Between Law and Code: Challenges and Opportunities for Automating Privacy Assessments

Mugdha Khedkar [1*], Michael Schlichtig [1], Nihad Atakishiyev [3], Eric Bodden [1,2]

[1]Heinz Nixdorf Institute, Paderborn University, Paderborn, Germany.
[2]Fraunhofer IEM, Paderborn, Germany.
[3] Paderborn University, Paderborn, Germany.

*Corresponding author(s). E-mail(s): mugdha.khedkar@uni-paderborn.de;
Contributing authors: michael.schlichtig@uni-paderborn.de; nihad@mail.uni-paderborn.de; eric.bodden@uni-paderborn.de;

Abstract

Android apps collecting data from users must comply with legal frameworks to ensure data protection. This requirement has become even more important since the implementation of the General Data Protection Regulation (GDPR) by the European Union in 2018. Moreover, with the proposed Cyber Resilience Act on the horizon, stakeholders will soon need to assess software against even more stringent security and privacy standards. Effective privacy assessments require collaboration among groups with diverse expertise to function effectively as a cohesive unit.

This paper presents an interview-based study (N=16) exploring the challenges these experts encounter during privacy assessments and their views on automation as potential support. To ground the discussion, we use *Assessor View*, a prototype developed for this work that integrates static analysis to extract privacy-relevant information directly from Android Application Packages (APKs), as a research probe. Its design provides dedicated views for both technical and non-technical stakeholders, enabling reflection on how automation can enhance assessment practice.

Our study identifies key challenges in conducting privacy assessments, including knowledge and communication gaps between experts, the privacy–innovation trade-off, delayed involvement of privacy professionals, and the lack of source code analysis-based tools. The user study conducted alongside the interviews reveals that the GDPR warnings and guidance provided by Assessor View are valuable to Data Protection Officers and privacy experts, and its design is particularly well suited for these stakeholders. Overall, our findings indicate that Assessor View represents a significant step toward improving communication between legal and technical experts and automating privacy assessments.

Keywords: static analysis, data collection, data protection, privacy-aware reporting

1 Introduction

Any software that reaches the European market needs to adhere to the General Data Protection Regulation (GDPR) [1]. Moreover, with the European Union's Cyber Resilience Act (CRA) [2] on the horizon, software developers will soon face the challenge of writing code that complies with even more stringent security and privacy standards. These regulations extend to Android applications that gather data from users within the European Union.

The GDPR defines personal data as *"any information relating to an identified or identifiable natural person, a data subject"*, and imposes several obligations on the access, storage and processing of such data. Under the GDPR, data protection violations can result in severe financial penalties [3]. If these violations cause vulnerabilities and data leaks, additional and similarly severe fines may be levied under the CRA.

Advocating for transparency, Article §13 [4] of the GDPR mandates that the app publishers disclose the collection and processing of personal data to the user by providing documents such as privacy policies. However, these privacy policies are often very long and vague, and may not even be authored by the app developers themselves. Several studies [5–9] have consistently shown significant discrepancies between privacy policies and the actual source code, undermining their accuracy and misleading users.

To address these inaccuracies, Google launched the Data Safety Section [10] in 2022, shifting the responsibility of privacy-related reporting to app developers. This necessitates the completion of a form on Google's Play Console, outlining how apps collect, share, and secure user data.

Google's Data Safety Section (DSS) form consists of three main sections: *data sharing, data collection, and security practices.* User data is categorized into different types within these sections. Before creating a store listing for an Android app, developers must complete this DSS form. This requires manual effort which may lead to inaccuracies in reporting [11], providing users with a false sense of privacy. In 2022, Google introduced Checks [12], a paid service that assists app developers with privacy compliance, and completing the Data Safety Section. However, its cost could hinder startups or independent app developers from using its services. Recently, open-source alternatives [13, 14] have been introduced to assist developers in accurately completing the DSS form.

While accurately filling out the DSS form is the developers' responsibility, ensuring GDPR compliance through assessments requires collaboration among groups with a diverse expertise. Under Article §35 [15] of the GDPR, the data controller (e.g., a software provider or organization) is responsible for ensuring that a Data Protection Impact Assessment (DPIA) [16] is carried out whenever processing operations may pose high risks to individuals' rights and freedoms. However, the task of conducting the DPIA typically falls on privacy professionals such as Data Protection Officers (DPOs) or legal experts, who must evaluate data protection risks in collaboration with technical teams. This process involves a systematic analysis of data flows and mitigation measures, and the results must be documented and, in some cases, shared with the European Data Protection Board [17].

Prior research has extensively examined how developers handle privacy requirements [5–9], including how they document data collection and complete Google's Data Safety Section [11, 18–20]. However, a clear **research gap** remains in understanding the challenges faced by DPOs, privacy experts, and legal experts—those responsible for conducting DPIAs [16] and ensuring organizational compliance. Understanding their perspective is essential, particularly to automate parts of the assessment process. Throughout this paper, we use the term *privacy assessors* to collectively refer to DPOs, privacy experts, and legal experts. While some DPOs have legal training, others come from compliance or technical backgrounds; therefore, we distinguish DPOs from legal experts but include both under the broader category of privacy assessors.

To address this gap, we conduct an interview-based study (N=16) with the **goal** to *explore how privacy assessors conduct privacy assessments in practice, the challenges they encounter, and how they perceive the role of automation in this process.* In a previous workshop paper [21], we proposed the idea of *Assessor View*, a tool designed to address this knowledge gap among the various parties involved in privacy assessments. As part of this work, we implement an initial prototype of *Assessor View* that serves as a research probe for exploring how automation can support privacy assessors. Assessor View is based on static analysis and visualizes source code components through a data privacy vocabulary [22], offering multiple visualization modes tailored for legal and technical experts.

While Assessor View is designed to support multiple stakeholders involved in privacy assessment—namely static analysis experts, app developers, and privacy assessors—this paper specifically focuses on the perspective of privacy assessors, including DPOs, legal, and privacy experts. Our goal is to understand their assessment practices and expectations for automation, as they play a central role in ensuring compliance and conducting DPIAs. Complementary to this work, our prior study [23] evaluated the tool's usability and usefulness with 12 developers, providing insights into how technical users interpret and act upon privacy-related analysis results. Together, these studies offer a more comprehensive understanding of how different user groups interact with and benefit from Assessor View.

To achieve the above goal, we design our study in two parts: (i) *semi-structured interviews* to understand real-world assessment practices and challenges; and (ii) a *hands-on session with Assessor View* to explore experts' experiences and concerns with Assessor View, and its potential usage in conducting DPIAs. We aim to answer the following research questions:

RQ1. *How are privacy assessments conducted in a real-world setting?* We identify the key stakeholders responsible for these evaluations, their specific roles, and whether they rely on any tools or frameworks for support. Understanding these aspects will provide insight into the practical challenges and best practices within privacy assessment workflows.

RQ2. *To what extent does the source code analysis-based Assessor View support privacy assessors in conducting DPIAs?* We examine to what extent Assessor View can act as a bridge between technical and non-technical stakeholders, enhancing communication and collaboration. The aim is to assess its effectiveness in reducing knowledge gaps and ensuring smoother privacy assessments through better alignment of technical and legal perspectives.

Our study identifies key challenges in conducting privacy assessments, including knowledge gaps and poor communication between legal and technical experts, the absence of source code analysis-based privacy tools, and the delayed involvement of privacy assessors. The hands-on session conducted using Assessor View revealed that the GDPR warnings and guidance provided by Assessor View are valuable to DPOs and privacy experts, and its design is particularly well suited for these stakeholders.

Overall, our findings indicate that Assessor View represents a promising first step toward improving communication and partial automation in privacy assessment processes. Participants found the tool valuable for bridging the communication gap between stakeholders, but emphasized the continued importance of manual review in final risk evaluations. We use these insights to discuss implications for designing tools that complement, rather than replace, expert judgment. We make all artifacts available at https://doi.org/10.5281/zenodo.15085611.

To summarize, this work makes the following contributions:

- It provides an empirical account of how privacy assessors conduct privacy assessments and collaborate with technical teams.
- It identifies recurring challenges and points of friction in the current assessment process, highlighting opportunities for partial automation.
- A user study with 16 privacy experts explores the effectiveness of Assessor View and its potential to support and automate aspects of privacy assessment.

The remainder of the paper is organized as follows. Section 2 uses a motivating example to introduce the problem, and discusses related work. We present our approach in Section 3. We discuss the implementation details in Section 4. We explain the methodology of our interview study in Section 5, and discuss the study results in Section 6. We explain the limitations of our approach in Section 7, and conclude in Section 8.

2 Problem and Related Works

Consider Alice, the newly appointed Data Protection Officer of a company, who initiates a Data Protection Impact Assessment [16]. To complete the assessment, she requires app developers' assistance in answering the DPIA questions. She seeks help of Bob, a Java programmer unfamiliar with legal privacy frameworks. They use the sample DPIA template [24] and aim to address questions regarding the source of the data collected; the use, storage, and deletion of data; its processing and sharing; and whether the data is pseudonymized.

Without tool support, Alice and Bob must manually sift through extensive lines of code to find answers. While Bob navigates the code proficiently, Alice struggles due to her limited programming expertise. Conversely, Alice provides legal insights to Bob, but she struggles to correlate legal terms with source code components. This leads to an extensive discussion with the legal team, during which Bob must convey technical concepts to non-technical experts and persuade them regarding the implementation of data protection measures. After weeks of manual labor and discussion, they eventually complete the assessment successfully.

Conducting a DPIA involves identifying and mitigating data protection risks, requiring collaboration between developers and privacy assessors who may be DPOs or legal experts. App developers may often spend considerable time convincing the team of appropriate data protection measures.

Alice and Bob could have been better supported with a tool to systematically assess privacy-by-design concepts [25] in source code. This highlights the need to design and implement an automated approach to bridge the knowledge gap between app developers and privacy assessors, ensuring seamless privacy assessments.

While existing research provides foundational knowledge to address this issue and better support developers and privacy assessors in conducting the DPIA, several challenges still need to be addressed.

There has been some exploratory research in the area of automated GDPR compliance. Ferrara et al. [26] discussed main components needed to develop GDPR reports, and static analysis techniques in the form of source to sink analysis, to support GDPR compliance- but not specifically for DPIAs. They also observed that information flow analyses only report the program point where data is leaked, not the complete flow, arguing that these were not sufficient for GDPR compliance. Martin et al. [27] recommended privacy by design [25] and proposed tools for GDPR compliance, and Li et al. [28] explored automated GDPR compliance in continuous integration workflows. However, these approaches primarily target developers, without addressing the broader collaboration challenges faced by privacy assessors and legal experts—establishing a **research gap** in supporting multiple stakeholder roles during privacy assessments. Our study addresses this gap by using Assessor View to explore how automation can support collaboration across these diverse roles.

Existing static analysis tools [9, 29–31] aid app developers in assessing whether their app source code and privacy policies fulfil data protection requirements. Nachtigall et al.'s study on usability criteria of such static analysis tools [32] highlighted that usability criteria are often partially addressed, particularly when the tools lack interfaces for non-technical users, suggesting that supporting privacy assessors requires attention to communication and accessibility. A survey conducted by Franke et al. [33] showed that developers frequently cite the lack of automated support as a major obstacle to GDPR compliance. Yet, a **research gap** remains in understanding how collaboration between different stakeholders can be effectively supported. Our study addresses this gap by interviewing privacy assessors to identify practices and challenges that may inform better collaborative tools.

Feiyang Tang et al.'s analysis [34] was among the first to address this gap, aiming to assist app developers and DPOs in evaluating privacy compliance and simplifying the DPIA process. They introduced an automatic software analysis technique that presents the results as a graph of privacy flows and operations. While their approach aligns with our objectives, we build on our previous work [21, 35] which proposed additionally using the Data Privacy Vocabulary (DPV) [22] to make the analysis results more understandable to privacy assessors.

Pandit et al. [36] identified the absence of standard vocabularies to describe personal data, data handling purposes, and categories of processing. They introduced the Data Privacy Vocabulary (DPV), which describes various components and relationships between them [22]. However, there remains a **research gap** in mapping software components to such vocabularies, underscoring the persistent knowledge and representation challenges in privacy assessments. Our study addresses this gap by introducing Assessor View, which automatically maps code elements to DPV-based representations to support shared understanding among assessors.

Some commercial tools such as Privado [14] aim to automate compliance checks by analyzing program code or data flows. Such tools typically operate on local source code and provide structured reports to assist developers and compliance teams. While such tools demonstrate the feasibility of automated privacy analysis, a **research gap** remains in examining how privacy assessors actually engage with such automation in practice. Our study addresses this gap by using Assessor View to explore this interaction space. Assessor View operates on compiled applications and represents code behavior through a privacy-oriented vocabulary [22], enabling us to examine how such representations influence assessors' reasoning and collaboration.

3 Approach

Our approach to address the problem discussed in Section 2 comprises three key components: the user interface (Module Ⓐ in Figure 1), the Privacy Slice Visualizer (Module Ⓑ in Figure 1), and the DPV Transformer (Module Ⓒ in Figure 1). While Module Ⓑ reuses components from previous work [18, 23], Modules Ⓐ and Ⓒ are unique contributions of this paper.

Initially, the *user interface* Ⓐ accepts an Android app as input, and forwards it to the *Privacy Slice Visualizer* Ⓑ for further analysis. The Privacy Slice Visualizer [23] is an existing component that statically slices the source code from privacy-related data sources, and visualizes the resulting program slices. Subsequently, the *DPV Transformer* Ⓒ accepts these program slices as input and translates them into graphical (DPV) representations familiar to the different stakeholders involved in a privacy assessment, enhancing their comprehension of the code.

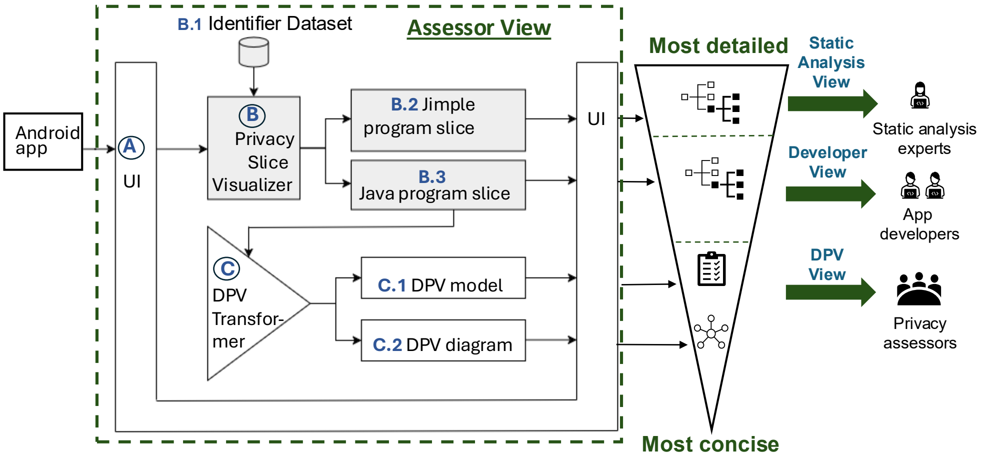
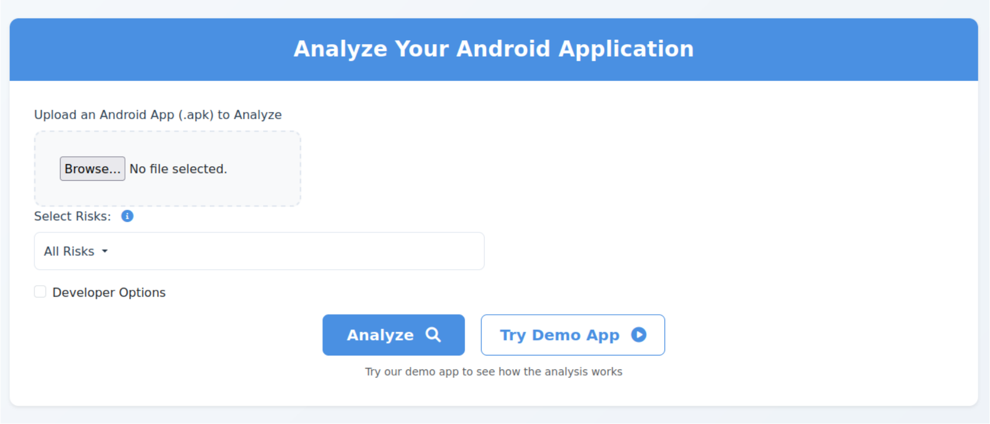


Fig. 1: Our approach. Module Ⓑ (shaded with gray) builds on previous work [18, 23].


Analyze Your Android Application

Upload an Android App (.apk) to Analyze

Browse... No file selected.

Select Risks:

All Risks

□ Developer Options

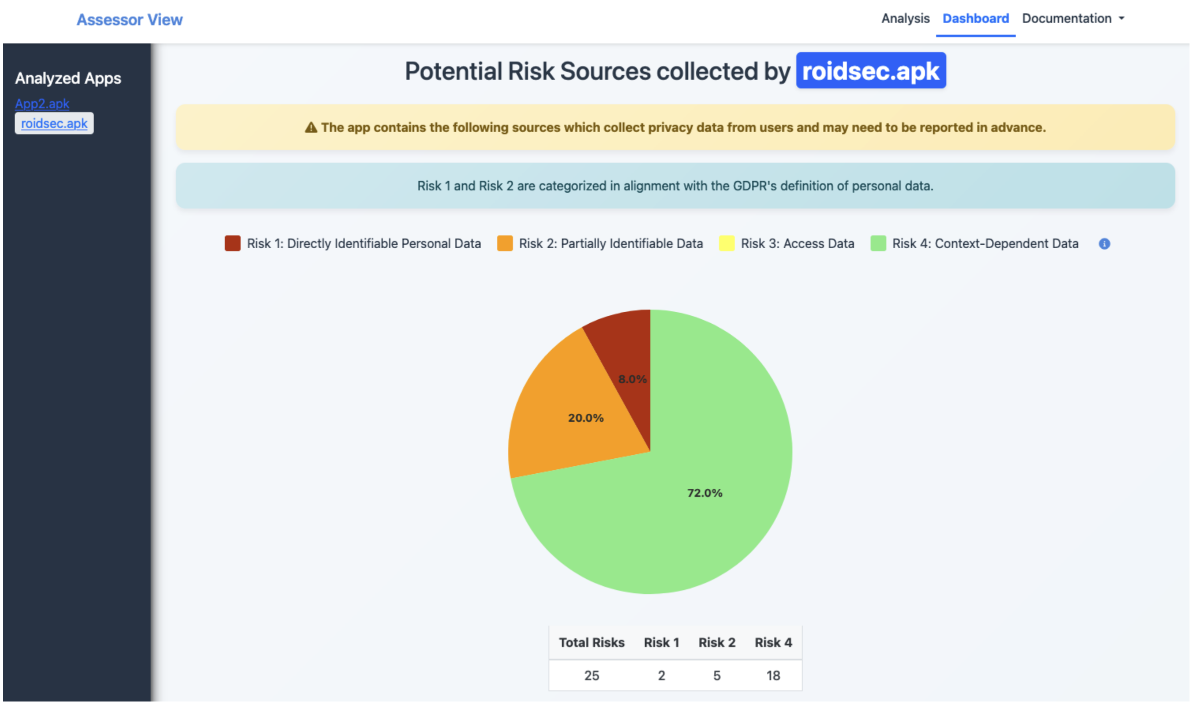
Try our demo app to see how the analysis works

Fig. 2: Uploading an APK to Assessor View

3.1 User Interface

The *user interface* (Module Ⓐ) accepts any APK as input (cf. Figure 2) and feeds it to the Privacy Slice Visualizer. For convenience, users can also click *Try Demo App* to analyze a sample APK provided by the back-end. This option is included solely to make the system easily accessible to users who may not have an APK readily available. It proved particularly useful during the user study (cf. Section 5), as it eliminated the need for a prior process to provide users with a sample APK.

Once Assessor View has been executed, the UI displays Assessor View's results on the dashboard (cf. Figure 3). The dashboard gives an overview of all risky sources detected in the APK (cf. Figure 3a), and the slices that originate from these sources (cf. Figure 3b). We explain these results in the following sections.



Assessor View

Analysis Dashboard Documentation ▾

Analyzed Apps

App2.apk

roidsec.apk

Potential Risk Sources collected by **roidsec.apk**

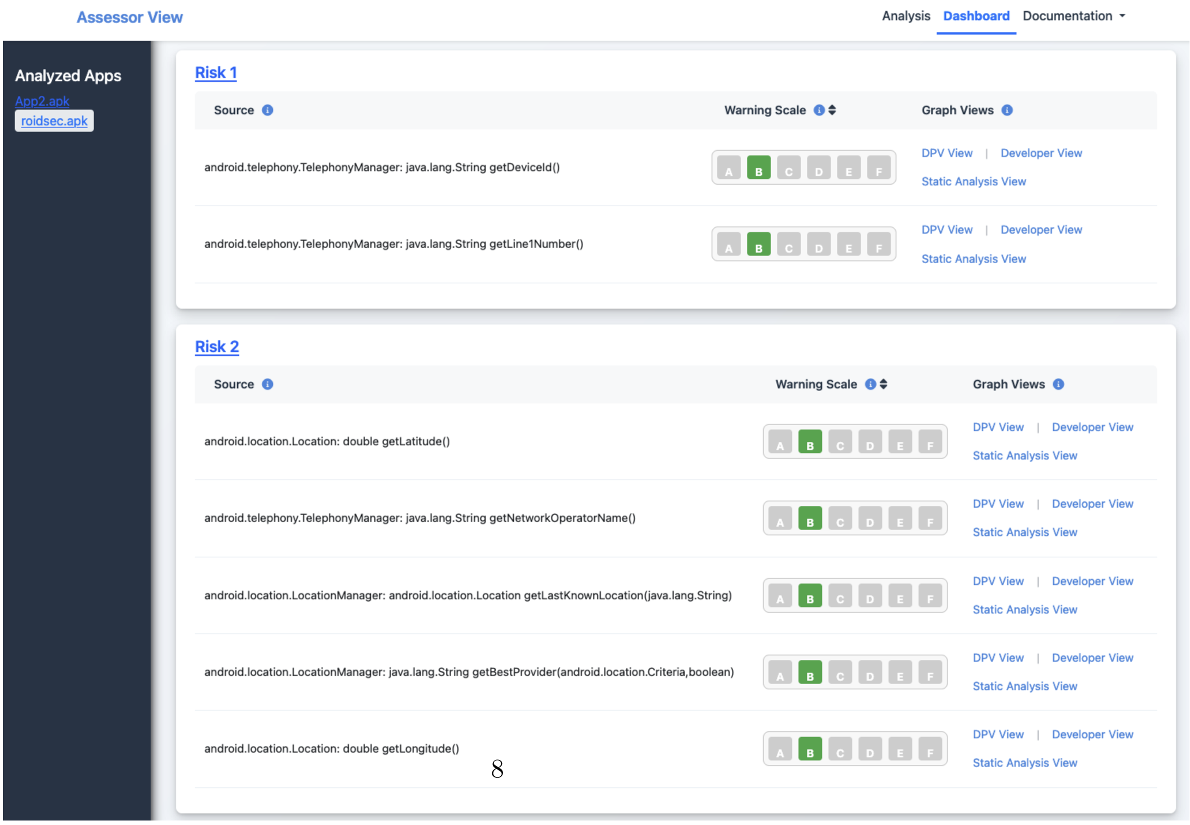
⚠ **The app contains the following sources which collect privacy data from users and may need to be reported in advance.**

Risk 1 and Risk 2 are categorized in alignment with the GDPR's definition of personal data.

🟥 Risk 1: Directly Identifiable Personal Data 🟧 Risk 2: Partially Identifiable Data 🟨 Risk 3: Access Data 🟩 Risk 4: Context-Dependent Data ⓘ

Total Risks	Risk 1	Risk 2	Risk 4
25	2	5	18

(a) Visualizing risk sources detected in the APK.



Assessor View

Analysis Dashboard Documentation ▾

Analyzed Apps

App2.apk

roidsec.apk

Risk 1

Source ⓘ	Warning Scale ⓘ ⇕	Graph Views ⓘ
android.telephony.TelephonyManager: java.lang.String getDeviceId()	A B C D E F	DPV View | Developer View Static Analysis View
android.telephony.TelephonyManager: java.lang.String getLine1Number()	A B C D E F	DPV View | Developer View Static Analysis View

Risk 2

Source ⓘ	Warning Scale ⓘ ⇕	Graph Views ⓘ
android.location.Location: double getLatitude()	A B C D E F	DPV View | Developer View Static Analysis View
android.telephony.TelephonyManager: java.lang.String getNetworkOperatorName()	A B C D E F	DPV View | Developer View Static Analysis View
android.location.LocationManager: android.location.Location getLastKnownLocation(java.lang.String)	A B C D E F	DPV View | Developer View Static Analysis View
android.location.LocationManager: java.lang.String getBestProvider(android.location.Criteria,boolean)	A B C D E F	DPV View | Developer View Static Analysis View
android.location.Location: double getLongitude()	A B C D E F	DPV View | Developer View Static Analysis View

(b) List of risky slices detected in the APK.

Fig. 3: Assessor View dashboard for Roidsec, a real-world app from the TaintBench suite [37].

3.2 Privacy Slice Visualizer

The *Privacy Slice Visualizer* (Module Ⓑ) builds upon our prior work [23] and performs static analysis to identify and analyze privacy-relevant data sources within Android applications. A privacy-relevant data source refers to any API call or data access point in the source code that collects personal data—that is, information that can directly or indirectly identify an individual or device. This includes both explicit identifiers (e.g., device ID, email address) and contextual data (e.g., location, network information) that may lead to identification when combined with other data.

To extract these data sources, the Privacy Slice Visualizer extends Jicer [38], a static program slicer that operates on Jimple, an intermediate representation of Java code [39]. While program slicing has been extensively studied across platforms [40], Jicer was chosen because it is the only open-source static slicer that addresses Android-specific challenges, such as the absence of a main method, extensive use of callbacks, and inter-component communication [38].

The Visualizer first constructs a system dependence graph (SDG) of the given app, capturing all control and data dependencies. It then cross-references each system API call in the code against a manually annotated *Identifier Dataset* (**B.1**) introduced in our prior work [18]. This dataset provides a risk-based definition of privacy-relevant data, allowing fine-grained labeling of APIs as data sources:

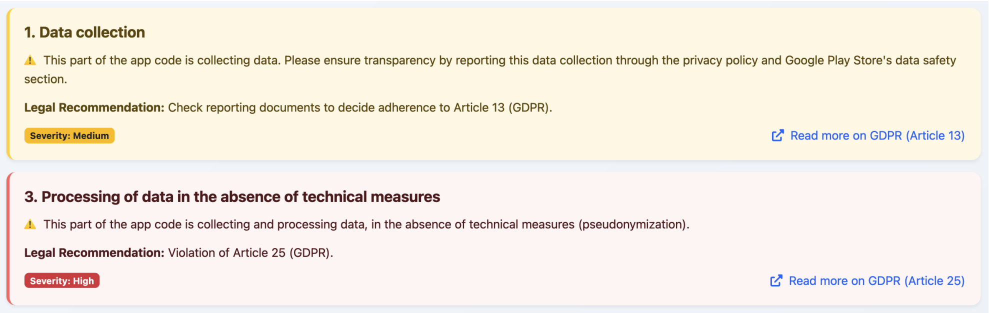
- *Risk 1*: Data that can identify a user or device without additional information.
- *Risk 2*: Data that requires supplementary information to enable identification.
- *Risk 3*: Data that can grant access to an individual's device but cannot identify an individual's device.
- *Risk 4*: Data that may identify an individual or a device depending on its content and the context in which it is used.

The data categories in the dataset are aligned with Google's Data Safety Section, as developers are required to disclose these same data types when publishing apps on Google Play Console [41]. This alignment ensures that the tool's output directly supports developers' compliance and reporting obligations, making the analysis results immediately actionable for completing the Data Safety Section.

After identifying privacy-relevant sources, the Privacy Slice Visualizer performs forward slicing to retain all program statements affected by these data flows. It then converts the extracted slices into graph-based representations that illustrate control and data dependencies in the source code (cf. Figure 5a and Figure 5b). These slices are represented both in Jimple (**B.2**) and Java (**B.3**), supporting different technical audiences. We explain Figure 5 in greater detail in Section 3.4.

To enhance interpretability, the tool runs additional static analyses on the slices, labeling API calls related to pseudonymization, analytics, advertisements, authentication, networking, input/output, email, image processing, data serialization, and location—matched against custom datasets developed in our prior work [23].

Finally, the Privacy Slice Visualizer provides two tailored outputs: **Static Analysis View** which visualizes Jimple-level program slices for static analysis experts, and the **Developer View** which translates those slices into Java-like statements accessible to Android developers. Both views are represented as graphs, where nodes correspond



1. Data collection

⚠ This part of the app code is collecting data. Please ensure transparency by reporting this data collection through the privacy policy and Google Play Store's data safety section.

Legal Recommendation: Check reporting documents to decide adherence to Article 13 (GDPR).

Severity: Medium

Read more on GDPR (Article 13)

3. Processing of data in the absence of technical measures

⚠ This part of the app code is collecting and processing data, in the absence of technical measures (pseudonymization).

Legal Recommendation: Violation of Article 25 (GDPR).

Severity: High

Read more on GDPR (Article 25)

Fig. 4: GDPR Warnings for Roidsec. These warnings correspond to rules 1 and 3 in Table 1.

to source code or Jimple statements labeled with privacy-relevant information (e.g., analytics, pseudonymization, string manipulation), and edges capture control and data dependencies between nodes.

User studies conducted to assess the usability of these program slices in our prior work [23] have confirmed the need for a simplified view that will facilitate non-technical stakeholders' comprehension of the source code.

3.3 DPV Transformer

To address the knowledge gap between different stakeholders involved in privacy assessments, Assessor View uses the *DPV Transformer* (Module Ⓒ) to convert Java program slices into a privacy-oriented view with multiple abstraction levels. This module maps code-level properties observed in the Java slices to corresponding privacy concepts defined in the Data Privacy Vocabulary (DPV).

The DPV Transformer is currently Java-dependent, as it operates on statically generated Java program slices (**B.3**). It performs a direct, rule-based translation of Java statements and API calls into DPV concepts such as data categories and processing operations. While the approach could in principle be extended to other languages, doing so would require adapting the parsing and mapping rules to their respective syntax and semantics. The details of this mapping are discussed in Section 4.

DPV Transformer accepts the Java program slices as input, and renders DPV information using two high-level DPV specifications, the DPV model (**C.1**), and DPV diagram (**C.2**).

The final step involves analyzing data processing activities within a program slice to identify potential GDPR compliance issues based on predefined risks. To provide warnings and suggestions for GDPR compliance, the DPV Transformer establishes these guidelines using clear and recognizable patterns. Currently, the transformer addresses five specific GDPR rules (cf. Table 1), which the user interface displays to the user (cf. Figure 4). Each box in the UI visualization represents a unique compliance risk and is color-coded semantically: light green for potential adherence warnings (rule 4 in Table 1), light yellow for warnings (rule 1 in Table 1), and pale red for high-severity GDPR violation risks (rules 2, 3, and 5 in Table 1). Section 4 further details how these compliance checks are implemented.

Rule	Program Slice Property	Warning	Legal Recommendation
1	Data collection.	This part of the app code collects data. Ensure transparency by reporting this data collection through privacy policy and Google's Data Safety Section.	Check reporting documents to adhere to GDPR Article §13 [4].
2	Data collection but no processing operations.	This part of the app code collects data but may not be processing it. Ensure that data collection is necessary.	**Potential violation** of GDPR Article §5 [42].
3	Processing of data in the absence of technical measures.	This part of the app code collects and processes data, in the absence of technical measures (pseudonymization).	**Potential violation** of GDPR Article §25 [43].
4	Processing of data after implementing technical measures.	This part of the app code collects and processes data, but only after implementing technical measures (pseudonymization).	**Potential adherance** to GDPR Article §25 [43].
5	Processing of data before implementing technical measures.	This part of the app code collects and processes data, before the implementation of technical measures (pseudonymization).	**Potential violation** of GDPR Article §25 [43].

Table 1: GDPR rules and warnings addressed by the DPV Transformer.

3.4 Example

Throughout this section, we discuss a particular program slice from from Roidsec[1], a real-world app from the TaintBench suite [37].

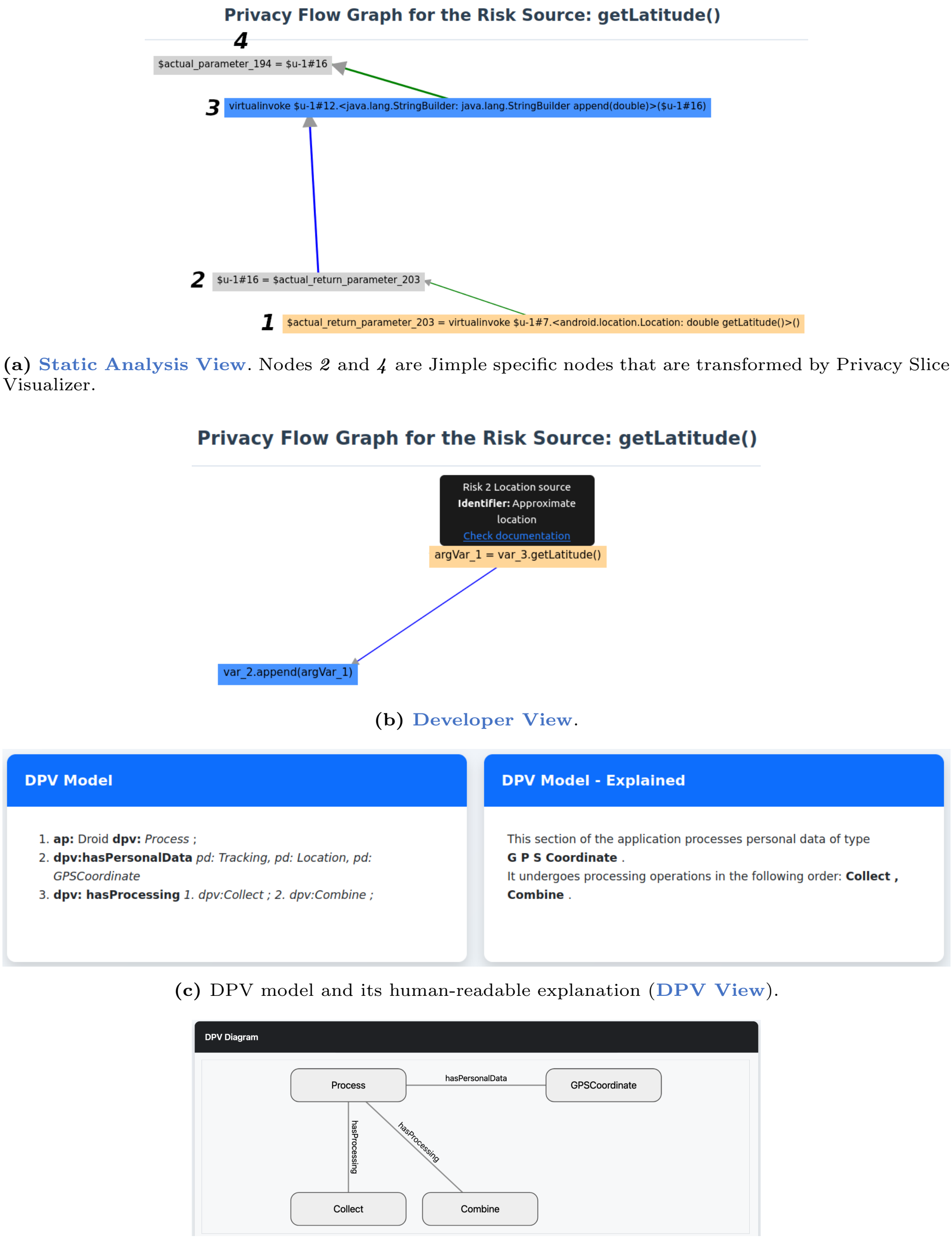
We begin with the **Static Analysis View**, the most detailed representation produced by the Privacy Slice Visualizer. This view shows the program slice in Jimple intermediate representation as extracted from the slicer, preserving low-level Jimple-specific nodes (eg., Nodes *2* and *4* in Figure 5a).

Next, the Privacy Slice Visualizer transforms the Jimple-based representation into a higher-level graph containing semantically correct, Java-like statements instead of raw Jimple code. This transformation, developed by us in our prior work [23], enables developers to analyze privacy-relevant behavior without requiring Jimple expertise. The resulting output is the **Developer View** (cf. Figure 5b), where nodes represent source code statements and edges represent control and data dependencies.

In this slice, *latitude* (labeled as location data) is collected and combined with a string variable *var_2*. Since the DPV treats the *combine operation* as data processing, Assessor View recognizes that data collection and processing have occurred but notes that the source code lacks pseudonymization.

For a higher-level overview, the DPV View represents the Java slice as DPV specifications (**DPV View**), presenting the app in terms of the Data Privacy Vocabulary (DPV) [22]. The DPV provides a detailed, machine-readable taxonomy for privacy-related information, based on GDPR terminology.

[1] https://github.com/TaintBench/roidsec


Privacy Flow Graph for the Risk Source: getLatitude()
4
$actual_parameter_194 = $u-1#16
3
virtualinvoke $u-1#12.<java.lang.StringBuilder: java.lang.StringBuilder append(double)>($u-1#16)
2
$u-1#16 = $actual_return_parameter_203
1
$actual_return_parameter_203 = virtualinvoke $u-1#7.<android.location.Location: double getLatitude()>()

(a) Static Analysis View. Nodes *2* and *4* are Jimple specific nodes that are transformed by Privacy Slice Visualizer.

Privacy Flow Graph for the Risk Source: getLatitude()
Risk 2 Location source
Identifier: Approximate location
Check documentation
argVar_1 = var_3.getLatitude()
var_2.append(argVar_1)

(b) Developer View.

DPV Model

1. **ap:** Droid **dpv:** *Process* ;
2. **dpv:hasPersonalData** *pd: Tracking, pd: Location, pd: GPSCoordinate*
3. **dpv: hasProcessing** *1. dpv:Collect ; 2. dpv:Combine ;*

DPV Model - Explained

This section of the application processes personal data of type **G P S Coordinate** .
It undergoes processing operations in the following order: **Collect , Combine** .

(c) DPV model and its human-readable explanation (DPV View).

(d) DPV diagram (DPV View).

Fig. 5: Different views for Roidsec.

Figure 5c shows a concrete instance of this DPV model for Roidsec. It uses DPV terms such as personal data categories, and processing categories, with relationships established through predicates such as *hasPersonalData* and *hasProcessing*. In Line 1, the placeholder app name *Droid* is used to represent any application, and the keyword *Process* defines the program slice as a new use case. Line 2 suggests that the app collects location, more specifically the GPS coordinate. The terms *pd:Tracking*, *pd:Location*, and *pd:GPSCoordinate* show that the personal data falls under the GPS coordinate category, which is a descendant of the DPV class *pd:Location*, itself a descendant of *pd:Tracking*. Line 3 shows that the app first collects the data and then combines it with another variable. To assist users in correctly interpreting the DPV model, a human-readable explanation is provided alongside the model: *This section of the application processes personal data of type GPS Coordinate. It undergoes processing operations in the following order: Collect, Combine.* (cf. Figure 5c)

Assessor View also converts the DPV model to a concise DPV diagram, summarizing the app at a high level (cf. Figure 5d). The DPV diagram represents the highest level of abstraction and intentionally omits the order of processing operations. This high-level view displays potential GDPR violations, GDPR compliance warnings, and suggestions to the assessors (cf. Figure 4). The **DPV View** will flag the absence of technical measures such as pseudonymization, encouraging DPOs to question app developers about the source code. The lack of pseudonymization will be mapped directly to a **potential violation** of GDPR Article §25 [43], providing actionable insights for assessors.

During privacy assessments, assessors can easily switch between different views to ask developers specific questions and understand the technical measures implemented to protect data. Alice and Bob can now use Assessor View for discussions regarding privacy assessments. Alice (DPO) can first examine **DPV View**, which, although concise, lacks detailed information regarding privacy-related data flow. If she needs to discuss the intricacies of the source code with Bob (app developer), she can seamlessly switch to **Developer View**, which provides the most detailed information.

4 Implementation

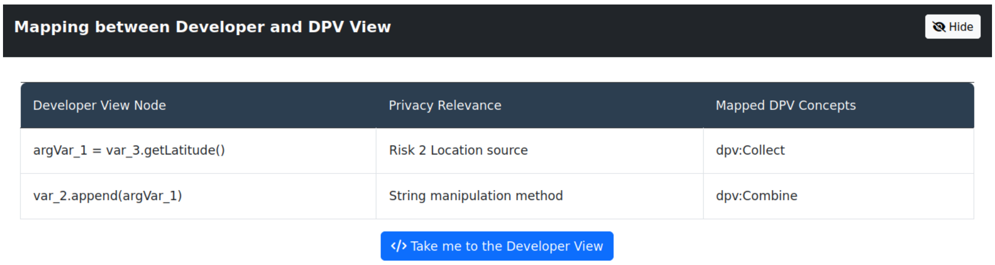
We now discuss the implementation of Assessor View (cf. Figure 1).

Users begin by uploading an APK in Assessor View, and its *user interface* (Module Ⓐ) processes user requests and executes the *Privacy Slice Visualizer* (ModuleⒷ) with selected command-line options. The user interface uses D3.js [44], a JavaScript library for visualizing data. Since the Privacy Slice Visualizer was developed and evaluated in a separate study [23], our focus will be on the implementation of the *DPV Transformer* (Module Ⓒ).

The DPV Transformer serves as the bridge between the technical program slices and the higher-level privacy vocabulary used by assessors. It accepts Java program slices (**B.2**) as input and translates individual Java statements into corresponding DPV concepts. This translation is achieved through a mapping algorithm that aligns each program statement from the **Developer View** with a relevant DPV concept.

Source code statement	Mapped DPV concepts
API methods from the Identifier Dataset (B.1)	dpv:HasPersonalData → <*pd:identifier*> dpv:HasProcessing dpv:Collect
Pseudonymization method	dpv:hasTechnicalMeasure <*encryption OR others*>
Assignment statement	dpv:HasProcessing dpv:Copy
Input output retrieval method	dpv:HasProcessing dpv:Retrieve
String manipulation method (eg., append(), concat())	dpv:HasProcessing dpv:Combine
Input output access method	dpv:HasProcessing dpv:Access
Database traversal method	dpv:HasProcessing dpv:Query
String replacement method (eg., replace())	dpv:HasProcessing dpv:Modify

Table 2: A small subset of the mapping from source code statements in **Developer View** to DPV processing operations in **DPV View**.


Mapping between Developer and DPV View
Hide
Take me to the Developer View

Developer View Node	Privacy Relevance	Mapped DPV Concepts
argVar_1 = var_3.getLatitude()	Risk 2 Location source	dpv:Collect
var_2.append(argVar_1)	String manipulation method	dpv:Combine

Fig. 6: An example mapping between **Developer View** and **DPV View** for the Roidsec example from Figure 5.

At present, this mapping process involves a one-time manual curation step. Each Java source method in the Identifier Dataset (B.1) is assigned an appropriate DPV term and stored in a reference database. During transformation, the DPV Transformer consults this database to automatically convert recognized source statements into DPV concepts. While this initial mapping was created manually, it is fully customizable and extensible, allowing new mappings to be added or refined as the DPV evolves or as new APIs are introduced.

The **Developer View** for each program slice provides structured details about the kinds of programming statements present, such as data manipulations, assignments, or I/O operations. The DPV Transformer uses this information to infer the type of processing activity being performed and map it to a corresponding DPV Processing concept (e.g., Collect, Store, Combine). A small subset of this current mapping is illustrated in Table 2. The full mapping is available in the source code included in the accompanying artifacts[2].

An example of this mapping for Roidsec (cf. Figure 5) is presented in Figure 6. This mapping is also shown to the users of Assessor View so they can inspect every node in detail if required. The resulting output is a DPV model (C.1) that captures both the types of data and the sequence of processing operations. This model is then transformed into a DPV diagram (C.2).

[2] https://doi.org/10.5281/zenodo.15085611

The DPV diagram represents the highest level of abstraction and intentionally omits some details, including the order of processing operations. Additionally, if multiple labels can be assigned to the collected personal data, the DPV diagram retains only the most specific one. For instance, in the example from Figure 5c, the DPV diagram would preserve only a *hasPersonalData* edge between Process and GPSCoordinate (cf. Figure 5d).

To analyze GDPR compliance, DPV Transformer applies a set of predefined rules that link observed data-handling patterns in the code to GDPR obligations. Each program slice extracted by the Privacy Slice Visualizer is represented as a sequence of processing events—such as collect, process, or pseudonymize—which are mapped to corresponding DPV concepts (cf. Table 2).

The rule engine then evaluates these event sequences to identify conditions reflecting potential compliance or violation scenarios. For instance, to detect *processing of data in the absence of technical measures*, the tool checks whether a slice contains both data collection and processing events but lacks any pseudonymization operation. When this condition is met, Assessor View issues a high-severity warning linked to GDPR Article 25 (3 in Figure 4).

Each rule similarly encodes a combination of DPV terms, logical conditions, and corresponding legal references. While the current implementation relies on manually defined rules, the modular design allows the addition or modification of rules as new DPV concepts or regulatory interpretations emerge.

Correctness and Performance Check

To ensure the correctness of Assessor View's results, we first conducted sanity checks by manually verifying its output. The experiments were performed on an Ubuntu 20.04.2 machine with an Intel(R) Core i7-10850H processor (6 cores). In total, we evaluated Assessor View on 51 benchmark applications.

We first confirmed the correctness of the slicing algorithm on which the DPV mapping is based. To this end, we applied Assessor View to 18 DroidBench test cases [45] and manually compared the generated program slices against the documented program behavior. To ensure coverage, we selected at least one test case from each category of the microbenchmark suite. No discrepancies were observed. We then verified that the detected DPV concepts (personal data and processing operations) were correctly mapped from the **Developer View** to the **DPV View**. Out of 293 detected DPV concepts, 281 (95.9%) were verified through manual inspection. The remaining cases involved return statements and other complex statements, which are conservatively matched to *dpv:Processing* by design, resulting in intentional over-approximations. On average, Assessor View processed each test case in **34.38 seconds**.

To further assess runtime and verify the correctness of detected risk sources, we evaluated the tool on 33 applications from TaintBench [37]. We chose all apps from the TaintBench suite (33 out of 39) that had code size less than 100K lines. This LOC threshold ensured that the detected sources remained comparable against the benchmark's ground truth—something not feasible with larger real-world apps. The sanity check revealed that all the sources detected by Assessor View matched with the sources present in the TaintBench ground truth. On average, Assessor View processed each app in **39.10 seconds**.

Currently, no benchmark exists to systematically assess the correctness or precision of GDPR-specific rule detection (e.g., the five GDPR rules listed in Table 1) for Android APKs. Consequently, while we verified the correctness of the underlying slicing mechanisms through DroidBench and TaintBench, quantitative accuracy measures (such as false positive/negative rates) for GDPR rule detection remain an open direction for future work. Developing or adopting a GDPR-focused benchmark would enable a more comprehensive evaluation of DPV View's accuracy in identifying compliance risks.

5 Interview-based User Study

In this section, we discuss the study we conducted with Assessor View to understand its possible role in conducting privacy assessments. We first discuss the research questions we aim to answer with our study, and then explain the methodology of our study. Given the exploratory nature of our research, we employed a semi-structured interview approach to gather insights while allowing flexibility for participants to share their perspectives.

5.1 Research Questions

In this study, we aim to answer the following research questions:
RQ1. *How are privacy assessments conducted in a real-world setting?*

This question identifies the key stakeholders responsible for these evaluations, their specific roles, and whether they rely on any tools or frameworks for support. Understanding these aspects will provide insight into the practical challenges and best practices within privacy assessment workflows.

RQ2. *To what extent does the source code analysis-based Assessor View support privacy assessors in conducting DPIAs?*

This question examines to what extent Assessor View can act as a bridge between technical and non-technical stakeholders, enhancing communication and collaboration. The aim is to assess its effectiveness in reducing knowledge gaps and ensuring smoother privacy assessments through better alignment of technical and legal perspectives.

5.2 Participant Recruitment

As interview subjects, we were seeking to recruit DPOs or other legal and privacy experts. The first author directly contacted 175 privacy experts on LinkedIn. We received 38 responses (21.71% response rate), out of which 15 agreed to participate in the study. We also recruited 4 participants by sending an email to the Data Privacy Vocabulary maintenance group [46].

Those interested in participating in our study were sent to a landing page informing them about the study's purpose. Then, they were shown our consent form where we informed them about how we would handle their data, their right to terminate the interview at any time, and where they could give consent to audio-recording of the interview. Finally, they could schedule their interview.

In total, we recruited 19 privacy assessors, but three participants had to cancel due to unforeseen circumstances. Therefore, 16 participants took part. We refer to our participants as P01–P16 (cf. Section 6).

5.3 Study Procedure

We next explain how we conducted the study.

5.3.1 Interview Process

We used Zoom for all interviews. All interviews were audio-recorded via Zoom's functionality. Five interviews (P01-P05) were conducted by the first and third author, and the others by the first author. All interviews were conducted remotely between February and June 2025. We expected the interviews to take 45 to 60 minutes, and scheduled one-hour appointments with all interviewees. Interviews lasted between 26 to 85 minutes; the median duration was 47.5 minutes.

5.3.2 Interview Structure

At the beginning of each interview, we introduced ourselves and provided details about the interview process. We explained that participation was voluntary, and we could skip any questions with no penalty (interview briefing in Appendix A). Interviewees could ask questions, and we took additional verbal consent for interview recording. Given the exploratory nature of our research, we employed a semi-structured interview approach to gather insights while allowing flexibility for participants to share their perspectives. The interview had 4 sections:

1. **Pre-questionnaire**: We first asked participants to self-report their knowledge about data visualization, GDPR, GDPR compliance, Data Protection Impact Assessment (DPIA), and Data Privacy Vocabulary (DPV). The pre-questionnaire also asked them about their experience, their highest degree, their additional privacy certifications (if any), and their role within their organization.
2. **Current state of the art**: In this section, we asked participants questions about their experience with DPIA, and other GDPR compliance assessments. This step was to gain understanding on how privacy assessments are conducted within an organization, who is involved in them, and to what extent they are currently automated (*RQ1*).
3. **Tool-based questionnaire**: We introduced Assessor View with a short 3-minutes demo video[3]. We then asked participants to try out Assessor View, and complete two tasks to assess whether given source code of the demo application complies with GDPR guidelines. This part of the interview allowed participants to explore the various views of Assessor View, and report on their experiences.
4. **Tool usability and effectiveness for DPIA**: Finally, we asked for some qualitative feedback about Assessor View, and tried to understand its relevance for

[3] Available at https://doi.org/10.5281/zenodo.15085611

Topic (#)	Example questions
Privacy assessments/ DPIAs (5)	Can you walk us through a privacy assessment process? Who are the stakeholders? Which parts of the assessment are manual? What is your role in the assessment process? How do privacy assessors navigate understanding source code to successfully complete such an assessment?
DPV (2)	Is DPV or any other vocabulary actively used in the assessment process? Any plans of long-term support version of DPV?
GDPR compliance (3)	Have you used any tools to verify whether a software complies with GDPR? Which documents do data controllers need to complete to adhere to GDPR? Do your clients consult you to help complete these documents? Is this a manual process?
Privacy training (2)	Are there any training programs specific to GDPR that are mandatory for the stakeholders involved in a privacy assessment process? Do these training programs cover automation?

Table 3: Summary of interview protocol. The first column is the topic and the number of questions for that topic.

automating DPIA. Here we also asked the participants some specific questions about the design of the DPV model, and DPV diagram. After these tasks, participants answered several 5-point Likert-type questions from the System Usability Scale (SUS) [47], a questionnaire designed to measure the effectiveness and efficiency of a system, and rated Assessor View using a Net Promoter Score (NPS) [48], that measures their likelihood of recommending the tool to a colleague. This final part of the interview study was aimed at understanding to what extent Assessor View, in its current state, can automate the process of DPIA (*RQ2*).

The interview questionnaire is available in Appendix B.

The interview protocol was refined following a pilot study with a colleague, ensuring clarity. A summary of the final interview protocol is presented in Table 3.

The last two sections of the study—the tool-based questionnaire, and tool usability evaluation—employed a mixed-methods approach, combining a cognitive walkthrough with a usability-focused questionnaire. During this phase, participants were encouraged to share their experiences freely, and as a result, we did not follow a strictly predefined protocol for this part of the study.

5.4 Participant Experience

In the pre-questionnaire, participants answered questions on a Likert scale from 1 (Beginner) to 5 (Expert) about their experience (cf. Table 4). Participants have an average of approximately **10.78 years** of experience in privacy and/or law. Participants self-reported familiarity with Data visualization (median 3/5), GDPR (median 5/5), GDPR compliance (median 5/5), Data Protection Impact Assessment (median 5/5), and Data Privacy Vocabulary (median 4.5/5).

6 participants have a doctoral degree. P03, P07, and P14 have a doctoral degree in Computer Science. P10 and P13 have a doctoral degree in Law, and are actively

ID	Job title	Country	Exp in privacy /law (yrs)	Familiarity				
				Data visualization	GDPR	GDPR compliance	DPIA	DPV
P01	Jr. Data Protection Counsel Legal	Germany	3.5	●●●○○	●●●●●	●●●●○	●●●○○	●●●●●
P02 (∗)	Researcher	Germany	7	●●○○○	●●●●○	●●●●○	●●●●○	●●●●○
P03 (∗)	Assistant Professor	Ireland	11	●●●○○	●●●●●	●●●●●	●●●●●	●●●●●
P04	Privacy Counsel	Germany	8	●●●○○	●●●●○	●●●●○	●●●●○	●○○○○
P05	Managing Director, External DPO	UK	15	●●●○○	●●●●●	●●●●●	●●●●●	●●●●○
P06	Manager	India	12	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●
P07	Researcher	Germany	15	●●●○○	●●●●○	●●●○○	●●●○○	●●○○○
P08	DPO	Cyprus	7	●●●○○	●●●●●	●●●●●	●●●●●	●●●●●
P09	Corporate DPO	Germany	9	●●●●○	●●●●●	●●●●●	●●●●○	●●○○○
P10 (∗)	Associate Researcher	Greece	14	●●●○○	●●●●●	●●●●●	●●●●●	●●●●●
P11	Regulatory Compliance Analyst	Nigeria	3	●●●○○	●●●○○	●●●○○	●●●●●	●●●●●
P12	DPO	Saudi Arabia	5	●●●●●	●●●●○	●●●●○	●●●●●	●●●●●
P13	DPO	Germany	12	●○○○○	●●●●●	●●●●●	●●●●●	●●●●○
P14 (∗)	Associate Professor	Spain	20	●●●○○	●●●●○	●●●●○	●●●●○	●●●○○
P15	Legal and Policy Counsel	Netherlands	4	●●●○○	●●●●●	●●●●●	●●●●●	●○○○○
P16	Legal Researcher	Belgium	7	●●○○○	●●●●●	●●●●●	●●●●●	●●●●●

Table 4: Summary of participant experience. (∗) denotes members of the DPV maintenance group [46]. Filled circles represent expertise levels on a Likert scale from 1 (Beginner) to 5 (Expert). P12 and P15 only participated in the non-tool based study (Sections 1 and 2 of the study).

working in data protection and compliance. P06 has a doctoral degree in Information Law.

P01, P04, P05, P08, P09, P12, and P13 (N=7) are DPOs. P01, P04, P05, P06, P08, and P09 (N=6) have additional privacy certifications from the International Association of Privacy Professionals [49]. P01, P04, P09, and P13 (N=4) have German DPO certifications. P02, P03, P14 are Computer Science researchers from the Data Privacy Vocabulary maintenance group [46], and P10 is a legal researcher from the DPV group. P11 is an auditor.

Participants have a versatile experience in attending and conducting DPIAs. P01, P04, P05, P08, P09, P12, and P13 (N=7) have been DPOs in many DPIAs. P03 and P07 have experience in leading the writing of DPIA, drafting documents and conducting the necessary research. P10, P15, and P16 provide legal insights in a DPIA. P02 and P14 are privacy researchers who have studied the process of DPIAs without being actively involved in its implementation.

5.5 Data Analysis

We recorded the audio of all interviews, transcribed them using NoScribe [50] and reviewed the transcripts for potential mistakes. We used an iterative open coding approach [51] to perform thematic analysis [52] of all interview transcripts. The first author carried out inductive coding to derive emergent themes, which were then compiled into a codebook. The second author independently reviewed the codebook, and disagreements were discussed and resolved through refinement of the final coding scheme. During this process, the authors reached full agreement on all sub-themes but refined the classification by renaming 10 of the 95 codes. The final codebook is available in Appendix C.

5.6 Ethics and Data Protection

The study design was approved by the Ethics committee of our institution. Our consent form, privacy policy, and data handling practices were approved by the Data Protection Office of our institution. NoScribe [50] was selected to transcribe the interviews following our DPO's recommendation. Some DPO participants also reviewed the study materials before participation to confirm compliance with data protection principles.

6 Results

In this section, we present the results of the analysis of the 16 semi-structured interviews.

6.1 RQ1. How are privacy assessments conducted in a real-world setting?

We now discuss the processes conducted as part of a DPIA, the stakeholders involved, tool support for DPIA, the challenges faced in conducting a DPIA, and the processes and support required for GDPR compliance.

6.1.1 Processes in a DPIA

The interview study revealed that a DPIA is primarily an **information-gathering process** *(P05)*. It is often conducted as an online call, with a mixture of forms and questionnaires where the team implementing a system provides details on the system's purpose, the necessity of its approach, the technical and organizational secure measures implemented within the system, and a discusses possible risks i.e. what would happen if the system went rogue, or breached and data would be lost or made public *(P04)*. While P01 described the process as taking stock, making inventory, and documenting it within the Record of Processing Activities (RoPA) [53], P03 summarized it as **GDPR-specific investigations of systems**. According to P12, the primary goal of the DPIA is to map where the personal data goes, how it is used, and what risk is attached at each stage. P05 emphasized on the importance of talking to their clients and exploring alternative approaches that could reduce risks to the personal data.

Participants also highlighted the different perspectives regarding the DPIA process. P01 explained how legal teams verify the security measures implemented in a system: *"We (the legal team) ask for confirmation from cybersecurity team, if they are happy with these measures that are implemented. If they say yes, we are happy with it. The legal team doesn't go and check the technical implementation."* P05 stressed on the ethical dimension of a DPIA: *"Something I add into a DPIA (which is not a part of the standard process) is making them think if they would be happy if this was their data that all this processing was happening to. In my mind, if that answer is "no", well then you need to ask yourself why anyone else would be happy (with this happening to their data)."* P02 highlighted the importance of source code analysis in DPIAs: *"(On why source code analysis is important in a DPIA) For example if there's a logging mechanism which is not necessary which logs personal data, then that is something that needs to be taken into consideration. If I do a bank account transfer, the bank would log when I transferred how much money to which other company (in addition to the account) then this is a second location where personal data is effectively stored. Then these internals of the software (where is this stored etc) are highly relevant. Otherwise IT administrator of the bank can see what I did. I don't want that."*
P09 explained their job in a DPIA: *"I can translate between technical and legal requirements. The core reason why I have a job is because that is what I'm good at that. I can talk with the developer and understand what they say and then I can talk with legal people or business people and understand what they say. So this is basically my core competency to bridge this gap of understanding."*

6.1.2 Stakeholders in a DPIA

Participants identified four main categories of stakeholders in a DPIA:

1. **Business functions**: These include entities responsible for collecting and processing personal data, such as data controllers, data processors, application owners, technical teams (IT, cybersecurity), management teams (human resources, marketing), and project sponsors.
2. **Privacy functions**: The Data Protection Officer (DPO) plays a counselling role, with the privacy team in the operational role. P01 (a DPO) said, *"I prepare the questionnaires, go after the stakeholders, walk them through the questionnaire, interview them to understand what activities they are doing and to collect the data flow from them from the moment they collect it. For which purpose they are using data, based on which legal ground, where are they hosted, if there are technical measures in the document, and other important topics of that conversations. I then document it on the Record of Processing Activities."*

3. **Data subjects**: DPIAs should involve representatives of the individuals directly affected by the system. As P10 observed, *"As the European data protection supervisor puts it, we do not protect the data. We protect the people behind the data. So our primary consideration in data protection law are those people and their vulnerabilities."*
4. **Third parties**: Business partners or third parties involved with the system under assessment. DPIA allows for consultations with different experts. In rare cases, DPOs may involve the data protection supervisory authority at later stages (*P03*).

6.1.3 Tool support for a DPIA

All participants with DPIA experience reported that most organizations rely on **manual processes** for conducting DPIAs. Furthermore, the degree of automation varies based on the level of digitisation of the company. Some mid-level digital organizations use software solutions to automate inventory of processing activities and stocktaking activities. More digitally advanced companies may apply process mining to analyze event logs.

P02 explored the use of a tool [54] by the French Data Protection Authority [55] that guides organizations in conducting DPIAs, resulting in a report that you can use to demonstrate compliance with GDPR and the risks, how you can mitigate risks, and to properly inform data subjects. P04 noted that some organizations automate meeting transcript generation and summarization. P09 and P12 used data governance tools such as Caralegal [56] and Collibra [57] to structure their DPIA. P06 explained, *"Some of these clients do have the money to invest in tools, like OneTrust, etc. where all DPIA questionnaires are just uploaded onto the tool and then is shared with the application owner."* P08 said, *"So, a lot of it is manual, but because of the type of environment within which we work. We have, like, standard forms that can just be triggered. But then someone still has to do the manual data entry and someone still has to do the manual risk review so it's not like it's powered by AI or something."*

P16 explained where manual assessment can be the most useful: *"As a strong lawyer and privacy oriented person, I would say the more manual the process, the better to a certain extent, especially in this phase of assessing the risks. While I see the automatic tools being more important for maybe having the catalog and mapping of data processing activities, but while I also understand that there are certain limitations as budget and resources, I believe that at least this final assessment of are we accepting these risks should be done manually."* Thus, automation remains uncommon in DPIA workflows.

6.1.4 Vocabularies used in a DPIA

The use of Data Privacy Vocabulary (DPV) or other standard vocabularies for structuring DPIA content is **rare**. In some cases, highly digital organizations develop their own internal vocabularies, shared across departments (e.g., IT, privacy, information security, and data governance) (*P04*). P08 explained, *"We are familiar with it (DPV), but I think that we don't really use it specifically for our DPIA process."* P16 further explained, *"These common vocabularies, standardization practices are completely relevant and used exactly because then it's easier to make this communication among* *different areas. So, for example, technical developers are more keen on understanding certain standards and vocabularies. So using these is really important. But it's also important to understand the limits of certain initiatives. Since they are not part of the regulation per se, they are not necessarily officially recognized. There are certain limits because they will not necessarily translate everything that is needed for a DPIA. And at the same time, you cannot just rely on that as to translate everything because the authorities will not necessarily see that as the final result or the final tool to be used. So there are limits, even though they are very used and very, very useful."*

6.1.5 Challenges faced in a DPIA

Participants identified **knowledge gaps** and **lack of communication** between the different stakeholders as the biggest challenge in conducting DPIAs: P01 observed, *"As my professional experience is concerned, that (privacy assessors understanding source code to complete a DPIA) is the biggest void in the market. There is no communication between developers and lawyers, they are in two completely different worlds."* Similarly, P03 emphasized the difficulty of collaboration due to differing expertise: *"Legal team drafting DPIA (liability or contractual rights or data subject rights: legal topics) would have to discuss stuff with a developer. A developer is not going to give them legal answers. There has to be some common ground, which doesn't exist."* P04 explained how the difference in conceptual models between legal and technical teams further complicates collaboration: *"We (legal experts and technical teams) have a difference in vocabulary and I'd say mental model and concepts, which makes it sometimes hard. This leads to frustrations on both sides, by the way."*

Beyond communication barriers, conflicts can arise due to the **tradeoff between privacy and innovation** of a software. P06 observed, *"It's not just legal versus technical as well, because you also need to understand the business interests. So you need to understand how the business functions, why they're using this application for, you know, exactly what purpose and what are they trying to achieve even fiscally, monetarily. Because you have to then, because it's not your job to just go and say no to everyone, right? You have to understand why they're trying to achieve a certain goal and then what might be the best way to actually protect data while achieving that goal."* P16 added, *"Sometimes technical developers are more inclined into going to innovation and first testing and training a certain tool and then thinking about compliance while it is really important to have privacy by design and by default. So even before we know the whole potential of a certain software, it is important to already have these privacy solutions embedded solutions. So this clash of interests is always a challenge."*

Participants also pointed to **delays in considering privacy principles**, with privacy teams often being involved too late in the process. Ideally, DPIAs should be conducted before processing activities begin, but this is rarely the case in practice. P01 noted, *"Best case (in conducting a DPIA) is before starting processing activities, but in practical world that does not happen because people get privacy involved later. According to privacy by design, developers should think of privacy principles. They should keep requirements of the law (such as data minimization, pseudonymization) in mind when they develop a new software. But it's not a rule. And technical knowledge of the lawyers is very limited."* P15 added, *"Sometimes we are involved a bit from* *the outset. Other times we are really a bit of an afterthought. There is no point in lying: most GDPR is an afterthought."* P05 stressed the importance of involving DPOs earlier in the development process: *"If DPOs have important conversations (with the IT team) on day one, it saves a lot of time and conflict. Reasons for such conflicts vary. In some cases, it is the ignorance and lack of knowledge (of what the law says they can and can't do). In my experience some know that this is wrong. They hope (in a DPIA), the DPO would be light on them and they would get away with doing anything they want."*

Another key challenge is the **lack of source code analysis-based tools** for assessing source code from a privacy perspective. P04 noted the absence of automated code scanning tools for GDPR compliance, despite their prevalence in security assessments: *"I have never seen automated code scanning from privacy point of view (so far). From security point of view, yes, which takes some privacy issues but it is more focused on security, not on GDPR compliance part apart from Article §32 [58] (security of processing)."* P09 observed, *"So I remember looking at tools like that would look through code and then give you certain information but nothing specifically tailored for DPIA. I remember some early prototypes, like 2020 maybe, but nothing that ever got into production that say like consistent use."*

P01 pointed out that privacy experts and DPOs are not actively engaged in guiding privacy-conscious development: *"Majority of recommendations of DPOs or privacy experts stay superficially on the principles of GDPR in Article §5 [42]: data minimization, limitation of storage, and things like that. At least I don't know anyone that goes inside of it in the development, in trying to influence and recommend Privacy Enhancing Technologies."*

Finally, P03 noted that many stakeholders lack awareness of the information required for a DPIA. As a result, privacy professionals must actively seek out relevant details: *"Usually the challenge is that most people don't know all the information needed for a DPIA, so you have to hunt for people who have the information, and figure out a way to convert their answers into what exactly you want in a DPIA."* P16 added, *"I see a challenge in defining roles in, for example, a software developing company. They can then act as processors or controllers of data. Exactly defining these roles is quite challenging."*

6.1.6 GDPR-Specific Documentation

Participants identified **18 documents** (cf. Appendix C) that data controllers may need to complete to adhere to GDPR. Fourteen participants identified **Record of Processing Activities (RoPA)** [53] as the most important document. RoPA, a legally mandated inventory detailing how personal data is processed, is often used as starting point for DPIAs. As P10 explained: *"Completing the RoPA is a data mapping exercise which allows stakeholders to envisage potential threats."*

Clients, typically data controllers, seek legal help to complete other GDPR-related documents: privacy notice, standard agreements and contracts, assessment of partner activities, rule of conduct, DPIA report, informed consent, report of data breach, data access request, and data protection by design assessment.

Most participants help their clients by manually completing and reviewing these documents. P04 noted that the approach depends on the level of sophistication in the company: *"If we're talking to a start up that basically has little financial resources then we work with Word and Excel templates that we provide to them that have comments on how something should be filled out. If we have a more financially sophisticated organization that is also capable to do more, then we use so-called legal tech tools that are decision trees in a way where for instance for legitimate interest assessment under Article §6.(f) you will be asked a couple of questions.. and in the end it'll give you the balancing test or the outcome of the balancing test created if you wish in a way by the court of justice. We have these types of document generators (that don't check source code) that help clients to fill out this documentation without me as a legal advisor being there all the time. We train them so they can use it."* P09 explained, *"But the last five years, I would say the focus was mainly getting people to enter the data rather than getting it out of systems directly."*

Moreover, 10 out of 16 participants mentioned using standard templates to complete GDPR-specific documentation. P07 noted: *"(Completing such documents) is mostly based on spreadsheets and word tables and word documents, because that's what most people are comfortable in using."* P10 elaborated, *"Most of the templates are available. So there are templates by the European Commission for processing agreements or international transfers. There are templates for deep data protection impact assessments. What you need is to put the expertise on how to fill it in to know what every section requires."*

6.1.7 Compliance Support

Participants claimed that GDPR compliance is a **manual process**. P03, P04, P08, and P12 mentioned using cookie compliance tools to verify website dialogue boxes but had not used any automated code-scanning tools for GDPR compliance. P07 admitted to using privacy policy scanners. P04 noted: *"I played around with a Privado [31] product in my spare time. Code scanner to identify what type of processing is done. But I've not seen this in organizations, or in their CI/CD pipelines. So I have little experience with it."* P12, who has used security and code analysis tools such as SonarQube [59], OneTrust [60], Salesforce Privacy Center [61], said, *"The tool is helping through the process, but I have in my position to check everything manually. I am the point of contact. I explain every section in the document and help them to complete it."*

6.1.8 Training Programs

P01, P05, and P10 pointed out the lack of training programs designed specifically for GDPR compliance.

In contrast, many participants observed different kinds of training programs and certifications that are encouraged. P04 observed: *"All organizations I've seen have this base level compliance for all employees (who use a computer) and then dedicated training programs for different roles depending on the intensity of handling personal data."* P09 said, *"For the business units, we usually have onboarding for the entire RoPA process, which is part of the DPIA."* P06 added, *"We recommend mandatory* *awareness sessions simply because if they don't know what they're doing, they will just treat it as a box-ticking exercise. So we recommend it, but obviously as consultants, we don't mandate awareness exercises."* P16 explained, *"There are no trainings that are mandatory. Any training offered is seen as a good practice and can be used to further show that all employees were part of a training, but none is mandatory. So there is no specific course that should be done, but any course can be used as a proof of compliance of making the whole company aware of privacy issues."* P12 emphasized different formal certifications (CIPP/E, CIPM) that DPOs can take to enhance their knowledge.

RQ1 summary:

- DPIA is a manual, information-gathering process requiring coordination among privacy, legal, and technical experts.
- Key challenges include knowledge and communication gaps between experts, the privacy–innovation trade-off, delayed involvement of privacy professionals, and the lack of source code analysis-based tools.
- Data controllers rely on privacy assessors for manual GDPR compliance support.
- RoPA is the most important documentation required for GDPR compliance.

6.2 RQ2: To what extent does the source code analysis-based Assessor View support privacy assessors in conducting DPIAs?

To answer RQ2, the final part of the interview involved a hands-on session with Assessor View. P12 could not access the tool at their location, and P15 declined participation in this session, and hence we have data from 14 participants.

6.2.1 DPV View and GDPR Warnings

After using the tool, almost all participants (N=13) agreed that the **DPV View helps in identifying potential GDPR violations** or areas requiring further investigation (cf. Figure 7). P09 remained neutral: *"I probably would end up having to look at the documentation of the specific function to understand the scope of the risk"*.

80% of all participants agreed that **GDPR compliance warnings and suggestions are useful in guiding their analysis**. P05 added, *"The compliance warnings and suggestions are good and would help in guiding analysis. I will caveat that by saying that of course without having sight of the application being reviewed itself, it is not possible to judge how accurately the issues have been detected."* P04 further remarked, *"They already hint at potential problems that would need further analysis and to be linked back with what was initially planned."*

80% of all participants agreed that the **warnings effectively highlight potential GDPR compliance issues**. P06 commented, *"The warnings do a good job of flagging issues that non-privacy folks may not have thought of."* While 70% of all participants felt that the warnings simplify the assessment of GDPR principles such as data minimization, purpose limitation, or security measures, P01 disagreed because

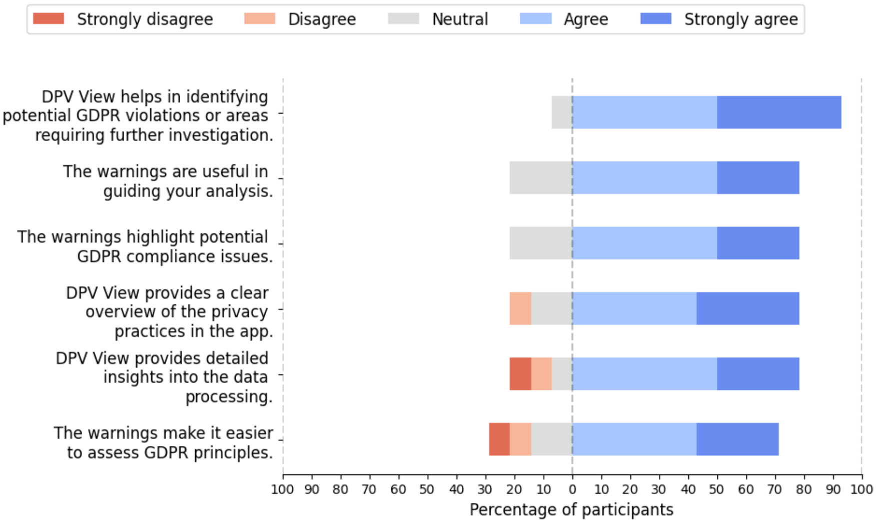


Fig. 7: Participants' responses to the tool-based questions about DPV View and GDPR compliance warnings.

they believe that *the thorough analyses of the compliance of the processing should be connected to an end purpose, which is difficult to be detected from the source code.*

80% of all participants agreed that the **DPV View provides a clear overview of privacy practices in the program slice**. P06, however, disagreed: *"The DPV View needs to work on explaining itself more. The flow itself may be useful, but the tags are not."* P04 commented, *"The DPV View for the demo app tells me that there is a combination of data, but I lack information on what data it was combined with."*

80% of all participants pointed out that the **DPV View provides detailed insights into specific data processing activities and flows**. However, P04 remained neutral, noting that the connections among different program slices are missing, making it challenging to fully understand the broader data flow within the application.

6.2.2 Usability and User Satisfaction

Participants answered several 5-point Likert-type questions from the System Usability Scale (SUS) [47], a questionnaire designed to measure the effectiveness and efficiency of a system.

Notably, over 80% of participants found Assessor View **easy to use**, and said they would like to use it for their needs. However, 20% felt that people would struggle to learn the tool quickly and, as a result, lacked confidence when using it. Additionally, over 40% indicated that they would need support from a technical expert to use the tool effectively—which is understandable, given that the tool is designed for both technical and non-technical users. Moreover, 20% identified a significant learning curve associated with using Assessor View.

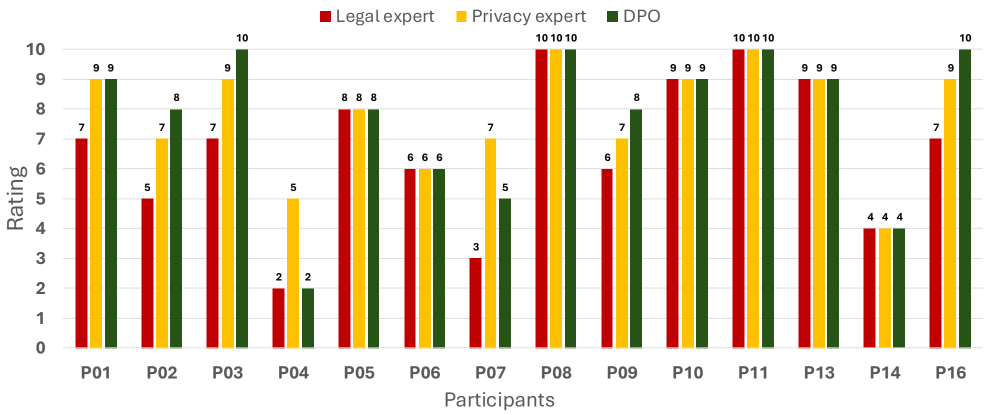


Fig. 8: Assessor View's rating for legal experts, privacy experts, and DPOs. (0: not recommended, 10: highly recommended)

The calculated SUS score for Assessor View was **75.41**, exceeding the industry benchmark of 68 and suggesting generally acceptable usability. Net Promoter Scores (NPS) indicated that the tool is better suited for privacy experts (28.57) and DPOs (21.43) than for legal professionals (-14.29) (individual ratings in Figure 8).

6.2.3 Potential Use for Assessments

All participants agreed that Assessor View can be used for conducting **DPIAs**. P01 said, *"It will help to get answers from the technical team and developers. But as I mentioned, connecting relevant slices to a final purpose would help a lot."* They added, *"It would be great for tech teams to self-assess their tools, and use the output to defend their decisions in a DPIA."* P05 commented, *"I think the tool is good. I think it could form part of an assessment very well, actually."*

Participants highlighted Assessor View's use during the **pre-DPIA screening phase**. P08 said, *"I think it might be very helpful at the information gathering part of the privacy assessment. Where we are trying to understand the processing activity and the context. I think it might be very, very helpful. And maybe where the company is in the process of building a software. They can also use this to reveal and have a better understanding of their codebase."* P16 noted, *"The tool has an enormous potential to facilitate the data mapping as it is based on the source of development, which mitigates the risks of not addressing certain activities that could not be mentioned in a manual process."* They added, *"It makes it even easier for a DPO or a legal expert to evaluate a software, which sometimes can be very challenging exactly because of this issue of translation between the developer and the person conducting the DPIA. So it's really, really relevant to make sure that the entry points and the exit points are mapped. So exactly to understand the data flow within an app."*

Participants emphasized on its importance for **bridging communication gap** between stakeholders. P11 said, *"Definitely it will ease communication because with this, it was like a lot more simple for me to see the issues in relation to the app.*

It kind of dumbed it down for me because I'm not technical." P09 further noted, *"I think it could actually help a little bit both sides to understand each other better. It basically opens the use for two groups, one who understands the things, who can drill down on it; and the other one to get a broad overview of what's going on."* P06 added: *"It's a good place to start because the biggest issue I see in privacy is because DPIAs generally have to be triggered by the application owners. A lot of them, by default treat it as a box-ticking exercise. So they just mark everything as low risk so that it never gets flagged for a DPO."* P04 said, *"Assessor View would not enable legal people to talk more in a technical way, but the other way around (technical people can talk more legally). With what we have now, it will analyze a piece of technology and hint at potential issues that you would have to have a look at. It will draw attention, and that could be helpful."* P02 remarked, *"The DPO is not that involved in the development, and isn't the legal expert. The DPO can ask the developer: What do you think about this warning? Can you apply pseudonymization? Does it justify leaving boundaries? Or. . . . is storing something on Cloud necessary? I think this is a good tool for DPOs."*

At the same time, participants emphasized the **importance of retaining manual review for final risk evaluation**. P07 observed, *"There are other parts, especially, you know, regarding the question of risk assessment and access, like talking to stakeholders, understanding their individual concerns, which I think is important that I don't think can be, should be automated."* P06 remained neutral about tool usage, explaining, *"They (tool-based warnings) may be useful, but only if they are a precursor to a full-blown investigation."* P16 added, *"As a strong lawyer and privacy-oriented person, I would say the more manual the process, the better to a certain extent, especially in this phase of assessing the risks. While I see the automatic tools being more important for maybe having the catalog and mapping of data processing activities, but while I also understand that there are certain limitations as budget and resources, I believe that at least this final assessment of are we accepting these risks should be done manually."*

This perspective reinforces the idea that automation should **complement** rather than **replace** expert judgment, providing structured insights and traceable mappings of data processing activities while leaving ultimate decisions about risk acceptance to privacy assessors.

6.2.4 Additional Insights and Suggestions

P01 emphasized that processing personal data is only lawful if linked to a legitimate business purpose or legal ground, suggesting that GDPR compliance should be assessed in context rather than in isolation. For example, storing and processing personal data without pseudonymization could be GDPR compliant if justified by a legitimate business need. P02 supported this view, adding, *"This makes me skeptical if it is sufficient to look at the software without considering the context."* P04 recommended adding hints to show where analyzed functions are located in the source code to facilitate collaboration between privacy professionals and developers. P05 noted that while the tool identifies risks, it lacks actionable guidance on resolving them. They proposed adding hints such as *"this data should be encrypted at this point"* and *"users should be encouraged to use multifactor authentication to enhance security"*.

P04 recommended enhancing data flow transparency by providing information on the type of collected data, its movement within the system, and whether it is combined with other data. P07 suggested making the dashboard more interactive: *"It would be great if you could follow your thoughts, maybe leave notes somewhere and say, okay, I need to follow up on this."* P08 recommended adding pseudonymization related hints to the relevant GDPR warnings.

Participants suggested the following tool extensions: P01 recommended documenting dashboard visualizations to present as evidence of GDPR compliance to supervisory authorities. P02 noted, *"Dataflow diagram is the most relevant extension of this tool. You are showing me a graph which connects nodes within the software, but specially the case where data leaves the boundary of your software, things become interesting and relevant, and this is something I would like to see."* P03 and P05 suggested converting Assessor View's output to RoPA. P03 and P04 proposed displaying the tool's output as a list or report for easier discussion in DPIAs. P04 observed, *"A list would be more natural for a DPO. The way it's presented now is more natural to a developer."*

RQ2 summary:

- Participants found Assessor View well suited to assist DPOs and privacy experts in conducting DPIAs and pre-DPIA screenings.
- They suggested automatically converting DPV View to RoPA, or reports that DPOs can use to conduct DPIAs.

6.3 Discussion

The interview study identified a diverse range of stakeholders involved in a DPIA, with expertise spanning privacy, technology, and law. This highlights the need for effective communication and a platform such as Assessor View that caters to multiple target groups.

Further discussions revealed that DPIAs are still largely conducted manually, with minimal use of standard vocabularies. While this presents an opportunity for automation, the lack of active use of DPV in DPIAs poses a challenge for making Assessor View relevant in this context. Interestingly, although P01, P04, and P05 were unfamiliar with DPV, they were acquainted with other ontologies, which helped them understand the DPV View. Notably, P04, who identified as a beginner in DPV (cf. Table 4), remarked, *"DPV... oh, that looks familiar to me"* upon reading the DPV documentation in Assessor View. This suggests that while DPV is not yet widely used in DPIAs, privacy assessors may find the DPV specifications more accessible than source code-based representations. P07 supported the use of DPV for bridging knowledge gap: *"I wouldn't rely on it (DPV) to be the perfect dictionary, but it's a very good start, and I don't know of any better (standard vocabulary) at the moment."*

The biggest challenges in conducting a DPIA include knowledge and communication gaps between legal and technical experts, absence of automated privacy-focused tools, and delayed involvement of privacy professionals. These findings illustrate that the problem scenario (cf. Section 2) is true. The study also revealed that data controllers often rely on legal advice to comply with GDPR, with privacy assessors providing manual support.

Moving forward, our findings suggest several design directions for automation in privacy assessments:

- UI-based tools that allow partial document generation (e.g., pre-filling sections of the DPIA or RoPA from technical analyses) to reduce manual effort while maintaining expert oversight.
- Interactive dashboards that guide privacy assessors through compliance tasks, translating technical outputs into regulatory language suitable for DPOs and consultants.
- Automated prompts for business owners or developers when new data processing activities may require initiating or updating a DPIA.
- Integration with existing compliance ecosystems, since current tools such as OneTrust [60] do not process source code, and static analyzers such as SonarQube [59] overlook privacy-specific dimensions such as data minimization, purpose limitation [42], and Record of Processing Activities (RoPA).

Assessor View demonstrates how such automation can be used as a reflective probe—supporting structured discussions, improving mutual understanding, and helping privacy professionals translate code-level insights into compliance documentation. As several participants (P03, P04, P05) suggested, extending Assessor View to automatically generate reports such as the RoPA would be a natural next step.

Overall, these insights illustrate how automation can complement human expertise, providing structured information to support decision-making rather than replacing expert judgment.

Takeaway. Assessor View represents an initial step toward facilitating communication between legal and technical stakeholders, and automating privacy assessments. It is currently well suited for privacy experts and DPOs.

7 Limitations and Threats to Validity

In this section we discuss the limitations of the study performed in this paper and consider implications to the validity of the results.

As with any qualitative study, our findings are influenced by the sample of participants and their perspectives. To improve representativeness, we recruited a diverse group of privacy assessors through multiple rounds of outreach, including direct contact with 175 experts from 43 countries. While we aimed for a diverse participant pool, some potential participants outside the EU had limited familiarity with GDPR and redirected us to their EU-based teams, who did not respond. Despite these efforts, we acknowledge that experts with strong privacy concerns may be underrepresented, as they are typically less likely to participate in interview-based studies [62]. Similarly, we encountered a low response rate among DPOs and members of the DPV maintenance group, which may limit the breadth of viewpoints captured. However, we successfully recruited participants from 11 countries, out of which 4 participants were from outside EU.

To ensure that our analysis accurately represented participants' experiences with privacy assessments, we employed an iterative open coding process [51] following thematic analysis principles [52]. The first author performed inductive coding to derive emergent themes. To improve coding consistency and reliability, the second author independently reviewed the codebook, and disagreements were resolved through discussion, ensuring coder alignment. This systematic refinement process, along with the public availability of the final codebook (cf. Appendix C), enhances the transparency and construct validity of our findings by ensuring that the themes faithfully reflect the underlying data.

While our main goal was to capture experts' experiences with Assessor View, we also gathered quantitative feedback using the System Usability Scale (SUS) [47] and Net Promoter Score (NPS) [48]. Even though the study itself is primarily qualitative, the sample size is reasonable for calculating these scores. Prior work indicates that as few as six interviews can be sufficient when participants are relatively homogeneous and study goals are well-defined and qualitative in nature [63]. Expanding this into a larger-scale quantitative evaluation is an interesting future direction of research.

We used Assessor View to explore experts' perceptions on automation in privacy assessments. Ongoing efforts are required to optimize the static program slicing algorithm at the back-end and enhance the precision of the resulting program slices. Since Assessor View uses keyword matching to detect the presence of data sources including those from some third-party libraries, its analysis will not be able to handle code obfuscation in cases where the obfuscation renames also calls to these third-party APIs. While static analysis is useful for creating an initial prototype, future research could build on this foundation by exploring hybrid or dynamic approaches to further streamline privacy assessments.

Although Assessor View can generate slices for real-world apps, our preliminary experiments showed that these slices—often exceeding 100 nodes—quickly become too large for meaningful manual interpretation. This complexity highlights the need for manageable high-level views that summarize the source code, making it a strong motivation for our approach. However, scalability remains an open challenge when automatically visualizing such large slices across multiple views. Consequently, in this work we focus on smaller, controlled benchmarks: the hand-crafted micro-apps from DroidBench [45] and the apps from TaintBench [37], which are comparable in size and structure to typical Google Play applications.

The mapping between the Developer View and DPV concepts was performed manually and is therefore subject to human interpretation. We mitigated this threat to construct validity by engaging in detailed discussions among the authors and grounding our decisions in the official DPV documentation and definitions. We also sought feedback from the DPV maintenance group, whose suggestions helped refine the mapping. The implementation will be updated accordingly to include their recommendations in future revisions of the codebase.

Currently, no benchmark exists to systematically assess the correctness or precision of GDPR-specific rule detection for Android APKs. While we verified the correctness of the slicing algorithm of Assessor View through DroidBench and TaintBench, quantitative accuracy measures (such as false positive/negative rates) for GDPR rule detection remain an open direction for future work.

Assessor View currently supports DPV 2.0, and its mapping is designed to be modular but not fully self-updating. Any structural or conceptual changes in future DPV versions—such as updates to processing operations or personal data categories—will require manual review and adaptation of the Developer View to DPV mapping. When asked about any plans to introduce stable versions of DPV, P03 commented, *"In applying DPV to resources, you want the latest legal implications and interpretations, and that is only possible if you advance the vocabulary. So we plan to release a new version approximately every 6 months. Any fixes or such will get adopted into that version."* Continuous maintenance is therefore essential to ensure Assessor View remains aligned with the most recent legal and technical interpretations embedded in DPV.

8 Conclusion

In this paper, we presented an interview-based study (N=16) exploring the challenges privacy assessors encounter during privacy assessments and their views on automation as potential support. We introduced *Assessor View*—a static analysis-based academic prototype—as a probe to explore the potential benefits and limitations of automating parts of privacy assessments.

Our study identified key challenges in conducting privacy assessments, including knowledge gaps, poor communication between legal and technical experts, the absence of source code analysis-based privacy tools, and the delayed involvement of privacy professionals. The user study results reveal that the GDPR warnings and guidance provided by Assessor View are valuable to DPOs and privacy experts, and its design is particularly well suited for these stakeholders.

Overall, our findings indicate that Assessor View represents a significant step toward improving communication between legal and technical experts and automating privacy assessments.

Appendix A Interview Briefing

Thank you for participating in our study. We are going to be asking you some questions today about privacy assessments and GDPR compliance. Thank you for signing the pre-study consent form. I just wanted to emphasize that what you say to us during the study will be kept confidential, and you can stop participating at any time. If you feel uncomfortable answering any question, we can skip them with no penalty. We will be recording the study. Before we start recording and begin the study, would you like to ask us any questions? We will be sharing an online study questionnaire with you, that you can open in a new browser window. Would you be comfortable sharing your browser window? That will help us guide you through the process. We will sign another consent form before starting the study. Consent form: <obtain consent from participant>

Appendix B Interview Questionnaire

In this appendix, we detail the survey questions from Section 5. The anonymized answers are available at https://doi.org/10.5281/zenodo.15085611.

B.1 Section 1: Participant Information

1.1. What is your designation?
  – Free-text field.
  – Mandatory.

1.2. What is your primary role?
  – Free-text field.
  – Mandatory.

1.3. How many years of experience do you have in privacy and/or law? (in years)
  – Free-text field.
  – Mandatory.

1.4. What is the highest degree you have received? (If you're currently enrolled in school, please indicate the last degree you received)
  – Multiple choice question with an "Other.." field.
  – Choices:

  ∗ Bachelor's degree (LL.B., BSc., BTech, etc)
  ∗ Master's degree (LL.M., MSc., MTech, etc)
  ∗ Doctoral degree (Ph.D., Ed.D., etc.)
  ∗ Other...

  – Mandatory.

1.5. Have you received any additional privacy-related certifications? If yes, which ones?
  – Free-text field.
  – Optional.

1.6. Which country do you currently work in?
  – Free-text field.
  – Optional.

1.7. Rate your knowledge on the following:

A. Data visualization:
  – Likert scale from 1 (Beginner) to 5 (Expert).
  – Mandatory.

B. GDPR:
  – Likert scale from 1 (Beginner) to 5 (Expert).
  – Mandatory.

C. GDPR compliance:
  – Likert scale from 1 (Beginner) to 5 (Expert).
  – Mandatory.

D. Data Protection Impact Assessment (DPIA):
  – Likert scale from 1 (Beginner) to 5 (Expert).
  – Mandatory.

E. Data Privacy Vocabulary (DPV):

– Likert scale from 1 (Beginner) to 5 (Expert).
– Mandatory.

B.2 Section 2: Privacy Assessments and GDPR Compliance

2.1. Can you walk us through a privacy assessment or a DPIA process?
2.2. Who are the stakeholders in such a privacy assessment/ DPIA?
2.3. Which parts of the assessment are manual? Which parts aren't, and how are they assisted?
2.4. What is your role in the assessment process?
2.5. Do privacy assessors have to understand source code to successfully complete such an assessment/ to advice your clients? How do they navigate the process of understanding source code?
2.6. Is DPV or any such vocabulary actively used in the assessment process?
2.7. Any plans of long-term support version of DPV? (specifically for the DPV maintenance group)
2.8. Have you used any tools to verify whether a software complies with GDPR? Which tools?
2.9. Which documents do data controllers need to complete to adhere to GDPR?
2.10. Do your clients consult you to help complete these documents? Is this a manual process?
2.11. Are there any training programs specific to GDPR that are mandatory for the stakeholders involved in a privacy assessment process? Do they cover automation?

B.3 Section 3: Evaluating a Privacy-Relevant Program Slice

3.1. Assessor View's DPV View provides a clear overview of the privacy practices in the program slice.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Optional text-field for comments.
– Mandatory.
3.2. The DPV View helps in identifying potential GDPR violations or areas requiring further investigation.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Optional text-field for comments.
– Mandatory.
3.3. The DPV View provides detailed insights into the specific data processing activities and flows.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Optional text-field for comments.
– Mandatory.

B.4 Section 4: Understanding GDPR Compliance Warnings

4.1. The warnings highlight potential GDPR compliance issues.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Optional text-field for comments.

– Mandatory.

4.2. The warnings make it easier to assess GDPR principles such as data minimization, purpose limitation, or security measures.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Optional text-field for comments.
– Mandatory.

4.3. The GDPR compliance warnings and suggestions are useful in guiding your analysis.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Optional text-field for comments.
– Mandatory.

4.4. What additional compliance-related insights would you like to see?
– Free-text field.
– Optional.

B.5 Section 5: Usability and Interactivity

5.1. I would like to use Assessor View for my needs.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.2. I found Assessor View unnecessarily complex.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.3. I found Assessor View easy to use.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.4. I would need the support of a technical person to be able to use Assessor View.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.5. I found the various functions in Assessor View to be well integrated.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.6. I thought there was too much inconsistency in Assessor View.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.7. I would imagine that most people would learn to use Assessor View very quickly.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.8. I found Assessor View very cumbersome to use.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.9. I felt confident using Assessor View.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.10. I needed to learn a lot of things before I could get going with Assessor View.
– Likert scale from 1 (Strongly disagree) to 5 (Strongly agree).
– Mandatory.

5.11. How likely is it you would recommend Assessor View to a legal expert?
- Likert scale from 0 (Not at all) to 10 (Go for it!).
- Mandatory.

5.12. How likely is it you would recommend Assessor View to a privacy expert?
- Likert scale from 0 (Not at all) to 10 (Go for it!).
- Mandatory.

5.13. How likely is it you would recommend Assessor View to a DPO?
- Likert scale from 0 (Not at all) to 10 (Go for it!).
- Mandatory.

B.6 Section 6: Feedback and Recommendations

6.1. Do you think Assessor View would be the first step to reducing reliance on developers (for privacy experts)?

6.2. Do you think Assessor View can be used in privacy assessments or for conducting conversations about privacy assessments?

Appendix C Codebook

In this section, we present our codebook.

1. GDPR Compliance

1.1. Documents required

1.1.1. RoPA
1.1.2. Privacy notice
1.1.3. Privacy policy
1.1.4. Contracts

1.1.4.1. Service provider contract
1.1.4.2. Joint controller contract
1.1.4.3. Controller contract

1.1.5. Agreements

1.1.5.1. Service level agreement
1.1.5.2. Non-disclosure agreement
1.1.5.3. Data processing agreement
1.1.5.4. Works council agreement (Germany)

1.1.6. Assessment of partner activities
1.1.7. Rule of conduct
1.1.8. DPIA report
1.1.9. Consent
1.1.10. Report of data breach
1.1.11. Data retention policy
1.1.12. Data access request
1.1.13. Data protection by design assessment

1.2. Documentation support

1.2.1. Document generators
1.2.2. Decision tree-based legal tech tools
1.2.3. Standard templates
1.2.4. None

1.3. Compliance support

1.3.1. Cookie compliance tools
1.3.2. ChatGPT
1.3.3. Data classification tools
1.3.4. Source code analysis for security
1.3.5. Data protection management system
1.3.6. Privacy policy scanner
1.3.7. None

1.4. Training programs

1.4.1. Organization-specific base level compliance
1.4.2. Training programs for specific roles
1.4.3. Formal privacy certifications
1.4.4. Country-specific programs
1.4.5. Tabletop training for clients

2. DPIA (Article 35 GDPR)

2.1. Processes

2.1.1. Pre-DPIA Screening
2.1.2. Privacy by design
2.1.3. Data mapping
2.1.4. Information gathering
2.1.5. Identification of personal data processing
2.1.6. Stocktaking and recording in RoPA
2.1.7. Risk assessment
2.1.8. Discuss risk mitigation measures
2.1.9. Technical discussion
2.1.10. Interdisciplinary dialogue
2.1.11. Data transfer assessment
2.1.12. Document assessment
2.1.13. App review
2.1.14. Stakeholder consultation
2.1.15. Stakeholder approval
2.1.16. Regular monitoring and updates

2.2. Support

2.2.2. Charting tools
2.2.3. Automated DPIA summarization

2.2.4. Automated inventory of processing activities
2.2.5. Automated process mining
2.2.6. Data governance tools
2.2.7. Metadata management tools
2.2.8. Automated security control tools
2.2.9. Questionnaire-based DPIA tools
2.2.10. None

2.3. Roles

2.3.1. Documenter
2.3.2. Advisor
2.3.3. Organizer
2.3.4. Researcher
2.3.5. Assessee
2.3.6. Liaison

2.4. Stakeholders

2.4.1. Business functions

2.4.1.1. Management board
2.4.1.2. IT team
2.4.1.3. Data processors
2.4.1.4. Data controllers
2.4.1.5. Marketing team
2.4.1.6. Human Resources Department
2.4.1.7. Project sponsors
2.4.1.8. Works council (Germany)

2.4.2. Privacy functions

2.4.2.1. DPO
2.4.2.2. Privacy team
2.4.2.3. Legal counsel

2.4.3. Data subjects
2.4.4. Third parties

2.4.4.1. Business partners
2.4.4.2. Data protection supervisory authority

2.5. Vocabularies used

2.5.1. Organization-specific vocabulary
2.5.2. DPV
2.5.3. Law
2.5.4. Other standard vocabularies

2.6. Challenges

2.6.1. Knowledge gap

2.6.2. Communication gap
2.6.3. Delay in privacy considerations
2.6.4. Superficial legal recommendations
2.6.5. Absence of source code analysis
2.6.6. Lack of effort
2.6.7. Lack of awareness
2.6.8. Defining privacy roles
2.6.9. Privacy-innovation tradeoff
2.6.10. Lack of standard rules

References

[1] GDPR. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[2] Cyber Resilience Act. https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act

[3] GDPR Penalties. https://gdpr-info.eu/issues/fines-penalties/

[4] GDPR Article 13 (2018). https://gdpr-info.eu/art-13-gdpr/

[5] Yu, L., Luo, X., Liu, X., Zhang, T.: Can we trust the privacy policies of android apps? In: 2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 538–549 (2016). https://doi.org/10.1109/DSN.2016.55

[6] Zimmeck, S., Wang, Z., Zou, L., Iyengar, R., Liu, B., Schaub, F., Wilson, S., Sadeh, N., Bellovin, S., Reidenberg, J.: Automated Analysis of Privacy Requirements for Mobile Apps. Proceedings 2017 Network and Distributed System Security Symposium. Korea Society of Internet Information, Korea, Republic of Korea (2017). https://doi.org/10.14722/ndss.2017.23034

[7] Wang, X., Qin, X., Hosseini, M.B., Slavin, R., Breaux, T.D., Niu, J.: Guileak: Tracing privacy policy claims on user input data for android applications. In: Proceedings of the 40th International Conference on Software Engineering. ICSE '18, pp. 37–47. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3180155.3180196 . https://doi.org/10.1145/3180155.3180196

[8] Slavin, R., Wang, X., Hosseini, M.B., Hester, J., Krishnan, R., Bhatia, J., Breaux, T.D., Niu, J.: Toward a framework for detecting privacy policy violations in android application code. In: Proceedings of the 38th International Conference on Software Engineering. ICSE '16, pp. 25–36. Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2884781.2884855 . https://doi.org/10.1145/2884781.2884855

[9] Tan, Z., Song, W.: Ptpdroid: Detecting violated user privacy disclosures to third-parties of android apps. In: 2023 IEEE/ACM 45th IEEE International Conference on Software Engineering (2023). https://figshare.com/s/e6f5ff6b7478d571a9a9

[10] Data Safety Section. https://blog.google/products/google-play/data-safety/

[11] Khandelwal, R., Nayak, A., Chung, P., Fawaz, K.: Unpacking Privacy Labels: A Measurement and Developer Perspective on Google's Data Safety Section (2023)

[12] Castelly, N., Hurley, F.: Introducing Checks: Simplifying Privacy for App Developers. https://blog.google/technology/area-120/checks/

[13] Li, T., Cranor, L.F., Agarwal, Y., Hong, J.I.: Matcha: An ide plugin for creating accurate privacy nutrition labels. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **8**(1), 1–38 (2024) https://doi.org/10.1145/3643544

[14] Privado.ai (2022). https://www.privado.ai/data-safety-report

[15] GDPR Article 35 (2018). https://gdpr-info.eu/art-35-gdpr/

[16] Data Protection Impact Assessments. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/

[17] European Data Protection Board. https://www.edpb.europa.eu/edpb_en

[18] Khedkar, M., Mondal, A.K., Bodden, E.: Do android app developers accurately report collection of privacy-related data? In: Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering Workshops. ASEW '24, pp. 176–186. Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3691621.3694949 . https://doi.org/10.1145/3691621.3694949

[19] See No Evil: Loopholes in Google's Data Safety Labels Keep Companies in the Clear and Consumers in the Dark. https://foundation.mozilla.org/en/campaigns/googles-data-safety-labels/

[20] Girish, A., Reardon, J., Tapiador, J., Matic, S., Vallina-Rodriguez, N.: Your Signal, Their Data: An Empirical Privacy Analysis of Wireless-scanning SDKs in Android (2025). https://arxiv.org/abs/2503.15238

[21] Khedkar, M., Schlichtig, M., Bodden, E.: Advancing android privacy assessments with automation. In: Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering Workshops. ASEW '24, pp. 218–222. Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3691621.3694953 . https://doi.org/10.1145/3691621.3694953

[22] Pandit, H.J., Esteves, B., Krog, G.P., Ryan, P., Golpayegani, D., Flake, J.: Data Privacy Vocabulary – Version 2 (2024). https://arxiv.org/abs/2404.13426

[23] Khedkar, M., Schlichtig, M., Mohan, S., Bodden, E.: Visualizing Privacy-Relevant Data Flows in Android Applications (2025). https://arxiv.org/abs/2503.16640

[24] Sample DPIA Template. https://gdpr.eu/wp-content/uploads/2019/03/dpia-template-v1.pdf

[25] Cavoukian, A.: Privacy by design: The 7 foundational principles (2009). https://privacy.ucsc.edu/resources/privacy-by-design---foundational-principles.pdf

[26] Ferrara, P., Spoto, F.: Static analysis for gdpr compliance. (2018)

[27] Martin, Y.-S., Kung, A.: Methods and tools for gdpr compliance through privacy and data protection engineering. In: 2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 108–111 (2018). https://doi.org/10.1109/EuroSPW.2018.00021

[28] Li, Z.S., Werner, C., Ernst, N.: Continuous requirements: An example using gdpr. In: 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW), pp. 144–149 (2019). https://doi.org/10.1109/REW.2019.00031

[29] Arzt, S., Rasthofer, S., Fritz, C., Bodden, E., Bartel, A., Klein, J., Le Traon, Y., Octeau, D., McDaniel, P.: Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. In: Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation. PLDI '14, pp. 259–269. Association for Computing Machinery, New York, NY, USA (2014). https://doi.org/10.1145/2594291.2594299 . https://doi.org/10.1145/2594291.2594299

[30] Zhao, K., Zhan, X., Yu, L., Zhou, S., Zhou, H., Luo, X., Wang, H., Liu, Y.: Demystifying Privacy Policy of Third-Party Libraries in Mobile Apps. arXiv (2023). https://doi.org/10.48550/ARXIV.2301.12348 . https://arxiv.org/abs/2301.12348

[31] Privado open source scanning tool (2022). https://www.privado.ai/open-source

[32] Nachtigall, M., Schlichtig, M., Bodden, E.: A large-scale study of usability criteria addressed by static analysis tools. In: Proceedings of the 31st ACM SIGSOFT International Symposium on Software Testing and Analysis. ISSTA 2022, pp. 532–543. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3533767.3534374 . https://doi.org/10.1145/3533767.3534374

[33] Franke, L., Liang, H., Farzanehpour, S., Brantly, A., Davis, J.C., Brown, C.: An Exploratory Mixed-Methods Study on General Data Protection Regulation

(GDPR) Compliance in Open-Source Software (2024). https://arxiv.org/abs/2406.14724

[34] Tang, F., Østvold, B.M.: Assessing software privacy using the privacy flow-graph. In: Proceedings of the 1st International Workshop on Mining Software Repositories Applications for Privacy and Security. MSR4P&S 2022, pp. 7–15. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3549035.3561185 . https://doi.org/10.1145/3549035.3561185

[35] Khedkar, M., Bodden, E.: Toward an android static analysis approach for data protection. In: Proceedings of the IEEE/ACM 11th International Conference on Mobile Software Engineering and Systems. MOBILESoft '24, pp. 65–68. Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3647632.3651389 . https://doi.org/10.1145/3647632.3651389

[36] Pandit, H.J., Polleres, A., Bos, B., Brennan, R., Bruegger, B., Ekaputra, F.J., Fernández, J.D., Hamed, R.G., Kiesling, E., Lizar, M., Schlehahn, E., Steyskal, S., Wenning, R.: Creating a vocabulary for data privacy. In: On the Move to Meaningful Internet Systems: OTM 2019 Conferences: Confederated International Conferences: CoopIS, ODBASE, C&TC 2019, Rhodes, Greece, October 21–25, 2019, Proceedings, pp. 714–730. Springer, Berlin, Heidelberg (2019). https://doi.org/10.1007/978-3-030-33246-4_44 . https://doi.org/10.1007/978-3-030-33246-4_44

[37] Luo, L., Pauck, F., Piskachev, G., Benz, M., Pashchenko, I., Mory, M., Bodden, E., Hermann, B., Massacci, F.: Taintbench: Automatic real-world malware benchmarking of android taint analyses. Empirical Software Engineering (2021) https://doi.org/10.1007/s10664-021-10013-5

[38] Pauck, F., Wehrheim, H.: Jicer: Simplifying cooperative android app analysis tasks. In: 2021 IEEE 21st International Working Conference on Source Code Analysis and Manipulation (SCAM), pp. 187–197 (2021). https://doi.org/10.1109/SCAM52516.2021.00031

[39] Vallée-Rai, R., Co, P., Gagnon, E., Hendren, L., Lam, P., Sundaresan, V.: Soot: a java bytecode optimization framework. In: CASCON First Decade High Impact Papers. CASCON '10, pp. 214–224. IBM Corp., USA (2010). https://doi.org/10.1145/1925805.1925818 . https://doi.org/10.1145/1925805.1925818

[40] Xu, B., Qian, J., Zhang, X., Wu, Z., Chen, L.: A brief survey of program slicing. SIGSOFT Softw. Eng. Notes **30**(2), 1–36 (2005) https://doi.org/10.1145/1050849.1050865

[41] Google Play Console. https://play.google.com/console/signup

[42] GDPR Article 5 (2018). https://gdpr-info.eu/art-5-gdpr/

[43] GDPR Article 25 (2018). https://gdpr-info.eu/art-25-gdpr/

[44] D3.js Library (2024). https://d3js.org/

[45] DroidBench (2018). https://github.com/secure-software-engineering/DroidBench/tree/master

[46] Data Privacy Vocabularies and Controls Community Group. https://www.w3.org/community/dpvcg/

[47] Grier, R.A., Bangor, A., Kortum, P., Peres, S.C.: The system usability scale: Beyond standard usability testing. Proceedings of the Human Factors and Ergonomics Society Annual Meeting **57**(1), 187–191 (2013) https://doi.org/10.1177/1541931213571042 https://doi.org/10.1177/1541931213571042

[48] Reichheld, F.: The one number you need to grow. Harvard business review **81**, 46–54124 (2004)

[49] IAPP (2018). https://iapp.org/

[50] NoScribe (2025). https://github.com/kaixxx/noScribe

[51] Corbin, J., Strauss, A.: Grounded theory research: Procedures, canons and evaluative criteria. Zeitschrift für Soziologie **19**(6), 418–427 (1990) https://doi.org/10.1515/zfsoz-1990-0602

[52] Clarke, V., Braun, V.: In: Teo, T. (ed.) Thematic Analysis, pp. 1947–1952. Springer, New York, NY (2014). https://doi.org/10.1007/978-1-4614-5583-7_311 . https://doi.org/10.1007/978-1-4614-5583-7_311

[53] Records of Processing Activities. https://gdpr-info.eu/issues/records-of-processing-activities/

[54] CNIL PIA Tool. https://www.cnil.fr/en/PIA-privacy-impact-assessment-en

[55] CNIL. https://www.cnil.fr/en

[56] Caralegal (2025). https://caralegal.eu/en/

[57] Collibra (2025). https://www.collibra.com/

[58] GDPR Article 32 (2018). https://gdpr-info.eu/art-32-gdpr/

[59] SonarQube. https://www.sonarsource.com/sem/products/sonarqube/

[60] OneTrust. https://www.onetrust.com/

[61] Salesforce Privacy Center. https://www.salesforce.com/platform/

privacy-center/

[62] Schechter, S.E., Dhamija, R., Ozment, A., Fischer, I.: The emperor's new security indicators. In: 2007 IEEE Symposium on Security and Privacy (SP '07), pp. 51–65 (2007). https://doi.org/10.1109/SP.2007.35

[63] Guest, G., Bunce, A., Johnson, L.: How many interviews are enough?: An experiment with data saturation and variability. Field Methods **18**(1), 59–82 (2006) https://doi.org/10.1177/1525822X05279903 https://doi.org/10.1177/1525822X05279903