Privacy Shielding by Design: A Strategies Case for Near-Compliance

Michael Colesky, Sepideh Ghanavati
Digital Security
Radboud University Nijmegen
Nijmegen, The Netherlands
{mrc, s.ghanavati} @cs.ru.nl

***Abstract*—Changes to the EU-US agreements on transatlantic data transmission are accepted. With the updates leading to an adequacy decision for the Privacy Shield, the European Commission further advances US adherence to the General Data Protection Regulation. The regulation comes with increasing territorial scope for the processing of personal data of persons in the EU, and includes the risk of substantial fines. Soon, a Privacy Shield self-certification will be necessary for US organizations which process EU data. Compliance with these requirements may be assisted by privacy by design. In particular, a recent approach to this uses privacy design strategies. Our paper takes this approach and applies it to the Privacy Shield and its suggested changes. It then explores a case study within scope of the Privacy Shield to demonstrate how to apply privacy by design using strategies.**

Keywords—GDPR; data protection legislation; privacy by design; privacy design strategies; privacy shield

I. INTRODUCTION

The introduction of new data protection laws, in particular the General Data Protection Regulation (GDPR) [1] and the EU-US Privacy Shield [2], overrides earlier legislation with more comprehensive protection. These legal documents are important components of the EU's approach to protecting informational privacy, even for companies outside the EU. The GDPR facilitates the free movement of data, while respecting fundamental rights, and the Privacy Shield allows companies in the US to claim an adequate level of privacy protection. The organizations which are adhering to older legislation are not the only ones which need to update their practices, policies and implementations. Since the new laws apply on a much greater scope, many entities are looking towards their compliance.

Legislation in general does not often specify direct software design requirements. Rather, certain elements are prescribed. One such element is data protection or privacy by design (PbD) [3]. Specifically, GDPR Art. 25 mentions data protection (DP) 'by design and by default' as a requirement, given costs and the state of the art. PbD (or DPbD) is an approach to software development which protects privacy (ensures DP) from the early/concept stages of the software development life cycle [4]. Some tools exist for PbD, including privacy design strategies.

'Strategies' use the GDPR to formulate software design goals, and then suggest 'privacy patterns' (recurring solutions to common design problems in privacy) for implementation [5]. These were initially designed to assist in adhering to the EU Data Protection Directive [4]. The updated strategies provide a direct path from concept analysis (where PbD starts) to design and implementation through the privacy patterns. This paper explores a case study concerning the Privacy Shield in particular, by implementing these privacy design strategies.

The research methodology for this paper includes a literature review primarily of the legislation, and supporting tools of the PbD approach – the strategies and patterns in particular. We build our work upon previous advances in the field, namely on that of the strategies, and of PbD therein.

In this paper, we aim to demonstrate how to achieve what we call near-compliance. Specifically, software *designed with compliance in mind from the beginning, resulting in less legal consultant work.* An engineering effort primarily, which aims at full compliance, prior to legal review. In this case, to the Privacy Shield, with the help of our strategies.

The hypothetical example for our case study, Company X, should apply to the Privacy Shield without falling under direct GDPR scope – as this overrides the need for the agreement. Company X is a large US organization which features in many markets, one of which is deep learning. It pursues this out of public interest in their software based labs. Its subsidiary in the EU uses medical data from EU-based health services for deep learning and preventative research purposes. We explore the hypothetical transfer of this personal data to the US for further compatible processing, where Company X has better facilities.

This paper first introduces relevant background regarding the GDPR and Privacy Shield, as well as some recurring terms. It then discusses the strategies and their related concepts including 'tactics' and the patterns. We introduce and elaborate on our case study, and then provide our conclusions.

II. THE EU-US PRIVACY SHIELD AND GDPR

In this section, we discuss the new EU DP regulation, its differences with the Directive, the EU-US Privacy Shield agreement, and the amendments to the agreement.

The authors would like to thank all participants for their contributions, and in particular the Netherlands Organization for Scientific Research (NWO) for their financial support. We would also like to thank Merel Koning for her continued sound and comprehensive legal advice.

A. The General Data Protection Regulation (GDPR)

The GDPR will be in force from May 25th 2018 [1]. It replaces the EU Data Protection Directive 95/46/EC, and will apply to all member states without having a distinct national implementation. Article 4 (1) and (2) of the GDPR [1] provide definitions for 'personal data' and 'processing' respectively. Personal data in the GDPR refers to any information which relates to an identifiable natural person. Processing thereof is any usage of that personal data, from collection to erasure and anything in between. The GDPR and original Directive provide very specific examples, such as (non-exhaustively) making available, adaptation, alteration, or alignment (Art. 4(2)).

The new measures of the upcoming Regulation include more than its predecessor. For example, the GDPR introduces fines of ten million euros, or 2% of previous annual worldwide turnover (Art. 83(4)) – double in some cases (Art. 83(5)). It also requires data controllers to designate a Data Protection Officer in some circumstances (Art. 37). Notice and consent are well known principles for which the GDPR enforces strict requirements. Contact and data retention information clauses (e.g. within Art. 13, 14) feature, as well as data breach notification (Art. 34). Furthermore, explicit, freely given, and informed consent must be provable by data controllers (Art. 7). It is also withdrawable. The right to be forgotten is refined into the right to erasure and to data portability (Art. 17 and Art. 20).

The GDPR includes greater territorial scope, notably and importantly, international application to those who process personal data of individuals in the EU (Art. 3). Exceptions to this include legitimate national security or law enforcement purposes concerning detection, investigation, prevention, or prosecution. The GDPR has more rules for transferring personal data to international organizations or third countries than internally to the Union (Art. 44). One of these rules is based on an adequacy decision, that is a sufficient level of protection as per Art. 45(2) – assessing the entity's laws, supervision authorities, and international commitments.

If a country lacks adequate privacy law, a legal agreement may give grounds for adequacy, e.g. Safe Harbor – the original attempt at facilitating trans-Atlantic data flows between the EU and US. The Schrems case [6], where Facebook was taken to court in Ireland over privacy violations, revealed a flaw in the use of the Safe Harbor agreement – notably surveillance. This resulted in a renegotiation leading to a new adequacy decision, fulfilling Art. 45 of the GDPR – the EU-US Privacy Shield [8].

B. The EU-US Privacy Shield

The Privacy Shield is an effort to enable protected data transfers out of the EU and through the US for commercial purposes. It does not merely replace Safe Harbor, however, it provides a base on which to introduce oversight and redress, as well as stricter justification requirements for repurposing.

The main objective of the Privacy Shield is to provide 'essential equivalence' to the GDPR [7]. However, as noted below, the agreement initially fell short of this. This had led to an expectation of future versions, and as of July 12th 2016 a revision has been accepted [8][9][2]. Prior to this the European Data Protection Supervisor (EDPS) provided an opinion which suggested changes to the Privacy Shield [7]. It built upon the opinion of the Article 29 Working Party (WP29) [10], which when combined with the EDPS opinion make an authoritative overview. We include these opinions in our analysis. The next section highlights their suggestions and the resulting changes to the Privacy Shield.

C. Privacy Shield Amendments

Both authorities welcome the improvements made by the Privacy Shield, though each noted significant and necessary room for improvement. For one, the aspects covered by the Privacy Shield are (still) spread out beyond a single document, including the adequacy decision and its annexes [2][7][9].

The overall flaws mentioned by the EDPS and WP29 feature the omissions and insufficient specifications of *data minimization / retention* and *automated / bulk processing*, poor clarification of *purpose specification / limitation*, *exceptions / derogations*, as well as room for improvement for *onward transfer*, *redress*, *oversight*, and *rights to access/object* [7]. Of these, clarifications are provided in the new revision for some items, with improvement on several previously lacking items [2]. However, the risk based approach to re-identification has been seen to contradict WP29's opinion on identifiability [9].

These flaws (italicized) are mostly in the Privacy Principles section, Annex II – the main body of the agreement, where in some cases the lack of precise language, or specific purposes and requirements have been seen as failing to limit interference with rights to data protection and privacy [7]. Both the WP29 and EDPS aim for clear purposes and precise requirements for exceptions to the Privacy Shield's protection. While somewhat alleviated in the adopted text [9], some imprecision still exists.

For example, the adopted text still includes 'has been' vs. 'to be' transferred, where the EDPS notes a lack of accuracy may result in misunderstandings. For instance, in "...limitations on the access and use of personal data transferred…" in recital 55 of the draft decision [7]. Onward transfers now need contractual binding to take place [9]. Another example is variance on purpose 'consistency' or '[material] difference' in purposes [7]. These might have allowed false interpretations. We expected the use of '[in]compatible' purposes to appear in coming revisions regularly. Instead clarification was given. This includes both examples of compatible processing activities, and the notion that material difference is possible while 'compatible', but requires new consent [2][9].

In particular, the reuse of human resource or medical data for marketing purposes was troubling – we thus believed that Privacy Shield Annex II.II 9(b)(i) and II.II 14(b)(i) would be removed. Instead 9(b)(i) (human resource data reuse) gained compatible purpose consideration [9]. Medical data reuse still merely requires notice, which is fortunate for Company X in our case study. We had also expected more automation safeguards against harmful performance, credit, reliability, or conduct related personal data usage. While the US offers credit, mortgage, and employment protections, the area is to be monitored by the Commission [2]. Furthermore, despite 'relevance' not being considered sufficient in the context of data minimization by the EDPS [7], the usage is retained. The EDPS suggests that this term be adapted to "adequate and not excessive" or "limited to the information that is necessary."

Less clear cut are the exceptions granted to certain types of processing in the interest of free speech or expression, such as journalism, which contest the equal importance of privacy. These contradict key GDPR requirements, and still need to be balanced. On-going suggestions for this include more precise language to outline the scope of these exceptions and ensure that they only are used for proportionate free speech purposes. Some exceptions regarding initially commercial purpose data, including US laws still allowing public interest access while affecting personal data may eventually be reassessed. The EDPS has recommended international commitments for access to these by authorities [7].

Additional means to pursue redress by individuals in the EU would follow suggestions to reduce the current complexity. Presently redress is possible as per recital 43 of the adequacy decision [8]. The EDPS had suggested additional prevention through compliance monitoring by US authorities, as well as certification supervision by Data Protection Authorities (DPA), noting that the Judicial Redress Act does not have sufficient scope [7]. Further meaningful review may feature spot verifications on both commercial and authority access.

Having discussed the flaws still present in the Privacy Shield, which may result in future changes, it is important that a company which self-certifies does so expecting to eventually provide near GDPR level protection. The further monitoring of the situation by the Commission may result in additional changes. However, we find that medical data use for research receives special exemptions under the Privacy Shield, especially in the context of automated processing. Considering need for data protection by design and by default (Art. 25), we use the PbD tool, privacy design strategies, in our case study.

III. Strategies & Tactics

A privacy design strategy "specifies a distinct architectural goal in privacy by design to achieve a certain level of privacy protection" [5]. Privacy design strategies are a technique for achieving privacy by design first noted by Hoepman [4]. They previously translated the original data protection principles into design goals more accessible to software engineers. These strategies have been extended to make them more relevant in the GDPR context, and provide a direct path from concept to design implementation [5]. This is achieved through privacy patterns, common solutions to recurring design problems in privacy by design. Also introduced were privacy design tactics, a spin on software architecture's system quality attribute 'tactics' which connect the strategies to the patterns.

Both the strategies and tactics use capitalized verbs which more or less summarize their intent. They are shown in TABLE I. Separate for instance can be approached through both processing in isolation, and in distributing the data, both logically and physically [5]. The ongoing collection of privacy patterns which the strategies and tactics connect to are found in [11]. Our case study uses a number of these concepts to demonstrate an example in which the Privacy Shield applies.

TABLE I. Strategies by tactics (adapted from [5])

MINIMIZE	**HIDE**	**SEPARATE**	**ABSTRACT**
AGGREGATE EXCLUDE SELECT STRIP DESTROY	RESTRICT SEPARATE MIX OBFUSCATE DISSOCIATE	DISTRIBUTE ISOLATE	GENERALIZE GROUP
INFORM	**CONTROL**	**ENFORCE**	**DEMONSTRATE**
PROVIDE NOTIFY EXPLAIN	CONSENT CHOOSE UPDATE RETRACT	CREATE MAINTAIN UPHOLD	AUDIT LOG REPORT

IV. Case Study

Our example, Company X, is a receiving data controller, and its EU-based subsidiary is the initial data controller. In particular, we focus on its preventative deep learning re-use of medical data initially held by the subsidiary.

In most cases Company X must follow the GDPR directly. The GDPR's scope generally includes it based on Art. 3(1) [1] (i.e. intent to provide products or services to individuals in the EU). However, in some contexts the GDPR is not directly applicable. Our case study explores one such context.

The GDPR's scope is subject to the context of processing by an establishment. Where Company X processes the personal data of a person in the EU, it falls under scope since it has premises in the EU (the subsidiary research institute) with processing context, or intends to provide products and services to persons in the EU (also the subsidiary). This applicability fades where processing is not in those contexts, where the controller becomes self-certified with the Privacy Shield as grounds for processing. In those cases, the Privacy Shield applies to personal data transfers out of the EU to the US, and any further third party flow. This case study explores Company X self-certifying under the Privacy Shield in order to do this.

A. Repurpose of Medical Data

In our example, Company X's subsidiary research institute receives EU medical data under research purposes, following no request for new consent from data subjects by the medical institution. This can happen in the case of research.

Research purposes are awarded special exemptions in the GDPR, as stated by Art. 89 (i.e. derogations on certain rights for research), and further described in e.g. recital 157 (justifying coupling registries for enhancing knowledge). Predicting disease can be seen as a legitimate interest, though a transfer might also be done on the basis of national legislation until the GDPR comes into force. For example in the UK Data Protection Act, §4(3) Schedule 3(8) or Schedule 4(4) may allow further transfers and processing of medical data so long as only medical (including research) professionals are using it [12]. Of course even when countries are outside the EU, they are affected by the GDPR – the UK may require an adequacy decision like the Privacy Shield.

For Company X, recital 33 suggests that consent may play some role even in these research purposes (specifically distinction between consent for different research areas). Though, as per Art. 89(1) and recital 52, 62, 65, a data subject might not retain their rights to access, erase, rectify, restrict, or object to processed data. At best, Art. 14(5)(b) and recital 113

include notification if it is not considered too much effort. The nature of the data, Art. 9(2)(j) affirms (processing necessary for research), is exempt from the prohibition of special category data processing in Art. 9(1). This data may be kept identifiable longer as necessary for that purpose – Art. 5(1)(e).

Company X desires to use their local US infrastructure to further their EU research efforts. They wish to continue using the personal data acquired by their EU subsidiary. They choose to receive the data as a controller instead of a processor, as processors are subject to more responsibilities (especially regarding obligation to the controller, and standard contractual clauses) – Art. 28. As a controller, it determines purposes or means, within the boundaries of the purpose it gives to its subsidiary: e.g. *disease prevention and detection through deep learning and artificial intelligence*.

In this context, Company X is a controller not bound by the GDPR directly, but by the Privacy Shield. If it infringes upon any obligations within, it will be accountable under US law. This may lead to intervention from the ombudsperson or a fine.

B. Software Considerations

The strategies are designed to facilitate a certain level of privacy protection, allowing engineers to attempt compliance. Each strategy and its contained tactics and patterns are applied where necessary. The patterns and their sources may be found in [11]. In this case study we recommend the following.

1) MINIMIZE

Company X is under obligation to minimize the data as much as possible while still fulfilling the research purpose (recital 156). Though more data may translate to better results, and allows them to warn patients and their doctors – doing so is incompatible with their purposes. Therefore, it will be expected to reduce identifiability as much as it can without jeopardizing the learning (Privacy Shield II.II (14)(a)(i)) [2]. Training the learning system with data types which cannot correlate to symptoms, like names and identity numbers, may even provide poorer predictions. It may wish to choose the types of data that are more relevant, and remove those which are not. The SELECT and STRIP tactics should be used, including for example Partial Identification [13], Select Before You Collect [4], and Strip Metadata [14].

SELECT: *decide on a case by case basis on the full or partial usage of personal data, akin to whitelisting or opt-in.*
STRIP: *removing unnecessary personal data fields from the system's representation of each user.* [5]

2) HIDE

Company X is under strong obligations to protect their highly sensitive data from unauthorized access, use, sharing etc. as their research purpose does not exempt them from this – e.g. recital 53 (meriting higher protection). In particular, the transferring of data from their subsidiary should benefit from state of the art security safeguards. Furthermore, applied as a tactic instead of strategy, there are some cases where data (such as confidential data (Privacy Shield II.III (8)(c)), or pseudonym details (GDPR Art. 4 (5))) should be separated from other data.

The above considerations could benefit from all HIDE tactics. Company X could utilize [Purpose-based] Access Control [14], Pseudonymous Identity [15], K-anonymity [4], Local/Network Encryption, and finally Link Padding [15], with one pattern per HIDE tactic respectively below and in TABLE I.

RESTRICT: *preventing unauthorized access to personal data.*
SEPARATE: *preventing the correlation of personal information to reduce the likelihood of privacy violations.*
MIX: *processing personal data randomly within a large enough group to reduce correlation.*
OBFUSCATE: *preventing understandability of personal data to those without the ability to decipher it.*
DISSOCIATE: *removing the correlation between different pieces of personal data.* [5]

3) SEPARATE

Many cases benefit from ISOLATE and DISTRIBUTE, though these approaches to privacy protection do not directly facilitate the principles featured in the Privacy Shield. Not centralizing data, however, will make correlating data into meaningful predictions tougher. Company X does not need to utilize this strategy or its tactics, but doing so may afford its data subjects better privacy protection. It should try to find a good balance between personal data separation and effective deep learning.

4) ABSTRACT

Like with SEPARATE, Company X's use of deep learning to draw connections between data may be hindered by this strategy. In some cases, it may be able to generalize data without this drawback. However, their unique case does not really justify this. They are not obliged to use this strategy or its tactics, though in some cases the simplification of details can result in better predictive systems – by abstracting away less meaningful information.

5) INFORM

As mentioned in Section A and Section II.C, at best there is a proportionate need to NOTIFY data subjects that their data is being used for research. Data Breach Notification [4] is especially necessary, though not specifically to data subjects. When self-certifying, a public Privacy Policy will EXPLAIN usage. Also as per Privacy Shield II.II (8)(i), data subjects have the right to request access to verify that usage, and have it corrected where it is inaccurate or processed in violation [2].

SUPPLY: *making available extensive resources on the processing of personal data, including policies, processes, and potential risks.*
NOTIFY: *alerting data subjects to any new information about processing of their personal data in a timely manner.*
EXPLAIN: *detailing information on personal data processing in a concise and understandable form.* [5]

6) CONTROL

Also mentioned in Section A, the use of medical data which already acquired consent allows Company X to overlook data subject rights to reasonable levels of control. In fact, doing otherwise would likely constitute an incompatible purpose.

7) ENFORCE

Company X is not entirely without responsibility here, it still needs to CREATE, MAINTAIN, and UPHOLD their policies. These include their legal responsibilities, their procedures and their

practices. For the first two tactics, this will relate to the privacy policy in particular. For UPHOLD, there are Sticky Policies, and Usage Control Infrastructure [14] patterns. In addition, Company X is obliged to allow for recourse with clear and usable mechanisms, response within 45 days of a complaint is also necessary (Privacy Shield II.II (11)(d)) [2].

CREATE: *acknowledging the value of privacy and deciding upon policies which enable it, and processes which respect personal data.*
MAINTAIN: *considering privacy when designing or modifying features, and updating policies and processes to better protect personal data.*
UPHOLD: *ensuring that policies are adhered to by treating personal data as an asset, and privacy as a goal to incentivize as a critical feature.* [5]

8) DEMONSTRATE

Finally, it must show adherence to its responsibilities to the Department of Commerce by verifying its attestations, through self or external compliance REPORT. This must be signed annually by an authorized representative. All records pertaining to this must be kept. For Non-repudiation [16], it should LOG relevant usage, and perform a regular AUDIT.

LOG: *tracking all processing of data, without revealing personal data, securing and reviewing the information gathered for any risks.*
AUDIT: *examining all day to day activities for any risks to personal data, and responding to any discrepancies seriously.*
REPORT: *analyzing collected information on tests, audits, and logs periodically to review improvements to the protection of personal data.* [5]

Company X is not subject to repurpose justification as it is not repurposing. Its use of the research exception is only valid if it stays within that purpose. Doing so is not very difficult, though, as new research purposes will not be incompatible. This is like the earlier free speech argument for journalism, where public interest is balanced against privacy. As this data was owned by a medical institute, it was also commercial data. This is an aspect which we hope revisions will better specify to ensure proportionality.

The minimization and retention aspects are justified by the purpose as well, much of which is bulk and automated. This as we mentioned earlier will receive a few more safeguards, but due to purpose limitation, it will encounter difficulty if it influences data subjects' lives. An item perhaps worth further exploring is the use of 'key-coded data' (Privacy Shield II.II (14)(g)) where a transfer of this from the EU 'would not [be] subject to the Privacy Shield Principles' [2].

Company X is subject to oversight, but not so easily to data subject rights to access/object – as their consent was obtained prior to the transfer. This may be improved by way of spot checks, and better supervision by DPAs. In this situation particularly, as it features very sensitive information, Company X would be watched closely.

V. CONCLUSIONS

We explored the Privacy Shield and the criticisms by the Article 29 Working Party and Data Protection Supervisor. We provided a context in which the agreement applies after the GDPR comes into effect, to a recent and relevant hypothetical case in the medical research sector. The situation we describe in this paper was used to demonstrate how privacy and or data protection by design can be considered using privacy design strategies and the tactics and privacy patterns therein.

We determined that the considerations Company X would need to act upon or those it would benefit from, in medical data transfer from its subsidiary to the US, are considerably lower than most situations. Their research purpose permits a number of liberties, while restricting them from providing much information and control to data subjects. This results in a focus on HIDE, ENFORCE, and DEMONSTRATE – a bare minimum. The unique focus of deep learning also justifies less minimization than usual. Aside from using directly identifiable data (full names, identity numbers, etc.) their only need for MINIMIZE is to prevent wasting time and resources on useless data – their research purpose does not permit commercial opportunities.

Some suggestions for future work include different cases, those with more comprehensive applicability. The strategies themselves could also receive validation or evaluation in a practical commercial setting, with feedback from developers and legal teams regarding near-compliance.

REFERENCES

[1] European Parliament and Council of the European Union, "General Data Protection Regulation," *Official Journal of the European Union*, 2015.

[2] US Department of Commerce, "EU-US Privacy Shield Annexes," 2016.

[3] A. Cavoukian, "Operationalizing Privacy by Design : A Guide to Implementing Strong Privacy Practices," pp. 1–72, 2012.

[4] J.-H. Hoepman, "Privacy Design Strategies," *IFIP SEC 2014*, 2014.

[5] M. Colesky, J. Hoepman, and C. Hillen, "A Critical Analysis of Privacy Design Strategies," in *IWPE*, 2016.

[6] Opinion of Advocate General Bot, *C-362/14 Maximillian Schrems v Data Protection Commissioner*, vol. 1, no. September. 2015.

[7] European Data Protection Supervisor, "Opinion on the EU-U.S. Privacy Shield draft adequacy decision."

[8] The European Commission, "EU-U.S. Privacy Shield Adequacy Decision," 2016.

[9] G. Maldoff, "We've got a finalized Privacy Shield agreement: What's new?," *iapp*. [Online]. Available: https://iapp.org/news/a/weve-got-a-finalized-privacy-shield-agreement-whats-new-2/. [Accessed: 13-Jul-2016].

[10] Article 29 Data Protection Working Party, "Opinion 01/2016 on the EU–U.S. Privacy Shield draft adequacy decision," 2016.

[11] "privacypatterns.eu - collecting patterns for better privacy." [Online]. Available: https://privacypatterns.eu/. [Accessed: 20-Oct-2015].

[12] Her Majesty's Stationery Office, *Data Protection Act*. 1998.

[13] H. Baraki et al., *Towards Interdisciplinary Design Patterns for Ubiquitous Computing Applications*. Kassel, Germany, 2014.

[14] C. Bier and E. Krempel, "Common Privacy Patterns in Video Surveillance and Smart Energy," in *ICCCT-2012*, 2012, pp. 610–615.

[15] M. Hafiz, "A Pattern Language for Developing Privacy Enhancing Technologies," *Software - Practice and Experience*, vol. 43, 2013.

[16] L. Compagna, P. El Khoury, F. Massacci, R. Thomas, and N. Zannone, "How to capture, model, and verify the knowledge of legal, security, and privacy experts : a pattern-based approach," *ICAIL '07*, 2007.