Pragmatic Online Privacy: the SftE Approach

Vitor Jesus
PrivDash
vitor@privdash.com

***Abstract*—This position paper presents and proposes new requirements for Privacy and Data Protection. We first raise misalignments of current Privacy regulations and argue that current regulatory approaches do not benefit individuals as much as expected to the point that it primarily shields large organisations from ethical management of personal data.**

From this assessment, we propose the Start-from-the-End (SftE, pronounced "soft") approach to online Privacy. It puts the focus on the later stages of the lifecycle of Personal Data (such as the Right to Erasure), while removing focus from the points of collection of personal data. The ultimate goal is to reclaim straightforward enforcement and re-empower individuals in a way that is meant to be feasible and practical.

***Index Terms*—Privacy, Data Protection, Consent, Personal Data Receipts, Consent Record**

I. Introduction

Privacy on the Internet is now a much debated topic. There are perhaps two dimensions that drive online Privacy practices. On one hand, we have different regional individual attitudes towards Privacy which is, quite surprisingly, a subject of unexpected contention [1]; on the other hand, we have local or jurisdictional approaches that can be very diverse. There is no doubt, however, that, with a data-driven society firmly established, Privacy is now a top item in the political and societal agenda worldwide.

The key event for the state of things arrived about 20 years ago when computer technology was able to vastly scale, for example with cloud computing. Social networks such as Facebook, or very large pure digital services such as Google, were able to develop technology that, in unseen scale, made an anonymous experience by default become personal and unique. If before there was an implicit assumption that a server could not handle hundreds of millions of individual profiles, now it is a given.

Mobile phones are also a large part of the process not only because of the rich set of sensors (such as location or motion) but also because of the underlying business model that is able to, at a glance, offer for free feature-rich Operating Systems (namely Android). This trend is further breaking out to the physical domain. We see the proliferation of the Internet-of-Things (IoT) as seen in smart voice assistants (e.g., Alexa from Amazon), smart locks (e.g., Ring), inexpensive indoor cameras, connected cars, etc. Combined with recent advances such as Artificial Intelligence (AI) for image recognition, particularly in public spaces, Privacy concerns have now escaped the browser or the phone to become a truly immersive experience, for the worst reasons.

Even if not the first, EU GDPR is, fairly consensually, seen the most comprehensive and mature Privacy regulation at the moment. Adopted in 2016 and enforceable from 2018, it broke ground for a number of factors. For example, it was a regulation that immediately impacted 20+ countries and the third largest economy (contrary to a simpler directive); it also clearly set out a basic set of principles for defining the structure of governance of Data Protection up to quick breach notification requirements. We are here deliberately ignoring jurisdictions without strong protections of Data Protection. This is the case of most states in the US where Data Protection seems to be more positioned at commercial agreements or specific to sectors (such as HIPAA's for healthcare data).

In this position paper, we will argue that a new approach to Privacy and Data Protection is needed. Ultimately, we see that regulations are not working as expected and, to a large extent, are promoting misaligned incentives when compared to the original mission of empowering and protecting online individuals.

In Section 2 we review the essence of Privacy, from a socio-technical vantage point, which is followed by reviewing the key technical elements supporting Privacy; we then close with thoughts raising the fact that Privacy is not necessarily an absolute value. In Section 3 we share lessons learned from two projects: CASSIOPEIA, focused on shared smart homes and IoT, and Privacy-as-Expected, focused on Personal Data Receipts. In Section 4 we offer for discussion our Start-from-the-End (SftE) framework to Online Privacy: a pragmatic proposal focused convention, revocation, traceability and new approaches to software – altogether offering a pragmatic approach by acknowledging, we argue, a number of established facts and practices both from Individuals and Online Services.

II. The Problem of Online Privacy

In this section, we review the socio-technical problem of Privacy.

A. The Value of Privacy

It is important to revisit the human drivers for Privacy so to understand, from a technical system perspective, what requirements should look like. The Privacy debate is very often framed around the right to intimacy, private life or the right to image [2]. This is, perhaps, where the largest discrepancies exist between peoples. Anecdotally, the author recently asked an attendant of a conference how much, personally, privacy was worth to which the person answered €20. From a blatantly utilitarian perspective, this amount is probably far exceeding

what the typical internet user is willing to pay for all free services such as email, social networking, etc. Of course, this is not the general case but a final answer about whether Privacy is important or not seems to be difficult to establish – see, e.g., [1] that, despite being from 2005, clearly shows that there is a fairly conscious trade-off between Privacy and financial value. Perhaps a valid route is to explore user-tailored Privacy [3]. To a certain extent, this paper does argue for such but at a more fundamental level.

The fact is that Privacy, or rather the care with personal data, is more than a simple perception. It has practical implications with two clearly identified.

The first practical implication motivating better privacy practices is state surveillance. It should be stressed that this aspect must be framed under a cultural light. On one hand, one can argue that state surveillance offers a level of security and coordination which may be desired. On the other hand, it opens the door to abuse and raise conflict with other values, not the least with personal freedom.

Secondly, data breaches seem to be becoming more common, despite the focus on cyber security. Paradoxically, this seems to not be ideally communicated to individuals [4]. Rather gloomy, we feel that we are poised to reach a point where everybody's data will be public – name, address, national identifying numbers, etc. The danger with this is crime: from identity theft to common crime to being targets of cyber security attacks, notably those whose first vector is phishing. A data breach can have drastic implications and we argue we will see the severity increasing. For example, a breach of medical records can dramatically change the life prospects of a person. Finally, there are anecdotal reports of personal data being used to personalise, e.g., product quotations which, in itself, will lead to inequality. One should keep aware that a single, partial, data breach is as dangerous as a major one. Multiple breaches with partial, or perhaps even anonymised, personal information can be linked together to generate detailed records of someone. Above all, once the information escapes, it will stay accessible forever. It is well-known, despites decades of research, that maintaining unlinkability is a very hard problem (likely impossible) when a data set is exposed to a wide context. This particular aspect is a strong motivator of SftE.

B. The Technical Element and Focus on Point of Collection

It seems that the effort in tackling Privacy and Data Protection challenges online is focused at the point of collection, an idea similar to what Murmann & Fischer-Hübner [5] call "ex ante transparency". Two lines can be identified.

The first is Privacy-Enhancing Technologies (PET) which, in itself, is a family of technologies. We here understand PETs mainly as the set of techniques that are offered by the service-provider to better meet Privacy expectations. As such, we ignore user-side techniques such as ad blockers or anonymisation network overlays. Very briefly, PETs cover [6]

- Secure and anonymity-enabled communications
- Obfuscation at source supported by the server side
- Secure computation and storage with homomorphic encryption at the forefront
- Data storage and sharing models such as Differential Privacy

Whereas PETs are a critical element in any data-oriented architecture, it still relies on the management and business models of the Data Controller making the individual a passive subject.

A second line is the attempt to limit data collection and improve default anonymisation. We encompass here automatic tracking and profiling of individuals both on the browser and mobile apps. We are here speaking of the common tracking, either by using "browser cookies" or, far more invasive, gathering data directly tied to a reliable and accurate identifier as is the case of mobile phones. To this end, new approaches are being proposed such as Apple's privacy tags or Google's FLoC [7], perhaps combined with Do-Not-Track mechanisms. It is apparent that such initiatives cannot solve as such the problem. A key problem, for example, is that they are a self-regulated exercise – akin to a code of honour – and, particularly with FlOC, it may make the problem worse [8]. In essence, users are asked to blindly trust the good intentions of the same parties whose business models rely on extracting value from Personal Data at scale and using confidential, close-source, unauditable technologies.

A third line of research that, we argue, has been underexplored, is the notion of Personal Data Receipts (PDR). The key concept is similar to shopping receipts that essentially exist to keep the service provider accountable – thus empowering the individual. On sharing of Personal Data, the individual should have proof of such in the form of a receipt. Strictly speaking of GDPR, and the requirement of demonstration of consent for Data Controllers of the legal basis, with Consent at the forefront, PDRs offer simple and straightforward means for demonstrability. On a dispute, both parties psimply have to roduce their copy of the receipt. We find this concept extremely powerful and, if designed with auditability properties and as an actionable digital artefact [9], it can dramatically re-empower the user. To note that a receipt can be made anonymous, similar to a shopping receipt, but, yet, is sufficient means to allow exercising of rights – notably the Right to Erasure when the legal basis turns weak as (we argue) is the case for all personal data given sufficient time.

C. Online Harms, Business models – but Privacy

The problem of online privacy is further made complicated, often unnecessarily, by conflating different types of problems. Online harms, and often security and law enforcement, are common reasons to remove some priority from Privacy as a societal goal. The UK has recently planned for all adult content providers to require proof of age which, not unexpectedly, seems to be difficult to implement and even less to enforce. Also in the UK, there is an ongoing campaign about sharing medical records with third-parties on a opt-out basis for, e.g., research purposes. Whereas the objectives can be of praise, and the medical records are anonymised (yet necessarily with

weak assurances, by nature of anonymisation), this initiative has been poorly communicated.

On the other hand, but that justifies privacy as a trade-off rather than an absolute value, there is the commonly known idea that, if a product is free, the user is the product. We have nowadays available high-quality services for free: email, social networking, free multimedia repositories, instant messaging and communication, etc. These services, within a certain context, can even be seen as essential in the current days. The cost is, of course, privacy, commonly in the form of tracking and targetting. Major businesses were created with the data-driven economy, offering appealing products, but that are still businesses needing revenue. As discussed, when confronted with the two radical views, many (most?) individuals would not be willing to pay for, e.g., email, just to protect their privacy. As such, privacy must, naturally, take into account this delicate balance. To a sense, we speculate that, unless the Web is technically redesigned, it will never be possible to stop substancial tracking with technical tools. As much as we regret this situation, entirely due to how the Internet and the Web was designed, we argue that this fact needs to be accepted to some degree. The only way to stop tracking (up to surveillance in many cases) seems to rely on codes of honour (such as the "Do-Not-Track") whose effect, quite blatantly, is close to nothing. We argue privacy assurances must come from a different approach that can not rely on codes of honour or agility of regulators.

D. Regulations: Sounds Good, Doesn't Work

GDPR changed the game and brought great order and incomparable protection to Individuals. It is a key source of practice in different jurisdictions such as California/US, Brazil or Thailand. Designed in the early 2010s, and 3 years after being applicable, we now see its limitations.

One of its key problems is lack of enforcement. Whereas there are already a number of (large) fines imposed, they are too few when compared to expectations. To a large extent, it protects more Data Controllers than Individuals who either exercise their legal power and/or overload the user with information, or by finding loopholes. Even if not directly from GDPR, the so-called "cookie law" has significantly reduced the fantastic improvements in Web usability, during the 2010s, with the proliferation of consent banners. These are, arguably, not compliant with the law [10] but, nevertheless, there is no other known alternatives when balancing the (poor) means for online Consenting and keeping data collection needed for businesses to keep operating.

A further problem, particularly problematic, is third-parties. The common chain of value of personal data is that the Data Controller will not consume that data internally but, rather, will re-sell to (few and large) third-parties [11]. It is said that, anecdotally, such personal data hubs hold fine-grained details on hundreds of millions of people by aggregating personal data and context from many different sources. Since very quickly an individual lose track of who owns what data, this is virtually unaccountable, will live forever, and easily falls out of the range of any regulation.

Regulations also seem, in an unfortunate misalignment of incentives, to promote dark patterns – manipulating the user interface in way to direct their actions. Considering the high effort to conduct this type of investigations, there should be little hope that dark patterns will ever be made accountable. Business will always be one step ahead.

Finally, from a simplistic, yet practical perspective, one can see two broad groups of online services: ones for which personal data is a core component of the business, and another group for which personal data is unwanted liability. It is paradoxical that it is these small/medium businesses that incur most of the costs of aligning their operations with Data Protection and, indirectly but necessarily, cyber security. Whereas the indirect push for cybersecruity is a much welcome result, small businesses are nevertheless presented, essentially, with the same requirements as large organisations.

III. Start-From-The-End Approach

We argue that online privacy needs, broadly speaking, a realignment of incentives and methodologies. In this section, we lay out a proposal to revisit regulations concerning Personal Data.

The key goal is, on one hand, to be pragmatic and learn lessons. For example, it is well known that it is virtually impossible for all individuals to read all privacy notices and make truly informed decisions. On the other hand, we aim at revisiting the lifecycle of personal data and move the focus from the data collection point to the point of use and storage. This motivates the name SftE: Start-from-the-End. We propose to replace the effort of informed sharing with lean accountability for Data Controllers. We hope to see a better trusted ecosystem and a friendlier internet.

The key rules, to be detailed next, are

- Convention
- Revocation
- Traceability
- Middleware

A. Convention

We advocate simplification and accept the fact that users do not exercise, in general, informed consent. This should not substancially limit ownership of the personal data over time. Furthermore, even if users did read and scrutinise all the material made available before engaging with an online service, we argue it would change very little. Overall, the internet follows an unacknowledged model of Take-it-or-leave-it [12] in that, despite the abundance of services, there are little alternatives. Privacy, in general, is not a strong enough driver in the face of material needs or social pressure.

We propose we evolve to a model of convention and reasonable expectation. Whereas the details are difficult to pin down, we see a possibility that, when it comes to personal data, few doubts will exist when the rule is reasonable expectations connected to the service itself. For example, while

using a mobile app, one could expect the simple purposes of local/immediate advertising or for creating a social network.

To note that convention should be generalised to all angles of Data protection and Privacy. For example, Privacy Notices should follow community-practices (e.g., layered notices) instead of custom, legal-oriented, texts.

B. Revocation

Especially when the legal basis is Consent, individuals usually have the right to request revocation. This is closely aligned with the Right to Erasure, and akin to the right to be Forgotten. However, as observed by the author, it seems to be a difficult exercise in most cases – and when there is even a possibility that is not a direct contact form. Revocation, or requests for erasure, should be a cornerstone of Privacy.

One should note that this rule is much better enforceable, auditable and verifiable than today's regulations – even by automatic means. For example, whereas it is virtually impossible to automate the auditing of compliance practices, it is much simpler to detect whether personal data has been kept. In other words, detecting unlawful, and at scale, data retention should be a much more approachable problem than auditing internal compliance practices. As an indication supporting this rather optimistic statement, we note that, currently, the key source of fines come from (necessarily unintended) data breaches. Combined with the above, Convention, a set of computer interfaces could be made widely available and operated by different communities of engaged privacy advocates.

Furthermore, automatic deletion and short periods of data retentions should be imposed. The ideal retention period is an open question but there are informal suggestions that building profiles over months or years, for the purposes of advertising, has little more effect than targeting the individual based on the local and current context (such as the web page they are). We argue that the retention period should be short – perhaps days. Data Controllers should have to demonstrate the need for more. This period should also reflect a trade-off between user acceptance and business utility.

A far more difficult problem for revocation is (pseudo) anonymisation. If data is shared multiple times, pseudo-anonymised but likely re-identifiable, requesting erasure may not be formally possible. This is the worst combination of problems: the individual cannto prove ownership but profile data still is able to identify the person with high accuracy. Requests for Erasure require, quite rightly, identification and proof of ownership of data in order to prevent abuse and malicious activity. In this sense, users do need to be uniquely identified (not necessarily with a physical identity) and attached to means to prove identification. One could imagine an anonymous identifier so data can be indexed, timestamped and later referred to. This identifier should change very rapidly and kept in the records of the original Data Controller. Whereas this approach is currently implemented, in some form, a far better alternative is to use Personal Data Receipts (PDR, next subsection) that can act as a bearer token - something the user "holds".

C. Traceability

A key missing component of today's personal data is lack of user-side traceability. It is virtually impossible, today, to track personal data, particularly when it is re-sold or re-shared. We advocate Personal Data Receipts (PDR) to this effect. Upon any kind of exchange of personal data, a receipt should be generated capturing the object and context of the personal data transaction.

If well designed [9], PDRs can be made simple, non-invasive and with low overhead. Simple tools can be used to store and manage the receipts and, by combining all the information, allow the user to finely trace and observe how their personal information is used and by who. A PDR is able to capture all the relevant state of the transaction, and can be made fully anonymous. Similarly to a shopping receipt, when one returns an item no need for identification is needed as long as the receipt is produced. In technical terms, a PDR is a bearer token, or proof of possession – similar to cash, it can be made fully anonymous. To illustrate, Figure 1 sketches a browser add-on that generates PDRs automatically (adapted from [13]).

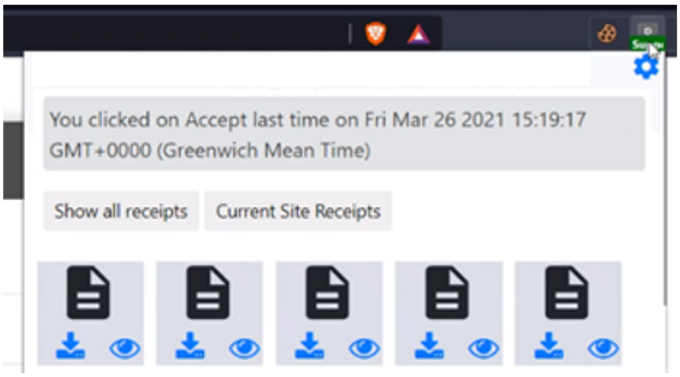

Figure 1: Automatic generation of Personal Data Receipts using a browser extension.

D. Middleware

The final component concerns the need to re-think software applications to keep personal data in mind. Software is still driven by functionality and privacy becomes a non-standard after-thought. Simply asking users to make better informed decisions before using software is of little effect. We argue the problem is not quite of information *a priori* but rather of choice. More information, in the form of generic tags or labels, will not restore control to users but continue to suffer the same fate of Take-it-or-leave-it.

We advocate that software applications must have explicit, embedded by external means, non-custom, auditable, information about Privacy. Consider the simplest example: a web page. There is no known information, embedded in the code of the page that reveals who the service provider is. Instead, metadata about Privacy is combined with, and diluted by, the content and function of the website. Furthermore, because this

is natural language, it is very difficult to design software tools to, for example, simply answer the question of who is the Data Controller. We show in Figure 2 a simple sketch of how such information could be embedded in the HTML of a webpage (adapted from [13]).

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="utf-8">
  <meta name="pisp" content="Consent Gateway">
  <meta name="lastPolicyUpdateDate" content="20
    <link rel="stylesheet" href="form.css">

 <script src="paecg.js"> </script>
  <script>
  var details={
    'info_for_receipt':{    'piicontrollers'
        "name": "Acme Inc.",
        "localid": "PIIC-A",
        "address": "Wonderland",
        "url": "http://example.com/",
        "contact": {
            "phone": "000",
            "email": "acme@example.com"
        },
        "policies": {
            "privacy": "http://example.com/p
            "termsconditions": "http://examp
        }
    }]},
    'consent_submission_elements':{'Submit':
    'user_inputs':['fname','lname','email'],
    'javascript':['http://3.10.208.186/updat
    'policyurl':['http://3.10.208.186/update
  };
  var paecg=new PaECG(details);
  paecg.setup();
  </script>
```

Figure 2: Sketch of embedding privacy information in the software application.

IV. Conclusions

This position paper discussed the current state of Privacy from the perspective of regulations. We show that, despite regulations being a great positive advance in protecting individuals online, the current paradigm suffers from a number of misalignments, the key of which, we argue, is the focus on the early stages of data collection instead of focusing on the later stages and rights exercise. We sketched a proposal, that we termed Start-from-the-End (SftE), that has the potential of alleviating these issues. SftE advocates that revocation, convention, user-managed traceability and privacy middleware should form the basis of new updates to Privacy and Data Protection Practices.

Acknowledgements

We wish to thank the partners of the EU H2020 NGI-Trust projects ”Privacy-as-Expected: Consent Gateway” and ”CASSIOPEIA” for their insightful thoughts.

References

[1] B. A. Huberman, E. Adar and L. R. Fine, *Valuating privacy*, in IEEE Security & Privacy, vol. 3, no. 5, pp. 22-25, Sept.-Oct. 2005

[2] Blank, Grant and Dutton, William H. and Lefkowitz, Julia, *Perceived Threats to Privacy Online: The Internet in Britain, the Oxford Internet Survey*, 2019 (September 6, 2019). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3522106

[3] B. P. Knijnenburg, *Privacy? I Can't Even! Making a Case for User-Tailored Privacy*, in IEEE Security & Privacy, vol. 15, no. 4, pp. 62-67, 2017

[4] Y. Zou and F. Schaub, *Beyond Mandatory: Making Data Breach Notifications Useful for Consumers*, in IEEE Security & Privacy, vol. 17, no. 2, pp. 67-72, March-April 2019

[5] P. Murmann and S. Fischer-Hübner, *Tools for Achieving Usable Ex Post Transparency: A Survey*, in IEEE Access, vol. 5, pp. 22965-22991, 2017

[6] Nesrine Kaaniche, Maryline Laurent, Sana Belguith, *Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey*, Journal of Network and Computer Applications, Volume 171, 2020

[7] Google, *Building a privacy-first future for web advertising*, January 2021, Accessible: https://blog.google/products/ads-commerce/2021-01-privacy-sandbox/

[8] Electronic Frontier Foundation, *Google's FLoC Is a Terrible Idea*, March 2021 accessible: https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea

[9] V. Jesus, *Towards an Accountable Web of Personal Information: the Web-of-Receipts*, in IEEE Access, vol. 8, pp. 25383-25394, 2020

[10] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. *Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective*, Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA,

[11] Andreas Claesson and Tor E. Bjørstad, Norwegian Consumer Council, *Out of Control – A Review of Data Sharing by Popular Mobile Apps*, Tecbnical Report, January 2021, accessible: https://www.mnemonic.no/news/2020/out-of-control/

[12] Zuiderveen Borgesius, Frederik and Kruikemeier, Sanne and Boerman, Sophie and Helberger, Natali, *Tracking Walls, Take-It-Or-Leave-It Choices, the GDPR, and the ePrivacy Regulation*, March 15, 2018, European Data Protection Law Review, Volume 3, Issue 3, p. 353-368.

[13] Harshvardhan J. Pandit, Vitor Jesus, Shankar Ammai, Mark Lizar and Salvatore D'Agostino, *Role of Identity, Identification, and Receipts for Consent*, Open Identity Summit, June 2021, Lyngby, Denmark