers to the neutrality of the data, ensuring that findings are based on participants' responses and not influenced by any potential bias or personal motivations of the researcher. This is achieved by quoting text to support the findings, reducing researcher bias. Additionally, an audit trail highlighting every step of data analysis made to provide a rationale for the findings is maintained to ensure confirmability. Recognizing the role of the researcher entails acknowledging the potential for researcher bias and taking steps to minimize it. In this study, the role of other researchers is included in checking the results and the research process, helping to recognize and mitigate potential bias. This comprehensive approach enhances the overall trustworthiness of the qualitative research.

3.8 Ethical consideration

In conducting this research, a set of ethical principles were adhered to, ensuring participants' protection and the study's integrity.

3.9 Informed consent

Before participating in semi-structured interviews, the participants were given a digital information sheet detailing the research aims, their involvement, and potential risks or benefits. Some examples of potential risks that developers and specialists may encounter during the interview include:

- Confidentiality risks: Participants may unintentionally disclose sensitive or proprietary information about their company's practices or technical strategies during the interview. This risk can be mitigated by reminding the participants not to share confidential information and assuring that any identifying information will be anonymized in the final report.

- Legal implications: There may be a chance where the participants show that some of their practices can potentially be non-compliance with data privacy regulations or other laws. In this case, the interview clarifies that the purpose of the study is not to expose non-compliant behavior but to understand the challenges faced by developers and specialists and identify the reasons behind such decisions.

- Emotional discomfort: Discussing challenges, dilemmas, or difficult decisions related to compliance and data privacy might cause participants to experience emotional discomfort or stress. The interviewer has to be attentive to signs of distress and address concerns or terminate the session if needed.

Conversely, developers and specialists may also experience benefits from participating in the interview:

- Self-reflection: The interview process can encourage developers and specialists to reflect on their experiences and practices in implementing Google Analytics, potentially leading to insights that may enhance their professional development.
- Knowledge sharing: Participating in the study allows developers and specialists to contribute to the broader understanding of compliance practices in the field, which may ultimately benefit other professionals and organizations grappling with similar challenges.
- Influence future regulations: The participant, when engaging in sharing their experiences and insights, can contribute valuable information to the body of knowledge regarding the legal climate surrounding the use of third-party plugins, which may potentially help shape future regulatory frameworks and decision-making processes within enterprises employing Google Analytics.

Participants will be informed of their right to withdraw from the study at any time and without penalty. Before the interview commences, digital written consent will be obtained from each participant.

3.10 Anonymity and confidentiality

Participants' identities are anonymized by a number as an indicator throughout the research to preserve their privacy. This practice is accomplished by giving a unique identification number to each participant that cannot be connected to their personal information. All interview transcripts and study data are securely preserved, and only the research team can access them. Any identifying material in the transcripts, such as the names of persons or corporations, are removed and replaced with generic descriptions.

3.11 Data storage and security

All data, including audio recordings and transcripts, are securely saved on password-protected devices or encrypted cloud storage, with access restricted to the study team. After the transcribing procedure is done, the audio recordings are destroyed. According to the institutional data retention policy, research data will be kept for a certain length of time before being safely deleted.

3.12 Researcher bias and reflexivity

Throughout the research process, the researcher remained mindful of the potential influence of personal biases and assumptions on data collection, analysis, and interpretation. To achieve this, the researcher practiced reflexivity by maintaining a research journal, which will be used to document any emerging personal ideas, feelings, or preconceptions that could potentially affect the study's findings. Regular reflection on these entries will help promote transparency and enhance the overall credibility of the research.

Here is a list of examples of biases that may emerge, including:

- Confirmation bias: The researcher might unintentionally focus solely on information that supports their pre-existing beliefs or expectations while discounting contradictory evidence. To mitigate this problem, the researcher actively seeks out and considers diverse perspectives and ensures that all viewpoints are fairly represented in the analysis. The researcher also re-checks with the reviewers regarding the conclusion of the finding.
- Social desirability bias: Participants may provide responses they believe the researcher wants to hear, which could skew the data. The researcher strives to create an open and non-judgmental interview environment, emphasizing the importance of honest and candid responses.
- Selection bias: The researcher might unintentionally select participants who share their views or experiences, leading to a homogeneous sample. To counter this, the researcher employs a diverse and representative participant recruitment strategy (via referrers and social media).
- Interpretation bias: The researcher may unconsciously interpret data in a way that confirms their preconceptions. To reduce this risk, the researcher engages in a systematic and transparent data analysis process, utilizing inductive and deductive approaches to identify themes and patterns.

In addition to practicing reflexivity, the researcher ensures that developers' contributions are accurately cited and represented in the research report. This approach helps mitigate confirmation bias and provides a fair, balanced representation of participants' perspectives, further enhancing the trustworthiness of the study's findings.

3.13 Respect for participants

"Netherlands Code of Conduct for Research Integrity" by NWO (2018) is used as a standard for the interview; this helps create an atmosphere where they may freely share their experiences and viewpoints. Throughout the interviews, the researcher is alert for any indicators of discomfort or distress and will swiftly address any concerns or cancel the session if required.

By following these ethical standards, the researcher guarantees that the study is carried out with integrity and that the rights and well-being of the participants are protected throughout the research process.

4 Results and Findings

4.1 Outline of the findings

Here are how the findings and results are structured in this Chapter. It represents the themes and the respective codes for each theme, followed by the description of the code.

1. **Migration to Google Analytics 4 helps improve privacy and effectiveness, despite GA4 still having limitations in privacy and features.**
   (a) **Nearly all participants have migrated to GA4**
      - Description: Most participants have migrated to GA4, driven by privacy concerns or forced migration.
   (b) **GA4 increases utility and privacy by providing more options and customizations.**

- Description: GA4 has the potential to increase utility and privacy by providing more options, customization, and integration capabilities with other services.

(c) **GA4 still has limitations regarding privacy and support.**

- Description: GA4 has stricter rules on data collection, lacks certain functions, contains bugs, and still faces privacy issues.

2. **The Low Adoption of CNIL measures due to the perceived negative impact on the utility of GA and Technical Difficulty.**

(a) **Proxy Server may improve user's privacy, but it is not necessary as GA does not collect IP addresses anyway.**

- Description: Proxy Server's impact on data quality, recommendations from different authorities, uncertainty, and waiting for legislation.

(b) **The replacement/absence of the User Identifier can improve the privacy of the user but severely impact the utility of understanding the user's behavior.**

- Description: Pseudonymization improves privacy but can affect understanding user behavior in GA.

(c) **Removing UTMs and External Referrers negatively impacts the tool's utility in marketing, while there are no identified privacy issues with them.**

- Description: UTMs and Referrers may have implications for marketing performance, but no identified privacy issues.

(d) **The impact of limiting the browser's data varies, but it shall be fine with legitimate interest and consent.**

- Description: Limiting browser data can increase compliance, reduce complexity, and enhance privacy, but there should be clearer guidelines.

(e) **Adoption of Alternative EU-based tracking plugins may be considered for better compliance, and it is cheaper than a Proxy Server if the legal landscape escalates.**

- Description: Adoption of alternative plugins to avoid non-compliance risks and cost-effectiveness compared to Proxy Servers.

3. **Server-side Tracking provides more control than Client-side Tracking, hence increasing the privacy and utility when using Google Analytics.**

(a) **Server-side tracking increases privacy, utility, and compliance compared to Client-side tracking.**

- Description: Server-side tracking enhances privacy, utility, and compliance in GA implementation, with easier integration into other services.

(b) **Server-side tracking may introduce financial and technical difficulties to implement and maintain.**

- Description: Server-side tracking can be easier or harder to implement compared to Client-side tracking.

(c) **Uncertainty regarding Proxy Server can be an alternative to the use of Proxy Server**

- Description: Proxy Server and Server-side tracking can be alternatives to enhance user privacy, but certainty is lacking.

4. **Cookie is still heavily utilized to obtain the user's consent for third-party tracking, and more techniques are adopted to increase the utility of the available data collected.**

   (a) **Cookie and Consent provide users with more privacy options, which decrease the amount of data collected. However, Dark Patterns are also commonly used to nudge users to consent to collect more data.**

   - Description: Using cookies and consent allows users more privacy options and reduces the amount of data collected. However, dark patterns are often employed to nudge users to consent to more data collection.

   (b) **Laws and browsers pose a challenge when implementing Cookie and Consent.**

   - Description: The implementation of cookies and consent faces challenges due to browser limitations on cookie lifespan and the legal impact of improper implementation.

   (c) **Practices to balance utility and privacy when using Cookie and Consent.**

   - Description: Various practices, such as allowing user withdrawal of consent, transparency in tracking via cookie banners, data sampling, documentation for legal compliance, and monetization of data, help strike a balance between utility and privacy in cookie and consent implementation.

5. **Measures to increase the privacy of users by preventing or post-handling data leakage to Google Analytics.**

   (a) **Data anonymization, generalization, the use of Server-side tracking to prevent personal data from leaking to Google Analytics**

   - Description: Checking systems to ensure no personal data is sent, using hashed and encrypted information, identifying business needs for tracking, storing non-anonymized data outside GA, and using Server-side tracking to prevent data leakage.

   (b) **Transparency, Troubleshooting, and Data Removal to post-handle incidents of personal data sent to Google**

   - Description: Communicating with users about data leakage, contacting GA to identify leakage sources, data removal from GA, and troubleshooting systems to prevent future data leakage.

6. **Other practices to enhance the privacy, utility, and compliance of Google Analytics**

   (a) **Advocate for changes from external parties (Google and Law) for better utility and privacy**

   - Description: Requesting changes from Google Analytics and laws, implementing new tech solutions, and waiting for clearer legislation to enhance privacy.

   (b) **Different setup of GA per sites/regions increases privacy and compliance**

   - Description: Tailoring GA setup based on public/private sites and regional differences improves privacy and compliance.

(c) **Implementing your own tracking system increases utility and privacy compared to using third-party plugins**

- Description: Implementing a proprietary tracking system offers more control while nudging users to create accounts can impact data collection.

(d) **Non-technical practices such as organizing training and documentation can help increase privacy**

- Description: Documenting GA use for privacy benefits and providing organization training to raise privacy awareness.

7. **Influence of Business Practice, Law, and External influence in adopting measures to strike a balance between Privacy and Utility.**

(a) **Big Tech Companies and Experts shape the way GA (Google Analytics) is implemented**

- Description: Big tech companies and experts, as well as blogs and other influences, play a role in shaping the implementation of Google Analytics (GA).

(b) **GA is more popular, well-documented, and effective than other tracking plugins, despite being less GDPR-compliant**

- Description: Despite GA's popularity and effectiveness, it may be less GDPR-compliant than other tracking plugins. However, GA is still preferred due to its utility, support documents, and integration capabilities.

(c) **Influence of technical difficulty and financial cost in balancing the privacy and utility when using GA.**

- Description: Technical difficulty and financial costs influence decision-making when balancing privacy and utility in GA implementation.

(d) **Influence of the company's business on the practice of using Google Analytics**

- Description: The company's structure, international operations, data collection needs, client preferences, and integration with other Google products influence the decision to use GA and the extent of data collection.

(e) **Legal requirements have a significant influence on balancing utility and privacy when using GA, but it is hard to keep up with them**

- Description: Legal requirements significantly impact the balance between utility and privacy in GA implementation. The challenges faced are the difficulty in keeping up with these requirements, limited access to privacy knowledge, and technical measures not meeting EU country requirements.

4.2 Co-occurence analysis

Findings from the co-occurrence table suggest that the practice of using Google Analytics is influenced by a range of factors that span technical, financial, and regulatory dimensions. The full table of the co-occurrence analysis can be found in the appendix.

First, it was found that Server-side tracking was strongly associated with increased privacy. As noted by participants, Server-side tracking offers direct integration, reducing reliance on browser activity and device cookies, thereby enhancing privacy. This practice avoids potential discrepancies and privacy issues related to Client-side tracking. Despite this benefit, Server-side tracking was also associated with technical and financial challenges. The implementation and maintenance of Server-side tracking require

a significant development workload, resulting in higher financial costs. This aspect of Server-side tracking is shown in the co-occurrence of the code "Server-side tracking may introduce financial and technical difficulty to implement and maintain" with the code "Influence of technical difficulty and financial cost to balance the privacy and utility when using GA."

Further analysis revealed practices associated with the CNIL recommendations, which include practices to balance utility and privacy, Proxy Server implementation, measures to prevent personal data leakage, and non-technical practices such as training and documentation. These practices were found to frequently co-occur with "No adoption of the practice" and "Decrease the utility", indicating potential challenges in adoption due to utility-related concerns.

The company's business approach influenced several practices, such as the different setups of GA per site/region, implementing one's own tracking system, and measures to post-handle personal data leakage. The co-occurrence of these codes with "Influence of the company business on the practice of using Google Analytics" indicates that the company's business strategy, resources, and priorities play a critical role in how Google Analytics is used.

The implementation of Proxy Servers was found to be associated with increased privacy but also with financial and technical challenges similar to Server-side tracking. The codes "Proxy Server may improve user's privacy, but it is not necessary as GA does not collect IP address anyway" and "Uncertainty regarding Proxy Server can be an alternative to the use of Proxy Server" co-occurred with "Influence of technical difficulty and financial cost to balance the privacy and utility when using GA," showing that the utility of a Proxy Server for privacy enhancement may not justify the implementation challenges and costs, especially given that GA does not inherently collect IP addresses.

A high number of occurrences is also observed between using Cookie and Consent for increasing user privacy options and the resultant decrease in utility (8 instances). This relationship suggests that as more privacy options are provided to users, which limits the scope of data collected, the utility or value that companies can extract from the data diminishes. These practices are also strongly associated with instances without impact on privacy (7 instances), indicating that the offered privacy options might not significantly alter the status quo in some cases.

Lastly, several practices, including adopting alternative EU-based tracking plugins, advocating for changes from external parties, and the transition to GA4, were found to co-occur with "Suggestion to adopt the practice". This suggests a consensus among participants that these practices could be potential solutions to enhance both privacy and utility when using Google Analytics, despite the challenges presented by legal requirements and technical difficulties.

4.3 Google Analytics 4 Migration

4.3.1 Nearly all participants have migrated to GA4

There is a consensus among the participants that they have migrated to Google Analytics 4 from the previous version of (Universal Analytics), except for Participant 5, who stated "No, I haven't migrated". Participant 7 points out that Google is discontinuing support for Google Analytics 3 (Universal Analytics), which made migrating to GA4 an essential step.

4.3.2 GA4 increases utility and privacy by providing more options and customizations

Regarding the privacy aspect, multiple participants (Participant 7, Participant 3, and Participant 8) highlighted GA4's enhanced privacy settings. For example, GA4 includes IP anonymization by default, unlike Universal Analytics which requires it to be specified manually. Also, as mentioned by Participant 4, GA4 uses an event-based data model, aligning with industry norms and making data interpretation more intuitive.

On the utility side, as described by participants, the benefits of Google Analytics 4 are substantial and varied. One of the standout features, according to Participant 7, is the change in the fundamental logic and functioning of the tool. They highlighted the shift from "a JavaScript-based data collection model to a data stream-based one", which offers more flexibility and options for data interpretation and analysis. [2] Further, Participant 7 also added, "Google Analytics four has a lot of customization, which we like. So let's say on the Google Analytics three, you have much more limitations in terms of what and how you can interpret the data in Google Analytics four you are much more flexible, how you're going to interpret data, what kind of dashboard you're going to build based on your needs, and result of this."

GA4 is also noted to be better in terms of integration with other Google services and improved customization and flexibility. Participant 4 praised GA4's integration with Google's BigQuery service [3], which simplifies data ingestion and analysis, making it more efficient than previous technologies. Additionally, According to Participant 7, GA4 offers more customization options, permitting users to interpret data in ways that suit their needs and build custom dashboards. GA4's flexibility, in contrast with the limitations of Universal Analytics, is appreciated by users who are more advanced with Google Analytics.

On the compliance facet, Participant 3 noted that with GA4, IP randomization was introduced as a mandatory feature, which is a significant improvement from Universal Analytics, where it was an optional feature.

4.3.3 GA4 still has limitations regarding privacy and support

Despite significant upgrades from its predecessor, GA4 is still perceived to possess limitations on various aspects, including privacy, compliance, and utility. In one case, Participant 4 expressed concerns about the privacy measures in GA4. They noted the shocking realization that "the data collected by the Google Analytics would sometimes come with emails, names, last names of people like real, you know, private information" and they emphasized that there needs to be a more proactive system in place to avoid storing personal user information in a plaintext format both from the one who implements them and Google Analytics in detecting them.

[2] As noted by Google (n.d.) in "[UA→GA4] universal analytics versus Google Analytics 4 data - analytics help," Universal Analytics (UA) adopts a JavaScript-based data collection model focused on sessions and page views. Data is gathered at the property level with a tracking ID using hit types such as page views, events, and screen views. It emphasizes predefined reports and provides insights into website performance using custom dimensions/metrics, content grouping, User ID, Client ID, and data collection settings.

On the other hand, Google Analytics 4 (GA4) adopts a data stream-based model centered around events and parameters. Instead of collecting data at the property level, GA4 collects data at the stream level via a unique data stream ID. It focuses on tracking user interactions through events, and it supports custom dimensions/metrics, parameters, and user property. Although it offers fewer predefined reports than UA, GA4 provides greater flexibility for custom analysis and delivers more customer-centric insights. Crucially, GA4 allows for cross-platform and cross-device user behavior tracking.

[3] According to "BigQuery Enterprise Data warehouse — Google cloud" by Google (n.d.), Google BigQuery is a cloud-based big data analytics web service designed to process huge read-only data sets. It helps you combine data from different data sources such as Google Analytics, Facebook Ads, and Google Ads

Regarding the compliance issue, Participant 6 highlighted challenges around the implementation and compliance strategies associated with migrating from Universal Analytics to GA4, noting that obtaining "useful data is getting more difficult". However, they admitted that such limitations could be viewed positively from a privacy perspective. Participant 3 discussed issues around data privacy regulation conflicting with data analysis. They voiced concerns about the need for improved transparency regarding how Google uses personal data and how that information could potentially be compromised. They said, "I'm not sure how Google will use every single user's personal data... I see so many complete compliance complaints about quotas on Google Analytics four because people cannot see enough data if they couldn't reach the specific quota." On the same note, Participant 8 raised concerns about the transparency and clarity of GA4, stating that it was "hard for me to figure out what exactly it does and doesn't do."

Utility-wise, Participant 8 found the help documents for GA4 limited, making it difficult to understand what the platform does and does not do. They also pointed out several missing features in GA4 in Universal Analytics, such as the lack of views for segregating data and the absence of specific reporting options "for example, I don't have any views anymore, so I can't segregate my data". They also reported experiencing bugs in GA4 and were dissatisfied with the transition from Universal Analytics to GA4, which they felt was premature given GA4's current limitations. However, Participant 8 also expressed their belief that GA4 will eventually become better than its previous version with improvement.

4.4 Low Adoption of CNIL Measures

This chapter discusses a range of measures suggested by CNIL, most of the measures are not adopted by the participants and are perceived to have a significant impact on the functionality of Google Analytics.

4.4.1 Proxy Server may improve privacy but not necessary as GA does not collect IP addresses

The findings from the interviews reveal that many interviewees demonstrated a lack of familiarity with the concept of using Proxy Servers in this context. Interviewees suggested they had not considered or would implement this approach (Participants 7, 4).

"From the proxy. No, because I haven't been familiar so far and this is the first time I read. Now I will check it from the alternatives." (Participant 7)

"No, honestly, this I mean, I know what the Proxy Server is for my quite general tech background, but I really never heard about it in the context of tracking..." (Participant 4)

Regarding the impact on privacy and utility, a few participants indicated skepticism about the significant impact of using a Proxy Server on the two aspects. They indicated that Google Analytics does not store or expose IP addresses, suggesting that using a Proxy Server might not have a meaningful impact and be redundant.

"Well, we are getting really technical, but I don't know I may be wrong, but I don't think that this will impact the Google Analytics too much because Google Analytics does not store or expose IP address anyway." (Participant 4)

The perceived potential impact on the quality of location data is also identified. Participants noted that if Proxy Servers are not properly implemented and users are inaccurately placed, this could have significant implications for businesses requiring location-based data. Also, the participants expressed concerns about the negative impacts of using Proxy Servers on the utility of Google Analytics when adopting other measures suggested by CNIL.

"Yeah, well as everything. If the proxy is not calculated correctly and and the users are not well-placed, then the companies that are using or the users that are using Google Analytics one have more accurate data and this can impact our search and markets that need more and location-based data because the competitiveness of their business." (Participant 6)

While some participants acknowledged the potential benefits of Proxy Servers for users' privacy, they also indicated that using Proxy Servers might significantly limit the utility of Google Analytics (Participant 2, 3).

"So for the user, obviously it's a good measure. I mean, they don't expose their IP." (Participant 4)

"Yeah, I think using if they can get correct precise location data, it doesn't have to be IP address to be honest, but I think there should be some reasons the technology with using it..." (Participant 3)

"If these measures only work because they no, sorry, these things exist because they allow Google to identify a device or a user. So if you remove all identifiable things about a user, then you cannot identify anything anymore. So. So so analytics or it would become quite useless I guess." (Participant 2)

The findings suggest that a Proxy Server might not be considered an option to adopt among participants, especially in the third-party Tracking context. As stated by Participant 5, "Definitely not. Well, in my experience, not for Google Analytics. It's to do with protecting yourself from the bad guys."

4.4.2 Replacement of User Identifier improves privacy but impacts understanding user behavior

Based on the interviews, there is a consensus among the participants that the methods of using the replacement/absence of the User Identifier, Cross-site or Lasting Identifiers by the Proxy Server suggested by CNIL will severely impact the customer journey and personalization. This was particularly noted by Participant 4, who said, "This is what those tools are about, that you can correlate the same person across different visits, across different devices to create a customer profile...". They further explained that without User Identifiers, businesses would no longer be able to monitor customer interactions over time, which could substantially impact machine learning-based product recommendations and customer personas. On the same note, Participant 6 stated, "That's data that really does have an impact because many websites and applications are tracking how many times a user goes into a webpage... not being able to identify the user... It has a big impact...". The same participant also admitted to being unsure about how to solve this issue.

Some participants saw the benefits of increased privacy and compliance with regulations but also acknowledged the decrease in utility. Participant 3 highlighted the potential privacy issues, saying, "User I.D. is definitely anonymized and that's I guess kind of mandatory...". However, when quoting the full requirement from CNIL recommended measures that the ID replacement should be implemented in a way that a user cannot be identified twice, they were concerned about how this could cause trouble for marketers as every user would appear as a new user, disrupting the ability to track users over multiple sessions. Similarly, Participant 5 stated, "It will enhance the privacy" while acknowledging that without User Identifiers, "it will be impossible to do the tracking.". Participant 1 agreed with the idea of anonymizing User ID if that still allows them to perform data aggregations. Participant 1 affirmed, "Definitely. I think I believe that that is an important, important way to, to, to, to have a more compliant data in our company... As long as we can do aggregations, it doesn't, we don't mind if the client I.D. or the user alias, and I'm not, I'm not."

4.4.3 Removing URL parameters and external Referrers negatively impacts marketing utility

Similar to the suggested measure for the replacement/absence of User ID, the analysis of the interviews revealed various insights into the perceptions of interviewees regarding the suggested measure of removal of UTM parameters and Referrer information from a website. The related concepts were: the impact on marketing strategies, the decrease in utility of analytics tools, the potential increase in user privacy, the influence of other big tech companies on the practice of using Google Analytics, and an overall reluctance or hesitance in adopting the practice.

Several interviewees highlighted the significant impact such practices could have on marketing efforts. Participant 7 said, "Huge impact because maybe 90% of the marketers are using everything besides UTM tags, all, for instance, a request from companies that we work because they want us to place UTM tags. This is how they measure the performance." Participant 3 also articulated, "That's really that. From that point I was thinking, oh, that means you should not use Google Analytics because without a parameter, without, uh, without UTM parameters, marketers cannot use Google Analytics at all because that's the most important being for them, because recognizing where they are, where they are from, and where we need to invest to get more or better performance of the website." Participant 6 stated the same idea that without UTMs and Referrers, "you cannot see which clients are coming from this channel, and then you cannot target this type of users because you know the demography, you know, the age, gender and stuff."

Participants acknowledged the potential benefits in terms of enhanced user privacy. Participant 6 pointed out, "From which the data is captured, I see the benefits of not being targeted like that. But yeah, it's like a compromise." Despite acknowledging the potential benefits, most participants were hesitant to adopt these practices due to their perceived impact on marketing and analytics. Participant 2 exemplified this perspective by stating, "It would severely impact our measuring because we want to know from which point people come and what impact would this have, What's a positive effect within that. So can you tell me."

4.4.4 Limiting browser data increases compliance, reduces complexity, and enhances privacy

The participants in this study had varying perspectives on reprocessing information that could be used to generate a fingerprint and the deletion of any other data that could lead to re-identification. Several participants emphasized the importance of balance when collecting user data. They were aware of the privacy risks data collection entails, yet also acknowledged the need for certain data to improve user experience and functionality. Participant 3 said, "The necessity and the kind of downsizing of collecting data can be a good idea, but there cannot be a baseline, or we are all strict guidance so that another hard thing." Similarly, Participant 1 also believed that more selective data collection would enhance privacy and improve data analysis, while others were concerned that it would inhibit the development and optimization of their products or services. Participant 1 conveyed, "Yeah, I definitely think that as well. Too many data is not good as well on my end because that's going to dilute the information that are going to make the complexity of the analysis even bigger."

Despite this, participants recognized the challenges faced when trying to implement privacy safeguards while maintaining utility, as there is no one-size-fits-all solution to balance data collection for utility purposes and privacy. Organizations must tread this delicate balance carefully, taking into account their specific needs, the privacy expectations of their users, and the regulatory environment in which they operate. Participant

3 stated, “That should be different by company, by company or industry, by industry and the size of company or something like that. So it’s really complicated to be decided by a legal agency. I think.” Participant also 2 expressed this dilemma, saying, “I need to do certain things and I can try to disable as much as I want, but I still need and at the end of the day to that’s my marketing to work. So that’s, that’s the challenge. Yeah.”

4.4.5 Alternative EU-based tracking plugins considered for compliance and cost-effectiveness

Based on the response, the participants expressed a lack of active considerations or plans to adopt alternative tracking plugins or Proxy Servers, either due to a lack of awareness or other factors. Participant 5 mentioned not having thought about using alternative tracking plugins and was unsure about considering them. “I don’t know. I wouldn’t know. I haven’t thought about that.” (Participant 5). On the other hand, although Participant 2 expressed a negative attitude towards Google and its practices, they stated the necessity of using Google Analytics due to its integration with other marketing channels, such as Google AdWords. “As much as I hate Google, it’s in the monopoly that we cannot afford. We have to use.” (Participant 2).

Participant 8 indicated a current focus on waiting for legislation and not implementing changes until further guidance is available. “We’re currently waiting for legislation. . . . and we don’t want to implement all kinds of things before we know what they’re actually going to stay and decide about that.” (Participant 8). Participant 4 when comparing Google Analytics and other tracking plugins, mentioned that viable alternatives are available, although they may not be as popular or well-documented. This might be the reason why Google Analytics is still a popular choice. They acknowledged that Google Analytics has the largest market share but mentioned the existence of other viable alternatives that are less popular and have fewer integration options.“There are other alternatives that I think are also viable. It’s just they are not so popular. The documentation is not so much, the companies that offer the integration with them is less so.” (Participant 4)

The other participants expressed the idea of considering alternative tracking plugins to replace Google Analytics with certain conditions, especially if there are concerns about privacy and compliance with regulations. Participant 7 said that they recommend exploring alternative tools if using Google Analytics becomes a big problem in multiple countries. They believe Google Analytics may need to change something to address privacy concerns. “If this becomes a big problem in a lot of countries and there is a like must you need to do for sure, I recommend alternatives. If Google Analytics four doesn’t change something to solve this because they also, I’m sure they will work on that direction as well.”

Participants mentioned the impact of legal authorities and regulations on using Google Analytics, indicating that considerations are being made to ensure compliance and avoid potential issues. Participant 4 discussed conducting due diligence for a client, considering whether to continue using Google Analytics or implement a GDPR-proof alternative due to concerns in the European Union. “We did like kind of a pseudo due diligence for one of our clients if we should stay with Google Analytics with them, given all the noise in the European Union, or whether we should maybe implement something else that we know is GDPR proof, and we know that it will not be blocked one day.” (Participant 4)

Finally, Participants acknowledged the importance of privacy and compliance with regulations, indicating a willingness to use tools that are more GDPR-compliant and prioritize user privacy. Participant 1 expressed the view that it is fair to use a more

GDPR-compliant tool instead of Google Analytics in their company."In my current company, yeah, it's fair to use another tool that is more compliant with GDPR instead of Google Analytics." Participant 3 recognized the necessity of balancing data collection with privacy considerations, understanding the reasons behind privacy concerns. "Yeah, definitely, there are some necessity, but definitely we understand why we need to consider the privacy." Participant 8 mentioned following advice from the legal authority and being open to making changes based on their guidance. "So now we're still following the advice of the (Dutch document name), and once they say we need to do something else, we look into it then."

4.5 Server-side Tracking

4.5.1 Server-side tracking increases privacy, utility, and compliance compared to Client-side tracking

Participants expressed a preference for Server-side tracking over Client-side tracking (browser-based tracking). They believed that Server-side tracking provided more accurate data, better privacy regulation compliance, and reduced discrepancies.

Participant 7 stated "For me Server-side tracking is the best so cookies you put it on the browser on the web for me are cookies that are going to be removed soon. I think it was the end of the year. That's why I said Server-side tracking is by far the best I know, and I trust."

Participant 4 said, "Server-side is the answer. And then you control each and every bit of information that you send to Google Analytics."

Participant 6 also favored Server-side tracking, claiming "Yeah. I think sometimes the server, it's better."

Participant 8 expressed "Well, it depends on what you want because, for me, it's easier to do the server side."

Participants concurrently emphasized that Server-side tracking provided more accurate data and offered more significant control over the collected data. They believed that Server-side integration allowed direct communication between the server and the analytics platform, leading to more reliable data.

Participant 7: "The data is much more accurate... What I understand, because I'm not sure whether I'm correct, it's much more privacy safe because you don't go to any browser that's a browser device, you don't link into the browser settings which I think is the one that causing problems."

Participant 4: "What I would recommend is probably still to use Server-side tagging because it also gives you more control over what happens after the data is collected."

Participant 6: "So I think that can be a good thing. Yeah. If you are taking into account the type of data that you are managing and if you have a lot of GDPR data, I think that having more control on it, it's better."

Participants believed that Server-side tracking was more privacy-safe and compliant with regulations such as GDPR. They perceived Server-side integration as a way to avoid linking data to browser settings, which could pose privacy risks.

Participant 4: "but which at the same time means an improvement for privacy."

Participant 3: "I definitely consider Server-side is better, especially for the users of the websites who are considering privacy because Server-side offers the opportunity to reduce the number of dimensions of the data."

Participant 5: "I would recommend Server-side. I believe that Client-side is an invasion of privacy... these tech companies, they just completely empty your iPhone. So they'll take everything they see, which is basically everything that's their other people's cookies."

Despite the overwhelming benefits, participants acknowledged that Server-side tracking implementation required more technical development and additional workload compared to Client-side tracking. They perceived it as a trade-off between improved privacy and increased development efforts. Some others

Participant 7: "The only negative side is a bit more technical, and you need a bit more development workload to deploy."

Participant 4: "Implementation of this obviously brought some additional work that had to be done for the clients to maintain some level of tracking that was possible before."

Participant 8: "Generally, I never saw the actual added value... for just the privacy issues."

4.5.2 Implementation and maintenance of Server-side tracking may pose financial and technical difficulties.

While the results suggest a preference among participants for Server-side tracking due to its perceived benefits for data privacy and accuracy, the barriers to its implementation, including financial, technical, and regulatory uncertainty, are significant considerations that need to be addressed. Participant 7, with a strong endorsement of Server-side tracking, noted its superior accuracy and regulatory compliance, "Server side tracking... we have much, much fewer discrepancies... the data is much more accurate." They highlighted that Server-side tracking results in direct integration, meaning one server communicates directly with the analytics server, bypassing the browser. This, according to Participant 7, provides a safer environment from a privacy perspective as there is no need to engage with the browser settings that could pose potential privacy issues. Nevertheless, they identified the requirement for a greater development workload and technical acumen as a disadvantage, "The only negative side is a bit more technical, and you need a bit more development workload to deploy."

Participant 4 echoed these sentiments. They observed a shift from Client-side to Server-side tracking due to the latter's improved feasibility, especially in light of the increasing privacy concerns associated with the former. While acknowledging the additional workload and somewhat limited capabilities compared to Client-side tracking, Participant 4 also suggested that this shift could be seen as an enhancement of privacy.

Participant 3, expressing a similar preference for Server-side tracking due to privacy benefits, added concern about the financial implications. They stated, "Only the downside would be the cost because it will require more budget from the companies collecting data."

Lastly, Participant 8 touched upon the ambiguity of current legislation, which impacts organizations' willingness to transition to Server-side tracking, "Also, because the legislation at this moment is very unclear, so if it says we have to do that, we will look into that, but not if we're not forced to do it."

4.5.3 No guarantee that Server-side tracking can replace the requirement of using a Proxy Server

The potential of Server-side Google Analytics as a replacement for setting up a Proxy Server seems to be met with uncertainty. When questioned on this matter, Participant 4 stated, "I don't know really. I don't want to create an answer, you know but I don't know." The lack of clarity on this subject suggests that there may be no guarantee that Server-side tracking can effectively replace the requirement of adopting a Proxy Server.

Further elaboration was provided by Participant 3, who touched on the potential issues with implementing Server-side analytics: "A French government agency sent suggestions for the company who really wants to keep using Google Analytics was using

Server-side and stripping out all the privacy-related stuff. But it's not that practical because it makes really important information out of the tool. So we cannot connect the users into where they are from or something like that."

Participant 3 further shared their experience of companies making decisions to use Google Analytics but implementing it Server-side. However, they stated that companies were "informed enough about their risk for that."

When discussing an article about the potential solution of using an EU Proxy Server, Participant 3 voiced a critical concern from the analyst's perspective: "the most important things from the analyst's perspective would be could we get the correct attribution data, so until the proxy can give us that feature, I'm not sure if it would be useful for practical reasons."

4.6 Usage of Cookie and Consent

4.6.1 Cookies and consent provide users with privacy options, but dark patterns may nudge users to consent to more data collection.

Participants discussed how user consent and the use of cookies could affect the quality of data collected through Google Analytics. They mentioned that users who reject tracking or configure their browsers to block cookies result in missing data, which can impact the accuracy and completeness of the analytics. Participant 7: "Not really... 90% of the websites already implemented this [cookie consent], and people are already familiar that if they continue using the website, they are already being tracked."

Participant 4: "Naturally, you lose between 10 to 25% of the visitors already because they click 'I do not consent' or have their browsers configured to block tracking and cookies."

Participants acknowledged that the rejection of tracking and the use of cookies by some users could decrease the utility of the collected data. They mentioned that the population opting out of tracking might not be large enough to impact the overall insights derived from the data significantly, but it still affects the accuracy and completeness of the analytics.

Participant 4: "It does affect the data quite a bit, but in many cases, the population that decides to reject tracking is not large enough to significantly lower the information... you only need a sample of something to derive knowledge about the population."

Participant 6: "If that many users didn't read properly what the cookies are gathering... many users want to accept all and get what they want."

Participants agreed that the design and ease of interaction with consent banners greatly influence user behavior. For instance, Participant 7 underscored the importance of equally sized and colored consent buttons while expressing skepticism towards the users' willingness to be tracked, stating that "most of the people don't want to be tracked, so they click reject by far."

On the other hand, Participant 4 highlighted a drop in site visits due to the introduction of cookies, suggesting a decrease of "10 to 25%" in visitors who either decline consent or have browser settings blocking cookies and tracking.

Participant 6 noted that implementing consent banners may not significantly impact user behavior, primarily when users focus more on accessing website content. They also observed that companies can use consent banners to subtly "hide cookies that are gathering some type of information behind an accept all button."

Participant 3 argued that detailed information on data collection might not necessarily enhance user choices concerning privacy, suggesting that long texts may cause people to accept without fully understanding the terms. Similarly, Participant 5 shared anecdotal evidence that most people simply accept the cookies because customizing preferences is "time-consuming".

Finally, Participant 2 was skeptical about the overall impact of consent withdrawal, arguing that most people will not withdraw consent unless they are fully aware and willing to spend time doing it.

As a result of the human behavior and psychology above, the participants largely agreed that companies often use dark patterns in the design of cookie banners to steer users toward providing consent. For instance, Participant 7 noted that the size and color of consent buttons could affect user decisions, suggesting that larger and more colorful buttons could inadvertently lead to more acceptance. They, however, stressed that the new regulation advocates for equal size and color of buttons to mitigate such influence.

Participants 6 and 2 mentioned the common practice of making the "Accept All" button more visible and accessible than the "Reject All" or "Customize" options. Participant 6 further argued that this pattern becomes more conspicuous in websites with a long list of services using cookies, making the process of manually disabling each service tedious and pushing users to simply "accept all."

In a similar manner, Participant 3 also pointed out that overwhelming users with detailed information might inadvertently lead to more acceptance, as people may not want to read through all the details. This sentiment was echoed by Participant 5, who cited the time-consuming nature of customizing cookie settings as a deterrent for many users.

4.6.2 Practices to balance utility and privacy when using cookies and consent

The practices to balance privacy and utility when using cookies and obtaining consent in Google Analytics revolve around transparency, user control, and creative solutions to continue data gathering when consent is not provided. Participant 7 shared experiences around the importance of transparency, specifically regarding what types of tracking are being utilized, and the duration of data storage. They stated,"When you don't have mentioned what you track... One thing is that you don't mention all of them also what is the duration they are placed and what is the duration of the data that your store is in 90 days, two years, one year, one day a session, or whatever." This highlights the importance of clear communication about data practices to the users. Furthermore, Participant 7 stressed the necessity of allowing users to withdraw consent or reject being tracked, pointing out that"this is a must."

Repeating Participant 7's sentiments, Participant 4 advocated for being straightforward with users about the data being collected. They suggested that users should be informed about their data being tracked and should receive some benefit or reward in return, adding "the biggest problem here for it always was that the companies were making use of users information that they were getting for free." Participant 4 also recommended that businesses promote users to create accounts for more personalized tracking instead of using external tracking platforms.

In contrast, Participant 3 expressed that while legal cookies should always be covered, it is unnecessary to go into detailed specifics of what is collected. Participant 1 shared an approach using proxies and sampling to analyze a subset of users in cases where users did not consent to data gathering, adding "if the number of samples is big enough, we can say that it's statistically significant for us to derive something from it."

However, despite these varying approaches, a common thread that emerged is the emphasis on respect for user privacy and the importance of clear, transparent communication about data practices. This was highlighted by Participant 8, who mentioned using a neutral cookie bar and ensuring the 'decline' button was not hidden.

4.7 Preventing Data Leakage to Google Analytics

4.7.1 Data anonymization, generalization, and Server-side tracking to prevent personal data leakage.

Participants 4 and 7 emphasized developers' instrumental role in data sanitation before its dispatch to Google Analytics. Participant 4 advocates for Server-side management, expressing that "Server side is the answer. And then you control each and every bit of information that you send to Google Analytics." Participant 7 echoes this sentiment, particularly emphasizing the necessity to use separate systems for sensitive data, advising, "But Google Analytics 4 is only for non-anonymized data, so make sure you don't connect both, but let them separate."

Participant 7 also underlines the importance of transparency, suggesting to "give the user option to reject being tracked and specifically say what kind of information you collect, how many days you're storage, where is your storage." Participant 3 underscores the importance of evaluating the necessity of data collection as the guiding factor to strike a balance between privacy and utility, declaring that, "People really need to think of the data they want to collect if they really need it."

Participant 5 also identifies"anonymization of data before it is sent to Google Analytics" as a pivotal practice in data protection, stating,"Yes, I am. I have examples in places where I worked where they anonymize the data so that it does not sit on a server," and clarifying that the anonymization process occurs "before sending it out." Additionally, data encryption was underscored as an important privacy measure. Participant 5 validates that "You could use a lot more encryption to transfer information from one place to another." This approach safeguards data during transit by rendering it unreadable without the correct decryption keys. Participant 5 also pointed out that generalizing personal data prior to its transmission to Google Analytics effectively protects user privacy. They noted, "Um, those three things [identifiable User ID, corporate data surrounding the person, and job title] are important to, to generalize make it more general."

When outsourcing analytics to a third party, as highlighted by Participant 2, represents another method for preserving personal data from being sent to Google Analytics. They explained their preference for using Google Analytics for public spaces while keeping sensitive data separate under strict policies to prevent it from being processed by Google Analytics or any external data.

In conclusion, strategies to prevent personal data from being sent to Google Analytics range from data anonymization and encryption to generalization and outsourcing analytics to a third party. Other methods include allowing users to opt out of tracking, educating them about data handling practices, and adhering to regional regulations like GDPR. The challenges that persist include determining the necessary amount of data for collection and striking a balance between user privacy and the effectiveness of marketing strategies.

4.7.2 Transparency, troubleshooting, and data removal to post-handle personal data leakage

Participant 7 highlighted the significant importance of preventing the exposure of personal data to Google Analytics. The participant admitted that encountering personal data in their analytics would be a serious issue requiring immediate attention. The strategy suggested was contacting Google to understand the situation better. The participant also acknowledged that specific steps for addressing the issue were not clear-cut and would depend on the particulars of the situation, as emphasized in the quote: "This is a serious question because if Google can recognize someone that's a big problem, prob-

ably we will get in contact with Google to understand what's going on because we might have serious problems regarding this. Yeah, if I need to be more concrete, how you can know the issues? I don't know. I will figure it out once I realize it."

Participant 4 expressed a proactive approach toward data privacy. In cases where personal data was unintentionally sent to Google Analytics, the strategy was to halt the tracking of such information by informing the responsible tracking implementation team, as noted in the quote: "And that is a challenge because I, as a data engineer already and data knowledgeable person, can tell that this should not have happened. We should not be able to correlate the site visits with the private information right away without any hashing done and stuff like that. So that was the challenge. Well, the overcoming of this was to notify the people responsible for tracking implementation, to stop tracking that information." Participant 8 mentioned a similar approach, suggesting the importance of understanding the cause of a data leak before implementing a solution. The strategy discussed was first to identify the source of the issue and then engage Google to remove the data. This is emphasized in the quote: "That would really depend on how I figured it out, because I should be able to technically trace the cause of that, see whether it's the fault of Google or the fall of me having set, uh something badly, or someone else having set up something badly or misuse something."

Participant 6 proposed a strategy based on damage control and transparency. When unintended data leakage occur, the approach is to make changes to prevent further data leakage and communicate the situation to users. However, the participant also noted the lack of interaction with Google in such cases, saying: "Usually what happens is that a change is done to avoid this type of leakage of data. Then you create a communicate. Yeah, you do a communication to your users and say this happened and tried to be transparent about what happened. I haven't seen a lot of interaction with Google in this case, like talking with them and ask them to not use the data."

Participant 3 reported a critical incident where form data containing personal information was sent to Google Analytics. The reaction to this incident involved downloading the log file, removing the personal data from Google Analytics and the downloaded file, and ensuring all original files were deleted. The quote from Participant 3 outlines this process: "So we were, uh, what we did was downloading the low file and removing the data from Google Analytics and removing the queries from the downloaded file and make sure all the original files were deleted and that until that part I can remember...". Similarly, Participant 1 referred to an example where the company removed specific data points to comply with GDPR. "The company is going to be more GDPR-compliant and then they kind of removed some data points that we thought that were available."

4.8 Other Practices to Enhance Privacy, Utility, and Compliance

4.8.1 Advocacy for change from legal authorities and Google

The participants expressed a range of opinions and recommendations for addressing the privacy vs. utility issue in implementing Google Analytics that should not be on their end only but need a concerted effort from the legal authority and Google themselves. Their suggestions included more visible warnings, stricter regulations, clearer guidelines, and the need for privacy measures to keep pace with technological advancements. These findings highlight the complexity of balancing user privacy and utility in the context of analytics tools like Google Analytics and the ongoing challenges faced by developers and analytics specialists.

From Google:

Participant 4 highlighted the issue of user privacy and the presence of private information in Google Analytics data. They suggested that Google should take immediate

action to address this concern. They recommended that it should make it inaccessible to anyone working on it whenever Google Analytics detects private information. This would involve stopping storing user information in plaintext format and ensuring it is appropriately hashed or encrypted. "I would expect, you know, that on your main Google page or wherever, you know, related to Google Analytics, you would have it like blazingly read, Hey, you are actually collecting user information in a way that is very accessible to possibly other people." They recommended that Google take immediate action to address the issue of private information being collected by Google Analytics and ensure that it is not stored in plaintext format (Participant 4).

Participant 2 also emphasized the need for Google to make improvements. They proposed that Google should provide clear warnings and notifications about collecting user information accessible to others. They suggested that Google should take privacy problems seriously and highlight potential privacy issues to users (Participant 2).

From the Law:

Participants acknowledged the challenging task of striking a balance between privacy and utility. Participant 6 mentioned the evolving nature of technology and the need for regulations to catch up with advancements. They highlighted the importance of laws and regulations to ensure that user data is handled appropriately and that privacy is maintained."We want a free internet where everyone is free to do whatever they want and at the same time, we want a secure space where we know that everything that is happening is according to the laws and nobody is getting harmed." (Participant 6)

Participant 3 discussed the conflict between data privacy regulations and data analysis practices. They mentioned the existence of US laws that allow government agencies to access data held by companies like Google. They noted the lack of major decisions or agreements to address these concerns, indicating the need for more substantial measures "The most important thing the countries and courts are asking is the existence of US law... to the government agency such as IP, FBI can reach out to Google and take the data outside of that." (Participant 3).

Participant 8 expressed that while they do not fully agree with all measures suggested by CNIL, they do acknowledge the importance of privacy, stating, "I think it's OK to care about people's privacy because you don't know...What people are using the data for, but it should not be over the top because yeah, there's no harm in measuring somethings as long it does not lead to a fingerprint like that." Besides, Participant 8 also expressed the view that the government should play a significant role in setting clear regulations and standards for privacy. They emphasized the need for agreements and adherence to those regulations by parties like Google Analytics to reduce the burden on individual users "So I think the government should add clear regulations like come to an agreement about that and the party said offer analytics should adhere to that. So don't bother me as a very low level user of that. That would be best practice for me" (Participant 8).

4.8.2 Non-technical practices like organization training and documentation

On the practice of documentation when using Google Analytics, participants' responses revealed a pattern of inconsistent practices.

Participant 7 indicated that while updates were documented, no formal documentation on privacy regulations had been established. They stated, "We do document any updates, but we haven't documented compliance issues yet at the moment because we if we do any new dashboard, we documented what we've done. But in terms of privacy regulations so far, we haven't done anything in this direction."

Participant 4 reflected a lack of awareness and nonchalance toward documentation about privacy, saying "from a data engineer perspective, I have never seen any documentation about it". Participant 3 shared they had only written documentation for

privacy-related tools but not legal documentation. “I yeah, I’ve documented, I’ve written a documentation for privacy related tools such as one or two or something like that. It’s Yeah. But that, that is more like actually the way of using of the tool. So it wasn’t necessarily related to legal documentation because I’m not a legal person at all.”

Participant 5 revealed a lack of personal experience with such documentation but acknowledged its potential utility. “No. I haven’t done it myself. Yes. Yes. But given the training.” Participant 8, too, acknowledged a lack of such documentation, mentioning it might be useful under certain circumstances – “Maybe if things set up gets complex, and you do actually use some personal identifiable information, and if multiple people are working on it, but at the moment I’m the only one who implements Google Analytics, so I know what I’m doing.”

On the other hand, the interviews also revealed that company practices around privacy and compliance documentation vary significantly, influenced by factors like the nature of their business and the level of perceived risk. Participant 2 described how their company had strict documentation practices for their customers but laxer practices for potential customers – “Our customers have a sort agreement, Do you know that term? It’s a privacy agreement between, you and so, so with our customers, we have a very strict written and signed document that says, I will treat your data like this. With the customer data in the marketing channel, we are, um, yeah, dealing with potential customers that we do not know. We have no contacts and to be honest, no interest in engaging in a legal, uh, yeah, records and even the cookie clicker that you have to accept.”

Participant 1 pointed out that documentation and compliance were the responsibility of a separate role in their organization – the data governance to: “I believe it is well documented. But then again, I didn’t go much into it there. So that’s another person that goes into the is a governance data governance manager.”

However, a pattern that emerged across several interviews was the recognition of the potential benefits of documenting privacy and compliance efforts. Participant 3 stated, “The most important thing I guess is getting, keeping documentation on what we collect and, what we Yeah, collect like dimensions or events. And because that makes it easy to review, uh, when privacy related incidents happen or registration changes or. Yeah. In case we want to improve the privacy related stuff.”

Participant 1 acknowledged Google Analytics as one of the better-documented tools concerning user privacy and GDPR, suggesting that it does enhance user privacy and makes data analysts more aware of privacy considerations – “I think GA it’s just one of the tools that have a better documentation of user privacy and which GDPR. Yeah. A compared to other tools. So yeah, I believe it is well documented.”

4.8.3 Different GA setup per sites/regions increases privacy and compliance

There was a consensus among the participants that there were notable differences in how data was handled and collected in Europe compared to other regions, which is mainly due to compliance purposes. Collectively, the findings in this Chapter indicate a range of practices and attitudes towards balancing privacy and utility in Google Analytics implementation, impacted significantly by regional regulations, particularly GDPR. The general consensus among participants was that it is a necessity to adapt and evolve practices to maintain compliance, though this is managed in various ways across different contexts.

Participant 7 offers a clear picture of the difference between data collection inside and outside Europe, highlighting the impact of GDPR regulations on American companies and the necessity of data localization, meaning data of EU users should be stored within the EU. They claimed, “When we work with American companies, they’re very scared

about the GDPR and all these regulations here in Europe. Even within California, which is the CCPA, they're like, it's nothing compared to what we do here in Europe. If you own data from users in Europe, it needs to be collected in Europe."

Participant 6 affirms this distinction, focusing on the stricter data collection rules enforced by GDPR and the changes it brought about in multinational companies. According to Participant 6, "Because I was working with a company that operated in five Latin American countries. And yeah, there the legislation is a little bit less on this type of data. And the headquarters was in Europe. So even though in, in the local countries they don't have any issue with the data that they are collecting, they have the GDPR laws in Europe."

Participant 1 suggests a potential solution to the privacy-utility challenge that is adopted in their company by combining multiple analytics tools instead of relying solely on Google Analytics. They believe this approach allows their company to have less dependency on Google Analytics and could help in the event of GA being banned in more regions, as they feel more compliant tools are available and can replace GA. They responded, "I believe that combination of tools is better to go forward. Um, replace GA... maybe, but currently we are using depending on GA to do campaign analytics."

Participant 2 discussed their company's differentiation in data collection between public and private domains. They explained, "We have a second channel that is our customer channel, that is we deliver software for um, associations to, to manage their, their, their member data. So we have very sensitive data in our systems and Google Analytics is banned there any, any external party or any external data has been very strict policies."

Lastly, Participant 8 presents a unique case where the same data collection approach, namely using a cookie bar, is applied universally, without differentiation between EU and non-EU users. This approach is due to the impracticality of maintaining separate systems. In their own words, "Now we just have a cookie bar that you can use accept or decline, and it's for everyone, not just European users."

4.8.4 Implementing your own tracking system increases utility and privacy compared to third-party plugins

From the data, it is clear that implementing a unique tracking system is an option that resonates with certain participants. However, the opinions vary, reflecting differing attitudes towards this approach.

Participant 7 perceived this strategy as disruptive to the free nature of browsing, viewing the obligation to create an account for tracking as potentially off-putting for users. They explained: "This is again, for me, not the right thing, because as I say, if you ask the user to create an account, you ask them to feel quite a lot of information, at least email and names. And for me that my results some people to leave the website."

On the other hand, Participant 4 viewed this as a feasible method to stay on top of changing trends. They shared an example where a company successfully implemented their own tracking system: "They already invested some years ago into this entire info platform to actually promote users to use their services and across devices...they managed to create really the platform that is really understandable for customers."

4.9 Influence of Business Practice, Law, and External Factors

4.9.1 Big Tech companies and experts shape GA implementation

Based on the interview participants, it is noticeable that entities such as Big Tech corporations and industry specialists exert considerable influence over the method in which Google Analytics (GA) is integrated. Participants are found to depend on an

array of sources to keep abreast with alterations in the legal landscape and best practices pertinent to privacy and GA application.

As conveyed by Participant 7, industry experts and updates from Google Analytics through online mediums such as blogs and community forums are pivotal to maintaining cognizance of privacy regulations. The participant underscored the significance of being well-informed about privacy rules and remarked on the prevalent provision by large websites for users to either consent to or refuse to be tracked. According to the participant, the implementation of cookie consent banners on websites has evolved into a convention, with users adapting to the notion of being tracked while utilizing websites. However, they acknowledged that complete adherence to privacy rules is not a universal practice among websites."Oh, I'm following certain people, the experts I'm following also Google ads, Google Analytics sorry, updates in announcement blog. And I've been in certain groups and circles that they're within the marketing experts talking about this."

Participant 4 recognized the vital role that privacy regulations like the GDPR have played in curtailing the number of individuals being tracked. They also highlighted the fact that tech giants such as Apple had introduced browser technologies that restrict the duration of cookie life, thereby creating challenges in tracking user paths over protracted periods. This limitation could potentially impact the analysis of customer journeys, particularly in the context of more drawn-out processes like car purchases."I mean, look, GDPR is the first point, right? Where where they ask people, you know, all the companies to start using the cookies, banners and all that, right. This already led to a smaller amount of people actually being tracked. And then you have those solutions that the tech giants are doing. Apple, to name the biggest one that's kind of created their browser technologies in a way that still kind of limit, you know, the lifespan of a cookie."

For Participant 8, maintaining an up-to-date understanding of the interplay between GA and privacy involved perusing blogs and seeking customized courses."I read blogs about Google Analytics and privacy things, yeah.","I tried to follow a course, a custom course to have explained to me how does it work".

4.9.2 GA is popular and effective despite being less GDPR-compliant

While GA remains a powerful tool, there are increasing concerns over its privacy implications, leading some to explore alternatives or adopt multiple tools to minimize dependencies on GA.

Participant 7 notes the dominance of GA, describing it as a "monster in this thing". They point out that they have experimented with other alternatives, such as Yandex, but due to the extensive capabilities of GA, they haven't completely transitioned. The participant suggests that if privacy issues escalate, especially within the EU, they would consider recommending alternatives if GA does not adapt to these changes.

Participant 4 expresses concerns about Google's monopoly in the market but acknowledges other alternatives. They mention a specific company advertising itself as completely GDPR-compliant, although they did not recall its name. They affirm the possibility of other viable options but point out that these are not as popular, and their integration and documentation are less comprehensive than GA.

Participant 6 emphasizes the issue of a lack of alternatives to GA. They explain that although there is a demand for alternatives, especially in light of stricter GDPR regulations, a robust alternative that covers all the requirements, like GA is still not existing in the market. They propose that this situation may stimulate the market to create new services that could be viable contenders to GA.

Lastly, Participant 1 suggests using a combination of tools for user behavior tracking rather than solely relying on GA. They point out that GA is mainly used for campaign

analytics in their company, while different tools handle other tracking and analytics tasks.

4.9.3 Technical difficulty and financial cost influence privacy and utility balance

Participants express that cost and technical difficulty are factors considered when implementing practice to balance privacy and utility when using the tool. Participant 7 indicated that utilizing a Proxy Server might require additional financial resources, making Google Analytics less of a free tool and raising the question of the cost-effectiveness of such a solution. They also suggested that this could prompt organizations to explore other paid tools that might provide a safer environment for user privacy. Furthermore, the participant noted the ever-evolving legal landscape, implying that frequent adaptations might be required to keep up with regulations.

"So Google Analytics turns out not being a free tool anymore, which there now is the question, maybe if this is a lot of free tool, maybe there is another tool that either you gonna pay which will give you a much more privacy safe environment because next month you might have the French government might have another request, which I need to do another thing to solve and so on and so on." (Participant 7)

Participant 4 provided insight into the shift from Client-side to Server-side tracking. This change is presented as having both positive and negative implications. On the one hand, Server-side tracking improved user privacy but on the other, it reduced the utility of collected data. "I only kind of heard people who are busy, you know, transitioning from client side to server side...it definitely brought a hassle in terms of the utility, but which at the same time means an improvement for privacy." (Participant 4)

Participant 6 recounted experiences working in both European and Latin American markets. They noted that privacy legislation, such as GDPR in Europe, and other laws in different regions led to a shift in data collection practices. As a result, this participant highlighted the cost implications and strategic adjustments required for the practice. "So this change, it's positive, but at the same time has some hidden costs for the companies to comply with. Yeah." (Participant 6)

Participant 3 voiced a preference for Server-side tracking due to privacy considerations, highlighting the opportunity it provides for reducing the number of data dimensions. However, the business cost implications were also mentioned as a potential downside. "I definitely consider Server-side is better... Only the downside would be the cost, because it will require more budget from the companies collecting data." (Participant 3)

4.9.4 Influence of the company's business on using GA

Participants' responses highlight the impact of the company's own business needs on the practices, and this can include the requirement from their clients to how their marketing team heavily relies on Google Analytics for market analysis or their business contract.

Companies' clients present a major influence over how the companies themselves implement certain practices. For example, Participant 7 highlighted the consideration of alternative tracking plugins, such as Yandex Metric, for measuring website analytics due to client pressure. They recommended exploring other software if Google Analytics does not address privacy concerns adequately. The participant also mentioned that the company's expansion into different countries and compliance with privacy regulations could influence the choice of adopting alternative tools.

Participant 2 also presented the same idea and explained how their company differentiated between public and private sites. They used Google Analytics extensively for the public domain but had strict policies and agreements to protect sensitive data

in the private domain where their clients are located. The company's business model influenced the choice to use Google Analytics selectively. Participant 2 also mentioned the challenge of maintaining privacy while running a website that involves advertising and marketing. They expressed a desire to minimize data collection and avoid tracking personal information, focusing on collecting only necessary attributes for marketing purposes. The participant highlighted the importance of user consent and voluntary provision of data.

Company culture is also a factor that needs to be taken into consideration. Participants 3 and 5 discussed the importance of the company culture when balancing data collection and user privacy. They mentioned that some individuals within the company might prioritize collecting more data while others value user privacy. Balancing these perspectives is crucial for achieving the desired utility without compromising privacy. Participant 5 stated that the companies they worked for prioritized collecting as much data as possible, emphasizing utility over privacy. However, they acknowledged that privacy considerations should be followed if mandated by the law. Participant 2 acknowledged the resistance from their business executives when rejecting invasive tools like Hotjar (a third-party tracking tool like Google Analytics). They emphasized that the perceived value of extensive data collection often does not align with the actual utility gained. The company preferred a selective approach to data collection, respecting user privacy.

How the company is structured also plays a role in the decision to adopt certain measures. For example, Participant 4, a data engineer, admitted to the limitations of their role in the process of implementing tracking systems. As they stated, "I didn't face any challenges, right? I'm not implementing tracking. This is this is completely different teams. I process whatever is made available to me." They did, however, identify an incident where user private information was directly observable in the raw Google Analytics feeds, an issue they rectified by notifying the tracking implementation team. In the company of Participant 1, they stated that The role of data analyst lack an involvement in monitoring privacy-related legal updates, stating that"there are the legal people that may keep track on it but as a data analyst we just being told that that that you cannot use this or that they can use." This implies segregation of responsibilities where data analysts focus on analysis, leaving legal matters to dedicated personnel.

When a company integrates GA with other Google services, this integration also raises privacy concerns and potential financial costs when considering adopting practices for maintaining the balance. Participant 2 exhibited a more skeptical view of Google, expressing concerns about its data collection practices. They noted their reliance on Google's AdWords, stating that withdrawing from GA would negatively affect their marketing efforts due to the interconnectedness of these Google services. "And my business model is to serve my clients...And so AdWords needs analytics or the same data to, to function...And unfortunately, Adwords is our number one marketing channel. We spend many thousand euros a month on Google AdWords. So if we would withdraw from analytics, our marketing would be then would vanish."

Participant 7 outlined the need to adapt GA's settings based on unique business needs, considering factors such as traffic sources. However, they mentioned the limitations related to data transfer when integrating GA with other Google products, such as Google Data Studio or Looker Studio, revealing Google's strategy of charging for such services. "Yes, for sure. What you need to do first, you need to migrate all the events that you're tracking, depending on the is it a web or app?...Second, it's I know that it's integrated very easily with Google Data Studio or now so-called Looker Studio, but it has some limitations on the amount of data you can transfer because, again, Google wants to charge for this service." (Participant 7). Participant 4 also added the improvement in their job as a Data Engineer when GA4 comes with integration with Big Query,

which makes the tool more efficient than previous technology. Similarly, Participant 4 also noted the improved integration with BigQuery, making data ingestion and analysis more efficient than previous technologies "integration of the service called big query, which is the most important Google's query engine, and it simplifies the data ingestion and analysis a lot compared with the previous technology"

4.9.5 Legal requirements have a significant influence, but it is challenging to keep up with them

The influence of the legal authority on the practice of Google Analytics appears to be significant in shaping both individual and company perspectives, thereby affecting the decision to adopt and use this tool. Several participants expressed their perspectives on this matter, and the impact of the law is reflected in different ways:

Some major changes brought by the legal authority have been reflected in the change in Google's product. Participant 3 highlighted how the GDPR's influence has led to changes in Google Analytics, with features such as IP randomization becoming mandatory in GA4. The participant also mentioned concerns about data privacy shared between the US and EU, demonstrating the influence of international data protection laws on adopting and using analytics tools. The main concern here was data being accessed by US government agencies like the FBI. "Depending on the which was the concerns on Google Analytics, I would say because the use your GDPR-related stuff, Google actually is that the most important part was such as IP randomization because that was kind of optional one for universal analytics, but it was introduced to mandatory ones for GA4 forward. But looking inside deeply, the core question regarding the banning of Google Analytics was not that because the most important thing the countries and courts are asking is the existence of US law." (Participant 3). Participant 6 expressed the evolving nature of privacy regulations and the challenges development teams face when migrating to new versions of Google Analytics, such as GA4. They mentioned stricter tagging requirements and difficulties in obtaining data from certain devices, such as iOS 14 devices, are the result of the requirement. The constant changes in privacy policies and technology create a challenging environment for maintaining the balance between privacy and utility. "So yeah, getting useful data is getting more difficult. Yeah, I don't know if that's better or not. Yeah, from the user perspective, I think that limiting how the data is gathered and having clear rules of that is good. But on the other side, I have been on the developer side, and not being able to gather this data hinders the development process as well. So yeah, I don't know if it's better or not" (Participant 6)

On an industry level, Participant 5 forecasted the potential impact of future regulations, suggesting that global tech giants like Google might face stricter regulations to ensure consumer protection. The participant saw this as a positive development. "Um, well, I think on a global level, you guys are going to get hit with the regulation that, that the EU threatens. In other words, you have too much power over the consumer. So, um, there will be regulation to, to keep things fair." (Participant 5) Participants highlighted the impact of privacy regulations such as the General Data Protection Regulation (GDPR) on data collection and analysis. Participant 4 mentioned the implementation of cookie banners and browser technologies that limit the lifespan of cookies, thereby reducing the ability to track users across more extended periods. This limitation affects the ability to understand customer journeys, especially for longer processes such as purchasing a car. These challenges demonstrate the influence of legal regulations on data privacy and utility in Google Analytics. "GDPR is the first point, right? Where they ask people, you know, all the companies to start using the cookies, banners, and all that, right. This already led to smaller amounts of people actually being tracked. And then you have those solutions that the tech giants are doing. Apple, to name the

biggest one, that's kind of created their browser technologies in a way that still kind of limits, you know, the lifespan of a cookie" (Participant 4)

The legal requirement also shaped how companies' vision regarding the use of such plugins, Participant 1 expressed that their company was willing to use a tool more compliant with the GDPR than Google Analytics, demonstrating the direct influence of legal regulations on the choice of analytics tools. "Um, we well, in my current company, yeah. It's fair to use another tool that is more compliant with GDPR. Instead of Google Analytics." Participant 8 conveyed a sense of waiting for more explicit guidelines from the legislation before deciding to use alternative tools or methods, such as Proxy Servers, to address legal issues around Google Analytics. "I think we're currently waiting for legislation because there's so much to do about it and we don't want to implement all kinds of things before we know what what they actually going to stay and decide about that." Participant 1, as a data analyst, acknowledged the challenges in balancing data utilization and compliance with privacy regulations. They mentioned the amount of data exposed due to GDPR, which sometimes limits the information available for analysis. This balancing act requires collaboration with data governance managers and compliance teams to ensure the collection of legal and relevant data for analysis. "As a data analyst, one of the challenges is the amount of information being exposed... Sometimes I feel that I didn't get enough information of what we're trying to achieve, of the hypothesis we're trying to test, the questions that we're trying to ask. So there are some balancing acts definitely from data utilization and data compliance" (Participant 1)

The influence has not always been perceived as positive. Participant 8 highlighted the challenge of measuring campaign effectiveness while complying with privacy regulations. They mentioned the importance of tracking results to optimize advertising spending. Balancing the need for data with the financial implications of ineffective campaigns becomes crucial in the decision-making process. "We don't want to blow all money on advertising, not knowing if it has any effect... We really hope we can still track the results of our campaign. And that way, we do have the link between Google Analytics and Google Ads because we think that's just necessary in other cases" (Participant 8). Apart from that, Participant 8 mentioned the limitations and challenges arising from a lack of privacy-related knowledge and limited guidance from legal authorities. Participant 8 expressed frustration about the limited knowledge of legal experts regarding Google Analytics and the difficulty in receiving proactive advice. They emphasized the need to figure things out independently and the challenges of understanding the changing landscape of privacy regulations. "I've tried to understand the rules as well as I can, but I'm not a legal person... It's really hard to communicate without each other... It's possible that I made a mistake, and I find that very frustrating" (Participant 8)

5 Discussion

This chapter will present the analysis and arguments supporting or contrasting the related work and framework. The chapter revolves around the purpose of the study, answering the research question and proposing the direction for future research, as well as the limitations.

5.1 Understanding the Regulatory Challenges and Opportunities for a sustainable approach to Privacy Protection

Current privacy regulations like the European Union's GDPR have marked considerable progress in securing user privacy, with evidence of awareness among the participants.

Nevertheless, the precarious equilibrium between privacy and utility poses a constant challenge. The researcher argues that these regulations are often deemed complicated and insufficiently address the challenges between the two aspects, leading to diverse interpretations and implementations among businesses, potentially influencing both privacy and utility. Moreover, gaps are evident in the application and enforcement of these laws. Many organizations struggle with the effective implementation of privacy strategies and transparency, and regulatory authorities often lack the resources to solve these problems on a case-by-case basis.

Furthermore, the researcher also notices that rapid technological advancements are often one step ahead of the legislation, thus creating potential loopholes and ambiguities. For example, the emergence of Server-side tracking or the utilization of AI and machine learning for data analysis suggested by participants presents new privacy challenges not fully covered by the current regulations. While current policies and regulations have created a robust foundation for privacy protection, the researcher believes their effectiveness could be enhanced through improved legal enforcement, clearer business guidance, and constant updates to tackle emerging privacy concerns. The researcher argues that focusing on more precise definitions in upcoming regulatory updates, especially concerning cutting-edge technological advancements and data utilization, could bring additional benefits. For instance, "anonymous data" and "user consent" are concepts that could benefit from further elaboration to ensure a consistent application across the board.

In the search for a sustainable approach to the Privacy-Utility Dilemma, the research identifies multifaceted variables such as the nature of collected data, objectives of data collection, and various legal, cultural, and societal norms in balancing privacy and utility. Thus, the researcher believes a singular "one size fits all" solution appears unlikely to address this immense spectrum of considerations comprehensively. Organizations should instead be encouraged to mold their strategies to suit their specific contexts. For instance, the necessity of privacy protection for a healthcare provider handling sensitive medical data could be more stringent compared to a retailer examining customer shopping patterns. Furthermore, compliance with different privacy norms, such as GDPR in the European Union, necessitates contextual adjustments. Hence, while a universal solution may not be practicable, disseminating best practices and guiding principles can offer substantial benefits across various contexts.

The ongoing discourse surrounding privacy and utility is poised to have several long-term ramifications for businesses and consumers alike. A potential shift towards more privacy-focused models and technologies may be necessary for businesses, requiring investments in privacy-enhancing technologies, robust consent mechanisms, and data minimization practices. These changes could impact their data collection and analytical capabilities and grant competitive advantages in an increasingly privacy-conscious market. This shift could simultaneously spur demand for privacy expertise, fostering the growth of privacy-centric roles and teams.

For consumers, heightened privacy awareness by legal authorities could precipitate changes in online behavior. Users may become more judicious about their shared data or gravitate towards platforms and services prioritizing privacy. However, there is also a risk of 'consent fatigue', potentially leading to a less meaningful interaction with privacy options.

Lastly, the privacy-utility debate will continue to shape regulatory norms and standards, influencing both businesses and consumers. Moreover, the rising prevalence of anti-tracking measures by browsers and operating systems could diminish the efficacy of some current practices over time. Therefore, despite the viability of current privacy practices, significant modifications or overhauls may be needed in the future to ensure sustainability and compliance.

5.2 Google Analytics: Striving for a Balance between Utility-Privacy issue and Navigating the Controversies of Server-side Tracking

The transition to Google Analytics 4 (GA4) marks a notable shift in the landscape of data tracking and processing. As this research indicates, GA4, supported with a range of features such as enhanced data models, AI-driven insights, and improved cross-device tracking, presents significant utility. Nevertheless, this enhanced utility has not deterred criticisms concerning perceived privacy limitations and transition-related support. The duality of GA4's utility and privacy is strongly contingent on its implementation and utilization by businesses. Consequently, the researcher believes that the perceived benefit of the GA4 migration is largely subjective among the participants, with dependencies on individual business requirements.

The interpretation of GDPR requirements is somewhat subjective, leading to varying notions of compliance among the answers from participants. Google Analytics has adopted a more proactive stance towards GDPR compliance by anonymizing IP addresses and availing data deletion tools. However, reservations persist regarding the sufficiency of these efforts. One such concern pertains to the effectiveness of IP anonymization in entirely disidentifying user data, given data recombination's potential for individual re-identification. Consequently, the researcher argues that the need for more effective anonymization methods and user data usage transparency is mandated from Google's side.

An interesting remark by Denmark's DPA notes that "For Google Analytics 4, it is apparent from Google's documentation that IP addresses are used to determine the approximate location of the visitor, after which the address is discarded before the data is logged to a server. As with Universal Analytics, the same issue is also relevant for Google Analytics 4, as – depending on the location of the data subject – there can be a direct connection to, among others, American servers before the address is discarded" (Datatilsynet, 2022) This practice is concerning as it shows that despite not collecting IP addresses in the EU region, data may still be potentially exposed. The researcher is of the opinion that Google Analytics has not done enough for the privacy issue and could have done more; Google should apply privacy-focused practices globally, transcending regional regulations and compliance requirements. After all, privacy should be a fundamental user right rather than a compliance checkbox for each region.

As a prominent data-driven service, Google Analytics should be more transparent about user data usage as participants in this study have expressed their lack of knowledge of whether Google will use the data for other purposes, also when presented with a scenario of data leakage, the practice of "data deletion request" through Google was not widely known among participants, suggesting a knowledge gap within the Google Analytics user base. The researcher's personal experience when browsing through the help documents to identify the privacy-enhancing strategies also found that how it is organized and scattered around may be one of the possible contributors to this knowledge deficit. The researcher asserts that how Google communicates with Analytics users still needs improvement in terms of presenting clear and concise information on data usage and security measures in a well-ordered manner in the support document. This could potentially increase trust and confidence in the platform, leading to better adoption of practices that enhance privacy within the companies.

A perfect equilibrium between utility and privacy in Google Analytics is a daunting yet achievable prospect. Considering Google's position as a global tech giant, the expectation for significant privacy advancements within Google Analytics is high. However, it is essential to remember that 'enough' is a subjective term in the dynamic field of privacy. Google can further bolster privacy within Google Analytics by offering more

comprehensive data control to users, augmenting data usage transparency, and providing ample resources for businesses seeking privacy compliance.

On the other hand, Server-side tracking, one of the recent advancements, has brought enhanced privacy and utility to the table by bypassing Client-side limitations and restrictions. However, these benefits are juxtaposed with financial and technical challenges that cannot be ignored. In the researcher's view, the potential privacy and utility gains might indeed outweigh the difficulties associated with implementing Server-side tracking. However, it is essential to recognize that these benefits aren't uniformly distributed among firms and businesses; organizations with the necessary resources and expertise are likely to gain more than those lacking such capabilities and stay ahead of the trend.

The feasibility of transitioning to Server-side tracking from the Client-side one, on the other hand, significantly depends on the developers' technical proficiency and the resources the companies can provide. A clear pattern can be comprehended from the result from Participants 3, 4, 6, and 7 that Server-side tracking, indeed, requires a more technical capability and development workload compared to Client-side tracking. As a result, this also implies additional costs associated with the maintenance of the feature. Even so, the researcher believes that for developers committed to enhancing privacy and willing to invest in the necessary training and resources, transitioning to Server-side tracking is not just feasible but highly advantageous. Firms can have more accurate insights and are less likely to be fined due to more privacy control will eventually outweigh the cost of implementation.

Server-side tracking does spark some controversy, primarily because it is viewed as a way to circumvent browser-based anti-tracking measures (the case is brought up by Participant 7). It operates by shifting data collection and processing from the client's browser to the server, effectively replacing browser cookies with server cookies. This approach can obscure tracking activities from end-users and anti-tracking tools, raising questions about transparency and consent. Furthermore, there is an ongoing debate about whether Server-side tracking should require user consent. While Server-side tracking can enhance privacy by providing better data control, it might be considered infringing user privacy if employed without transparent user consent. The researcher advocates that regardless of the tracking technique, consent should always be sought and respected, underlining the ethos of privacy by design and aligning with regulatory requirements. This statement also supports the suggestion by Papadogiannakis et al. (2022) that this type of tracking (referred to by the authors as stateless tracking/ cookieless tracking) needs to raise more awareness when it comes to this issue, and there needs to be a change in how the consent request banner is linguistically presented so that it can reflect what the site is collecting, inform the users with clarity and being GDPR-compliant.

5.3 Rethinking about the relevance of using Cookies Banner to obtain user consent for third-party plugins

While cookies and consent mechanisms are typically aimed at augmenting user privacy options, their efficacy is a matter of ongoing discussion. Cookies and consent dialogs can potentially offer users increased control over their data. Users can choose to accept or decline cookies and express their data collection preferences via consent dialogs. Nonetheless, the application of manipulative design techniques, known as "dark patterns," to secure user consent for data collection remains a controversial issue in the data analytics sphere. Deploying such mechanisms inherently contradicts the principle of informed consent, subverting user autonomy. From an ethical standpoint, the researcher agrees with the view of participants on this matter that users should be provided with clear, understandable, and unambiguous options when it comes to their data

privacy. They should be able to make informed decisions about their data without being manipulated or misled. This research calls for not only a regulatory response but also urges companies to self-regulate and foster ethical design principles that respect user autonomy.

Conversely, the over-utilization of cookies and consent mechanisms may also lead to a decrease in data utility. Users often block or clear cookies, leading to potentially incomplete or skewed data. Furthermore, consent dialogs can prompt 'consent fatigue,' a phenomenon where users indiscriminately accept dialogs without understanding the associated implications, undermining the intent behind obtaining informed consent. Therefore, while cookies and consent mechanisms can serve as essential tools for enhancing user privacy, their effectiveness largely hinges on their implementation.

Even though users play a crucial role in managing their data-sharing preferences, answering whether tech companies should be subject to stricter data privacy guidelines or the user should proactively manage their data-sharing preferences may be seen as a complex question. The researcher takes the stance that the responsibility for data protection should not be fully entrusted to users. The majority of users do not possess the technical know-how to comprehend the consequences of their data-sharing decisions fully. Furthermore, the available options are often not transparent or user-friendly. Consequently, tech companies, due to their technical expertise and the potential impact of their services on privacy, should be subjected to more stringent data privacy guidelines. These guidelines should accentuate transparency, data minimization, and meaningful user consent for the users to be equipped to make informed decisions, and tech companies should hold the primary responsibility for data protection.

The relationship between cookie consent practices and data utility is complex; however, this research suggests that these practices may lead to a perceived decrease in data utility. When users are offered an informed choice, many opt out of non-essential cookies. This leads to lesser data collection and potentially impacts the volume and richness of analytics data accessible to businesses. This could potentially hinder businesses' ability to conduct comprehensive user behavior analyses, personalize user experiences, or optimize their services based on data-driven insights. However, it is noteworthy that this relationship may not always be disadvantageous. Enhanced user trust through transparent consent practices could potentially boost user engagement over time. Furthermore, the shift towards obtaining consent may stimulate innovation in privacy-enhancing technologies and analysis methods that maintain utility while respecting user choices.

5.4 Evaluating EU-Based Tracking Plugins as a Potential Solution

Alternative EU-based tracking plugins, like Matomo, claim to offer better privacy safeguards, considering they are designed to adhere to stricter privacy norms like GDPR. However, their effectiveness as privacy solutions is not a simplistic conclusion. While these alternatives indeed offer potential privacy enhancements, their adoption hinges on factors such as functionality, ease of integration, and user-friendliness, as noted by the participants when comparing Google Analytics with its competitor. Businesses might find certain features offered by Google Analytics missing in these alternatives or encounter difficulties in data migration or employee retraining. Thus, the researcher is of the opinion that EU-based tracking plugins could contribute to the privacy solution but are unlikely to serve as a universal remedy. The decision to employ these alternatives necessitates a thorough understanding of their privacy benefits and potential limitations.

In the context of stricter privacy regulations like GDPR, the utilization of alternative EU-based tracking plugins could bolster privacy compliance. Designed to comply more closely with European privacy norms, these plugins could potentially help businesses

avoid legal entanglements related to international data transfers and consent norms. However, banning Google Analytics to encourage developers to adopt these alternatives may not be appropriate. The choice between Google Analytics and alternate plugins should be determined on a case-by-case basis, considering the unique requirements, resources, and constraints of each organization. Taking Participant 6's company for example, it may not be practical to force them to switch to privacy-enhancing plugins that are EU-based when they are functioning on different continents. For other companies that integrate Google Analytics with Google Ads to run advertisements as their business model, Participant 2 reported Google Analytics being banned also means companies who utilize those business models will get hit financially. Therefore, it is crucial to consider the specific circumstances of each organization before implementing any new regulations or policies related to data privacy. Additionally, it is important to provide support and resources for companies to make the necessary changes in order to comply with new regulations and maintain their business operations.

While some businesses might gain from the enhanced compliance offered by EU-based plugins, others might find it more practical to use Google Analytics while implementing supplementary measures to ensure privacy compliance. Furthermore, privacy compliance transcends the choice of tracking tools—it fundamentally necessitates a privacy-centric approach to data collection and utilization, achievable with a variety of tools, including Google Analytics and EU-based tracking plugins. However, factors limiting their widespread adoption persist. Google Analytics' comprehensive features and market dominance make it a tough competitor to substitute. Businesses and developers are deeply entrenched in Google Analytics and may be disinclined to dedicate time and resources to familiarize themselves with a new tool. The compatibility and integration capabilities of these alternatives may not match Google Analytics, which can effortlessly integrate with a wide spectrum of platforms and services. This could restrain their utility for businesses with intricate tracking needs or those heavily reliant on Google's suite of products. Further, issues related to cost, technical complexity, and availability of support could also inhibit the widespread adoption of these alternative tracking plugins.

5.5 The Implications of Business Influence on Privacy-Utility Balancing

Throughout the results, it can be seen that business influence has a significant role in shaping the practices of Google Analytics, acting as a demand driver for specific features and capabilities. The impact of this influence on privacy varies and cannot be classified merely as positive or negative. It relies heavily on businesses' privacy attitudes and how responsibly they employ Google Analytics. If businesses prioritize user privacy, their influence can foster more privacy-friendly practices. Conversely, if businesses utilize Google Analytics to infringe on privacy, their influence can be detrimental. The researcher believes that patterns in the finding show that businesses still have a tendency to gear the scale toward utility, and privacy in most participants are usually manifested through how well they are aligned with the legal requirements. Hence, the business influence, in this case, is believed to be more harmful than beneficial to privacy from the researcher's perspective. The researcher advocates that businesses must acknowledge the importance of both user privacy and data utility and leverage their influence to promote practices that respect and protect these aspects.

As major stakeholders affected by regulatory and public scrutiny, businesses bear considerable responsibility in shaping privacy practices. Yet, this responsibility should be executed in a manner that respects user privacy rights and regulatory demands. They should strive to maintain a balance between data utility and privacy assurance. This involves adopting privacy-by-design approaches, ensuring transparency about data

practices, providing robust user controls, and continually updating privacy practices in alignment with evolving standards. However, the ultimate onus of shaping and enforcing privacy standards should not rest solely on businesses. Other stakeholders, including regulatory authorities, technology providers, civil society organizations, and users, are also integral to this process.

From this research findings, it is evident that the capacity of businesses to handle the financial and technical challenges associated with their size and structure. Notably, smaller businesses may struggle more than larger enterprises due to limited resources and expertise. Larger organizations often have dedicated teams for legal and implementation, like Participants 1 and 3. While in Participant 8's case, the implementation, maintenance, and legally compliant handling are all handled by them, and they expressed their confusion and frustration on keeping up with the legal aspect. Moreover, the costs associated with maintaining compliance, implementing privacy-by-design methodologies, or transitioning to privacy-enhancing technologies like Server-side tracking can be substantial.

The researcher believes that smaller companies may not be adequately equipped with the financial and legal acumen to handle the challenge and requirements of privacy regulation. Any decisions or regulations should pay specific attention to these businesses, and provide support or resources to help them navigate the complex landscape of privacy compliance. It is important to ensure that these companies are not left behind or penalized for their inability to keep up with the legal aspect of privacy regulations.

Technical solutions alone are insufficient to address privacy concerns. This research indicates an emergent trend toward a hybrid approach, combining technical and non-technical measures, which can be the key to moving forward with this dilemma. This approach provides an additional answer to the study by Kröger (2022), in which the author also presented a range of similar problems when implementing privacy measures. The researcher is of the opinion that technologies like Server-side tracking and privacy-enhanced analytics tools are crucial, but still solely adopting them as checklists should not be enough. They should be supplemented with non-technical practices like organizational training, privacy policies, and documentation. As privacy concerns continue to garner prominence, businesses will need to adopt this holistic approach, viewing privacy not just as a technical issue to be solved but as a philosophical shift that must permeate all organizational levels.

5.6 Practices to maintain a balance between Utility and Privacy in Google Analytics

In light of the research question, the research has formed a list of several practices that can be employed to achieve this delicate equilibrium.

Transitioning to Google Analytics 4 (GA4): With the advent of GA4, the data analytics landscape has seen a notable shift that impacts both privacy and data utility. GA4 promises improved privacy measures, default IP anonymization, and an event-based data model that aligns with the ever-evolving industry standards and is compliant with privacy regulations such as GDPR. To take full advantage of these advanced privacy measures and analytical capabilities, transitioning to GA4 should be a high-priority undertaking. This updated version's customizable dashboards and integration with Google's BigQuery contribute further to its utility for businesses.

Appraising the Role of Proxy Servers: The use of Proxy Servers, although a seemingly attractive solution for anonymizing users' data, is largely unfamiliar to many industry professionals and is not broadly adopted. This unfamiliarity is one of the initial obstacles to their implementation. Skepticism regarding the actual impact of these servers on privacy is another area of concern, as Google Analytics does not store or

expose IP addresses in the first place, limiting the privacy benefits that a Proxy Server could theoretically offer. Further, the implementation of Proxy Servers introduces a significant challenge regarding data quality, particularly concerning accurate location data. Inaccurate data can have profound consequences for businesses requiring precise location-based information, undermining their competitiveness. These concerns are compounded by mixed opinions about how Proxy Servers might impact privacy. While the servers could potentially enhance user privacy, they might significantly limit the analytic capabilities that form the foundation of Google Analytics' value proposition.

Embracing Server-side Tracking with Google Tag Manager: The research findings indicate that Server-side tracking, as opposed to Client-side tracking, is a crucial consideration for enhanced privacy in Google Analytics. Although this approach demands technical know-how and a budget for implementation, the opportunity to control data collection to a greater extent is identified as a significant advantage, mainly when dealing with General Data Protection Regulation (GDPR) sensitive data. Server-side tracking can also be the answer to the future of browsers blocking the device's cookies. However, it should be kept in mind that the potential of Server-side tracking to effectively replace the need for a Proxy Server, as mandated by some legal authorities, seems uncertain.

Promoting Transparency through Consent Mechanisms: The practice of using cookies and obtaining user consent forms a crucial part of online data collection. Google Analytics' Consent Mode is a critical component that helps GA adjust its behavior accordingly to the choice made by the user. Additionally, the design and execution of consent banners can significantly shape user behavior; the prevalence of dark patterns in consent banner designs, however, can obscure this transparency and manipulate user choices. Although effective in garnering consent, these tactics may jeopardize trust and raise ethical concerns. The finding suggests that sites should present users with sufficient information about what is being tracked and presents them in a clear and concise manner, with options to reject certain cookies. It will thereby empower them to make an informed decision regarding their privacy and foster trust with the site.

Moreover, consent withdrawal and cookie rejection could impact the quality of data collected, potentially reducing its utility. Nevertheless, the decrease in data volume due to such rejections may not be substantial enough to hinder the derivation of insights, especially since the ones who reject are likely to be a very specific domain of users and may not represent the customer base. Companies should shift their focus to make use of the available data from those who consent to track by leveraging advanced data science techniques.

The Importance of Documentation: Creating and maintaining thorough documentation of Google Analytics use is fundamental. This practice aids in audits, educates stakeholders about data collection methods, and aids in identifying potential issues during data leakage incidents. Comprehensive documentation of tracked dimensions and metrics, as well as any specific privacy-enhancing configurations, should be maintained. Complementing this with organizational training can raise awareness among personnel about Google Analytics' usage and implications.

Limiting browser data collection and data minimization when using Google Analytics: Achieving a balance in this practice is a complex effort due to the absence of clear, universally applicable guidelines. The varied nature of businesses, industry-specific demands, organizational size, and other context-specific factors make it challenging to establish a one-size-fits-all policy for data collection and privacy protection. Businesses should first identify the operational needs that can be supported by specific data before engaging in its collection. Defining and documenting the dimensions of collected data can avoid excessive data accumulation, thus circumventing needless complexity in subsequent analysis. This can also help in case of an audit or a legal incident to prove

the "legitimate interest" reasons for the collecting. Overreliance on data collection may introduce noise and complexity and dilute actionable insights. Additionally, prioritizing user privacy demonstrates a commitment to ethical practices, fostering trust and loyalty among customers.

Effective Management of Personal Data Leakage Incidents: It is essential to note that prevention measures are more critical than post-handling ones, as personal data leakage is considered a serious incident among participants and can have severe consequences for organizations that involve the legal authorities, end-users, and other stakeholders to damage-control the situation. To prevent this incident from happening, the use of Google Analytics should be reviewed in a way that it should not be in contact with this information in the first place. One of the few places where personal data is usually presently mentioned by the participants is an open form, private dashboard like account information, and search bar where they can accidentally input their personal information, ... Data anonymization techniques, such as making sure data that could lead to the identification of a natural person (in User IDs, URLs or Referrers, ...) should be employed. Other than that, Server-side tracking with Google Tag Manager is also a good option that can become a checkpoint that monitors and anonymize the data before passing it to the Google server. Keeping a document about the use of Google Analytics and performing regular audits with the supported documents also helps lower the chance of having incidents.

For the post-incident handling process, taken from the experience of the participants, these can be taken step-by-step:

- Halt the data collection from Google Analytics.
- Identify where the problems come from.
- Notify the engineers who implemented the features to fix the issue.
- It is followed by submitting a data removal request to Google to remove the personal data from Google Server as soon as possible
- For compliance, perform a risk analysis regarding the impact of the issue and notify the legal authority within 72 hours of the incident.

It is important to note that a data breach does not automatically result in a fine; relevant supervisory authority of a data breach within 72 hours of becoming aware of it unless the breach is unlikely to result in a risk to the rights and freedoms of individuals. Failure to report a data breach within this timeline can result in significant penalties. Being transparent with the issue can help lower the chance of getting substantial fines from legal authorities and also build trust with the customers by addressing how the companies handle the situation.

Exploring Regional Tracking Plugins: The use of alternative tracking plugins located in Europe as potential substitutes for Google Analytics should be contemplated. Designed to adhere to stricter privacy norms like GDPR, these plugins may offer improved privacy safeguards for organizations mainly operating in the EU. However, a decision to adopt these plugins should account for functionality, integration ease, and user-friendliness. Balancing the benefits and limitations of each tracking plugin by assessing its compatibility with specific business requirements and integration with other resources is crucial.

Regional Tracking Plugins: If the company has not adopted Google Analytics yet, or the reliance on Google Analytics is still minimal (the use of GA does not involve the integration with other Google Products, or how the company generates profit does not require Google Analytics), considering an alternative tracking plugin located in

Europe may avoid the participants from getting hit by the privacy regulation. When comparing Google Analytics, it is essential to keep in mind that while Google Analytics is initially a free product, the optimal configuration to make Google privacy-safe needs to include the cost of implementation of Server-side tracking, additional anonymization, and different data structure based on regions. A comparison also includes limitations of each plugin, their document and support, integration capability, and development cost needs to be taken into context for a comprehensive analysis.

5.7 Limitations

This research, while offering in-depth insights into the intricate nuances of balancing user privacy and utility during Google Analytics implementation, acknowledges some inevitable limitations that have the potential to influence the findings and subsequent interpretations. The study's representativeness is tempered by a modest sample size, covering a limited array of participants with varying backgrounds and perspectives. Consequently, the breadth of experiences and viewpoints pertaining to the subject might not be wholly encapsulated within the findings. Future research should aim for a more substantial and diversified participant base to fully encapsulate the intricacies associated with balancing privacy and utility in the context of Google Analytics.

Another limitation of the study is that the self-reported data from interviewing introduces certain degrees of subjectivity that could potentially skew the findings and may not fully represent the actual picture. Likewise, personal beliefs or preexistent societal norms may influence the participants' perspectives rather than objective observations or experiences. Furthermore, another factor is that the potential for social desirability bias, where participants may provide responses that they believe are socially acceptable rather than their true thoughts or behaviors, could impact the accuracy of the data collected. This bias may lead to an underreporting of certain incidents or behaviors, thereby affecting the overall reliability and validity of the study's results.

Furthermore, the research predominantly revolves around participants with experience as developers, site owners, marketing specialists, and data analysts. While these insights are undeniably vital, a comprehensive understanding of the privacy-utility dynamics necessitates the inclusion of various stakeholders, including end-users, regulatory authorities, and legal personnel within the companies. Future investigations should aim to incorporate these broader perspectives for a holistic understanding.

Moreover, the context-specific nature of the research findings warrants consideration. The influence of the prevailing regulatory environment, technological surroundings, and cultural norms potentially confounds the generalizability of the findings to regions with different regulatory structures, technological capabilities, or societal attitudes toward privacy. Therefore, it is important to interpret the study's results within the specific context in which they were obtained. This contextual understanding can help researchers and policymakers make informed decisions about the applicability and potential impact of the findings in different settings.

Also, the focus of the research was chiefly on Google Analytics; thus, the distinctive challenges associated with alternative analytics tools and platforms might be overlooked. The conclusions and recommendations offered by this study are contingent on the knowledge and information available until June 2023. Given the rapid and perpetual evolution of privacy regulations, technological advancements of Google Products, and industry practices, some conclusions may be subject to obsolescence or incompleteness when these above factors change. Readers must remain vigilant in the evolving landscape and update their strategies accordingly.

5.8 Future research

Future investigations should direct attention toward emerging and transformative concepts such as "Pre-Transmission Hashing and Encryption of Data in Google Analytics" or "Bridging Consent Deficit through Data Sampling Techniques." These methodologies, representing the forefront of strategic privacy enhancement, call for a comprehensive examination to fully discern their utility. Indeed, there is a pressing need to scrutinize whether data sampling methodologies and "Google AI conversion modeling"[4] can efficiently compensate for the prevalent data deficit induced by the non-consensual user landscape. Research in this sphere could generate key insights into the practicality of these pioneering methodologies, potentially revealing their capacity to augment privacy safeguards without compromising the functional efficacy of web analytics tools like Google Analytics. As the utility of cookies dwindles, the urgency to explore and validate viable alternatives for consent collection intensifies, underscoring the question of legality in a post-cookie tracking milieu and stimulating anticipation for the advent of novel technologies capable of addressing this evolving dynamic.

Another promising trajectory for future research involves examining the types of data to be relayed to Google Analytics. This process is predicated on the specific needs of individual companies, and it would be crucial to probe into how the "legitimate Interest" of these data can be justified, given that this is likely to differ substantially across industries. Such analysis could provide a recommendation for businesses and legal authorities alike, shaping regulations around third-party plugins and nudging companies towards a data minimalism approach, thereby mitigating the threat of non-compliance with evolving data privacy legislations and norms.

Finally, a comparative study on Google Analytics versus other privacy-centric third-party plugins is merited. This comparative endeavor should be multidimensional, examining aspects such as cost-benefit analysis, feature comparative studies, and integration simplicity. A rigorous analysis within this domain could potentially demystify prevailing fears and misapprehensions that shroud lesser-known plugins, fostering an ecosystem wherein businesses are empowered to make informed decisions when selecting plugins. Presently, the tendency towards popular or cost-free plugins, such as Google Analytics, heavily dominates the selection process. Nonetheless, a comprehensive comparative study could inspire a paradigm shift towards a holistic approach where businesses select plugins based on their unique privacy requirements and the specific functionalities offered, rather than a simplistic preference for the most popular or economically reasonable solution.

6 Conclusion

In search for the answer to the question "In implementing Google Analytics, how do developers and analytics specialists strike a balance between user privacy and utility?" it has indeed been addressed through semi-structured interviews with eight experienced participants. These participants shared invaluable insights into adopting the practices to manage the two delicate aspects. From their perspectives, this balance can be maintained through careful data management, keeping updates and rigorous compliance with privacy regulations, the strategic use of new technologies like Server-side tracking, and maintaining transparency with users through informed consent mechanisms. All these

[4]Conversion modeling is a feature in Google Analytics utilizing machine learning to model the behavior of users who reject analytics cookies based on analyzing the behavior of similar users who accept analytics cookies. This feature allows marketers to attain valuable insights from Analytics reports while respecting the users' choice for privacy – Google (n.d.) in "Consent Mode on websites and mobile apps" – Analytics Help

efforts aim to uphold user trust and maximize the utility of data collected while adhering to an ever-evolving regulatory landscape.

Through the research process, it has become apparent that the nuance between responses from the participants suggested that this balance is not static, but rather requires constant refinement. Various strategies and tactics adopted by the participants reflect a keen awareness of this evolving context and a commitment to navigating it responsibly rather than leaving that responsibility to the users. In light of the insights shared, the researcher believes that a one-size-fits-all approach for data collection and privacy protection does not seem feasible due to the diverse nature of businesses and other context-specific factors. As a result, a tailored, balanced, and thoughtful approach toward data collection and management emerges as the most viable path, one that is capable of addressing compliance requirements, reducing analytical complexity, and enhancing user privacy sustainably.

While the research provides in-depth insights into the current practices and strategies, it also underscores the dynamic nature of the field and the continuous need for vigilance, creativity, and adaptability in response to regulatory changes and technological advancements. Therefore, further studies are encouraged to continue exploring this area in the future, ensuring that the delicate balance between user privacy and utility remains a priority in the fast-paced, data-driven world of digital analytics.

Finally, this research has highlighted the significance of preventive measures (for instance, against personal data leakage) rather than damage control in the use of third-party plugins, the value of proper documentation and organizational training, and the potential for regional tracking plugins has been noted as complementary to the technical measures. With this hybrid approach to this issue, The researcher hopes that these findings can serve as a helpful guide for practitioners navigating this complex landscape, fostering a data culture that is both responsible and productive.

7 Appendix

7.1 Co-occurence table

Table 1: Practices to Balance Utility and Privacy in Google Analytics

Practice	**Decrease the utility Gr=38**	**Increase compliance to privacy regulation Gr=20**	**Increase Privacy of the user Gr=34**	**Increase Utility of GA and tracking Gr=22**	**Influence of technical difficulty and financial cost to balance the privacy and utility when using GA Gr=9**	**Influence of the company business on the practice of using Google Analytics Gr=22**	**Legal Requirement has a big influence on balancing utility and privacy when using GA but it is hard to keep up with them Gr=29**	**No adoption of the practice Gr=39**	**No impact on privacy Gr=15**	**No impact on utility Gr=15**	**Suggestion to adopt the practice Gr=29**
Adoption of Alternative EU-based tracking plugins may be considered for better compliance and it is cheaper than Proxy Server if the legal landscape escalates Gr=8	0	4	0	0	2	4	3	3	0	0	3
Advocate for changes from external parties (Google and Law) for better utility and privacy Gr=12	2	1	1	0	1	1	4	1	1	0	2
Big Tech Companies and Experts shaping the way GA is implemented Gr=16	1	2	0	2	0	3	5	1	2	1	2

Continued on next page

Table 1 – continued from previous page

Practice	Decrease the utility Gr=38	Increase compliance to privacy regulation Gr=20	Increase Privacy of the user Gr=34	Increase Utility of GA and tracking Gr=22	Influence of technical difficulty and financial cost to balance the privacy and utility when using GA Gr=9	Influence of the company business on the practice of using Google Analytics Gr=22	Legal Requirement has a big influence on balancing utility and privacy when using GA but it is hard to keep up with them Gr=29	No adoption of the practice Gr=39	No impact on privacy Gr=15	No impact on utility Gr=15	Suggestion to adopt the practice Gr=29
Cookie and Consent provides user with more privacy options, and this decrease the amount of data collected, but Dark Pattern is also commonly used to nudge user to consent to collecting more data Gr=26	8	1	0	5	0	0	1	1	1	7	0
Different set up of GA per sites/regions increase privacy and compliance Gr=6	0	2	3	0	1	2	2	2	0	0	2
GA is more popular, well-documented, effective than other tracking plugins despite being less GDPR-compliant Gr=8	1	3	1	1	0	3	4	1	0	0	3
GA4 increases the utility and privacy by providing more options and customizations Gr=8	1	0	2	4	0	3	1	0	2	1	0

Continued on next page

Table 1 – continued from previous page

Practice	Decrease the utility Gr=38	Increase compliance to privacy regulation Gr=20	Increase Privacy of the user Gr=34	Increase Utility of GA and tracking Gr=22	Influence of technical difficulty and financial cost to balance the privacy and utility when using GA Gr=9	Influence of the company business on the practice of using Google Analytics Gr=22	Legal Requirement has a big influence on balancing utility and privacy when using GA but it is hard to keep up with them Gr=29	No adoption of the practice Gr=39	No impact on privacy Gr=15	No impact on utility Gr=15	Suggestion to adopt the practice Gr=29
GA4 still has limitations regarding privacy and support Gr=9	4	1	3	0	1	0	1	2	0	0	2
Implementing your own tracking system increase utility and privacy than using third-party plugins Gr=4	1	0	4	2	0	1	2	0	0	0	3
Influence of the company business on the practice of using Google Analytics Gr=22	3	6	6	4	1	0	6	2	2	2	7
Legal Requirement has a big influence on balancing utility and privacy when using GA but it is hard to keep up with them Gr=29	6	7	3	1	3	6	0	5	1	1	5
Measures to post-handle personal data leakage to Google Analytics Gr=7	0	2	1	0	0	1	0	0	0	0	1

Continued on next page

Table 1 – continued from previous page

Practice	Decrease the utility Gr=38	Increase compliance to privacy regulation Gr=20	Increase Privacy of the user Gr=34	Increase Utility of GA and tracking Gr=22	Influence of technical difficulty and financial cost to balance the privacy and utility when using GA Gr=9	Influence of the company business on the practice of using Google Analytics Gr=22	Legal Requirement has a big influence on balancing utility and privacy when using GA but it is hard to keep up with them Gr=29	No adoption of the practice Gr=39	No impact on privacy Gr=15	No impact on utility Gr=15	Suggestion to adopt the practice Gr=29
Measures to prevent sending personal data to Google Analytics Gr=21	1	3	9	4	1	4	2	1	0	0	7
Nearly all participants have migrated to GA4 Gr=10	2	0	2	1	1	1	1	0	2	1	4
Non-technical practices such as organization training and documenting can help increase privacy Gr=11	0	3	4	2	0	3	1	6	3	3	3
Practices to balance utility and privacy when using Cookie and Consent Gr=21	3	1	7	6	0	1	1	0	0	3	7
Proxy Server may improve user's privacy but it is not necessary as GA does not collect IP address anyway Gr=13	4	1	3	1	2	1	2	9	2	1	2

Continued on next page

Table 1 – continued from previous page

Practice	Decrease the utility Gr=38	Increase compliance to privacy regulation Gr=20	Increase Privacy of the user Gr=34	Increase Utility of GA and tracking Gr=22	Influence of technical difficulty and financial cost to balance the privacy and utility when using GA Gr=9	Influence of the company business on the practice of using Google Analytics Gr=22	Legal Requirement has a big influence on balancing utility and privacy when using GA but it is hard to keep up with them Gr=29	No adoption of the practice Gr=39	No impact on privacy Gr=15	No impact on utility Gr=15	Suggestion to adopt the practice Gr=29
Removing UTMs and External Referrers negatively impacts the utility of the tool in marketing while there is no privacy issue identified with it Gr=9	6	0	1	0	0	0	0	8	2	0	0
Server-side tracking increases Privacy, Utility and Compliance than Client-side Tracking Gr=15	2	0	9	4	2	0	1	1	1	1	5
Server-side tracking may introduce financial and technical difficulty to implement and maintain Gr=4	1	0	3	2	2	0	0	1	1	1	0
The impact of limiting browser's data are varied but it shall be fine with legitimate interest and consent Gr=9	4	2	4	3	0	1	1	5	1	0	3

Continued on next page

Table 1 – continued from previous page

Practice	Decrease the utility Gr=38	Increase compliance to privacy regulation Gr=20	Increase Privacy of the user Gr=34	Increase Utility of GA and tracking Gr=22	Influence of technical difficulty and financial cost to balance the privacy and utility when using GA Gr=9	Influence of the company business on the practice of using Google Analytics Gr=22	Legal Requirement has a big influence on balancing utility and privacy when using GA but it is hard to keep up with them Gr=29	No adoption of the practice Gr=39	No impact on privacy Gr=15	No impact on utility Gr=15	Suggestion to adopt the practice Gr=29
The replacement/absence of the User Identifier can improve the privacy of the user but severely impact the utility of understanding user's behavior Gr=10	7	1	2	0	0	0	0	7	2	1	2
Uncertainty regarding Proxy Server can be an alternative to the use of Proxy Server Gr=4	1	0	0	0	0	0	1	0	0	0	0

7.2 Code Book

This is the code book used to analyze the transcribed interview. The code hierarchy starts with themes and then codes, to sub-codes. The last two groups denoted with a double-plus sign (++) are created as filter codes. They are only used for co-occurrence analysis and filtering the other principal codes.

- **Cookies are still heavily utilized to obtain the user's consent for third-party tracking, and more techniques are adopted to increase the utility of available data collected**
  - **Code: Cookie and Consent provides the user with more privacy options, and this decreases the amount of data collected, but Dark Pattern is also commonly used to nudge the user to consent to collect more data**
    - Consent and Cookie limits the amount of data collected and reduce the utility

    - Cookie and Consent allows more users to reject being tracked Code: Cookie banner is a standard in obtaining consent from the user
    - Dark pattern in Cookie Banner is commonly used to nudge users to provide consent to more tracking data
  - **Code: Laws and Browsers pose a challenge when implementing Cookie and Consent**
    - Browser limiting the cookie lifespan poses a challenge to the use of cookie
    - Legal impact when a cookie is not properly implemented
  - **Code: Practices to balance utility and privacy when using Cookie and Consent**
    - Allowing the user to withdraw from consent to allow the user more privacy option Code: Being transparent with users about what and how they are being tracked via the cookie banner
    - Creating user accounts to collect data instead of using GA and cookie/ consent
    - Data sampling can be used to increase the utility of the limited data from the consent
    - Documenting the legal use of cookies to ensure compliance with data privacy regulation
    - explicitly presents tracking info to the user
    - Monetization of data to obtain consent from the user Code: Provide users with the option to withdraw from being tracked and cookies to increase their privacy Code: Using a neutral cookie bar to increase privacy and build trust with the users
    - Using Server-side Tracking is a future alternative to The use of Cookie

- **Influence of Business Practice, Law and External Influence in adopting measures to Strike a Balance between Privacy and Utility**
  - **Code: Big Tech Companies and Experts shaping the way GA is implemented**
    - Blogs and Experts impact how GA is implemented
    - Influence of the other big tech companies on the practice of using Google Analytics
  - **Code: GA is more popular, well-documented, and effective than other tracking plugins despite being less GDPR-compliant**
    - GA is more popular than other tracking plugins
    - GA may not be as GDPR-compliant as other tracking plugins
    - Google Analytics is less privacy pervasive than other tools like Facebook
    - Google Analytics is superior to other tracking plugins in terms of Utility and support documents
    - Google Analytics is used in conjunction with other tracking plugins, and one cannot replace another
  - **Code: Influence of technical difficulty and financial cost to balance the privacy and utility when using GA**
  - **Code: Influence of the company business on the practice of using Google Analytics**
    - Companies collect fewer data to avoid privacy non-compliance

        * Companies collect more data to make an informed decision in a different department
        * Company operates on an international level and influences how GA is set up
        * company prefers third-party plugins rather than implementing their own
        * Company wants as much data as it can when tracking user
        * Company's structure affects the decision-making of using GA
        * Influence of the clients on how the company sets up GA
        * Influence of the integration with other Google Products when using Google Analytics
    - **Code: Legal Requirement has a big influence on balancing utility and privacy when using GA, but it is hard to keep up with them**
        * Influence of the legal authority on the practice of Google Analytics
        * Lack of access to Privacy knowledge hampers the ability to maintain a balance between utility and privacy
        * Technical Measures are not able to fulfill the requirements of EU countries

- **Measures to increase the privacy of users by preventing or post-handling data leakage to Google Analytics**
    - **Code: Data anonymization, generalization, the use of Server-side tracking to prevent personal data from leaking to Google Analytics**
        * Checking the system to make sure not to send personal data when using Google Analytics
        * Hashed and encrypted information before sending to GA
        * Identify your business need for tracking
        * No way to prevent personal information sent to Google Analytics
        * Storing data outside of Google Analytics for non-anonymized data
        * Using Server-side tracking to monitor data sent to GA to prevent data leakage
    - **Code: Transparency, Troubleshooting, and Data Removal to post-handle incidents of personal data sent to Google**
        * Communication with users about data leakage
        * Contacting with GA to identify how the data leakage happened
        * Data removal from Google Analytics
        * Troubleshooting the system to prevent data leakage from happening in the future

- **Migration to Google Analytics 4 helps improve privacy and effectiveness despite GA4 still has its limitation in Privacy and Features**
    - **Code: GA4 increases the utility and privacy by providing more options and customizations**
        * GA4 has the potential to increase the utility by improving with more functionalities in the future
        * GA4 increases privacy by providing more privacy options
        * GA4 increases the utility by providing more customization
        * GA4 increases the utility with easier integration with other services

        * Implementing custom measurements is required for GA4 to increase its utility
        * there is no decrease in performance and utility of GA4 compared to the previous version
    - **Code: GA4 still has limitations regarding privacy and support**
        * GA4 increases privacy and decreases utility by having stricter rules on how to collect data
        * GA4 reduces the utility with the lack of certain functions and has lots of bugs compared to UA
        * Privacy issues are still present with GA4
        * There is a lack of help/support documents in GA4
    - **Code: Nearly all participants have migrated to GA4**
        * Forced to move to GA4 because of privacy issue
        * Participant has migrated to Google Analytics 4
        * Participant has not migrated to GA4

- **Other practices to enhance the privacy, utility, and compliance of Google Analytics**
    - **Code: Advocate for changes from external parties (Google and Law) for better utility and privacy**
        * Demand changes from Google Analytics and Google Product
        * Demand changes from Laws and Legal authorities to enhance privacy when using tracking plugins
        * Implementation of new tech solutions for privacy protection
        * Waiting for legislation to have clearer guidance
    - **Code: Different set up of GA per sites/regions increase privacy and compliance**
        * Difference set up of GA on public vs. private site
        * Different Tracking Practices between Europe vs. Others Area helps increase compliance when using GA
    - **Code: Implementing your own tracking system increases utility and privacy than using third-party plugins**
        * Implementing your own tracking system allows more precise tracking data and better utility
        * Nudging people to create an account for your own tracking system can cause missing out of data
    - **Code: Non-technical practices such as organizing training and documenting can help increase privacy**
        * Documentation of the use of Google Analytics is rarely adopted but believed to benefit the privacy of the tool
        * Organization training raises awareness about privacy vs. utility issues in GA
    - **Server-side Tracking provides more control than Client-side Tracking, hence increasing the privacy and utility when using Google Analytics**

    * **Code: Server-side tracking increases Privacy, Utility, and Compliance than Client-side Tracking**
        · Participant has no awareness of Server-side tracking Server-side item tracking helps increase Compliance with Privacy Regulation Server-side item tracking helps increase the privacy Server-side item tracking helps increase the Utility of Google Analytics
        · Server-side tracking is recommended to adopt rather than Client-side tracking
        · Server-side tracking provides easier integration into other services
    * **Code: Server-side tracking may introduce financial and technical difficulty to implement and maintain**
        · Server-side tracking is easier to implement than Client-side tracking in Google Analytics
        · Server-side tracking is harder to implement than Client-side tracking
    * **Code: Uncertainty regarding Proxy Server can be an alternative to the use of Proxy Server**
        · Server-side tracking can be an alternative to using Proxy Server
        · Server-side tracking is not certain to be an alternative to using Proxy Server to send data to Google Analytics

– **The Low Adoption of CNIL measures due to the perceived negative impact on the utility of GA and Technical Difficulty**
    * **Code: Adoption of Alternative EU-based tracking plugins may be considered for better compliance, and it is cheaper than Proxy Server if the legal landscape escalates**
        · Proxy Server may improve user's privacy, but it is not necessary as GA does not collect IP addresses anyway
        · Bad Configuration of Proxy Server can lead to a negative impact on the data and utility
        · Follow the guide by autoriteitpersoonsgegevens instead of adopting Proxy Server by CNIL
        · No familiar with the concept of Proxy Server in User Tracking to increase the privacy of the user
        · Not exposing GA to IP Address can improve user's privacy
        · Proxy is not realistic as it strips all the meaningful information
        · Proxy Server is redundant as GA does not collect IP addresses anyway
        · Using Proxy to not expose GA to IP Address can improve user's privacy
        · Waiting for legislation instead of implementing Proxy Server
    * **Code: Removing UTMs and External Referrers negatively impacts the utility of the tool in marketing while there is no privacy issue identified with it**
        · Collecting UTMs and External Referrers as long as they have consent
        · No Privacy issue identified with external Referrer and UTMs
        · Not aware of the impact of removing UTMs and Referrers
        · Removing UTMs and external Referrer will heavily impact the performance of marketing
        · Removing UTMs and Referrers is too harsh from the law

    - **Code: The impact of limiting Browser's data is varied, but it shall be fine with legitimate interest and consent**
        - it is okay to collect some browser's data with the legitimate interest
        - Limiting Browser's data can increase compliance with privacy regulation
        - Limiting Browser's data can increase utility by reducing complexity in data analysis
        - Limiting the Browser's data increases the privacy of the user
        - Limiting the Browser's data can have a negative impact on the utility of GA
        - There should be clearer guidelines on how much data you can collect
    - **Code: The replacement/absence of the User Identifier can improve the privacy of the user but severely impact the utility of understanding the user's behavior**
        - ID pseudonymization can improve the privacy of the user in GA
        - ID pseudonymization is not necessary as it does not link to personal data
        - ID Pseudonymization recommended by CNIL can have a severe impact on the utility of understanding user's data
        - Adopting an alternative to GA can avoid non-compliance with privacy regulation risk
        - Adopting an alternative to GA can be more cost-effective than using Proxy Server
        - Considering adopting an alternative to GA if Google does not address the legal issue
        - Not considering moving an alternative to GA due to their business model
        - Not considering moving to another alternative to GA yet
        - Using a combination of third-party tracking tools together to replace GA
        - Waiting for clearer legislation instead of adopting another third-party plugin

- **++ Impact Assessment of Measures on Privacy, Utility, and Compliance**
    - Decrease the privacy
    - Decrease the utility
    - Increase compliance with privacy regulation
    - Increase Privacy of the user
    - Increase Utility of GA and tracking
    - No adoption of the practice
    - No impact on privacy
    - No impact on utility
    - Suggestion to adopt the practice
- **++ Legal Awareness, Attitude Toward Google Analytics**
    - Familiarity with legal issues surrounding Google Analytics
    - Negative attitude toward Google and its monopoly
    - Positive attitude toward Google

References

Alhazmi, A., & Arachchilage, N. A. G. (2021). I'm all ears! listening to software developers on putting gdpr principles into software development practice. *Personal and Ubiquitous Computing*, *25*(5), 879–892.

Art. 44 gdpr – general principle for transfers. (2018, Mar). Retrieved from https://gdpr-info.eu/art-44-gdpr/

Autoriteit Persoonsgegevens. (2022, Apr). *Handleiding privacyvriendelijk instellen van google analytics.* Retrieved from https://www.autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/handleiding_privacyvriendelijk_instellen_google_analytics_april_22.pdf

Ayalon, O., Toch, E., Hadar, I., & Birnhack, M. (2017). How developers make design decisions about users' privacy: The place of professional communities and organizational climate. In *Companion of the 2017 acm conference on computer supported cooperative work and social computing* (p. 135–138). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3022198.3026326 doi: 10.1145/3022198.3026326

Bednar, K., Spiekermann, S., & Langheinrich, M. (2019). Engineering privacy by design: Are engineers ready to live up to the challenge? *The Information Society*, *35*(3), 122-142. Retrieved from https://doi.org/10.1080/01972243.2019.1583296 doi: 10.1080/01972243.2019.1583296

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, *2*, 8-14. Retrieved from https://www.sciencedirect.com/science/article/pii/S2352900816000029 doi: https://doi.org/10.1016/j.npls.2016.01.001

Bogdan, R., & Biklen, S. (2007). *Qualitative research for education: An introduction to theories and methods.*

Burby, J., Brown, A., Committee, W. S., et al. (2007). Web analytics definitions. *Washington DC: Web Analytics Association*.

CJEU. (2020, Sep). *The cjeu judgement in the schrems ii case - european parliament.* Retrieved from https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/652073/EPRS_ATA(2020)652073_EN.pdf

Clifton, B. (2012). *Advanced web metrics with google analytics.* John Wiley & Sons.

CNIL. (2022a, Jul). *Google analytics and data transfers: How to make your analytics tool compliant with the gdpr?* Retrieved from https://www.cnil.fr/en/google-analytics-and-data-transfers-how-make-your-analytics-tool-compliant-gdpr

CNIL. (2022b, Feb). *Use of google analytics and data transfers to the united states: The cnil orders a website manager/operator to comply.* Retrieved from https://www.cnil.fr/en/use-google-analytics-and-data-transfers-united-states-cnil-orders-website-manageroperator-comply

Datatilsynet. (n.d.). *Google analytics - recent european practices.* Retrieved from https://www.datatilsynet.dk/english/google-analytics

Datatilsynet. (2023, Mar). *Varsel om vedtak i google analytics-saken.* Author. Retrieved from https://www.datatilsynet.no/aktuelt/aktuelle-nyheter-2023/varsel-om-vedtak-i-google-analytics-saken/

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., & Holz, T. (2019, Aug). *We value your privacy ... now take some cookies.* Springer Berlin Heidelberg. Retrieved from https://link.springer.com/article/10.1007/s00287-019-01201-1

Eckersley, P. (2010). How unique is your web browser? In M. J. Atallah & N. J. Hopper

(Eds.), *Privacy enhancing technologies* (pp. 1–18). Berlin, Heidelberg: Springer Berlin Heidelberg.

European Center for Digital Rights - NOYB. (2022a, Jan). *Austrian dsb: Eu-us data transfers to google analytics illegal.* Retrieved from https://noyb.eu/en/austrian-dsb-eu-us-data-transfers-google-analytics-illegal

European Center for Digital Rights - NOYB. (2022b, Jul). *Update: Further eu dpa orders stop of google analytics.* Retrieved from https://noyb.eu/en/update-further-eu-dpa-orders-stop-google-analytics

European Commission. (2022, Mar). *European commission and united states joint statement on trans-atlantic data privacy framework.* Retrieved from https://ec.europa.eu/commission/presscorner/detail/it/ip_22_2087

European Data Protection Board - EDPB. (2021, Mar). *Spanish dpa fines vodafone spain more than 8 million euros.* European Data Protection Board (EDPB). Retrieved from https://edpb.europa.eu/news/national-news/2021/spanish-dpa-fines-vodafone-spain-more-8-million-euros_en

European Data Protection Board - EDPB. (2022, Dec). *The portuguese supervisory authority fines the portuguese national statistics institute (ine) 4.3 million eur.* Retrieved from https://edpb.europa.eu/news/national-news/2022/portuguese-supervisory-authority-fines-portuguese-national-statistics_en

[ga4] introducing the next generation of analytics, google analytics 4 - analytics help. (n.d.). Google. Retrieved from https://support.google.com/analytics/answer/10089681?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-a). *About geographical data - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/6160484?hl=en&sjid=16910055291387163749-EU#zippy=%2Cin-this-article

Google. (n.d.-b). *Analytics tools & solutions for your business.* Author. Retrieved from https://marketingplatform.google.com/about/analytics/ (Last Accessed: 2023-06-29)

Google. (n.d.-c). *Bigquery enterprise data warehousenbsp; —nbsp; google cloud.* Author. Retrieved from https://cloud.google.com/bigquery (Last Accessed: 2023-06-29)

Google. (n.d.-d). *Consent mode on websites and mobile apps - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/9976101?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-e). *Data controls in google analytics 4 - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/13126616 (Last Accessed: 2023-06-29)

Google. (n.d.-f). *Eu-focused data and privacy - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/12017362?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-g). *[ga4] custom dimensions and metrics - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/10075209?hl=en&sjid=6074742843893540071-NA (Last Accessed: 2023-06-29)

Google. (n.d.-h). *[ga4] data collection - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/11593727?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-i). *[ga4] introducing the next generation of analytics, google analytics 4 - analytics help.* Retrieved from https://support.google.com/analytics/answer/10089681?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-j). *[ga4] measure activity across platforms with user-id - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/9213390?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-k). *[ga4] predefined user dimensions - firebase help.* Author. Retrieved from https://support.google.com/firebase/answer/9268042?visit_id=638169342991234723-4109228308&rd=1 (Last Accessed: 2023-06-29)

Google. (n.d.-l). *[ga4] set up cross-domain measurement - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/10071811?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-m). *Google analytics - campaign url builder.* Author. Retrieved from https://ga-dev-tools.google/ga4/campaign-url-builder/ (Last Accessed: 2023-06-29)

Google. (n.d.-n). *An introduction to server-side taggingnbsp; —nbsp; google tag manager - server-sidenbsp; —nbsp; google developers.* Author. Retrieved from https://developers.google.com/tag-platform/tag-manager/server-side/intro (Last Accessed: 2023-06-29)

Google. (n.d.-o). *Ip masking in universal analytics - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/2763052?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-p). *Measurement protocol parameter reference nbsp;—nbsp; analytics measurement protocol nbsp;—nbsp; google developers.* Author. Retrieved from https://developers.google.com/analytics/devguides/collection/protocol/v1/parameters#cid (Last Accessed: 2023-06-29)

Google. (n.d.-q). *Regional data collection - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/11598602?hl=en&sjid=14252130377373592178-NA (Last Accessed: 2023-06-29)

Google. (n.d.-r). *Tag manager overview - tag manager help.* Author. Retrieved from https://support.google.com/tagmanager/answer/6102821?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-s). *[ua→ga4] universal analytics versus google analytics 4 data - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/9964640?hl=en#zippy=%2Cin-this-article (Last Accessed: 2023-06-29)

Google. (n.d.-t). *Understanding pii in google's contracts and policies - analytics help.* Author. Retrieved from https://support.google.com/analytics/answer/7686480?hl=en (Last Accessed: 2023-06-29)

Google. (n.d.-u). *Why and when to use server-side tagging? — server-side tagging fundamentals — google developers.* Author. Retrieved from https://developers.google.com/tag-platform/learn/sst-fundamentals/3-why-and-when-sst (Last Accessed: 2023-06-29)

Jansen, B. J., Jung, S.-G., & Salminen, J. (2022, May). Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites. *PLoS One*, *17*(5), e0268212.

Krishnamurthy, B., & Wills, C. E. (2006). Generating a privacy footprint on the internet. In *Proceedings of the 6th acm sigcomm conference on internet measurement* (p. 65–70). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1177080.1177088 doi: 10.1145/1177080.1177088

Kröger, J. L. (2022). Technology cannot fix the privacy crisis. *Chapter adapted from: Kröger, JL (2022). Rogue Apps, Hidden Web Tracking and Ubiquitous Sensors. Doctoral dissertation, Technische Universität Berlin*, 216–219.

Kurose, J. F., & Ross, K. W. (2017). *Computer networking: A top-down approach (7th ed.).* Pearson. Retrieved from https://www.amazon.com/Computer-Networking-Top-Down-Approach-7th/dp/0133594149

Maso, I., & Smaling, A. (2004). Qualitative research: practice and theory. *Amsterdam:*

Boom.

Mayer, J. R., & Mitchell, J. C. (2012). Third-party web tracking: Policy and technology. *2012 IEEE Symposium on Security and Privacy*, 413-427.

MDN. (n.d.). *Referer*. Retrieved from `https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/Referer` (Last Accessed: 2023-06-29)

NWO. (2018, Oct). *Netherlands code of conduct for research integrity.* Retrieved from `https://www.nwo.nl/en/netherlands-code-conduct-research-integrity`

Papadogiannakis, P. P. K. N. M. E. P., E. (2021). User tracking in the post-cookie era: How websites bypass gdpr consent to track users. *Proceedings of the Web Conference 2021*. doi: 10.1145/3442381.3450056

Peters, F. (2018). On privacy and utility while improving software quality. *Electron. Commun. Eur. Assoc. Softw. Sci. Technol.*, *75*.

Quach, S., Thaichon, P., Martin, K. D., Weaven, S., & Palmatier, R. W. (2022, Mar). *Digital technologies: Tensions in privacy and data - journal of the academy of marketing science.* Springer US. Retrieved from `https://link.springer.com/article/10.1007/s11747-022-00845-y`

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., ... Jinks, C. (2018). *Saturation in qualitative research: Exploring its conceptualization and operationalization.* U.S. National Library of Medicine. Retrieved from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5993836/`

Shenton, A. (2004, 07). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*, 63-75. doi: 10.3233/EFI-2004-22201

Spiekermann, S., Korunovska, J., & Langheinrich, M. (2019). Inside the organization: Why privacy and security engineering is a challenge for engineers. *Proceedings of the IEEE*, *107*(3), 600-615. doi: 10.1109/JPROC.2018.2866769

Stöver, A., Gerber, N., Pridöhl, H., Maass, M., Bretthauer, S., Döhmann, I. S. g., ... Herrmann, D. (2023). *How website owners face privacy issues: Thematic analysis of responses from a covert notification study reveals diverse circumstances and challenges.* Retrieved from `https://doi.org/10.56553/popets-2023-0051`

Szymielewicz, K., & Budington, B. (2018, Jun). *The gdpr and browser fingerprinting: How it changes the game for the sneakiest web trackers.* Retrieved from `https://www.eff.org/deeplinks/2018/06/gdpr-and-browser-fingerprinting-how-it-changes-game-sneakiest-web-trackers`

Tahaei, M., Abu-Salma, R., & Rashid, A. (2023, Jan). *Stuck in the permissions with you: Developer amp; end-user perspectives on app permissions amp; their privacy ramifications.* Retrieved from `https://arxiv.org/abs/2301.06534v2`

Tahaei, M., Frik, A., & Vaniea, K. (2021). Privacy champions in software teams: Understanding their motivations, strategies, and challenges. In *Proceedings of the 2021 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3411764.3445768` doi: 10.1145/3411764.3445768

Tahaei, M., & Vaniea, K. (2021, 03). "developers are responsible": What ad networks tell developers about privacy.. doi: 10.1145/3411763.3451805

Tonyan, J. (2016). Measuring the success of your social media presence with google analytics. *Library Technology Reports*, *52*(7), 38–42.

Utz, C., Amft, S., Degeling, M., Holz, T., Fahl, S., & Schaub, F. (2022). *Privacy rarely considered: Exploring considerations in the adoption of third-party services by websites.* arXiv. Retrieved from `https://arxiv.org/abs/2203.11387` doi: 10.48550/ARXIV.2203.11387

Weber, J. (2015). *Practical google analytics and google tag manager for developers.* Springer. doi: 10.1007/978-1-4842-0265-4