Unpacking Privacy Labels: A Measurement and Developer Perspective on Google's Data Safety Section

Rishabh Khandelwal, Asmit Nayak, Paul Chung, and Kassem Fawaz,
University of Wisconsin-Madison

https://www.usenix.org/conference/usenixsecurity24/presentation/khandelwal

This paper is included in the Proceedings of the 33rd USENIX Security Symposium.

August 14–16, 2024 • Philadelphia, PA, USA

978-1-939133-44-1

Open access to the Proceedings of the 33rd USENIX Security Symposium is sponsored by USENIX.

Unpacking Privacy Labels: A Measurement and Developer Perspective on Google's Data Safety Section

Rishabh Khandelwal*
University of Wisconsin–Madison

Asmit Nayak*
University of Wisconsin–Madison

Paul Chung
University of Wisconsin–Madison

Kassem Fawaz
University of Wisconsin–Madison

Abstract

Google has mandated developers to use Data Safety Sections (DSS) to increase transparency in data collection and sharing practices. In this paper, we present a comprehensive analysis of Google's Data Safety Section (DSS) using both quantitative and qualitative methods. We conduct the first large-scale measurement study of DSS using apps from the Android Play store (n=1.1M). We find that there are internal inconsistencies within the reported practices. We also find trends of both over and under-reporting practices in the DSSs. Next, we conduct a longitudinal study of DSS to explore how the reported practices evolve and find that the developers are still adjusting their practices. To contextualize these findings, we communicate with the app developers to uncover the process they go through when working with DSS. We highlight the challenges faced and strategies developers employ for DSS submission and the factors contributing to changes in the DSS. Our research contributes valuable insights into the complexities of implementing and maintaining privacy labels, underlining the need for better resources, tools, and guidelines to aid developers. This understanding is crucial as the accuracy and reliability of privacy labels directly impact their effectiveness.

1 Introduction

Privacy policies have traditionally served as the primary method for conveying the privacy practices of a service to users. However, studies have demonstrated that privacy policies are often ineffective, mainly because users neglect them due to their length and vagueness [6, 13]. Introduced by Kelly et al. [19], privacy nutrition labels summarize the privacy practices of websites in a nutrition label format, making them easier to understand. Privacy labels have gained traction in the tech industry, with Google introducing Data Safety Sections (DSS) and Apple introducing Apple Privacy Labels (APL) for all new and updated apps on their app stores.

*Equal Contribution

Recently, researchers showed the utility of privacy labels for users, making privacy practices more accessible [34]. However, the utility of the privacy labels depends on the developers correctly filling out the forms. These forms should accurately reflect the developers' intentions regarding the privacy practices of their apps. Underreporting these practices can lead to inaccurate privacy labels, confusing and potentially instilling a false sense of security and privacy among users. Conversely, overreporting privacy practices can harm the developer by deterring users who will perceive the apps as less secure and private than they are.

Therefore, it is essential to investigate the developer's experience with reporting their practices via the privacy labels. Prior research has reported the responsiveness of developers in implementing Apple Privacy Labels and analyzed the data collection practices of apps according to these labels [2, 23]. Through small-scale studies, researchers have discovered that inaccurate APLs can exist due to the developer's knowledge gaps or resource limitations [22].

However, a considerable gap remains in comprehensively understanding the developer's interaction with privacy labels, particularly regarding Google's enforcement of data safety sections. This interaction includes determining intended privacy practices, overcoming platform challenges, accurately submitting privacy label forms, and updating them over time. Prior research has not accounted for this wide-ranging perspective. In this paper, we address this gap by focusing on Google's Data Safety Section, answering three key research questions:

RQ1: How do developers report their app privacy practices in Google's Data Safety Sections?

RQ2: How have these patterns of reporting practices evolved over the year following the implementation of DSS?

RQ3: What are the driving factors behind the changes in Google's DSS, the challenges, and the behaviors of app developers?

To answer these questions, we conduct a large-scale analysis of app privacy labels on the Google Play Store. We periodi-

cally scrape the Google Play Store to collect metadata, including permissions and DSS forms, for over 2M apps between June 2022 and May 2023. Subsequently, we communicate with 3,500 developers of these apps via emails to understand their decision-making process concerning privacy practices, the completion of DSS forms, and their subsequent updates. Our analysis of developers' responses helps answer the third research question by constructing an analytical framework to model their experience with DSS.

Our responses to the research questions offer new insights into Google's DSS:

- As of May 2023, privacy labels are only present for 46.8% of apps on the Google Play Store. Among those apps featuring DSS, we observe patterns of underreporting privacy practices, overreporting practices, and submitting inconsistent DSS forms.
- Our longitudinal analysis unveils a dynamic landscape for DSS, suggesting that developers are still refining their comprehension and implementation of DSS requirements. About 40% of the apps updated their DSS at least once over the past year, adding and removing high-level privacy practices and data categories.
- Our analysis of developers' responses indicates they confront challenges when aligning their intended practices with DSS forms. These challenges often lead them to prioritize successfully submitting DSS forms over accurately populating them.

Finally, based on our findings, we propose novel recommendations to enhance the developer experience with DSS. These recommendations include action points for platforms and regulators, such as better educational resources, multilingual support, form simplification, consistent feedback, improved support for third-party libraries, and mechanisms to solicit developer feedback. Lastly, we plan to release a large-scale dataset of the metadata for 1.14M Android apps spanning June 2022 to May 2023.

2 Background and Related Works

Privacy Nutrition Labels. Originally introduced by Kelley et al. [19, 20], privacy nutrition labels summarize the privacy practices of websites in a nutrition label format, making them easier to understand. They later designed the "Privacy Facts" display to allow the users to consider privacy while installing apps [21]. More recently, researchers proposed an Internet of Things (IoT) security and privacy label [9, 10] to surface privacy and security information about IoT devices to the users. Researchers have also studied the design and evaluation of privacy notices and labels [4, 7, 8, 11, 19–21, 26, 28].

In 2020, Apple adopted the privacy nutrition labels for the app store and mandated that app developers provide their

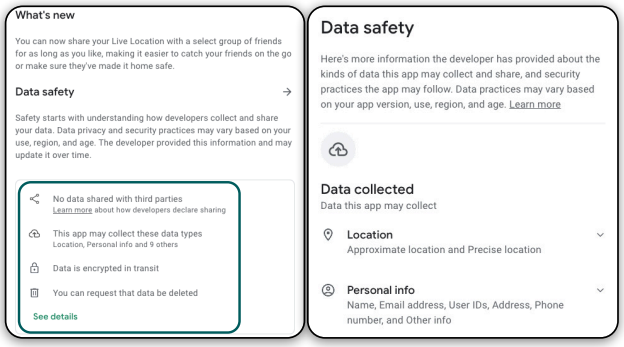


Figure 1: An example of the data safety section of an Android app.

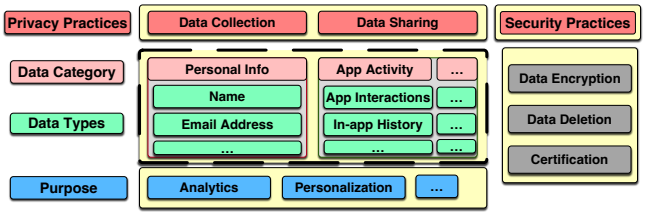


Figure 2: The hierarchy of the Google Data Safety Section showcasing the various layers and components.

apps' privacy information in the form of the Apple Privacy Label (APL). Later, in 2022, Google required developers to add a Data Safety Section (DSS) to the Google Play Store. Examples of Google's privacy labels are in Figure 1.

Google Data Safety Section The Data Safety Section (DSS) consists of four levels (Figure 2), where the first is high level *Privacy Practices*. The second and third levels consist of *Data Categories* and *Data Types*, and the fourth level consists of *Purpose*.

The first level includes three practices: *Data Collection*, which covers the details about the data that is collected and its intended use; *Data Sharing*, where the developers disclose what data is shared with third parties; and *Security Practices* that covers the data practices related to user choice and data security. *Security Practices* include three tags: *Encrypted in Transit*, *Data Deletion Option*, and *Review against Global Security Standards*.

In the second level, *Data Categories* includes 14 categories such as *App Info* and *Performance and App Activity*. Each *Data Category* can also have *Data Types*, which provide fine-grained information about the data used by the app. For example, *App Activity* includes *App Interactions and Installed App*, as shown in Figure 2. The final level of the Data Safety Section consists of *Purposes* that describe the reasons for collecting or sharing the data.

Google App Submission Review. Google reviews submitted apps to the Play Store to ensure their compliance with its

guidelines about design, content, and style [32]. As part of these guidelines, Google requires apps to comply with its data safety policies. In particular, all developers must specify the data collected and shared by their app, including the data handled by third-party libraries or SDKs [17].

To provide developers with better expectations towards the privacy criteria during the app review process, Google launched the *Checks* service. It allows developers to verify that their apps comply with the data safety policies before submitting the app for review [5]. The Data Monitoring feature of Google *Checks* monitors multiple channels of data collection and sharing, including with SDKs, via in-app permissions, and to external sites. The service provides developers evidence, such as permissions and network traffic, that the data safety form or the privacy policy is non-compliant.

Usability of Privacy Labels. Researchers have studied the usability of APLs from both users' [34] and developers' [22] perspectives. From the developers' perspective, Li et al. [22] interviewed 12 iOS developers and reported that developers err by under-reporting and over-reporting data collection in privacy labels. They further concluded that the label design is confusing for the developers either due to known factors (lack of resources, improper documentation) or unknown factors (preconceptions, knowledge gaps). Xiao et al. [33] characterize non-compliance of Apple privacy labels by studying data flow to label consistency of 5K iOS apps. They provide insights for improving label design based on their characterization. Researchers also built and evaluated a tool [12] that helps iOS developers generate privacy labels by identifying data flows through code analysis.

Studies on Privacy Labels. The works most similar to ours perform longitudinal measurement of privacy labels to understand the adoption and evolution of Apple privacy labels over time [2, 22, 29]. In particular, Scoccia et al. [29] conducted an empirical study of 17K apps to characterize how sensitive data is collected and shared for iOS apps. They found that free apps collect more sensitive data for tracking purposes. Li et al. [22] and Balash et al. [2] collected weekly snapshots of Apple privacy labels and characterized the privacy practices mentioned in privacy labels for 573k apps. Balash et al. [2] also perform additional correlation analysis with app meta-data like user rating, content rating, and app size.

Our Contributions. In this paper, we investigate the data safety sections of the apps listed on the Google Play Store, observing their evolution over the course of a year. Based on our measurement, we conduct a large-scale study by interacting with more than 3,500 Android app developers, using their apps as case studies. This analysis allows us to model the developers' engagement with the Data Safety Section ecosystem, gaining insight into their challenges, data practices, and the factors influencing their decision-making.

While prior works have explored challenges faced by iOS developers, it remains important to study Google's Data Safety Sections and challenges associated with them because APL and DSS cover different high level practices, as highlighted by Google [14]. Furthermore, this work reaches a broader audience, confirms prior works' findings and provide news insights into the life-cycle of updating the DSS. As such, our results complement prior works by showing that developers on the Apple and Google platforms share common problems when completing the DSS and APL. However, we go beyond prior work by substantiating the underreporting, over-reporting, and inconsistencies in data practices (Section 4); studying the evolution of DSS and practices that developers' report (Section 5); studying the process by which developers interact with the data safety sections; identifying strategies they use to bypass the review process; and revealing the factors that result in changes to the data safety sections over time (Section 6). Finally, we provide new recommendations based on our insights (Section 7).

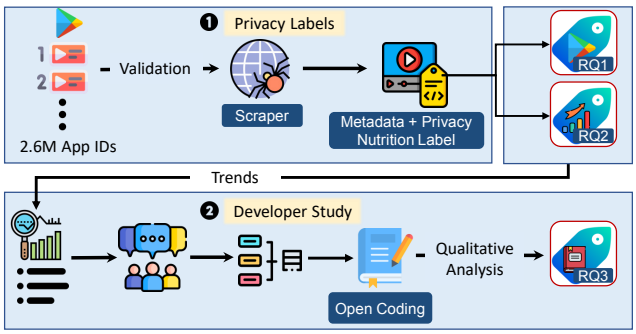


Figure 3: Our data measurement and analysis pipeline consists of scraping the DSS of Android apps to answer the first two research questions and interacting with app developers to answer the third research question.

3 Google Data Safety Dataset

We curate the Google Data Safety Dataset by scraping the metadata and privacy labels of the apps from Google Play. We contact the developers of some of these apps to understand their process of deciding on privacy practices, filling out the DSS form, and updating the DSS. We show an overview of the analysis pipeline in Figure 3.

Dataset Collection We took 10 snapshots of the Data Safety Sections for 2.46M apps present on the play store between June 20, 2022, and May 31, 2023. We captured eight weekly snapshots from June 20 to Aug 1 and three more snapshots on September 2022, November 2022, and May 2023. We chose to gather data more frequently around July 20, 2022, the date Google required app developers to complete the Data Safety Section. This allowed us to observe the developers' immediate response to this requirement. The additional three snapshots

enabled us to study the developers' long-term reactions. We refer to this dataset with 10 snapshots as the **DSS Dataset**.

We initiated data collection with the apk list provided by Androzoo [1]. This daily-updated list consists of Android app ids from various sources, including those from the Google Play store. We capture the metadata of each app, including its data safety sections, using the app IDs and a customized version of publicly available google play store scraper library `google-play-scraper` [18]. Across all the snapshots, we observed a total of 2.72M unique apps and 2.17M common apps. For the latest snapshot (May 2023), we retrieved metadata for 2.46M apps, which includes apps with very low download counts. To ensure that our statistical analysis is not skewed by these apps, we filter out apps that have fewer than 1000 downloads resulting in a total of 1.1M apps with 539k having privacy labels.

Dataset Statistics Our dataset shows that developers have been slow to add privacy labels to their apps, even after the hard deadlines have passed. Privacy labels are present only for 46.8% of the apps on the Google play store (as of May 31 2023). We also break down the DSS adoption rate according to app metadata, focusing on the apps' number of downloads, age rating, and pricing.

Number of Downloads: We examine the relationship of the DSS adoption rate and the number of app downloads. In particular, we categorize the apps based on their download numbers on a logarithmic scale, adhering to Google Play Store's binning methodology: (1000+, 5000+, ..., 5B+, 10B+). We measure the Spearman's correlation between the adoption rate and the downloads, which reveals a strong correlation with a correlation coefficient of 0.96 and the *p-value* of $7e{-}9$. Starting at 38.96% adoption rate of apps with 1000+ downloads, the adoption rate increases monotonically reaching 100% for apps with 10B+ downloads. This observation suggests that developers of highly downloaded apps tend to place more emphasis on the DSS, potentially due to access to greater resources and concern over public perception.

Age Rating: The Play Store categorizes apps based on age ratings, including Everyone, Teen, Mature 17+, and Everyone 10+.[1] We observed 51% of apps with the *Mature 17+* rating have a Data Safety Section (DSS), while the fraction of apps with a DSS in the other age ratings ranges from 45% (Everyone) to 54% (Everyone 10+).

App Pricing: We study the difference in practices based on whether the app is available for free, free with in-app purchases, or paid. We find that 65% of the paid apps, 59% of free with in-app purchases, and, for free apps, only 44% have DSS. We revisit this observation in the developer study where we highlight how developers report the privacy practices of Ad libraries in free apps.

[1] Google also has Adults 18+ rating, but we found less than 200 apps in this category and decided to filter it out for this analysis.

Ethical Considerations We collected data only from publicly available web pages and APIs. While our data collection scripts might load Google's servers, we were careful to not abuse these resources. In particular, we added back-off strategies in case of errors and waited for sufficient time before retrying for the failed cases.

4 Data Practices in Privacy Labels [RQ1]

We analyze the May 2023 snapshot of the DSS dataset to understand how developers report high-level privacy practices, data types, and purposes. Figure 4 depicts the percentage of apps reporting high-level privacy practices on DSS, with data encryption practices being the most reported. Figure 5 shows how app developers report collection and sharing for the different data categories. Developers report collecting *Location* and *Personal Information* at a higher rate than other data categories, primarily for *App functionality* and *Analytics*. They also report sharing *Location* and *Device Ids* more commonly for *Advertising or Marketing* purposes. The heatmap in Figure 2 (in Supplemental File[2]) provides a detailed breakdown of the declared purposes for collected and shared data types.

We also note that apps may declare sharing data without collecting it, as evident from Figure 5. This discrepancy arises from the definition of *Data Collection*, which covers developers retrieving user data from the device using the app [15]. Whereas *Data Sharing* denotes the cases when the data is transferred from the device to a third party. Thus, developers can share data without collecting it if their application employs third-party libraries that directly send data to their servers.

We further analyze this snapshot to understand how developers interact with the DSS forms. Our analysis reveals three key patterns: overreporting privacy practices, underreporting privacy practices, and submitting inconsistent practices.

Underreporting Practices: Analyzing Figure 4, we note that only 36% reported sharing at least one data type, suggesting that the majority of the apps on the play store do not report sharing data. This figure contrasts the findings from prior work [24, 31], which found that the majority of the apps use at least one third-party library collecting sensitive information [24]. We confirm this observation by analyzing the 15K most popular apps from our snapshot that report not sharing data with third parties. Specifically, we download the apk file from androzoo and use the LibRadar library [1, 25] to determine whether third-party libraries are used. Given an apk file, LibRadar [25] identifies the third-party libraries and tags them into categories like Advertisement, Analytics, etc.

Analyzing the apks, we find that 42% of the analyzed subset of apps used at least one third-party library for advertisement or analytics. This result clearly indicates that developers are

[2] https://osf.io/q3wv2?view_only=0b4aa040161a4b259c0a32c7fb3ae82c

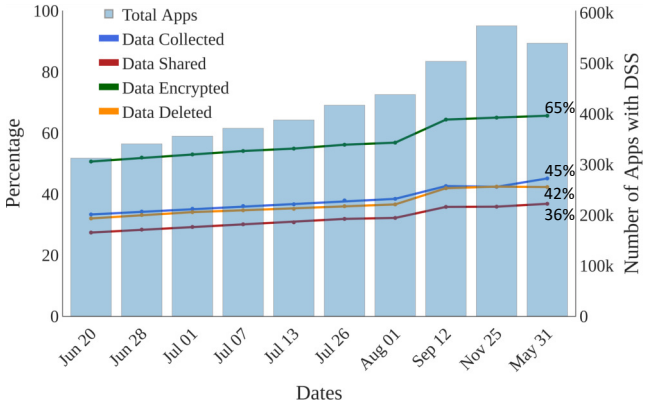


Figure 4: The evolution of Data Safety Section over the 10 snapshots in our dataset

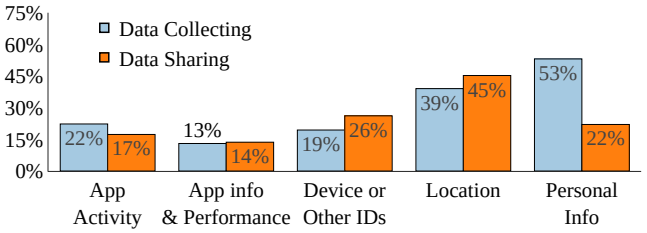


Figure 5: The distribution of Top-5 data categories for high-level practices

underreporting the sharing practices. One possible explanation is that the privacy practices of third-party libraries are often vague and developers find it hard to understand the collection and sharing practices of third-party libraries. We find evidence supporting this explanation in prior work [3, 22] and our developer study (Section 6).

Overreporting Practices: Analyzing the purposes for *Data Type* collection, we observe that many apps report a large number of purposes when listing datatypes. We note that out of the 7 possible purposes for collecting data, more than 3.5K apps list 6 or more purposes for every data type they collect, which may indicate that app developers list all purposes out of convenience. For example, *Workplace from Meta* with over 15M+ downloads, lists the same 6 purposes for all the data they collect, like access to *Installed Apps*, *SMS or MMS*, *Music Files*. We also note that while 3.5K is small compared to the dataset, it still has the potential to impact millions of users.

A possible explanation for this observation could be that app developers lack the knowledge required to fill the DSS and choose to select all options. Another possible reason is that they are unaware of the policies of the third-party applications that they use, and take a cautious approach by overreporting. Findings in prior work [22] and from our developer study (Section 6) align with this observation. Note that accurately determining the correctness of selected purposes is a challenging problem that is out-of-scope for this work. Our hypothesis for over-reporting stems from the observation that some apps select 6 purposes for all the data they collect. For example, the ARknet app (7K downloads) states that they collect Purchase History for Developer Communications, an unlikely data type for developer communications.

Inconsistent Practices: We observe the developers report practices that are inconsistent with other declared practices or with the app permissions. For example, we find that 40% of the apps state that they do not collect or share data, but encrypt the data in transit. We delved deeper into this observation by cross-verifying security practices with apps' network permission requests. 59% of apps do not request network permissions, yet state that they encrypt data in transit. It is not clear why apps need to encrypt transit data if they are not collecting or sharing data. We note that we only consider apps having the 'android.permission.INTERNET' permission as apps with network permissions. There may be other network permissions like 'android.permission.ACCESS_NETWORK_STATE,' which we do not count because this permission does not allow the app to access the internet. As *Encryption in transit* has important implications for privacy, we mark this trend and examine it in detail in our developer study in Section 6.

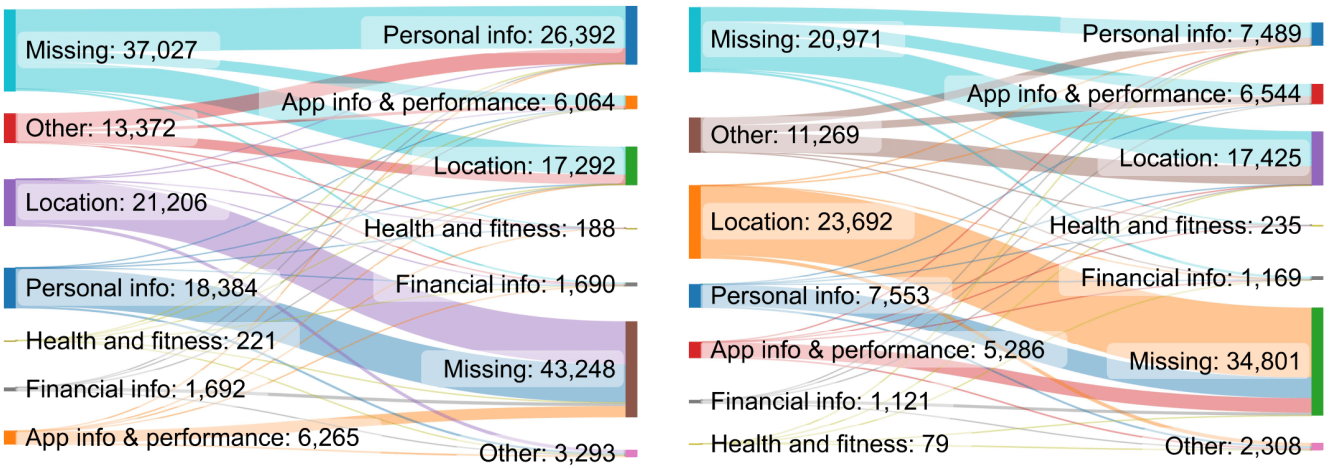
We also cross-verified the collected and shared data types from the DSS to the app permissions. Several apps report collecting or sharing several data types without even asking for the corresponding permissions. For example, 11.5% of the apps report collecting or sharing precise location data without obtaining location permissions. Another example is 23.7% of the apps report collecting or sharing files and documents without the "Photos/Media/Files" permissions.

In Section 6, we identify some of the reasons why developers provide inconsistent privacy labels. One reason is developers over-reporting their practices, as indicated above. Other reasons include developers choosing labels by mistake, misunderstanding the label definition, and not updating the labels after updating app features.

5 Evolution Of Data Safety Section [RQ2]

Next, we conduct a longitudinal analysis of the DSS dataset to understand how the data practices disclosed by developers evolved over time. As described in Section 3, our data collection spanned the timeframe before and after the hard deadline set by Google for app developers to comply with the new DSS requirements. This design allows us to understand not only the static state of app data safety disclosures at a given time point but also their evolution as developers navigated this significant policy change.

Looking at Figure 4, we observe that during the period spanning June 20, 2022 to May 31, 2023, the number of apps with DSS increased from 312K to 539K. The largest per-day change, of around 9K, happened between June 28, 2022, and July 1st, 2022. Interestingly, we find that from our initial



(a) Data Type evolution in Collection

(b) Data Type evolution in Sharing

Figure 6: Change in Data Categories between the first and the final snapshot for (a) Data Collection and (b) Data Sharing. The figure shows that at Data Category level, the reported practices change a lot, indicating that developers change the DSS frequently.

snapshot, 21% of apps removed their DSS over the course of our data collection. Specifically, 67K updated their play store page to remove the privacy label. For example, *Sport Prediction* with 1M app download had a DSS as of August 1, 2022, but did not have it by Nov 25, 2022.

Updates in DSS. We analyze the DSS dataset to understand the frequency of updates in DSS by comparing the first snapshot of the DSS for an app with subsequent snapshots. Figure 11 in Appendix A.1 shows the CDF of updates in DSS. We find that 40% (n=283K) of the apps updated their DSS at least once, while 4% (n=27K) updated it at least twice. Moreover, the frequency of updates is higher in the September and November snapshots. For example, *Adobe Acrobat Reader: Edit PDF*, with over 500M downloads, updated their DSS 3 times between Jun 20, 2022, and May 31, 2023. In Section 6, we discuss potential reasons for DSS changes. These reasons include app feature updates or the discovery of incorrect DSS.

Evolution of High-Level Practices. We also analyze the evolution of high-level practices present in the DSS. First, we observe a shift in data collection practices. 44K applications that initially reported collecting data have updated their DSS to state that they no longer collect data. Conversely, some apps (n=45K) that initially claimed not to collect data have revised their DSS, admitting to data collection.

We also observe similar trends in data-sharing practices. Several apps (n=42K) that initially reported sharing data with third parties later updated their DSS, indicating an end to such sharing. On the other hand, some apps (n=37K) that did not report data sharing in their initial DSS later added such practices. As changes in collection and sharing practices have implications on user privacy, we highlight these trends and analyze them in detail in our developer study in Section 6.

We observed a steady increase in the number of apps reporting *Encryption of Data in Transit*, an aspect crucial for data security. The count rose from 157K on June 20, 2022, to 353K on May 31, 2023, demonstrating a steady progression towards improved data encryption practices (Figure 4). Similarly, we observe an increase in the number of apps providing *Data Deletion Option* where the count rose from 99K to 227K. However, despite the overall increasing trend in *Data Deletion Option* and *Encryption in Transit* practices, a closer examination reveals that some apps are incorporating these practices, while others are withdrawing them. This observation implies that developers are still adjusting their Data Safety Sections.

The ongoing changes in reported data practices, even ten months past Google's deadline for DSS implementation, point to a dynamic landscape, suggesting that developers are still refining their understanding and implementation of DSS requirements. We investigate the challenges that developers face in more depth in our developer study (Section 6.3).

Evolution of Collection and Sharing Practices. Delving deeper into the evolution at the datatype level, we investigate three *Data Categories* collected most frequently: Location, Personal Identifiers, and App Activity and Performance. *Personal Identifiers* include *Personal Info*, and *Device Id and other identifiers*. We show the evolution collection of these *Data Categories* over two snapshots in Figure 6a. For example, if *Location* and *Personal Info* are removed and *App Info and Performance* is added, we add two pairs of transition: (Location -> App Info and Performance and Personal Info -> App Info and Performance). We find that a large number of

apps initially reporting collection of user location (n=21K) and personal information (n=18K) have revised their DSS to indicate no longer collecting these datatypes. Similarly, around 6.2K apps that were initially collecting app info and performance data have updated their DSS, indicating a halt in this data collection practice.

Conversely, we also identified apps initially not collecting location data or other specific information, have now started collecting these datatypes. This shows the bi-directional nature of these changes. We find similar trends by analyzing evolution of these data types for *Data Sharing* in Figure 6b.

These findings highlight the evolving nature of data practices which can impact users' trust and privacy. If an app shifts from not collecting to collecting certain datatypes, it may expose users to new privacy risks especially as they will not be notified to this change. The findings can indicate developers' difficulties or confusion in accurately understanding their apps' data practices as we highlight in Section 6.

Trends in Over-Reporting of Data Practices In Section 4, we observe that developers over-reported the purposes for data collection. Analyzing the DSS dataset, we find a persistent pattern in over-reporting over time, even after ten months. This suggests that a considerable number of developers continue to perceive an environment of low risk or consequence in over-reporting their data practices.

We uncover the potential reasons behind this trend in our developer study. We find that Google's current policy enforcement does not impose penalties for overstating data collection practices, likely due to the limitations inherent in compliance checks, that rely on static and dynamic code analysis. Consequently, some developers may be inclined towards a risk-averse strategy, choosing to over-report to prevent potential policy violations. Although the number of apps following such practices is low (n=4K) in our set, the impact on the wider privacy label ecosystem can be substantial as it could affect users' trust in privacy labels.

6 Developer Study [RQ3]

We conduct a study with Android app developers to understand their perspective when engaging with Data Safety Sections. We identified *interesting* patterns (described in Section 4 and Section 5) in apps' Data Safety Sections and contacted the app developers to inquire about the factors responsible for the patterns, and the challenges that they face.

6.1 Methodology

We first describe the study design and analysis methods.

Study Design. We reached out to app developers through emails, probing their experiences and asking specific questions about the privacy labels of their apps. We crafted these questions carefully to stimulate responses. We contact the developers via email and analyze their responses to answer our research questions. Prior work [22] has conducted in-depth interviews with iOS developers to understand the challenges developers face while working with the Apple Privacy Labels. In-depth interviews can be limited by the number of participants involved in the study. In this work, we opt to conduct the study via email questions. While an email study does not provide the depth or understanding that comes from an interview study, it allows us to reach a wider audience with diverse perspectives.

Ethical Considerations When emailing developers, we clearly identified ourselves as researchers and stated that we were studying their application and sought information related to their data safety section (Appendix A.3). We do not collect any personally identifiable data in our study. As such, the IRB certified our study as not "human subjects research," and we were not required to obtain consent before sending out the emails. However, to use the developers' responses in our qualitative study, we sent one follow-up email and obtained informed consent. The content of the email is shown in Appendix A.3. We emphasize that we did not send any follow-up emails to the developers, except for the one to obtain consent. We also note that while we elicited voluntary responses from 889 developers, our analysis reached saturation before reaching all email responses.

Developer Selection. Recall that in Section 4 and Section 5, we identified the following three trends. (A) apps stating that they encrypt data without collecting or sharing data, (B) apps changing their practice from not collecting/sharing data to collecting/sharing data, and (C) apps changing their practices from collecting/sharing data to not collecting/sharing data. Trend (A) points to inconsistency in reported security practice, whereas Trend (B) and (C) have implications of user data privacy. We identified the apps corresponding to these patterns, sorted them in order of real installs, and contacted the top 10,000 developers for each pattern, asking them about the trend. We emphasize that we made a conscious decision not to include probing questions beyond the DSS trends. We asked developers about their DSS and the trend, implicitly prompting them to think about the process. This allows us to obtain the developer responses that are contextualized in their experience instead of our questions.

In response to our emails, we received 3500 responses. To get a clean set for analysis, we filtered out automated replies and non-English responses, leaving us with 889 responses. We then qualitatively analyze these responses to explore how developers describe their privacy practices, the challenges they face while working with the privacy labels, and the factors that prompted them to update their data safety section. We also note here that although not prompted, many developers highlighted the challenges with third-party libraries (Section 6.3) and overstating collection (Section 6.4) while discussing Trend B and C, indicating that our analysis covers

under-reporting and over-reporting practices.

Qualitative Analysis Method. In accordance with qualitative research guidelines, we employed a strategy of random sampling and coding of developer responses until data saturation was achieved [27]. The coding process began with two authors evaluating an initial set of 50 responses and developing preliminary codes. Next, the research team discussed these codes, clarified differences, and established the initial codebook. The notable differences were in the granularity of the codes, and we opted for fine-grained codes as they can be consolidated later. Subsequently, the two authors independently coded a randomly selected subset of 25 responses each time, comparing their codes and iterating on the codebook until they achieved high inter-rater reliability ($\kappa = .87$)) by the 125th response. Using the refined codebook, the two authors continued to code independently until they stopped observing new codes by the 175th response. Then they coded two additional batches of 25 responses, reaching 225 responses, and marked it as the point of data saturation. After conducting a thematic analysis, we model the developers' engagement with the Data Safety Section ecosystem through an analytical framework as depicted in Figure 7. Our codebook is available online.[3].

Demographics and App Metadata We analyzed responses from developers of 225 apps. We did not collect demographic data from the developers because we did not want to: (1) ask for personal information, and (2) create friction for developers to respond. More importantly, we interacted with developer emails that potentially cover developer teams with varying demographics where inquiring about the demographics of the respondent would not be necessarily representative of the whole developer team. To understand the demographics of the responses, we collect the geographic location of the apps team, as listed on the Play Store. For apps that do not list their address (n=89), we manually visit their webpage and extract the address. We find that we covered apps from 59 countries, covering 5 regions (continents), showcasing the advantages of our approach in reaching diverse developers. The top 3 countries are *United States - 29*, *India - 25* and *Russia - 12*. A world map showing the locations (country level) of the apps is shown in Figure 1 (in Supplemental File[4]). We also find that the apps in the study cover a total of 27 categories, with the most common being *GAME* (n=66), which is also the most popular app category on Play Store. The distribution of the apps with the category is shown in Figure 12a. Treating real installs as a proxy for app popularity, we find that the study consists of popular as well as less popular apps, with 42% of apps having more than 1M downloads. We show the distribution of apps in the study with real installs in Figure 12b. Finally, we show the distribution of apps with Age rating in Figure 12c. We note that 88% of the apps belong to the *Everyone* rating. This is consistent with the distribution of all apps on the Play Store, with 86% of the apps having *Everyone* rating.

Findings Overview. Figure 7 describes the lifecycle of updating the DSS as perceived by the developers. App developers go through several stages while working with the Data Safety Section. First, they deduce their intended privacy practices (Section 6.2) based on their app implementation. While attempting to fill the DSS form, they face challenges (Section 6.3) mapping their intended practices to the form. The challenges result in strategies (Section 6.4) to initially fill out the form and get it accepted. After submitting the initial form, there might be inaccuracies highlighted to them either via Google or through their internal review system. Finally, they update the form (Section 6.5). In the subsequent sections, we present the findings of our analysis, unpacking the experiences and behaviors of app developers as they go through this process of uploading their Data Safety Sections.

6.2 Intended Privacy Practices of Developers

We first discuss developers' intended data collection and sharing practices. Our thematic analysis reveals three subthemes within the privacy practices of the apps: third-party data collection, encryption, and first-party practices.

Third-party Data Collection: We find that developers often reported involving third parties for various purposes. The main purposes were ads and analytics. We also find here that developers tend to use these libraries primarily for ad revenue. For example, one developer expressed: "*... to make profit from my DJ app, I use third party advertising, such as Google AdMob and AppLovin. They (third-party ads) are collecting users data to showing perfect ads to my DJ users*." This is consistent with our findings from Section 4 where we find that *Ads* and *Analytics* are the top two purposes for data sharing.

However, we observe developers do not tend to share data when they employ in-app purchases. For example, when asked about changes in their DSS, one developer noted: *"In the last update, I removed all the ads of the app, so I receive money from the few premium subscriptions. The Google AdMob SDK collected all that data, so now it doesn't get any type of information from the user."* This suggests developers of paid apps tend not to share data possibly as they do not rely on ads for revenue. In Section 3, we observed that the fraction of apps with DSS in paid apps is higher than in free apps. Analyzing the *Data Sharing* practices of paid apps (Figure 8), we indeed find that a lower number of paid apps report sharing data than free apps.

Encryption: We find that developers often secured data during transmission. One developer reported–"*we currently utilize the IPsec protocol to ensure secure transmission of data. IPsec is a widely adopted industry standard for VPNs and*

[3] https://osf.io/92sez/?view_only=0b4aa040161a4b259c0a32c7fb3ae82c

[4] https://osf.io/q3wv2?view_only=0b4aa040161a4b259c0a32c7fb3ae82c

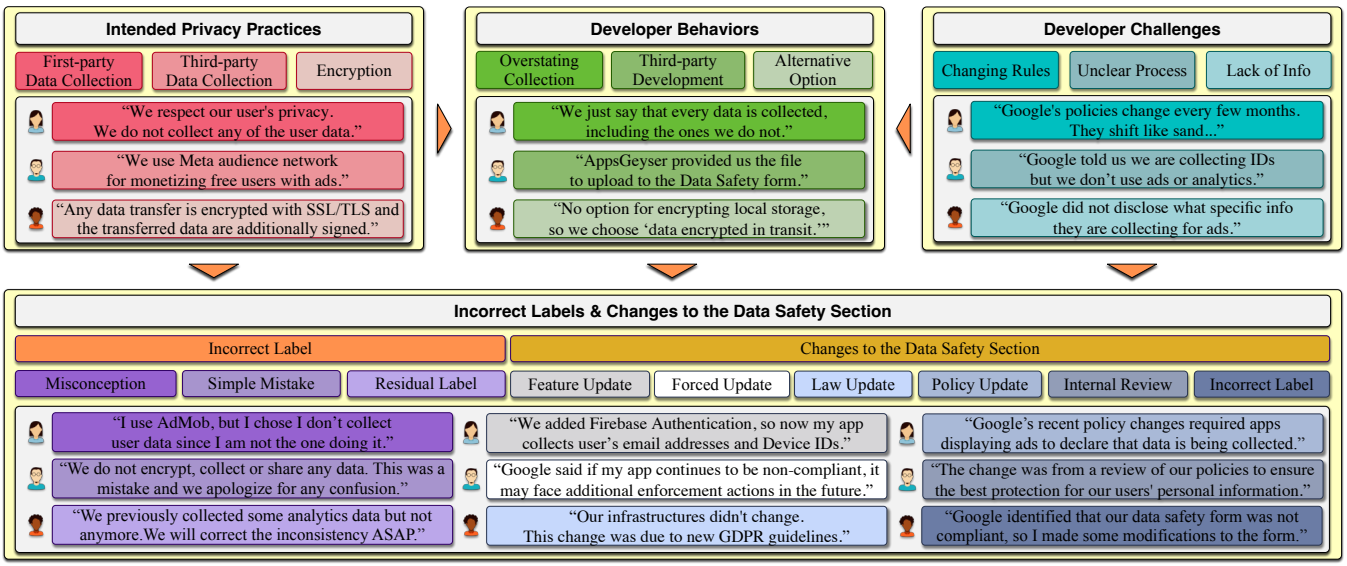


Figure 7: Our analytical framework with the high-level themes, secondary-level themes, and developer quotes. The framework describes the lifecycle of updating the DSS as perceived by the developers. We do not show all secondary-level themes due to a lack of space.

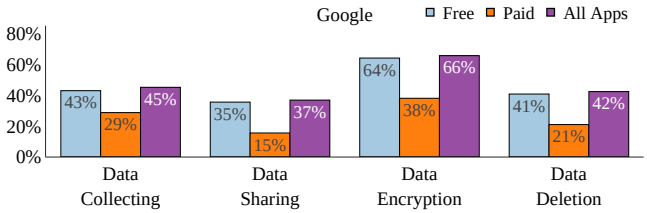


Figure 8: The distribution of high-level data practices based on price. We note that the fraction of paid apps that collect and share data is lower than free apps.

provides robust encryption, authentication, and integrity protection for transmitted data." Some developers also assumed that third-party libraries will provide encryption, although this might raise questions about their direct control over data security. For example, one developer mentioned, "... *I use Google API like Drive, YouTube, AdMob and Firebase. They do collect user data I guess and they state that data is encrypted*".

First-party Practices: In our analysis, developers also conveyed their own data practices. Many developers reported not intending to collect (n=60%) or share data (n=48%) themselves. As one developer stated, "... *we do not collect any personal data or share any personal data. Our app only collects anonymous analytics events (button clicks and screen impressions/load) and app crashes via Google's Firebase framework so that we can fix user issues and crashes*." Others asserted not integrating any ads or analytics services, while some stressed the importance of obtaining user consent before collecting data, underlining their commitment to respecting user privacy.

6.3 Challenges Faced by Developers

Next, we uncover the challenges that the developers face while mapping their intended privacy practices to DSS forms. These challenges are categorized into seven sub-themes with Figure 9 depicting the frequency of each sub-theme for each region. We observe that the theme covers apps from all the regions highlighting the diversity of the app developers included in the analysis. We also observe that Europe emerges as the region where developers most frequently report numerous challenges. Additionally, we also show the distribution of sub-themes with respect to app categories (Figure 13) and the number of downloads (Figure 15) in the Appendix. We find that the apps cover 20 categories present on the Play Store. Similarly, analyzing the apps along the download count, we find that the theme consists of apps with low download count as well as high download counts.

Lack of Information About Third-Party Libraries: Developers indicated uncertainty about the privacy practices of the third-party libraries used in their apps. One developer noted: "*I currently face a challenge in that I am unsure of what specific APIs are included in the app code provided by Appsgeyser.com that may be contributing to this issue of non-compliance, as said by Google. Additionally, I have been unable to find a reliable source of information on this matter.*" Notably, in the play console, Google shows the permissions required by popular libraries, as well as flags problematic SDKs, however, there is no help regarding the collection practices of these libraries.

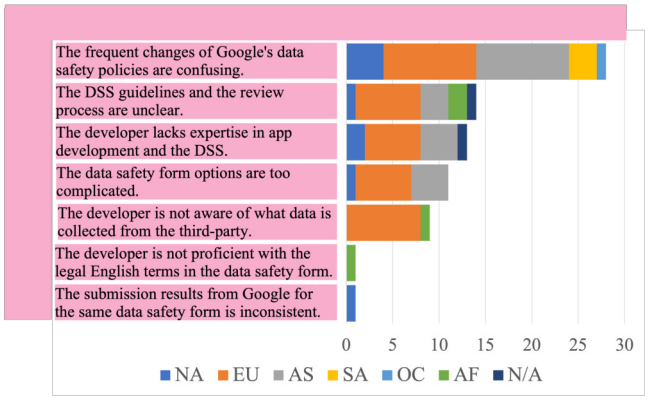


Figure 9: The frequencies of seven subthemes of developer's challenges with different geographic locations, as explained in Section 6.3. The region abbreviations are–NA: North America; EU: Europe; AS: Asia; SA: South America; OC: Oceania; AF: Africa. N/A corresponds to the apps whose region we were not able to identify.

Inconsistency: Some developers reported inconsistency in the acceptance of their data safety form. A developer struggling with the review process noted: *"... we submit the same form over and over again and often times Google rejects our answers with no or at most a vague explanation. Eventually, Google accepts it."* Such inconsistency can lead to confusion about what is required, and result in frequent revisions, and inaccurate DSS.

Lower English Proficiency: Non-native English speakers cited difficulties understanding the data safety policies. For example, a developer, whose native language is not English, was confused with an English word: *"... our administrator whose first language isn't English did not seem to understand the meaning of 'ephemerally' and ticked 'No, this collected data is not processed ephemerally'. So even though we declared the data collected/shared, Google play did not disclose this on the app's store listing."*

Complicated: Developers also complained that options provided in DSS form are complicated. For instance, a developer made mistakes in the data safety form because there were simply too many options, whereas another was confused with the complicated description of the DSS option. This complexity can lead to errors or misunderstandings when filling out the form.

Unclear Process: Our analysis also revealed that developers are confused regarding Google's review process for DSS, finding the process unclear and the explanations vague. For instance, a developer noted *"Google at one time prompted me that my app collects data of which we knew nothing about, the data they spoke about is the 'Devices and Other IDs Data' and that it's a compulsory data been collected by most apps on Play Store even if they do not collect user data. We don't see or collect any user data from my users and our app doesn't not even use firebase, one signal or ads."*

Lack of Expertise: Developers also acknowledged difficulty arising from a lack of expertise in both app development and data safety policies. For example, one developer noted–*"... due to my limited experience, I made this app with the help of third party app developer-AppsGeyser. I am not sure of how I should represent the third-party APIs in the app code and I have been unable to find a reliable source of information on this matter.."*. Another developer noted that they do not understand half of the DSS policies and so they just keep submitting answers in hopes of finding the right configuration. This gap in knowledge can happens due to vague definitions in data safety policies. Moreover, the lack of information can lead to developers having an incomplete understanding of the privacy practices, resulting in inaccurate representation of these practices in the DSS.

Changing Rules: Developers reported confusion stemming from changing DSS policies within the Google Play Store. They mentioned that the constant evolution of these rules made it difficult for them to ensure compliance. As one developer said: *"... so we haven't changed anything in <app_name> in almost a decade. What has changed, and seems to keep changing every few months is Googleś privacy policies. They are difficult to understand and they shift like sand ..."*. This indicates that developers have difficulty keeping up with the dynamic landscape of rules and regulations regarding DSS.

These challenges highlight the need for more developer support in areas such as policy comprehension, form simplification, and the handling of third-party libraries. Addressing these issues could significantly aid developers in enhancing data safety and privacy in their applications.

6.4 Developer Behaviors and Strategies in DSS Submission

We now discuss the various strategies adopted by the developers to navigate the challenges they face while submitting the DSS form. These behaviors and strategies are classified into five sub-themes. Figure 10 shows the frequency of each sub-theme in the coded responses with each geographic region. The theme consists of app developers from all regions. We note that the sub-themes *Future Policy* and *Overstating Collection* are found primarily in the *Asian* region. This could be reflective of practices in a specific region; we leave this exploration for future studies. Furthermore, the distribution of sub-themes with app category and download count are shown in Figure 14 and Figure 16, respectively in Appendix A.4. We observe that the theme covers 7 categories, including the top 5 app categories on the Play Store. The theme also has apps with diverse download counts, with apps having 1K+ downloads as well as 10M+ downloads.

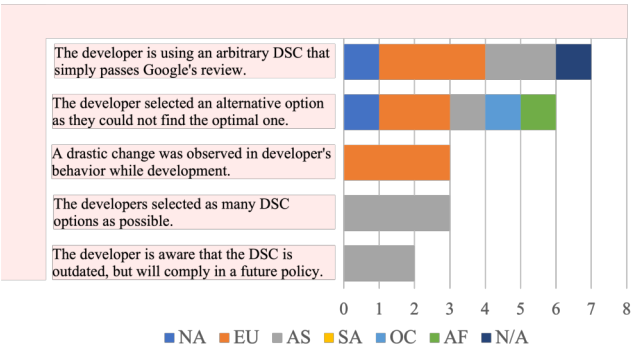


Figure 10: The frequencies of five subthemes of developer's behaviors with different geographical regions, as explained in Section 6.4. The region abbrevations are same as Figure 9.

Alternative Option: When developers were unable to find the optimal data safety card option for their app, they resorted to choosing an alternative one. For instance, as said by one developer–"*... we do not collect any data, everything user write on our app will be stored on their device. There wasn't any other option to select except 'data is in encrypted in transit ...*" . This highlights the potential need for more comprehensive or flexible options within the form.

Change in Developer Behavior: Our analysis also captured developers altering their behavior based on their experience while interacting with DSS. For instance, one developer that stopped the development of their app noted: *"... considering all the hassle I had been going through, it has been determined that it would not be practical to continue updating these apps on the Google Play Store, given the minimal profit generated and the effort required to maintain compliance."* This highlights the scenario where the complex policies of DSS can potentially discourage developers to update their apps, hurting the users as they might be forced to work with outdated versions of the apps.

Future Policy: Some developers acknowledged that they intentionally had an incorrect DSS to increase flexibility, and in case of future app updates. For example, one developer responded: *"It is true that we do not share or collect any kind of personal data. It means that if we intend to do so in the future then we will be encrypting the data. At the moment we are not sharing or collecting personal data."* This might convey incorrect information to the users. Conversely, if the developer does not indicate that they collect data, but utilize the same flexibility, it could pose a risk to users.

Overstating Collection: Some developers indicated that Google only checks for the data that is collected but not mentioned as collected but not the other way. As a result, they select all options for collection in the DSS form. For example, one developer noted –*"... it is safe to declare all possible data collected in Data safety section so, in our apps, even the information (e.g. phone numbers) that is not collected is mentioned as collected."* We also observed similar over-reporting of practices in Section 4 and Section 5. This observation highlights an issue with Google's compliance analysis tools.

Form That Works: Some developers reported using a data safety form that had previously passed Google's review, regardless of whether it fully aligned with the current data handling practices of the app. For example, one developer noted –*"... my app did not collect any data. but in the data safety section, I wrote that I collect data because other than that Google would not publish my app."* This approach suggests a pragmatic yet concerning response by the developer as it may result in inaccurate DSS.

The findings highlight that developers attempt to circumvent the review process either by over-reporting their privacy practices or by trying random configurations until their DSS is accepted. Combined with the challenges (Section 6.3) that the developers face, this indicates that the DSS submission tool might not be usable by the developers.

6.5 Inaccurate Labels and Driving Factors behind Changes in Data Safety Section

As the developers navigate the complex challenges, they may submit inaccurate labels. We used the developers' intended practices from Section 6.2 and compared them with the reported practices in DSS. Additionally, our analysis of the developers' responses revealed several key factors that influence developers' decisions to update their applications' Data Safety Sections (DSSs).

6.5.1 Inaccurate Labels and Type of Inaccuracies

We conducted a comparative analysis between the reported practices (*Data Collection* and *Data Sharing*) in DSS and their intended practices from the emails. We find a notable discrepancy; 41% of developers who stated in their emails not to engage in data collection were nonetheless reported as collecting data in their DSS. A similar inconsistency (42%) was observed for data sharing. For instance, the developer of an app with over 100M downloads stated in an email, "*In Google Play Console we pointed that we collect/share data because of Ad network SDKs integrated into the game. Not the app itself, but ad SDKs do collecting or sharing.*", but their DSS did not report any data collection or sharing. This discrepancy highlights the fact that the DSS sections might not be thoroughly monitored by Google and reiterates the need for more effective mechanisms for accurate reporting of privacy practices. Next, our analysis of the developers' responses also reveals factors responsible for inaccuracies.

Misconception: Some developers demonstrated misunderstandings of the data safety policies, which resulted in inaccurate DSS. Such developers have the misconception that they do not have to disclose practices for third-party libraries

even when they are using third-party SDKs. For example, one developer noted –*"... we do not collect any personal data or share any personal data. Our app only collects anonymous analytics info through Google Firebase. Hence, we chose that we do not collect data."*

Recall the *interesting* trend that we observed in Section 4 where we noted that many apps state they do not collect or share data, but encrypt data in transit. We find that this could be happening due to misconceptions about using third-party libraries. For example, one developer noted – *"I use ADMOB for banner ads at the bottom of my app, but Google collects Device ADID. I checked that data is encrypted for that part."* The developers report *Encryption*, but due to the misconception, do not report *Data Shared*.

Simple Mistake: Some Developers also acknowledged that they made a mistake and conveyed their plans to update their DSS during the communication with the authors. For example, one developer mentioned –*"previously, the process of filling out the Data safety section was interrupted and uncompleted due to one of the checkboxes left unchecked."*

Residual Label: Some developers failed to update their DSS after making feature updates. The developers conveyed their plans to update their forms, but this trend can potentially be detrimental to the DSS ecosystem, especially if the update adds *Data Collection* or *Data Sharing* flows.

6.5.2 Driving Factors for Change in DSS

After submitting the initial forms, developers may need to update their DSS. We now discuss the various factors responsible for the change in DSS. Our analysis revealed six distinct sub-themes for these factors.

Feature Update. One significant trigger for updating DSSs was the introduction of new features. Most commonly, developers updated their DSSs due to the addition of advertisement or analytics SDKs into their app. For example, one developer notes: *"... In the initial version there were no ads. But now I have put ads in the app for which I had to change my policy as well."*

Law Update. Changes in legal and regulatory frameworks also compelled developers to modify their DSSs. For example, one developer noted – *"Our methods and infrastructures didn't change. This change was made in light of new GDPR guidelines that are in effect right now."*

Policy Update. Changes imposed by the platform (like the Play Store), required adjustments to the DSS. For instance, one developer using a third-party ad service mentioned that: *"We don't collect or transmit any user data. However, some of our apps use Google's AdMob for ads. As per new directions from Google, we are required to reproduce Google data policies related to AdMob even if our app does not collect any data."*

Forced Update: We also observed instances where developers were compelled to update their DSS to avoid having their apps removed from the Play Store. For example, one developer noted – *"We detected user data transmitted off device that you have not disclosed in your app's Data safety form as user data collected. If your app continues to be non-compliant after August 22, 2022, your app updates will be rejected and your app may face additional enforcement actions in the future."* This showcases the influence of platform policies on developers' decision-making processes.

Internal Review. The results of internal reviews also surfaced as a factor for DSS updates. Specifically, a developer mentions that the updated DSS guidelines provided by Google enabled them to more accurately portray how they handle data. They also emphasize that: *"... the added features on the stores rightfully illuminated some areas that were unknown to users, which made a lot of apps appear to handle their data differently than they actually are."*

Incorrect Label. Finally, developers noted that the discovery of incorrect labels results in DSS updates. As noted by a developer–*"... Although I still do not collect any data, I was essentially required by Google some time ago to say that I collect data on behalf of the advertising services in order to remain compliant with Google's policy ..."*–most of the developers fixed the DSSs after they realized that they chose the incorrect data safety card option due to their confusion after being formally notified by Google.

7 Discussion and Recommendations

Our research has highlighted the complex landscape of developers' experiences and challenges with Data Safety Sections. We follow on our findings by presenting a set of recommendations aimed at improving the developer experience with DSS. Then, we discuss the roles of the platforms and regulators in addressing the deficiencies in DSS. Finally, we list the limitations of our methods.

7.1 Recommendations

We provide a list of recommendations to improve the developer experience with the DSS section in Google Play. These recommendations, summarized in Table 1, go beyond those presented in prior work [22].

Enhance Educational Resources. An important need emerged for enhancing educational resources surrounding data safety policies and the filling of data safety forms. Such educational resources include detailed explanations about data collection, storage, and sharing practices that are expected to be reported in the Data Safety Section. Providing developers with real-world examples, interactive tutorials, and guidelines to navigate data safety requirements can go a long way in

Recommendations
1. Enhance Educational Resources: Educational resources should be enhanced to provide developers with real-world examples, interactive tutorials, and guidelines.
2. Provide Multilingual Support: Resources should be available to developers in multiple languages
3. Simplify Data Safety Forms: Data safety forms should be simplified to improve comprehension.
4. Consistent Feedback from the Review Process: Developers should be given clear and transparent feedback on the approval or rejection of their data safety forms.
5. Improve Support for Third-Party Library Data Practices: Platform providers and third-party library developers should ensure transparency about their data practices.
6. Regular Consultations and Feedback Mechanisms: Consult with developers to ensure continued relevance of privacy labels; allows developers to share their feedback.

Table 1: Summary of the recommendations.

reducing misunderstandings and ensuring proper compliance.

Provide Multilingual Support. The global landscape of app development calls for the need to support diverse languages. Developers across the globe should be able to access, understand, and interpret guidelines without language acting as a barrier. Thus, providing guidelines, forms, and supporting services in multiple languages is crucial. Platform providers could consider deploying multilingual support teams and translation services to cater to this diverse community.

Simplify Data Safety Forms. Several developers have reported difficulties in understanding the complexity of data safety forms (Section 6.3, leading to inaccuracies in reporting. Addressing this concern would involve simplifying the forms to ensure they are easy to understand and complete. This could be achieved by using accessible language, clear terminologies, and unambiguous options. Additionally, redesigning the form layout to enhance readability and ease of use might also prove beneficial.

Consistent Feedback from the Review Process. In Section 6.3, we find that developers have little confidence in the approval process. As a result, developers resort to filling the form to just receive the approval, without the form actually being accurate. Developers expressed frustration as to understanding why a particular form is accepted or rejected. Developers need to have confidence in a consistent and transparent review process. The approval or rejection of data safety forms should be communicated clearly with an explanation of the reasons. This will not only help developers improve their subsequent submissions but also reduce confusion.

Improve Support for Third-Party Library Data Practices. The increasing use of third-party libraries poses challenges for developers to accurately represent data practices in their apps. To mitigate this challenge, platform providers should demand transparency from third-party providers about their data practices. Additionally, developing tools or mechanisms that help developers to track and represent these practices in their data safety forms would be beneficial. An existing effort from Google is the *Google Play SDK Index*, which aims to provide developers transparent information on all the third-party SDKs that is usable in Android app development [16]. While the index informs the users of the Android OS permissions these services are requesting, it still fails to provide information on the specific category of data being collected.

Regular Consultations With Developers and Feedback Mechanisms. To ensure the continued relevance and effectiveness of privacy labels, regular consultations with developers should be conducted. This would provide a platform for developers to share their experiences, voice their concerns, and give feedback on existing processes. Such feedback mechanisms could provide invaluable insights for platform providers to understand evolving challenges and adapt their policies and support systems accordingly.

7.2 Discussion

Impact of Industry Intervention. Google's recent introduction of the *Checks* service [5], providesing paid compliance analysis to developers, adds to the dynamics of the privacy labels. From the developers' perspective, popular apps with resources might find Checks to be an invaluable tool, simplifying the task of policy compliance and mitigating the likelihood of errors or misunderstandings. However, the service's cost may create hurdles for smaller or independent developers, possibly leading to disparity between reported practices by developers with more substantial means and those operating under more constrained circumstances.

As for the platform, the Checks service could fulfill several objectives for the Play Store. It might elevate the quality of reported practices and compliance of apps. However, Google could face criticism for monetizing a critical aspect of the app development process, particularly if this act is seen as establishing a 'pay-to-play' mechanism.

The implications for users, albeit less direct, are still significant. If the Checks service contributes to higher policy compliance and reduced errors by developers, users could get accurate data safety practices. Conversely, if the cost of the service results in a less diverse app marketplace due to financial barriers for smaller developers, users may face a reduction in their choices of apps.

Regulators Even though our analysis finds inconsistencies

between privacy labels and privacy practices, evidence [34] suggests that privacy labels have can carry specific information about the practices. This information can be very useful for privacy concerned users who want to ensure that they are only using privacy respecting apps.

However, as shown in this work, the accuracy of privacy labels is not guaranteed. While developers are required to disclose their data practices in order to obtain a privacy label, there is no guarantee that the information they provide is accurate or complete. Therefore, it is necessary to have systems in place to verify the accuracy of privacy labels and to hold developers accountable for any discrepancies. This is particularly important because the false labels can create a false sense of security among the users.

One potential model for regulating privacy labels is a system similar to the one used for food nutrition labels, which are regulated by the Food and Drug Administration (FDA). A regulatory body could be established to oversee privacy labels and ensure that they are accurate and consistent. This could help to build trust among users and encourage developers to be more transparent about their data practices.

Another solution could be providing the monitoring system that is used during the app submission reviews directly to the user's device. The system could perform real-time dynamic analysis on the app installed on the user's device and show the analysis results, such as network traffic and the data or sensor access logs. This way, the users are given a more detailed view of the privacy practices of the app and would be able to make a better-informed decision on whether to use the app. An example in production includes the Apple *App Privacy Report* introduced in iOS and iPadOS 15.2, which provides data and sensor access logs from apps installed on the device, and network traffic from the apps directly to the users [30].

7.3 Limitations

Data Collection Our data collection process was carried out over specific periods, and thus may not fully capture the dynamic nature of the app ecosystem. Apps can update their DSS at any time, and changes outside of our data collection windows would not be reflected in our analysis. The apps included in our study are a subset of all available apps on the Google Play Store. Although we made an effort to include popular apps and span various categories, our sample may not be fully representative of the entire app ecosystem. Finally, we noted discrepancies in data practices based on the presence of third-party libraries. However, some instances of third-party library usage might have been missed due to limitations in the tools used for detection, leading to potential underestimation of their prevalence and impact.

Developer Study Our developer study relies on email communication with app developers. While this allowed us to reach a larger number of developers, it may not have provided us with the depth or understanding that comes from an interview study. Additionally, it may have also biased our results toward those who were more willing to respond, potentially neglecting the perspectives of developers who did not reply.

The nature of self-reported data could also pose challenges to the reliability and accuracy of the information collected. There is the potential for bias in the developers' responses, as they may present information in a way that portrays their apps more favorably. Another limitation of our study is the lack of demographic data about the developers who responded to our email. As we prioritized respecting privacy and ensuring anonymity, we did not collect any personally identifiable information, including demographic details such as age, gender, nationality, or years of experience in app development. Such demographic data could potentially provide meaningful context and allow for a more nuanced understanding of developers' perspectives. For instance, a developer's experience level or geographical location might influence their understanding of privacy issues or their familiarity with privacy regulations. Despite the limitations outlined, our research sheds light on the challenges that the developers face, and it fills a gap in the literature about the factors affecting developers' making while working with the privacy labels.

8 Conclusion

In conclusion, our study takes a comprehensive approach to examine the landscape of Data Safety Sections (DSS) in Google Play Store apps. Through a large-scale analysis, we highlighted inconsistencies in reported practices, as well as instances of both underreporting and overreporting. Our longitudinal study emphasized the dynamic nature of DSS implementation. We observed the persistence of overreporting trends and revealed how developers are still adjusting to the requirements, even ten months after Google's imposed deadline. We investigate the developers' perspective by examining responses from them about their DSS. Our analysis uncovers the process developers undertake when navigating the DSS landscape. We outline the challenges developers face, the strategies they employ to comply with DSS policies, and the factors prompting changes in their DSS. Finally, based on the challenges and developer behaviors, we provide recommendations aimed at improving the developer experience with Data Safety Sections.

Acknowledgments

This work was supported by the NSF through awards: CNS-1942014 and CNS-2003129, and by gifts from Google, NVIDIA and Meta. Finally, we thank the reviewers for their fruitful discussions and recommendations.

References

[1] Kevin Allix, Tegawendé F. Bissyandé, Jacques Klein, and Yves Le Traon. Androzoo: Collecting millions of android apps for the research community. In *Proceedings of the 13th International Conference on Mining Software Repositories*, MSR '16, pages 468–471, New York, NY, USA, 2016. ACM.

[2] David G Balash, Mir Masood Ali, Xiaoyuan Wu, Chris Kanich, and Adam J Aviv. Longitudinal analysis of privacy labels in the apple app store. *arXiv preprint arXiv:2206.02658*, 2022.

[3] Rebecca Balebako, Abigail Marsh, Jialiu Lin, Jason I Hong, and Lorrie Faith Cranor. The privacy and security behaviors of smartphone app developers. *Citeseer*, 2014.

[4] Rebecca Balebako, Florian Schaub, Idris Adjerid, Alessandro Acquisti, and Lorrie Cranor. The impact of timing on the salience of smartphone app privacy notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 63–74, 2015.

[5] Nia Castelly and Fergus Hurley. Introducing Checks: simplifying privacy for app developers - Google: The Keyword. https://blog.google/technology/area-120/checks/, Feb 2022. Date accessed: 2023-06-01.

[6] Fred H Cate. The limits of notice and choice. *IEEE Security & Privacy*, 8(2):59–62, 2010.

[7] Lorrie Faith Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.*, 10:273, 2012.

[8] Lorrie Faith Cranor. Mobile-app privacy nutrition labels missing key ingredients for success. *Communications of the ACM*, 65(11):26–28, 2022.

[9] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. Ask the experts: What should be on an iot privacy and security label? In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 447–464. IEEE, 2020.

[10] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. Which privacy and security attributes most impact consumers' risk perception and willingness to purchase iot devices? In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 519–536. IEEE, 2021.

[11] Grace Fox, Colin Tonge, Theo Lynn, and John Mooney. Communicating compliance: developing a GDPR privacy label. In *24th Americas Conference on Information Systems*, 2018.

[12] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain, and Norman Sadeh. Helping mobile application developers create accurate privacy labels. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 212–230. IEEE, 2022.

[13] Joshua Gluck, Florian Schaub, Amy Friedman, Hana Habib, Norman Sadeh, Lorrie Faith Cranor, and Yuvraj Agarwal. How short is too short? implications of length and framing on the effectiveness of privacy notices. In *Twelfth symposium on usable privacy and security (SOUPS 2016)*, pages 321–340, 2016.

[14] Google. Provide information for google play's data safety section. https://support.google.com/googleplay/android-developer/answer/10787469, 2022. Date accessed: 2023-08-12.

[15] Google Play Help. Understand app privacy & security practices with Google Play's Data safety section. https://support.google.com/googleplay/answer/11416267?sjid=17006176392115416702-NA, 2022. Date accessed: 2023-06-01.

[16] Google Play Console Help. Make informed choices with Google Play SDK Index. https://support.google.com/googleplay/android-developer/answer/12034434?hl=en, 2022. Date accessed: 2023-06-01.

[17] Google Play Console Help. Provide information for Google Play's data safety section. https://support.google.com/googleplay/android-developer/answer/10787469?hl=en, Mar 2023. Date accessed: 2023-06-01.

[18] JoMingyu. google-play-scraper: Google play scraper for python. https://github.com/JoMingyu/google-play-scraper, 2022. Date accessed: 2023-06-01.

[19] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, New York, NY, USA, 2009. Association for Computing Machinery.

[20] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing privacy notices: An online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1573—-1582, New York, NY, USA, 2010. Association for Computing Machinery.

[21] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, page 3393–3402, New York, NY, USA, 2013. Association for Computing Machinery.

[22] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I Hong. Understanding challenges for developers to create accurate privacy nutrition labels. In *CHI Conference on Human Factors in Computing Systems*, pages 1–24, 2022.

[23] Yucheng Li, Deyuan Chen, Tianshi Li, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I Hong. Understanding ios privacy nutrition labels: An exploratory large-scale analysis of app store data. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7, 2022.

[24] Jialiu Lin. *Understanding and capturing people's mobile app privacy preferences*. PhD thesis, Carnegie Mellon University, 2013.

[25] Ziang Ma, Haoyu Wang, Yao Guo, and Xiangqun Chen. Libradar: Fast and accurate detection of third-party libraries in android apps. In *2016 IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C)*, pages 653–656, 2016.

[26] Aleecia M McDonald, Robert W Reeder, Patrick Gage Kelley, and Lorrie Faith Cranor. A comparative study of online privacy policies and formats. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 37–55. Springer, 2009.

[27] Benjamin Saunders, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity*, 52(4):1893–1907, Sep 2017.

[28] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Eleventh symposium on usable privacy and security (SOUPS 2015)*, pages 1–17, 2015.

[29] Gian Luca Scoccia, Marco Autili, Giovanni Stilo, and Paola Inverardi. An empirical study of privacy labels on the apple ios mobile app store. In *9th IEEE/ACM International Conference on Mobile Software Engineering and Systems 2022*, 2022.

[30] Apple Support. About app privacy report. https://support.apple.com/en-us/HT212958, Jul 2022. Date accessed: 2023-06-01.

[31] Haoyu Wang, Yao Guo, Ziang Ma, and Xiangqun Chen. Wukong: A scalable and accurate two-phase approach to android app clone detection. In *Proceedings of the 2015 International Symposium on Software Testing and Analysis*, pages 71–82, 2015.

[32] Google Workspace. About app review. https://developers.google.com/workspace/marketplace/about-app-review#areas, May 2023. Date accessed: 2023-06-01.

[33] Yue Xiao, Zhengyi Li, Yue Qin, Jiale Guan, Xiaolong Bai, Xiaojing Liao, and Luyi Xing. Lalaine: Measuring and characterizing non-compliance of apple privacy labels at scale. *arXiv preprint arXiv:2206.06274*, 2022.

[34] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. How usable are ios app privacy labels? *UMBC Faculty Collection*, 2022.

A Appendix

A.1 Changes in DSS

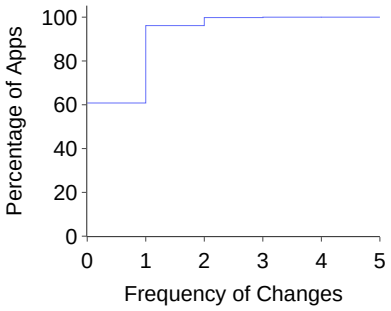


Figure 11: The Cumulative Distribution Function for the frequency of DSS updates made by apps throughout the snapshot timeline.

Figure 11 shows the frequency with which apps change their DSS between the period of June 20, 2022, and May 31, 2023. We find that a significant number, 283K (40%) changed labels at least 1 time. Moreover, there were over 1.3K apps that changed their DSS at least 3 times.

A.2 Data Practices in Privacy Labels

In this section, we look into the distribution of the high-level practices of apps based on their price, as shown in Figure 8. Next, we look at the heatmap shown in Figure 2 (in Supplemental File[5]) to see the distribution of datatypes across purposes. The heatmap is normalized by the total number of apps collecting or sharing a given datatype.

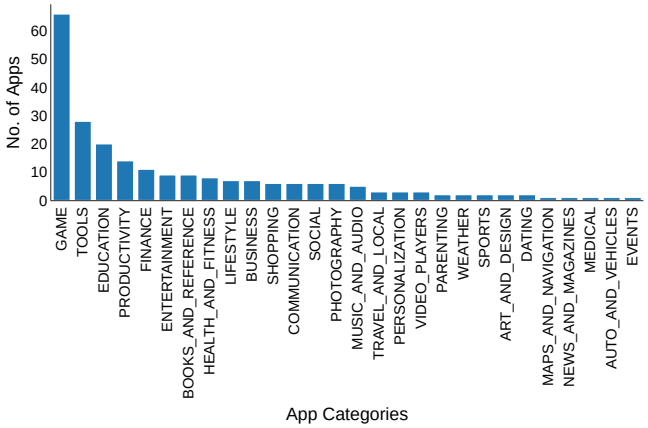
A.3 Developer Study

For the Developer Study (Section 6) we sent emails to developers in 3 different categories: (A) apps stating that they encrypt data without collecting or sharing data, (B) apps changing their practice from not collecting/sharing data to collecting/sharing data, and (C) apps changing their practices from collecting/sharing data to not collecting/sharing data.

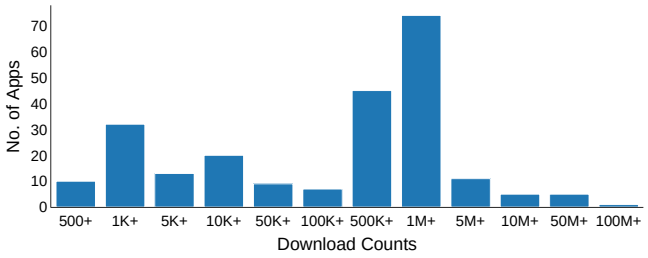
For category (A) we used the following template:
We hope this email finds you well. We are researchers at <LAB_NAME> and have been using your app, <APP_NAME>, in our recent studies. We have noticed that in the data safety section of your app, it states that you encrypt data. However, we have also noticed that your app does not collect or share data.

We are reaching out to ask if you could clarify this for us. We are trying to better understand the data safety section

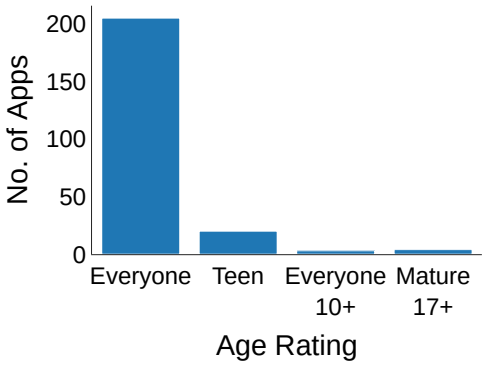
[5] https://osf.io/q3wv2?view_only=0b4aa040161a4b259c0a32c7fb3ae82c



(a) Apps per category (or genre)



(b) Apps per download counts



(c) Apps per age rating

Figure 12: Distribution of apps coded according to (a) download counts, (b) categories, and (c) age ratings

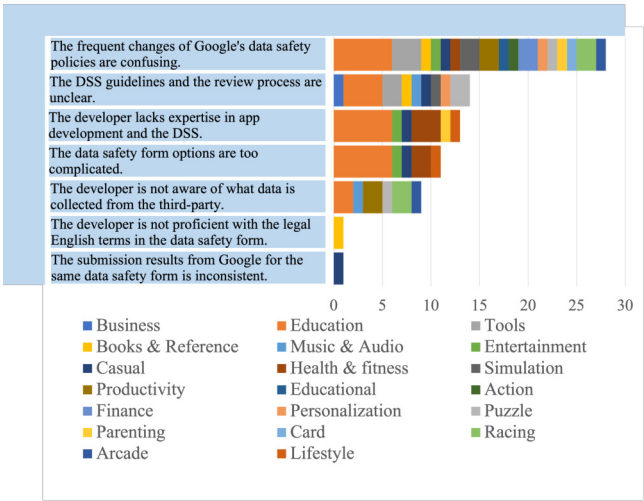


Figure 13: The frequencies of seven subthemes of developer's challenges with different app categories, as explained in Section 6.3.

implemented in your app. We appreciate any information you can provide.
Thank you for your time and we look forward to your response.

We also sent one additional follow-up email to obtain informed consent to use the developers' responses in our research. The content of the email was:
Thank you for contributing to our study. May we have your permission to use anonymized excerpts from your feedback in our research? These ***quotes will not be linked to you or your application*** *and solely represent a developer's viewpoint.*

A.4 Coded App Demographics

For the 225 apps we used for the coding task, we present the metadata distribution. Figure 13 and Figure 14 demonstrate the distribution of the 225 apps based on their categories, each respectively on the reported developer challenges and behaviors. Specifically, 24 out of the 77 reported challenge instances and 12 out of the 20 behavior instances originate from the *Education* category.

Figure 15 and Figure 16 demonstrate the distribution of the 225 apps based on the number of downloads, each respectively on the reported developer challenges and behaviors. The most frequent numbers of downloads were between [1K, 5K) for the challenge theme, with 20 out of the 77 reported instances. For the behavior theme, the most frequent numbers of downloads were between [5K, 10K), with 9 out of the 20 reported instances. In addition, 21 apps from the challenge instances and 5 apps from the behavior instances had a download count of 1M+.

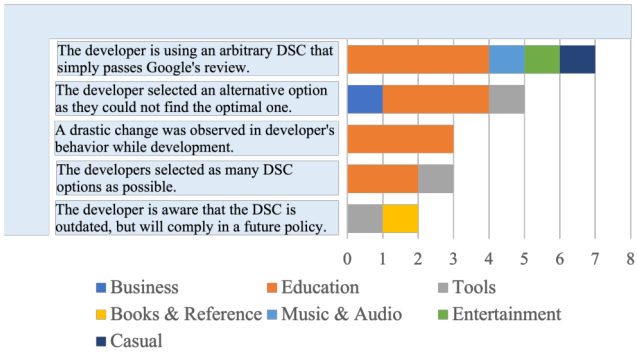


Figure 14: The frequencies of five subthemes of developer's behaviors with different app categories, as explained in Section 6.3.

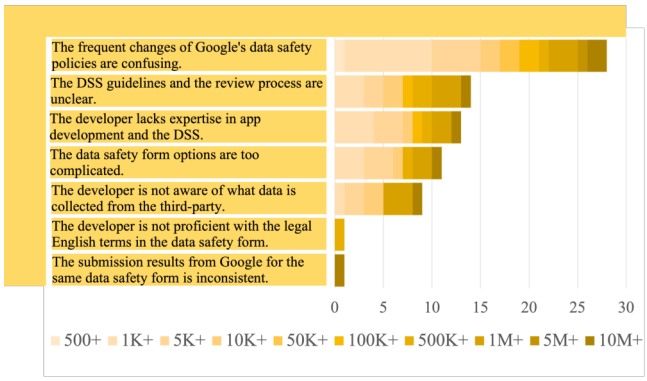


Figure 15: The frequencies of seven subthemes of developer's challenges per the number of app downloads, as explained in Section 6.3.

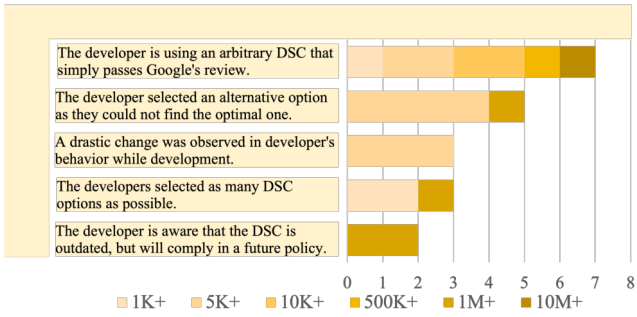


Figure 16: The frequencies of five subthemes of developer's behaviors with per the number of app downloads, as explained in Section 6.3.