On Lawful Disclosure of Personal User Data: What Should App Developers Do?

Yung Shin Van Der Sype
Interdisciplinary Centre for Law and ICT
KU Leuven
Leuven, Belgium
yungshin.vandersype@law.kuleuven.be

Walid Maalej
Department of Informatics
University of Hamburg
Hamburg, Germany
maalej@informatik.uni-hamburg.de

***Abstract*—The proliferation of mobile devices and apps together with the increasing public interest in privacy and data protection matters necessitate a more careful precaution for legal compliance. As apps are becoming more popular, app developers can expect an increased scrutiny of privacy practices in the future. In this paper, we focus on the problem of the disclosure of personal data to third parties and the role of app developers to enhance user privacy and data protection in the app ecosystem. We discuss the EU data protection principles and apply them to the mobile app context. We then derive requirements and guidelines for app developers on how to contribute to the protection of their users' data.**

***Index Terms*—Legal Requirements, Mobile App Development, Privacy by Design, Anonymisation, Big Data, EU Data Protection Legislation**

I. Introduction

With the recent advance in data mining technologies and the massive increase in computing power and data storage capacity, data processing became notably faster and cheaper than before [1]. A large amount of the processed data is retrieved from apps that run on mobile devices. Hundreds of thousands of apps access and process the data of their users. These data often reveal large quantities of personal information, which are sometimes used for other purposes than initially intended. For instance, app owners (also called app vendors) might disclose personal user data to third parties in order to get insight into the behaviour of their users or to improve their apps [2]. The same data can be used in ways different from what was initially foreseen, sometimes without the awareness and the agreement of users [3].

According to the EU Data Protection Directive, personal information can only be processed in ways that are compatible with a specific purpose (Art. 6, 1), b Directive 95/46/EC) [4]. New techniques of using big data have put the notion of "compatible use" under pressure by extending and stretching its meaning. As this endangers the privacy of users, additional or alternative privacy-preserving paths should be explored. This is a complicated exercise because the protection of privacy and personal data needs a certain level of abstractness to be technology neutral and sufficiently comprehensive. Moreover, privacy and data protection are often victims of many cultural [5] and nationally inspired obstacles and pitfalls. Privacy is a domain that is characterised by diverse requirements and dynamic expectations with typically substantial differences amongst the EU Member States and beyond [5].

In this paper, we focus on the role of app developers. As apps are becoming more popular, app developers can expect an increased scrutiny of privacy practices in the future. App developers have a major influence on how data processing is undertaken and how user information is presented within the app [3]. However, to achieve a sufficient level of protection, active involvement of all relevant stakeholders in the software and data management lifecycle is required [3]. App developers often have to collaborate with other parties in the app ecosystem and the final responsibility towards the data subjects (i.e. the individuals under protection) remains an obligation of the data controller (Art. 2, d) Directive 95/46/EC). Nevertheless, we argue that privacy and data protection have to grow as a default mode of design and operation [6] and that app developers can bring an important contribution to the protection of user privacy by taking privacy-friendly decisions already in the early stages of the development process.

Technical research on privacy and data protection is often reduced to research on anonymisation techniques. However, anonymisation cannot be the only safeguard for privacy and data protection [7]. In Section II we elaborate further on the strengths and weaknesses of anonymisation techniques to assess whether they provide a sufficient level of privacy and data protection. In Section III we introduce the main legal principles of EU data protection law: the principle of purpose limitation, the principle of data minimisation, the principle of data security, and the principle of transparency. Afterwards, in Section IV we delve deeper into the meaning behind those principles in order to acquire more concrete guidelines and requirements for app developers. In Section V we discuss the limitations of our guidelines on three different levels. First, privacy will always have a certain level of subjectivity coloured by many different perspectives such as the cultural backgrounds and the technical affinity of the users. Second, the role of app developers in the overall privacy and data protection chain is sometimes limited, as data protection remains an obligation of the data controller. Third, our discourse is also burdened with geographical limitations as data protection is a national competence and enforcement is limited ratione loci. Finally, we summarise the related work in Section VI and conclude the paper in Section VII.

II. Anonymisation as a Solution for Privacy and Personal Data Protection?

The evolution in big data analytics creates more drivers to disclose personal information of users to third parties, e.g., to get more insight in the app performance or to personalise the app services [2]. However, European data protection law limits further use of personal data to what is compatible with the specified purpose of the initial use (Art. 6, 1), b Directive). It is likely that the disclosure of personal data to third parties is incompatible with the initial purpose of the processing operation and this might raise data protection concerns.

One way to avoid these concerns follows from Recital 26 of the Data Protection Directive, which excludes anonymised data from the scope of EU data protection legislation [4]. This means that when personal datasets are *sufficiently* anonymised, the data can be further processed for other purposes and the datasets disclosed to third parties. However, the Data Protection Directive lacks a prescriptive technical standard and does not state which anonymisation techniques are sufficiently de-identifying under EU law. Recital 26 means that to sufficiently anonymise personal data, the datasets must be irreversibly stripped from all personal elements (such that the data subject can no longer be identified) by **all** means which are likely reasonable to be used by either the data controller or by any other third party [8]. Since sufficient anonymisation of data falls outside the scope of the Data Protection Directive, we wonder to which extent the use of anonymisation techniques should be considered as a sufficient protection of personal data. If considered as a sufficient protection, the use of anonymisation techniques provides a way to further process data about users and to disclose this information to third parties without breaching EU data protection law [8] [9]. For instance, when traffic data are effectively anonymised prior to their disclosure any concerns regarding incompatible processing are removed, as there is no risk of re-identification [9].

Even though EU Directive 95/46/EC on the protection of personal data excludes irreversibly anonymised data from its applicable scope, it is still possible that the user's privacy rights (Art. 8 European Convention on Human Rights) are endangered by the disclosure, e.g., in case of profiling [8]. Privacy and data protection do not completely overlap. Despite some similarities, there are multiple differences from the formally legal and the substantive legal points of view [10]. This means that compliance with privacy law does not necessarily imply compliance with data protection law and vice versa. Certainly, anonymised data cannot completely escape the data protection obligation, as anonymised data have to be processed from personal data before they become anonymised [8]. Therefore, anonymisation techniques cannot completely deprive users of their data protection rights as personal data always enjoy the initial protection before the anonymisation can be considered irreversible. Only once the data are sufficiently anonymised, the data protection principles are no longer applicable.

A pertinent question remains unanswered: From which point on should personal data be considered irreversibly unidentifiable. In order to determine whether a dataset is sufficiently anonymised, the Article 29 Working Party (which is an independent working group in the European Union that deals with privacy and data protection matters) uses the criteria of "all", "likely" and "reasonable" [8]. All personal identifiers must be filtered from the datasets to make it unlikely that the data controller or any other third party would reasonably be able to re-identify the data subject on the basis of the data at stake.

In addition, the Working Party highlights three technical risks that endanger the robustness of a given anonymisation technique: singling out, linkability, and inference [8]. Those technical risks are typical for anonymisation and can be used as criteria to assess the legal robustness of the technique or practice. "Singling out" refers to the question whether it is still possible to isolate personal identifiers in a dataset. "Linkability" refers to the ability to link a record to a data subject or a group of data subjects, e.g., by means of correlation. "Inference" refers to the possibility to deduce the value of an attribute or a set of attributes, with significant probability.

The Article 29 Working Party sets out the legal strengths and weaknesses of several anonymisation practices in Opinion 05/2014 [8]. Without delving deeply into the details of different anonymisation techniques, we can summarise that, for all anonymisation techniques, the common mistakes and failures impede the irreversible and sufficient anonymisation of personal data [7]. For example, as each record will still be derived from a single data subject, randomisation will not reduce the singularity of each record [8]. Combined with generalisation techniques, a better privacy protection can be guaranteed. Generalisation techniques have their flaws though. While generalisation "can be effective to prevent singling out, it does not allow effective anonymisation in all cases; in particular, it requires specific and sophisticated quantitative approaches to prevent linkability and inference" [8].

Besides the technical hindrances for robust data anonymisation, there are also more practical concerns. In particular, for the disclosure of personal data collected from mobile apps, irreversible anonymisation of datasets might decrease the value of the data for the third party to whom the data are disclosed to the extent that the disclosure itself would become irrelevant for the third party. The disclosure of irreversibly anonymised data would no longer suffice the purpose of the disclosure. For instance, the third party might need raw data or non-anonymised data to provide the app owner with the insights requested or a certain level of identification might be required to have value for marketing purposes.

The sole reliance on anonymisation techniques for data protection is hardly sufficient for now [7]. In the future, anonymisation techniques might provide a more solid solution for data protection once they become more robust against singling out, linkability, and inference [8]. Until then, other measures should be implemented. Thus the following question remains open: How can developers contribute to the protection of user data from two perspectives: (1) on the way to anonymisation as long as anonymisation is used as a personal data processing operation, and (2) when the anonymisation

technique is considered insufficiently irreversible and personal user data are still being processed.

III. Legal Principles of Data Protection

Even though the national implementation of the Data Protection Directive allows a certain differentiation amongst EU Member States, the Directive draws the general conditions for fair and lawful processing of personal information [4]. Those conditions are summarised below in four main principles: the principle of purpose limitation (Art. 6, 1), b Directive), the principle of data minimisation (Art. 6, 1), c-e Directive), the principle of data security (Art. 17 Directive) and the principle of transparency (Art. 10,- Directive).

A. Purpose Limitation

The principle of purpose limitation plays an important role in the protection of users against disclosure of personal data to third parties. It explicitly limits the secondary use of personal data for purposes incompatible with the specified purpose of the initial processing. The purpose limitation principle enables users to make deliberate choices to trust apps with their personal data. They will learn how their data will be used and they will be able to rely on the limitative purpose description to understand the purposes for which their data will be used [3].

The principle of purpose limitation has two aspects: purpose specification (specified, explicit, and legitimate purpose) and compatible use (no further processing of data in an incompatible way) [9]. The "reasons for the collection, use, and disclosure of personally identifiable information should be identified to the data subject at or before the time of data collection. Personal information cannot be used or disclosed for purposes other than those for which it was collected except with the consent of the individual or as authorised by law" [11]. For example, an app that is installed to enable e-mail communication between users cannot automatically link the e-mail addresses of the users to their telephone numbers or disclose this information to third parties without the renewed consent of the app users.

B. Data Minimisation

Even when the processing of personal data complies with a legitimate purpose, the processing must be necessary and proportionate. The principle of data minimisation requires that personal information can only be used or disclosed in order to achieve the specified purpose of the processing (which technically means collection, storage, or analysis). Thus, data can only be processed, if the processing is adequate, relevant, and not excessive in relation to the specific purpose for which the data was collected (Art. 6, 1) c Directive). This principle also implies that the processing is to be carried out in the least-intrusive way, considering, e.g., the risks at stake, the amount of data involved, and the purpose of processing. When the information is no longer accurate or necessary, it should be erased. For instance, imagine an app that identifies the title of the song the user is listening to. For this, the audio recording feature of the mobile device must be enabled. However, any monitoring or recording of audio at a moment not specifically asked for by the user is likely to be considered excessive and unlawful [3].

C. Data Security

The third principle relates to the security of the data. As malware and viruses for mobile devices are on the rise, data security measures should avoid the loss of large amounts of personal information processed on mobile devices [12]. One particular risk is the loss or the theft of the mobile device and the resulting illegitimate access to the data collected by the app.

Data security implies that "confidentiality, integrity, and availability should be safeguarded as appropriate to the sensitivity of the information" [9]. This implies the implementation of technical measures to limit data loss and data breaches. For this principle many guidelines, specifications, and best practices are available, for instance those of the ENISA [13].

D. Transparency

The last principle concerns the transparency of the processing and the empowerment of the user. Transparency means that users must be informed about the reasons behind the processing, the categories of data being processed about them, and the "destiny" of their data. User participation means: "Individuals should be empowered to play a participatory role in the lifecycle of their own personal data and should be made aware of the practices associated with its use and disclosure" [9]. This includes also concerns about the meaningful consent.

In order to be meaningful, consent must be given prior to a specific processing, based on a free and informed choice to accept or refuse the processing of personal data. For instance, for an app that provides information about nearby social activities, consent must be asked prior to the installation of the app and separately again each time the geo-location of the user is accessed [3]. This means also that the location of the user cannot be used for purposes incompatible with the goal to provide the user with information about nearby social activities. The location of the user cannot be transferred to third parties without an additional consent of the user.

IV. From Legal Principles to Technical Requirements

In this section we derive concrete recommendations for app developers to protect user privacy and to minimise the risks of an unlawful disclosure of personal data to third parties.

A. Purpose Limitation

Article 6, 1), b of the Data Protection Directive requires that personal data are only collected for specified, explicit, and legitimate purposes.

App developers should describe the purpose of their apps and the data processing operations carried out by their apps in a well-defined and comprehensible way for an average user without a legal or technical knowledge to understand [3]. For instance, the purpose might be described in the app store or in a special view in the app itself. A good practice is also to list the features of the app in the specific app page in the app store. In this page, developers can give information about the data collection and processing performed by the app as well as the

purpose of the collection and processing. **The purpose should be sufficiently specified** to enable the implementation of necessary data protection safeguards and the scope of the processing operation delimited [9]. The app store categorisation can be a first indicator for the description of the purpose and should therefore be carefully selected. A gaming app is likely to process personal information for the purposes of keeping competition records, while a restaurant finder might process personal information for the purpose of locating nearby restaurants. In case the processing of personal information by the app is ambiguous, more details on the purpose are required [9]. This means that if the purpose of the processing is clear within a certain context, usually less details are required. The same holds in case the processing of data goes beyond what is customary in a given context [9]. **The purpose should also be explicit**, meaning that it has to be clearly expressed, so that the user can understand the reasons behind the processing. Finally, the **purpose should be legitimate** and compatible with all legal principles of applicable law, such as non-discrimination principles. For instance, racial profiling where "white" customers pay higher prices than "Asian" customers is an illegitimate purpose [9].

Further processing of personal information cannot be incompatible with the purposes for which the personal data were collected and initially specified and must thus meet the requirement of compatibility (Art. 6, 1) b Directive). For instance, it is not allowed to transfer personal information of users to third parties for market research without the consent of the users. This means that such disclosure of data is not automatically allowed and that **additional user consent** is required. App developers might prevent incompatible use without prior renewed consent.

The Article 29 Working Party lists some key factors to be considered during the compatibility assessment [9]. As most obvious the Working Party points to the relationship between the purposes for which the data have been collected and the purposes of further processing. This refers to the substance of the processing and covers mainly situations where the further processing was already more or less implied in, or assumed as, a logical next step of the initial purpose. For example, a department for transport asked a telecommunications company to provide mobile phone location data in order to calculate the speed at which the phones are moving over various routes. The department wanted to take traffic-calming measures based on this information. However, it is unlikely that the purposes related to road traffic are compatible with the telecom information initially collected [9]. Other key factors are the nature of the data (e.g., content vs. metadata vs. interaction data vs. sensor data), the impact of the further processing on the data subjects, the measures applied by the controller to ensure fair processing, and the measures taken to prevent any undue impact on the data subjects.

B. Data Minimisation

Mobile devices typically have direct access to different sensors and data. Apps often abundantly collect personal information from smartphones without any meaningful or obvious relationship to the functionality of the app [3]. The challenge for app developers is to keep the collection, storage, and use of personal data to a minimum from the early stages of the development process. The collection and the use of **personal data should be limited to what is reasonably necessary to achieve the desired functionality**. In case specific data will be needed for a feature that is planned for future releases, we recommend either to explicitly state this or to handle the data collection and processing at the design time but not at the run time. Once the feature is implemented and the data collection and processing is "instantiated" at run time, a new consent must be collected from the user. The first step is to find out whether the app needs to process personal information at all. This means that app developers should carefully consider the relevancy of the personal data they tend to process. Collecting information that is out of scope of the app purpose might also discourage potential users [14].

Studies found that users care about the relevance of the information requested by the app. If they perceive a potential privacy violation, they would opt-out of the app rather than to change their privacy settings [14]. Also in case the processing of personal information is truly necessary, the processing of the data should be kept to the minimum for the app to function. One way to assess whether the processing is necessary is by **defining the reasons why certain information is related to the features of the app**. If this is not possible, it is likely that the information is not necessary for the app to function.

To define the reasons behind certain information, it is helpful to **categorise different types of data used in the app**. From a legal point of view a distinction can be made between non-personal information, personal information, and sensitive personal information [4]. Personal information includes all information that relates to a person such as the name, the birth date, or the credit card number. Sensitive personal data are the data "revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life" (Art. 8, 1) Directive). For instance, a health app might need sensitive health data about the user whereas it is questionable whether a music app needs such kind of sensitive data. The data minimisation principle also applies for non-sensitive information. The collection and use of data cannot go beyond what is necessary for the goals of the app.

To keep the processing of data to a minimum, the amount of information should be limited to what is suitable to achieve the processing objective. Excessive or disproportionate processing of personal data is not allowed and cannot be legitimised by user consent. For instance, in order for an app to collect audio data or to activate the camera, a specific permission must be asked from the user. However, for an optional feature of an alarm clock that enables users to give verbal commands to silence or snooze the alarm, voice recording that goes beyond this purpose (e.g., audio recording whilst the alarm is silent) is likely to be excessive and unlawful. For a music app to pause music and receive incoming calls, a permission to monitor the calls must be asked. But even when granted, this permission only applies for the time when

the music app is running. A continuous monitoring of all call activities at all times is likely unlawful and excessive.

Besides the legal categorisation, there are various technical categorisations of information. Those might be based on a distinction between the available types of data, implying a distinction between technical data, on-device data, user entered data, and cookies [15]. However, for the protection of personal data such technical functionality-based data categories cannot be addressed at once. Some technical data, such as the device name, might include personal information about users. In some cases, technical information, such as IP addresses, might even inherently include personal information about a user. Although the technical distinction between several data categories might have a functional advantage, in order to sufficiently protect personal data it is recommended to **categorise the data based on the legal categories of personal data instead of on technical categories** as the criterion for applicability of the principles cannot be generalised to very general categories but should be specified for each data element separately.

Data minimisation during storage can be achieved by **a limitation of the time for which the gathered data is kept**. Even after deletion requests, service providers seem to be very reluctant to delete the personal data. Service providers prefer to make the data inaccessible rather than physically purging the data from storage [16]. Legally speaking, data is only deleted once all identifiers are filtered (sufficient anonymisation) or when the data are erased from the databases. Even without a user's request to erase certain data, **personal data cannot be kept for longer than necessary**.

Determining for how much time the data can be kept is difficult. First, the **storage of data might be hard limited** or obliged for certain purposes and for a certain time. For instance, the discussed and rejected Data Retention Directive prescribed the storage of some traffic data for 6 months for law enforcement purposes [17]. However, the Court of Justice of the European Union recently declared the Data Retention Directive invalid with immediate effect, as it constituted a disproportionate restriction on the fundamental rights to privacy, data protection, and freedom of expression [18]. Another hard limit is the deletion of an account. When a user deletes his account, the personal information should be promptly erased immediately after.

In case no hard limit is applicable, the next step is to **define a reasonable retention period** to keep the personal information about the users. For example, the retention period for a calendar app is controlled by the user. Once the user erases the data, there is no need for the app to keep the information for a longer time. For a navigation app on the other hand, it might be necessary to store the locations visited in the past. This could be limited to a reasonable number of visits or by a certain timeframe in the past. In order to efficiently enforce the hard and soft retention limits, app developers should **develop procedures to delete the no longer needed personal data automatically**.

C. Data Security

Poor security measures for storage and transfer of user data cause additional risks for the protection of the users' personal data. As the value of personal data is raising drivers for malicious intrusions is raising as well. Insufficient information security often leads to unauthorised access or unauthorised processing of personal information. For instance, imagine an app that advices women about their menstrual cycle. If this app does not sufficiently secure the communication of data between the client and the server, and a woman requests a back-up, her sensitive health data are transferred to the server in an unencrypted form, in a way readable and accessible by other users of a shared Wi-Fi network [19]. **Information about users should thus be protected** during storage and during communication between the app components. There are multiple ways to enhance personal data security. For example, access to user information stored in the server should be restricted to a "need to know" basis in the organisation of the app owner. A user authentication should be required for the use of the app on the device.

The use of encryption methods also limits the risks of misusing personal data. **Encryption should be used by default** [11]. All files and all communication transactions should be encrypted. For instance, for the transmission of data, it is recommended to use transport layer encryption (HTTPS) by default [16]. All network communication should be encrypted, e.g., with SSL or stronger methods. For sensitive personal data even stronger encryption methods should be used [16]. Finally, user information should never be transmitted in clear text and passwords should never be stored in clear text.

Data security is an obligation under EU law, without being sufficient for the protection of personal information. It is thus important to note that the encryption of data is not equal to the anonymisation of data. Even when decently hashed and secured, personal data remain personal as long as they are not stripped from all personal identifiers and re-identification remains possible. The distinction between anonymised data and secured data is highly relevant as Recital 26 of the Data Protection Directive only excludes anonymised data from the scope of the data protection legislation [4]. This means that the data protection principles are fully applicable in case of insufficient anonymisation.

D. Transparency

Users must be aware of the processing of their data, including the disclosure of their data to third parties. However, users are often left in the dark. In 2012 only 61.3% of the top 150 apps provided a privacy policy that sufficiently explained to the users which data were collected through the app and for which purposes [20]. Even when apps have a privacy policy, sometimes the text is unclear or incomplete. For instance, some applications mention in their policies that no personal data are processed, while actually they do process information about users, e.g., the unique numbers of the users or their devices [21]. One problem that results from this lack of transparency is that users are not fully aware of the processing operations. Therefore they cannot grant their meaningful consent. Since the cognitive limits of users make consent almost a legal fiction large attention must be paid to a transparent communication [22]. Transparency is thus of great importance to obtain the meaningful consent of the user. Other

challenges should be mentioned in particular for mobile apps. For instance, the limited screen size makes the presentation of additional information to users harder. The overload of pop-ups and notifications causes a notice fatigue. Finally, the lack of a designated area in app stores or in the screen designs to inform users or a common way to communicate the necessary information, complicates the process to obtain user consent.

App developers should ensure that their **design does not conceal an actual or a potential disclosure of information** [23]. In addition, app developers should ensure that the system remains reasonably transparent and subject to **independent verification** by all relevant stakeholders, including customers, users, and affiliated organisations. When personal information is used for third party purposes such as advertising and analytics or when sensitive personal information about a user is collected, an enhanced notice is recommended. Finally, in order to provide the user with a comprehensive description of disclosure and transfer practices, app developers **must understand the third-party components and code** they are using to ensure that no more information is collected than communicated.

Privacy policies must be easy accessible, easy to find, concise, and precise. Long and complicated policies should be avoided to reduce notice fatigue and information overload, but the message must be sufficiently clear and comprehensive. It is recommended to **place a link in a prominent space pointing the user to the privacy policy and the terms of use** not only before the app is installed or downloaded but also while the app is being used. The policies themselves should be **easy to read and browse** on the device. They should be clear and written in a simple language appropriate to the audience, taking into account the respective age, special needs, nationality, and culture of the individuals targeted. For instance, a government website with advice for elderly is not the same as a gaming website aimed at teenagers.

In order to keep the policy simple, different types of privacy practices can be displayed in a grid or by a "nutrition label for privacy". However, over-simplification should be avoided since sufficient granularity is needed ensure that all purposes are clear for the users [24]. By **using a layered notice**, a policy can highlight its most relevant parts regarding privacy, while providing links to the details. The Article 29 Working Party considers the layered notice often as a "workable way to provide key information to data subjects in a very concise and user-friendly manner, while also supplying additional information on the next 'layer' for the benefit of those who require further clarification" [24].

Another way of drawing attention to the privacy policy is by using media, such as graphics, icons, labels, images, colours, and sounds. For instance, a selective use of sound could alert users when privacy crucial operations occur. Once the purpose of the processing is changing or for every additional processing operation that is not compatible with the initial use, an update in the policy must describe the changes or the additional use. Silent or secret updates must be avoided. **Renewed consent should be asked before the changes are enforced**. The effective date of the policy and the process of modifications should be visible on these updates. It is recommended to **provide a brief statement with a hyperlink to the old policy.** It is also a good practice to present the privacy policy in a browser and in an extra view within the app.

In addition, users must be able to exercise their rights of access, rectification, erasure, and to object to data processing. They also must be informed about the existence of those features. App developers should provide the **tools to report problems** and exercise the data subject rights.

Whenever necessary a freely given, informed, and specific consent must be obtained from the users before the app is installed. However, a simple "Yes" or "No" consent does no longer fulfil the complexity of privacy implications. We therefore encourage **a more active participation of the user** when giving their consent [16], e.g., in a multi-step process or reactive user interface. In any case, users must have the opportunity to choose differently and to change their choices at any time in the future. For this, app developers might provide a privacy dashboard that **displays the actions users can take to modify their privacy settings** including **activation/ deactivation buttons** for processing operations if suitable.

When reasonably possible, it is recommended to enable users a granular consent for each type of data the app will access: minimum for location, contacts, unique device identifier, identity of the data subject, identity of the phone, credit card, and payment data, telephony and SMS, browsing history, email, social networks credentials and biometrics [24]. In addition users should be able to revoke their consent, uninstall apps, and delete the data. For instance, a restaurants app using location information should seek consent before the app is installed and every time when geo-location data are accessed separately.

V. Discussion

A. Tension Between Law and Technology

User expectations even left aside, a large gap exists between legal and technological perceptions of privacy. The relationship between law and technology is highly complex. Law is inherently abstract, while requirements should be concrete. While law leaves room for interpretations in a specific context, requirements typically anticipate all technical scenarios supported by the system. Requirements go beyond the application of legal rules in a specific and known context. They should capture other potential contexts as well.

In addition, law has a typical *a posteriori* character, while technology has typically an *a priori* character. The *a posteriori* character of law is twofold. First, law is developed after real-life problems showed the inability of existing law to them. Second, the goal of law is to create and maintain legal certainty, and to be enforced in case an imbalance occurred, which is usually the case after an issue happens. As a consequence, legal assessments of reality are typically event and impact driven. This means that the legal evaluation of reality is based on the impact of a certain commission (or omission) on other parties (users), rather than on the goal to shape and establish innovative changes to reality. Technology, on the other hand, is functionality or goal driven. This implies

that technical innovation is evaluated during the development of the technology, before its impact is known. Lawyers and developers are often driven by different types of questions. Lawyers ask questions such as, "Did an interference occur?" "Was there a breach?" or "Can this breach be justified?" In contrast, developers ask questions such as, "What is missing" and "How to overcome this?" Legal questions are asked after an event occurred, while technical questions are asked before.

B. Role of the App Developers

Privacy-unaware app-developers may "create significant risks to the private life and reputation of users of smart devices" [3]. By taking privacy-friendly design decisions in the early stages of the app design, developers significantly increase the protection of personal data in later stages of data management lifecycle.

The importance of privacy by design is formally recognised by Article 23 of the proposed Data Protection Regulation. The familiar EU data protection landscape with the central Data Protection Directive is changing. The proposed data reform package consists of two legislative documents aiming at a more comprehensive and coherent policy on the fundamental right to personal data in the European Union. The first document, a Regulation, should harmonise the general data protection in Europe in general [25] [26]. The second, a Directive, covers the data protection issues in the context of law enforcement [27]. Article 23 of the proposed Data Protection Regulation introduces the principle of privacy by design as a formal and enforceable principle into EU data protection law [25]. It is the crowning of the momentum behind privacy by design that has been steadily growing for years [28]. Article 23 translates the idea of privacy as an outset instead of an afterthought [11]. It presents privacy by design as the mitigation of privacy concerns and the achievement of data protection compliance by embedding privacy-requirements and privacy-preserving solutions in the engineering of products and services [6]. Privacy should be embraced from within the system design to become an integral part of the system without constraining the system functionality [11]. Privacy should grow to a default mode of design and operation in which all relevant actors in the lifecycle of personal data are involved.

Even though the technical implementation of privacy by design is essentially the task of the app developer, the responsibility of the quality of compliance still remains with the data controller. Under European law, the data controller is the person who determines the goals and the means of the processing operation (Art. 2, d) Data Protection Directive). In some cases this might be the developer, but it can also be another legal person. Article 23 is presently not addressed to developers. Nonetheless, it is undisputed that the principles of privacy by design are of great importance for developers as controllers are bound by them and accountable for compliance. Contractual obligations are likely to create incentives for the market of relevant goods and services [28].

The process of translating and implementing privacy principles into system requirements is often called privacy engineering and can be considered as "the gathering and application of privacy requirements with the same primacy as other traditional feature or process requirements and then incorporating, prioritising, and addressing them at each stage of the development lifecycle, whether it's for a process, project, product, system, application, or other. [...] The intent of privacy engineering is to close the gap between privacy policy and the reality of systems or technologies or processes. The greater the mismatch between the two, the greater the opportunity for needless inefficiencies, risk, or both" [29]. The implementation of privacy by design principles into software apps is a complicated exercise. The heterogeneous nature of distinct social, legal, and ethical concerns around privacy complicates the concrete translation in terms of system requirements [30] [23]. The role of the developer can be clarified by an important distinction Solove made to describe the notion of information privacy [31]. He distinguishes the control and immediate access to oneself from reducing the risk that personal information might be used in an unwanted way. This distinction between access control and risk management has been seen to suggest "two distinct dimensions to building privacy-friendly technologies and information systems" [32]. In this regard, developers are responsible for ensuring users immediate control over access to their personal information, and also for minimising future privacy risks by the protection of the information about the users. Another approach is to choose the best way to meet user expectations. Clarity on the expectations of users should help to determine which domains of risk are the most important to tackle. However, even though privacy is considered as a main issue for modern software systems, there is disagreement concerning the concrete importance of the different concerns and concrete ways to address them [5]. Sheth et al. found that users were more concerned about data aggregation and data distortion than developers and that no consensus on the best technique could be found [5]. Although the importance of privacy is commonly accepted, "the measures to reduce privacy concerns are divergent" [5].

C. Geographical Limitations (Ratione Loci)

A final remark on the relativity of our work follows from the geographical limitations of the legal provisions on data protection. In this paper we spoke about EU principles of data protection law. Yet, some considerations should be made in this regard. On the one hand, we mention that the current Data Protection Directive leaves a wide discretionary power at national level. As a result, national implementations of EU data protection law differ largely amongst the EU Member States. Even though the proposed EU Data Protection Regulation aims to harmonise the data protection law of the different EU Member States, multiple differences will remain [25]. The Regulation allows national specifications for a limited list of data processing types. For example, Article 81 of the Proposed Regulation allows certain specifications for processing operations in the employment context. On the other hand, we mention that the scope of European data protection legislation is always limited to the processing of personal data with a link to the European Union. Article 3 of the proposed Data Protection Regulation states that the proposal applies to "the processing of personal data in the context of the activities of an

establishment of a controller or a processor in the Union" or when the processing concerns "data subjects residing in the Union by a controller not established in the Union, where the processing activities are related to: (a) the offering of goods or services to such data subjects in the Union; or (b) the monitoring of their behaviour". Thus, although the territorial scope of the proposed Regulation is very widely formulated, it cannot go beyond European sovereignty and therefore, processing of personal data with no connection to the Union falls outside the scope of the EU data protection law.

VI. Related Work

In addition to the legal perspective, much research on privacy and data protection has been done in different research communities. We summarise the important related work focussing on usability and economic aspects of privacy, anonymisation techniques, and work from the software and requirements engineering community.

Many recent studies on online social networks show that there is a (typically, large) discrepancy between users' intentions for what their privacy settings should be versus what they actually are. For example, Madejski et al. report in their study of Facebook that 94% of their participants were sharing something they intended to hide and 85% were hiding something that they intended to share [33]. Liu found that Facebook's users' privacy settings match their expectations only 37% of the time [34]. A recent longitudinal study by Stutzman et al. shows how privacy settings for Facebook users have evolved over a period of time [35]. These studies have focused on privacy settings in a specific online system, while our goal is to derive common guidelines for designing any privacy-ware app. Further, the main contribution of these studies is to show that there is a discrepancy between what the settings are and what they should be and how settings evolve over time.

Fang and LeFevre proposed an automated technique for configuring a user's privacy settings in online social networking sites [36]. Paul proposed to use a colour- coding scheme for making privacy settings more usable [37]. Squicciarini, Shehab, and Paci propose a game-theoretic approach for collaborative sharing and control of images in a social network [38]. Toubiana presents a system that automatically applies users' privacy settings for photo tagging [39]. All these papers propose new approaches to make privacy settings "better" from a user's perspective (i.e. more usable and more visible). Our work is complementary abd help development teams decide when and which of these techniques should be implemented together with other privacy features. We also focus more on a broader legal requirements perspective of privacy than on a specific technical perspective.

There has been a lot of recent work on the economic ramifications of privacy. For example, Acquisti conducted a number of field and online experiments to investigate the economic valuations of privacy [40]. There has also been a lot of work about data anonymisation and building accurate data models for statistical use [41] [42] [43] [44]. These techniques aim to preserve certain properties of the data (e.g., statistical properties like average) so they can be useful in data mining while trying to preserve privacy of individual records. Similarly, there has also been work on anonymising social networks [45] and anonymising user profiles for personalised web search [46]. The broad approaches include aggregating data to a higher level of granularity or adding noise and random perturbations.

There has been research on breaking the anonymity of data as well. Narayanan and Shmatikov show how it is possible to correlate public IMDb data with private anonymised Netflix movie rating data resulting in the potential identification of the anonymised individuals [47]. Backstrom [48] and Wondracek [49] describe a series of attacks for de-anonymizing social networks. Also in the software engineering community, recent papers on privacy mainly focused on data anonymisation techniques. Clause and Orso propose techniques for the automated anonymisation of field data for software testing [50]. They extend the work done by Castro using novel concepts of path condition relaxation and breakable input conditions resulting in improving the effectiveness of input anonymisation [51].

Taneja and Grechanik propose using k-anonymity [52] for privacy by selectively anonymising certain attributes of a database for software testing [53]. They propose novel approaches using static analysis for selecting which attributes to anonymise so that test coverage remains high. There have been some recent papers on extracting privacy requirements from privacy regulations and laws [54] [55].

A few recent papers have also discussed privacy requirements, mainly in the context of mobile applications. Mancini conducted a field study to evaluate the impact of privacy and location tracking on social relationships [56]. Tun et al. introduce a novel approach called "privacy arguments" and use it to represent and analyse privacy requirements in mobile applications [57]. Omoronyia et al. propose an adaptive framework using privacy aware requirements, which will satisfy runtime privacy properties [58]. Finally, authors in the software engineering and requirements engineering communities mention privacy in the discussion or challenges section of their papers [59]. But in most cases, there is little evidence about what, how, and in which context privacy concerns exist and what the best measures for addressing them are [5].

VII. Conclusion

There are hundreds of thousands of different apps that access and process personal user data. Complex advancements in data mining technology, together with a massive increase in computing power and data storage capacity made it more attractive for app owners to disclose the data of their users to third parties for multiple reasons. This situation has put the data protection principle of "purpose limitation" under pressure.

In this paper, we argued that app developers could play an important role in the protection of the personal data of app users when data are disclosed to third parties. State of the art anonymisation techniques can hardly provide sufficient solutions for the protection of personal data. We discussed the

four main EU data protection principles. These are: purpose limitation, data minimisation, data security, and transparency. We then interpreted the principles from the perspective of app developers and derived technical recommendations and requirements, which should be considered when developing privacy-aware apps. Important limitations of our work include the "legally undefined" role of app developers and the geographical limitations of EU data protection law. Finally, app developers should also consider additional perspectives when implementing privacy-aware apps such as the economic perspective, the cultural differences between users and the technical affinities of the users.

ACKNOWLEDGMENT

The research leading to these results has received funding from the EU IST Seventh Framework Programme (FP7) under the grant agreement number 318508, project MUSES (Multiplatform Usable Endpoint Security).

REFERENCES

[1] O. Tene, J. Polonetsky, "Privacy in the Age of Big Data: a Time for Big Decisions," Stanford Law Review, vol. 64, pp. 63-69, 2012.

[2] P. Olson, "Meet The Company That Tracks More Phones Than Google Or Facebook," 30 10 2013. [Online]. Available: http://www.forbes.com/sites/parmyolson/2013/10/30/meet-the-company-that-tracks-more-phones-than-google-or-facebook/. [Accessed 1 6 2014].

[3] Article 29 Working Party, "Opinion 02/2013 on apps on smart devices," 2013.

[4] EU Parliament and EU Council, Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OC L 281/31, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:pdf, 1995.

[5] S. Sheth, G. Kaiser, and W. Maalej, "Us and them - A Study of Privacy Requirements Across North America, Asia, and Europe," Columbia University Computer Science Technical Reports, New York, 2013.

[6] A. Cavoukian (Information and Privacy Commissioner of Ontario), Privacy Engineering: Proactively Embedding Privacy, by Design, Canada, 2014.

[7] P. Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization," UCLA Law Review, vol. 57, pp. 1701-1777, 2010.

[8] Article 29 Working Party, "Opinion 05/2014 on Anonymisation Techniques, WP216," 2014.

[9] Article 29 Working Party, "Opinion 03/2013 on purpose limitation, WP203," 2013.

[10] J. Kokott and C. Sobotta, "The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR," International Data Privacy Law, vol. 3, no. 4, 2013.

[11] A. Cavoukian (Information and Privacy Commissioner of Ontario), Privacy and Security by Design: a Convergence of Paradigms, Canada, 2013.

[12] Y. Altshuler, N. Aharony, A. Pentland, et al., "Stealing Reality: when Criminals become Data Scientists (or Vice Versa)," IEEE Intelligent Systems Magazine, vol. 26, no. 6, pp. 22-30, 2011.

[13] ENISA, "Smartphone Secure Development Guidelines," file:///C:/Users/u0085676/Downloads/Secure_Dev_Guidelines.pdf, 2011.

[14] N. Wang, H. Xu, P. Wisniewski, and J. Grossklags, Designing the Default Privacy Settings for Facebook Applications, Baltimore, US, 2014.

[15] E. Yildirim, Mobile Privacy: Is There An App For That?, Amsterdam, 2012.

[16] A. Stopczynski, R. Pietri, A. Pentland, et al., "Privacy in Sensor-Driven Human Data Collection: A Guide for Practicioners," http://arxiv.org/pdf/1403.5299v1.pdf, 2014.

[17] EU Parliament and EU Council, Directive 2006/24/EC on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC, 2006.

[18] European Court of Justice, Digital Rights Ireland, C-293/12 and C-594/12, 2014.

[19] D. Tokmetzis, "Dit gebeurt er allemaal onder de motorkap van je smartphone," 21 5 2014. [Online]. Available: https://decorrespondent.nl/1034/dit-gebeurt-er-allemaal-onder-de-motorkap-van-je-smartphone/50352698-785491ea. [Accessed 23 5 2014].

[20] Future of Privacy Organisation, "Mobile Apps Study," http://www.futureofprivacy.org/mobile-apps-study/, 2012.

[21] S. Thurm, Y.I. Kane, "Your Apps Are Watching You," 17 December 2010. [Online]. Available: http://online.wsj.com/news/articles/SB10001424052748704694004576020083703574602?mg=reno64-wsj&url=http%3A%2F%2Fonline.wsj.com%2Farticle%2FSB10001424052748704694004576020083703574602.html. [Accessed 25 5 2014].

[22] D. Solove, "Privacy Self-Management and the Consent Dilemma," Harvard Law Review, vol. 126, pp. 1879-2139, 2013.

[23] S. Lederer, J.I. Hong, A.K. Dey, and J.A. Landay, "Personal Privacy through Understanding and Action: Five Pitfalls for Designers," Personal and Ubiquitous Computing, vol. 8, no. 6, pp. 440-454, 2004.

[24] Article 29 Working Party, "Opinion 15/2011 on Consent, WP187," 2011.

[25] EU Parliament and EU Council, Proposal for a Regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data COM(2012) 11 final, 2012/0011 (COD), http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf, 2012.

[26] EU Parliament, Draft European Parliament Legislative Resolution on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with regard to the processing of Personal Data and on the Free Movement of Such Data, Brussels, 2014.

[27] EU Parliament and EU Council, Proposal for a Directive on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of [...], http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0010:FIN:EN:PDF, 2012.

[28] P. Hustinx, "Opinion on the Data Protection Reform Package," Brussels, 2012.

[29] M. F. Dennedy, J. Fox, and T. R. Finneran, The Privacy Engineer's Manifesto, Apress, 2014.

[30] S. Gürses, C. Troncoso and C. Diaz, "Engineering Privacy by Design," in Conference on Computers, Privacy and Data Protection, Belgium, 2011.

[31] D. Solove, "A Taxonomy of Privacy," University of Pennsylvania Law Review, vol. 154, 2005.

[32] S. Spiekermann, and L.F. Cranor, "Engineering Privacy," IEEE Transactions on Software Engineering, vol. 35, no. 1, 2009.

[33] M. Madejski, M. Johson, and S.M. Bellovin, "A Study of Privacy Settings errors in an Online Social Network," in IEEE Pervasice Computing and Communication, 2012.

[34] Y. Liu, K.P. Gummadi, B. Krishnamurthy, and A. Mislove, "Analyzing Facebook Privacy Settings: User Expectations vs. Reality," in SIGCOMM, 2011.

[35] F. Stutzman, R. Gross, and A. Acquisti, "Silent Listeners: the Evolution of Privacy and Disclosure on Facebook," Journal of Privacy and Confidentiality, vol. 4, no. 2, 2013.

[36] L. Fang, and K. LeFevre, "Privacy Wizards for Social Networking Sites," in Internaitonal Conference on World Wide Web, New York, 2010.

[37] T. Paul, M. Stopczynski, D. Puscher, M. Volkamer, and T. Strufe, "C4ps: Colors for Privacy Settings," in International Conference on World Wide Web, New York, 2012.

[38] A.C. Squicciarini, M. Shehab, and F. Paci, "Collective Privacy Management in Social Networks," in International Conference on World Wide Web, New York, 2009.

[39] V. Toubiana, V. Verdot, B. Christophe, and M. Boussard, "Photo-Tape: User Privacy Preferences in Photo Tagging," in International Conference on World Wide Web, New York, 2012.

[40] A. Acquisti, L. John, and G. Loewenstein, "What is Privacy Worth?," in Workshop on Information Systems and Economics, 2009.

[41] V.S. Verykios, E. Bertino, I.N. Fovino, et al., "State-of-the-Art in Privacy Preserving Data Mining," in SIGMOD, 2004.

[42] H. Polat and W. Du, "Privacy-Preserving Collaborative Filtering Using Randomized Pertubation Techniques," in International Conference on Data Mining, 2003.

[43] N Lathia, S. Hailes, and L. Capra, "Private Distributed Collaborative Filtering Using Estimated Concordance Measures," in ACM Conference on Recommender Systems, New York, 2007.

[44] D. Argawal and C.C. Aggarwal, "On the Design and Quantification of Privacy Preserving Data Mining Algorithms," in Symposium on Principles of Database Systems, New York, 2001.

[45] S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava, "Privacy in Dynamic Social Networks," in International Conference on World Wide Web, 2010, 2010.

[46] Y. Zhu, L. Xiong, and C. Verdery, "Anonymizing User Profiles for Personalized Web Search," in International Conference on World Wide Web, New York, 2010.

[47] A. Narayanan and V. Schmatikov, "How to Break Anonymity of the Netflix Prize Dataset," in CORR, 2006.

[48] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography," in International Conference on World Wide Web, New York, 2007.

[49] G. Wondraeck, T. Holz, E. Kirda and C. Kruegel, "A Practical Attack to De-Anonymize Social Network Users," in IEEE Symposium on Security and Privacy, 2010.

[50] J. Clause and A. Orso, "Camouflage; Automated Anonymization of Field Data," in International Conference on Software Engineering, New York, 2011.

[51] M. Castro, M. Costa, and J.-P. Martin, "Better bug reporting with better privacy," in Internation conference on architectual support for programming languages and operating systems, New York, 2008.

[52] L. Sweeney, "K-Anonymity: a Model for Protecting Privacy," International Journal on Uncertainty, vol. 10, no. 5, pp. 557-570, 2002.

[53] K. Taneja, M. Grechanik, R. Ghani, and T. Xie, "Testing Software in Age of Data Privacy: a Balancing Act," in SIGSOFT/FSE, New York, 2011.

[54] T.D. Breaux, and A.I. Anton, "Analyzing regulatory rules for privacy and security requirements," in IEEE Transactions on Software Engineering, 2008.

[55] T.D. Breaux, and A. Rao, "Formal analysis of privacy requirements specifications for multi-tier applications," in IEEE International Requirements Engineering Conference, 2013.

[56] C. Mancini, Y. Roggers, K. Thomas, et al., "In the Best Families: Tracking and Relationships," in CHI, 2011.

[57] T.T. Tun, A.K. Bandara, B.A. Price, et al., "Privacy arguments: analysing selective disclosure requirements for mobile applications," in IEEE Computer Society, 2012.

[58] I. Omoronyia, L. Cavallaro, M. Salehie, et al., "Engineering Adaptive Privacy: on the role of privacy awareness requirements," in ACM ICSE, 2013.

[59] M. Robillard, W. Maalej, R.J. Walker, and T. Zimmerman (eds.), Recommendation Systems in Software Engineering, Springer, 2014.

[60] R. van der Meulen, "Gartner Says by 2017, Mobile Users Will Provide Personalized Data Streams to More Than 100 Apps and Services Every Day," 22 1 2014. [Online]. Available: http://www.gartner.com/newsroom/id/2654115. [Accessed 23 5 2014].

[61] European Data Protection Supervisor, "Privacy and Competitiveness in the Age of Big Data," 2014.

[62] EU Parliament and EU Council, Directive 2002/58/EC as revised by 2009/136/EC, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0011:0036:en:PDF, 2009.

[63] F. Brunton, H. Nissenbaum, "Political and Ethical Perspectives on Data Obfuscation," in Privacy, Due Process and the Computational Turn: the Philosophy of Law meets the Philosophy of Technology, Routledge, 2013, pp. 171-195.