# Tracking in apps' privacy policies

**Please check the document version of this publication:**

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

# Tracking in apps' privacy policies

Konrad Kollnig

University of Oxford

firstname.lastname@cs.ox.ac.uk

August 2019 (Revised November 2021)

## Abstract

Data protection law, including the General Data Protection Regulation (GDPR), usually requires a privacy policy before data can be collected from individuals. We analysed 15,145 privacy policies from 26,910 mobile apps in May 2019 (about one year after the GDPR came into force), finding that only opening the policy webpages shares data with third-parties for 48.5% of policies, potentially violating the GDPR. We compare this data sharing across countries, payment models (free, in-app-purchases, paid) and platforms (Google Play Store, Apple App Store). We further contacted 52 developers of apps, which did not provide a privacy policy, and asked them about their data practices. Despite being legally required to answer such queries, 12 developers (23%) failed to respond.

**Keywords:** apps; Android; iOS; GDPR; privacy policies

## 1 Introduction

App developers often share data with numerous third-parties to serve ads and analyse usage behaviour [9, 8, 2]. This *tracking* produces incredibly rich and valuable data, which can pose risks if in the wrongs hands. This problem with data collection is exacerbated by a centralisation of data by a few tech companies, essentially creating a single point-of-attack. App users are often not aware that they are being tracked. However, so far, few studies exist that analyse data collection in apps, beyond free Android apps.

At the same time, most data protection and privacy laws, including the GDPR in the EU and UK, implement some form of *notice and consent*, that is, informing the data subject about data practices and asking them for consent. In this scheme of notice and choice, ideally, a data subject should be able to familiarise themselves with the data protection terms of an app by first reading the privacy policy – but without data collection

taking place – and only make a decision on whether to give consent, after having read this policy.

In this paper, we analyse whether users are actually informed *before* data collection happens, by focusing on *tracking in apps' privacy policy webpages*. While this does not give insights into the actual data collection within apps, this allows for comparison across devices (Android, iOS), payment models (free, paid), and countries (EU, non-EU). For a subset of apps without a privacy policy, we additionally contacted the app developers for reasons.

Based on the above observations, this study aims to investigate two, related research questions:

1. To what extent do app developers allow their users to read the relevant privacy policies before collecting personal data about them?
2. How does third-party tracking compared across devices (Android, iOS), payment models (free, paid), and countries (EU, non-EU)?

## 2   The GDPR and the need for a privacy policy

The GDPR came into effect on 25 May 2018, and imposes high standards for the protection of data that relates to individuals ('personal data'). It applies to all companies that do business in the EU or UK, or process data relating to individuals from these countries. Non-compliance faces very high fines.

Personal data, as protected under the GDPR, reaches beyond obvious identifiers such as name, address, or telephone number. Other means of identification, such as online identifiers, may already suffice to relate data uniquely to an individual, rendering such data personal (Recital 30 GDPR). The same applies to device or advertising identifiers, commonly used for user identification in mobile apps [7]. Even IP addresses can qualify as personal data, especially if combined with further data [4].

The GDPR differentiates three main actors: the *data subject*, the *data controller* and the *data processor*. The data controller decides how the personal data of the data subject is processed by the data processor. In the case of mobile apps, typically, the user is the subject, and the app developer the controller – sometimes jointly with other third-parties that facilitate tracking if such is used.

The GDPR grants data subjects various rights, including the *right to be informed* (Article 12–14). This right obliges data controllers to communicate the terms of personal data processing, at the time such is processed. For mobile apps, the data controller must typically provide a *privacy policy* before app installation, which should additionally be accessible from within the app [1]. Among other aspects, the privacy policy must clarify the purposes of data processing, the legal bases upon which data processing is based, and

the recipients of personal data. For a full account of what must be disclosed, see Articles 12–14 GDPR.

An app user has a *right to be informed* about the terms of personal data collection, "at the time" of collection (Article 13 GDPR). Such is necessary for the data subject to make an informed decision about the usage of an app or any other service that processes personal data. This view aligns with the judgement of the EU Article 29 Working Party, stating that

> the data subject must have the necessary information at his disposal in order to form an accurate judgement. In order to avoid any ambiguity, such information must be made available *before* any personal data is processed [1, Section 3.4.1: Consent prior to installation and processing of personal data, emphasis added]

In the following, we will now analyse whether app users can actually read apps' privacy policies, before being exposed to potentially unwanted data collection.

# 3    Methodology

In this study, we focus on the *number of tracking parties contacted, when opening the privacy policy URL of an app.* Every app developer has the option to provide such a privacy policy on the app store presence. On iOS, such provision has been compulsory as of 3 October 2018, for every new app or update submitted to the App Store; on the Google Play Store, this obligation has now also been introduced, but after our study.

Focusing on the number of tracking parties in privacy policy pages provides a consistent metric across countries, platforms (iOS and Android) and payment models (free, paid, and in-app purchases). Further, there exist ample tools to detect such web tracking, allowing for scale and representative insights.

We first compiled a representative list of apps on iOS and Android, from different countries (UK, US, Germany, France, Poland) and across different categories (top free, top paid, top grossing, top new free, top new paid) from the online service `42matters`. Top grossing included those top free and paid apps, generating most revenue (from direct sales or in-app purchases). The *top new* and the *top new paid* categories were only available for Android apps. The iOS platform was further divided into app charts for iPad and iPhone apps. Each pair of country and category comprises 540 apps on Android and $1,500$ apps on iPhone and iPad each, leading to a total of $26,771$ unique apps. The data collection was conducted on 11 May 2019, about one year after the GDPR came into force.

For every one of these apps, we browsed its privacy policy URL, if provided on the app store presence, and counted the number of embedded trackers. The privacy policies were

opened with Google Chrome, and the trackers detected with `Ghostery`, a prominent anti-tracking tool. `Ghostery` enables Chrome users to detect and block trackers on websites, by detecting network communications to known tracker domains.

We automated this process with the website testing tool `puppeteer`, that allows to control a Chrome instance from the command line. This allowed to iterate over all policy URLs on our list of apps without supervision. For every policy, we waited waited until all HTML elements of the page were loaded, plus a further 3 seconds to allow the website scripts to load further content from the web. A modified `Ghostery` would then store information on all detected trackers to a database.

Other researchers have used dedicated website analysis tools, such as PhantomJS or OpenWPM. Whilst these tools may allow for higher detection rates and better scalability, they often do not reflect the real webbrowsing experience of a user and are sometimes sporadically updated, being research software. In fact, PhantomJS is currently not further developed, and is known to render websites differently than end-user browsers. So, we decided to use Google Chrome and `Ghostery`, to reflect a realistic end-user browsing experience and reliable tracking detection.

Some apps were found not to provide a privacy policy. So, we asked app developers what personal data they process and why they do not provide a privacy notice, under the GDPR's *right to be informed*. This right was invoked by sending emails to the developers. Developers have one month to answer such requests (Article 12 GDPR). If we had had not received an answer one week before the deadline, we sent a reminder to the app developer, similar to previous such studies. We selected app developers of the top paid apps, which came more often without privacy policy than free apps. This practice was approved by our departmental ethics board.

# 4   Results

A total number of 26,771 unique apps was inspected with regards to tracker in their privacy policy pages.

Of the considered apps, 3,996 (15%) apps provide no privacy policy on the respective app stores, yielding a total of 22,775 policy URLs to analyse. 578 policy URLs returned error pages, some of which included trackers, and are regarded as valid policy pages, since they may still expose users to unexpected tracking. Two thirds of policies (15,552, 68%) are transferred using HTTPS encryption. The top three formats to deliver policy are HTML (21,741, 95.5%), PDF (564, 2.5%), and plaintext (107, 0.5%). The top three domains are from Google Sites (480, 2.2%); Google Docs (261, 1.2%); and Iubenda, a privacy policy generator (241, 1.1%). All these website providers may include trackers, such as Google Analytics.

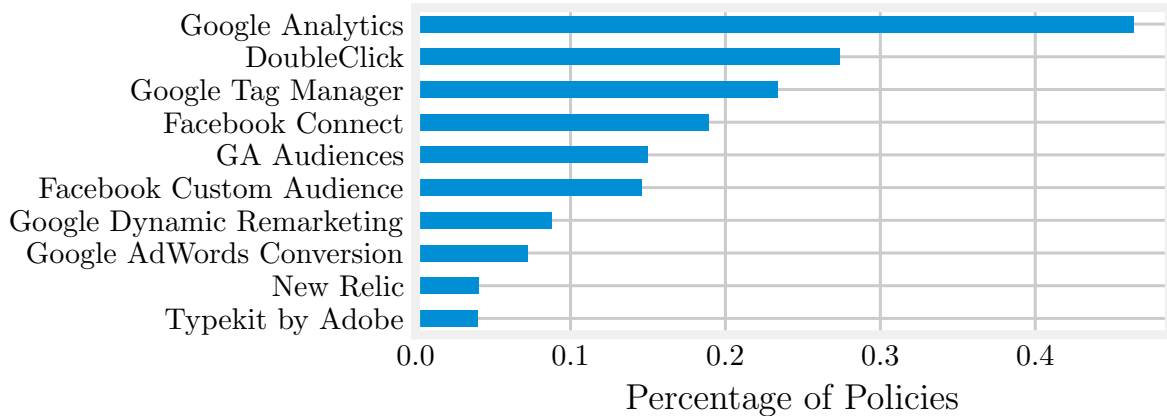Descriptive statistics about the trackers in the policies are provided in Table 1. 56%

Figure 1: Top 10 trackers on the 22,775 policy URLs.

| Count | Std | Mean | Median | 75th | 95th |
|-------|------|------|--------|------|------|
| 22 775 | 5.36 | 2.94 | 1 | 3 | 14 |

Table 1: Descriptive statistics about trackers in the 22 775 policy URLs.

of policies contain trackers, 53% analytics trackers, and 36% advertising trackers. The top 10 trackers are displayed in Figure 1; the top 8 trackers stem from either Google or Facebook.

Figure 2 compares the presence of privacy policies by store category, platform, and store country. Platform-wise, the provision of policies differs, with more Android (9%) apps providing a policy than iOS (14%). On both platforms, most *top free* and *top grossing* apps provide a privacy policy, whilst more than a fifth in the two paid categories does not (20% on Android, 37% on iOS). The same holds for new apps, which come less often with a privacy policy than *top free* and *top grossing* apps. The US store has least apps without a privacy policy (13%), whilst Poland has most (18%).

Figure 3 shows the presence of trackers in the policies, differentiating between analytics and advertising trackers. The presence of policy differs significantly between the categories, see Figure 3a. *Top free* apps show most trackers, closely followed by *top grossing* apps. The policies of paid apps contain fewer trackers than those of *top free* and *top grossing* ones, but more than of new apps. *New paid* apps come with least trackers, fewer than *new free* apps. Platform-wise, iOS apps show more trackers than Android (56%), both in analytics and advertising trackers. iPad (61%) apps show a slightly increased presence of trackers compared to iPhone (59%) ones. Country-wise, apps on the Polish app store show the smallest presence of trackers (53%), whereas the British app store has the highest (57%).
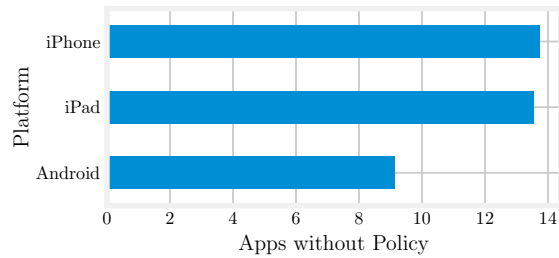
We contacted a total of 52 app developers who did not provide a privacy policy, and asked them about their data practices. Despite being legally required to answer such
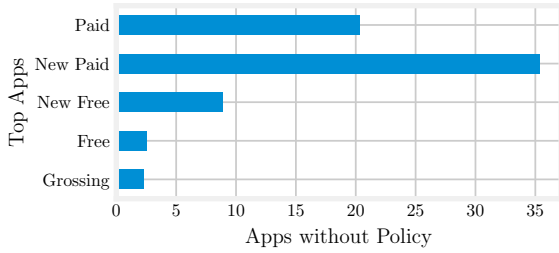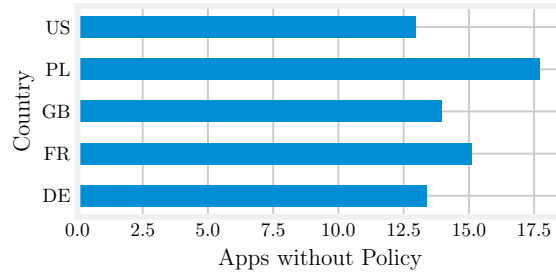
(a) Android Categories.
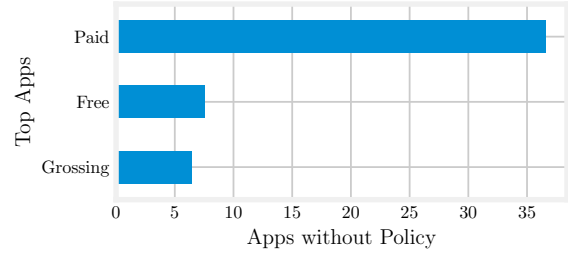
(b) iOS Categories.

(c) Platform.

(d) Store countries.

Figure 2: Percentage of top apps without privacy policy. Since iOS does not provide statistics about new apps, such are excluded from Figure 2c. Since there are fewer Android apps than iOS ones in the dataset, only the top 540 apps are considered for each triplet of country, category, and platform, in Figure 2c.

queries, 12 developers (23%) failed to respond.

## 5 Discussion

**Tracking in privacy policies.** Most app policies contained trackers. A user opening the privacy policy will thus share personal data with a third-party, before having had an opportunity to read and understand the terms of data collection. The top 5% of policies were even found to contain 14 trackers or more. These observations are in conflict with the right to be informed under GDPR.

**Use of insecure communication channels.** Only two thirds of privacy policies (68%) were transferred using HTTPS encryption. By contrast, 87% of all pages visited through Google Chrome on Android did use HTTPS[5]. Unencrypted communications can easily be intercepted by adversaries. If personal data is transmitted, a lack of encryption may even violate the principle of *integrity and confidentiality* of GDPR (Article 5 GDPR). The upcoming version of Android (API 28) will block unencrypted communication from apps, unless explicitly disabled by the developer.

**Widespread lack of privacy policy.** There have been identified significant differences in tracking, depending on the store category (e.g. *top grossing*) of the app. Compared to free apps, paid apps were both more likely to contain fewer trackers in their policy, and to come with no policy at all. This could indicate that paid apps tend
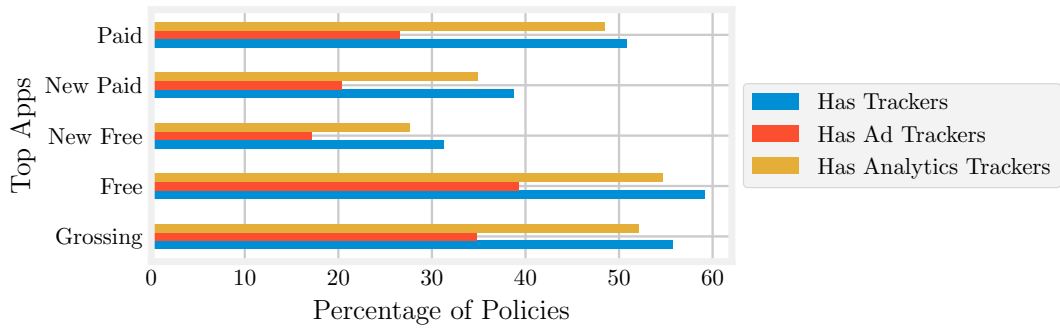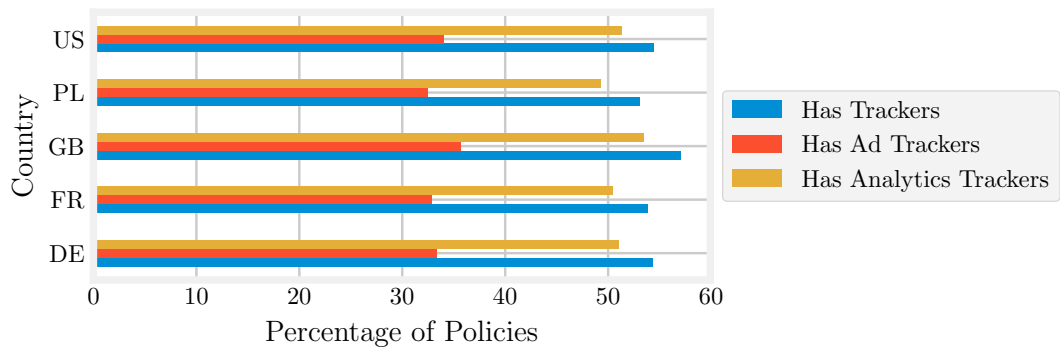
(a) Android Categories.
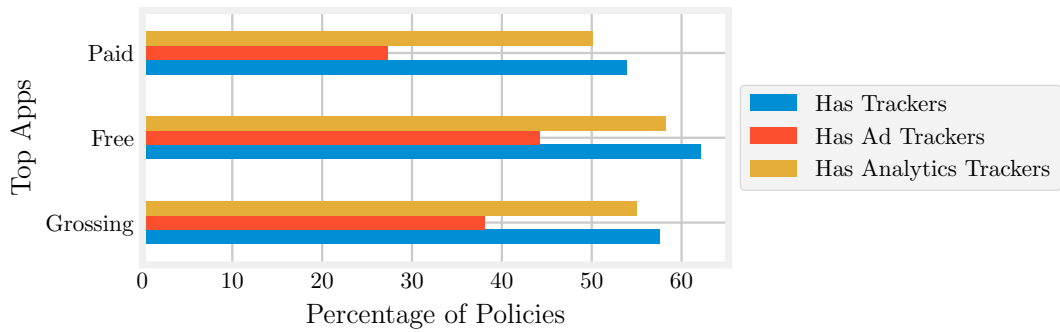


(b) iOS Categories.
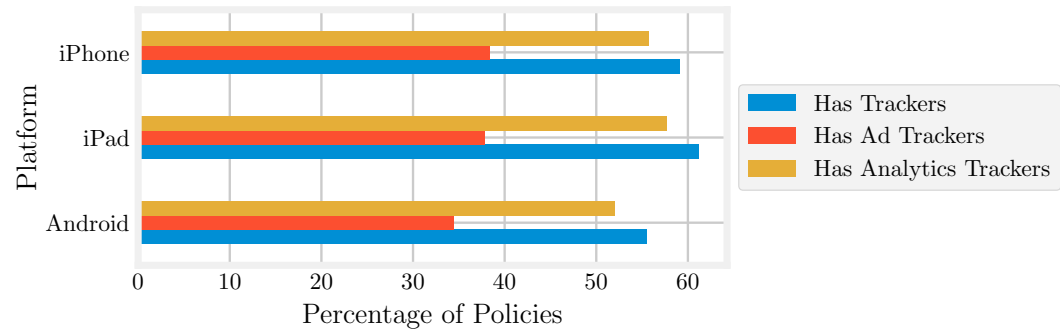


(c) Platforms.



(d) Store Countries.

Figure 3: Trackers in the policies of top apps. As in Figure 2, the number of iOS and Android apps are restricted for a fair comparison in Figure 3c.

to collect less data than free ones, because they do not pursue any data collection to disclose. There remains, however, an obligation for both paid and free apps to provide a privacy policy, if information is accessed or stored on the user's device, under the ePrivacy Directive. This will apply to most apps and highlights another potential violation of the GDPR.

iOS apps were found to provide a privacy policy less often than Android ones. This is despite the previous chapter having outlined that all iOS apps must provide a privacy policy, as opposed to Android ones (although this has been addressed by Google after our data collection). Apple only recently introduced the requirement of privacy policies in October 2018, applying to every new or updated app. iOS apps without a privacy policy have just not been updated on the store since.

**Limited differences between countries or platforms.** The analysis did not identify major differences between the platforms or countries with regards to tracking in privacy policies. There are two possible reasons for this. First, the policies use the same set of standard trackers, with Google Analytics being the most common one (see Figure 1). Second, the set of top apps is usually similar, irrespective of platform or country. Whilst one could analyse the unique apps per platform or country, such an approach would render the resulting set of apps unrepresentative of the day-to-day exposure of users. App users are exposed to a similar level of tracking in privacy policies, independent of country or platform.

**Paying for apps does not always improve privacy.** Our study also also highlights that paying for apps does not necessarily imply better data protection. *Top free* and *top grossing* apps were both found to contain similar numbers of trackers in their policies. In particular, *top grossing* apps contain a similar number of advertising trackers, despite generating significant revenue from user purchases. This illustrates how even apps, not relying on advertising revenues, may harvest personal data – especially when they rely on in-app purchases.

**Many developers did not respond to GDPR requests.** Despite legally binding, we found that many developers did not reply to our GDPR requests, even after having sent them a reminder. Some of the email addresses that the developers might just not be monitored, so the requests never actually ended up at the developers. However, there will also be emails that the developers received, but did decide not to answer. This points to a lack of understanding and awareness of their legal requirements with developers, as already found in previous studies.

**Limitations.** This analysis did not regard all apps on the app stores, but only the most popular and recent ones. This is to provide insights into the day-to-day exposure of app users, instead of analysing every app on the app stores. `Ghostery` relies on a manually curated list of trackers. This means that some trackers will be missed. Moreover, there might be a bias with regards to the physical location of the users of `Ghostery`, being

developed by a German company. For instance, Asian markets are dominated by other tech companies than in the West and the results might differ. The presented analysis does not provide any insight into trackers and GDPR violations within apps. Unfortunately, there currently exist no analysis for a comparison of the program code of free/paid and iOS/Android apps. As a solution, this thesis introduces the first scalable, cross-platform static analysis method in the following section.

# 6    Conclusions

This study analysed tracking of individuals when opening the privacy policy webpages of mobile apps. We found that most apps share data with third-parties, when a user accesses their privacy policy, which potentially violates the GDPR. Apps' privacy policies are more often transferred over unencrypted channels than regular websites, another potential GDPR violation. Despite being legally obliged to disclose the storage of data on the user's device, many app developers did not provide a privacy policy, and did not answer to our emails seeking clarification.

As a way forward, we believe that there should be more practical guidance on the technical implications of GDPR. Legal scholars have widely criticised the failure of GDPR to inform the underlying software development process [3, 6]. This is despite the GDPR explicitly prescribing this practice through the principle of Data Protection by Design, under Article 25 of the law. This paper has provided some technical evidence for this criticism of GDPR. Good guidance could potentially not only improve compliance with GDPR in practice, but also remove the pressure on the funding of the DPAs. We have created such a practical guide for app developers at `https://gdpr4devs.com`.

**Future work.** Our study analysed a large subset of popular apps. Follow-up work might consider an even larger dataset, trying to analyse the whole app ecosystem. This is absolutely feasible with our approach due to its immense and easy scalability. There is also a focus on privacy for EU users; follow-up work could analyse the practices in other jurisdictions. Furthermore, there is a emerging question around what responsibility app platforms themselves should face over apps' adherence to data protection and privacy laws. We leave this question for other academic studies.

# References

[1]    Article 29 Data Protection Working Party. *Opinion 02/2013 on apps on smart devices*. `https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp202_en.pdf`. Feb. 2013.

[2]    Reuben Binns et al. 'Third Party Tracking in the Mobile Ecosystem'. In: *Proceedings of the 10th ACM Conference on Web Science*. WebSci '18. May 2018, pp. 23–31. ISBN: 978-1-4503-5563-6. DOI: 10.1145/3201064.3201089.

[3]    Lee A Bygrave. 'Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements'. In: *Oslo Law Review* 1 (Aug. 2017), pp. 105–120. DOI: 10.18261/issn.2387-3299-2017-02-03.

[4]    European Court of Justice. *Judgement in Case C-582/14*. http://curia.europa.eu/juris/documents.jsf?num=C-582/14. June 2017.

[5]    Google LLC. *Transparency Report–HTTPS encryption on the web*. https://transparencyreport.google.com/https. 2019.

[6]    Lina Jasmontaite et al. 'Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR'. In: 4 (June 2018), pp. 168–189. DOI: 10.21552/edpl/2018/2/7.

[7]    Michael Veale, Reuben Binns and Jef Ausloos. 'When data protection by design and data subject rights clash'. In: *International Data Privacy Law* 8.2 (Apr. 2018), pp. 105–123. DOI: 10.1093/idpl/ipy002.

[8]    Haoyu Wang et al. 'Beyond Google Play: A Large-Scale Comparative Study of Chinese Android App Markets'. In: *Proceedings of the Internet Measurement Conference 2018*. IMC '18. 2018, pp. 293–307. ISBN: 978-1-4503-5619-0. DOI: 10.1145/3278532.3278558.

[9]    Sebastian Zimmeck et al. 'MAPS: Scaling Privacy Compliance Analysis to a Million Apps'. In: *Privacy Enhancing Technologies Symposium 2019* 72 (June 2019).