Written evidence submitted by Dr Inah Omoronyia, Professor Alice Miller and Professor Awais Rashid (DDA0041)

Authors
Dr Inah Omoronyia – (Senior Lecturer in Privacy, Bristol) PI for Innovate UK Cyber Security Academic Startup Accelerator Programme on privacy engineering techniques for software designers. In 2021, he was awarded a Research and Enterprise fellowship by Scottish Enterprise. He is also PI of a Scottish Enterprise High Growth Spinout project to help business organisations with Privacy Regulation and Compliance struggles. He leads a UKRI TAS Hub pump-priming research grant to investigate Consent Verification in Autonomous Systems.

Prof. Alice Miller – (Professor of Computing Science, Glasgow) Leads the Understandable Autonomous Systems theme at Glasgow. She is a Co-I on the UKRI Trustworthy Autonomous Systems Node in Governance and Regulation. She holds an EPSRC Case Award with The Defence Science and Technology lab (DSTL).

Prof. Awais Rashid – (Professor of Cyber Security, Bristol), has 20 years of expertise leading large, multi-partner, interdisciplinary projects. He is Director of the interdisciplinary EPSRC CDT: Trust, identity, privacy & security in large-scale infrastructures; and the National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online (REPHRAIN). He leads the Cyber Security Body of Knowledge, with a college of >110 world-leading experts (incl. privacy, online rights & law); interdisciplinary projects in national programmes: RISCS, RITICS and PETRAS.

About TAS Hub
The UKRI TAS Hub (EP/V00784X/1), comprises a team from the Universities of Southampton, Nottingham and King's College London. The Hub sits at the centre of the £33M Trustworthy Autonomous Systems Programme, funded by the UKRI Strategic Priorities Fund. The role of the TAS Hub is to coordinate and work with six research nodes to establish a collaborative platform for the UK to enable the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice.

About REPHRAIN
The National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online is funded by UKRI and consists of a team from the Universities of Bristol, Bath, Edinburgh and University College London. REPHRAIN explores how to keep people safe online while allowing them to fully participate in digital technologies and the contributions they make to an innovative, inclusive and healthy economy and society. Its different strategic projects address fundamental tensions and imbalances pertaining to protecting citizens online.

Summary
Data protection regulations are far from being a barrier to innovation or trade. They introduce regulatory certainty and data protection standards that allow businesses and consumers to thrive. This is the reason why countries around the world are seeking to forge data sharing alliances through 'data adequacy' agreements to maintain the free flow of personal data. Even adequacy agreements are not promised to be burden free. Hence, a data strategy that seeks to streamline data protection regulations may only help businesses in the short term and limited at the

national level. Whereas data use and sharing policies that also incentivise new privacy engineering solutions are more likely to support businesses in the longer term, as well as compete favourably at the international level.

We propose six guiding principles that should form the core of a UK data policy framework. These principles will foster new privacy engineering solutions to ensure appropriate privacy safeguards in the usage and sharing of data. An effective engineering solution would be agnostic to specific national regulation. It would enable organisations to free up valuable time and resources, improving how *data workers* manage risks, and allowing for a smarter and easier path to regulatory compliance. Indeed, data protection regulations also offer a market opportunity in the same manner data offers a pathway for a thriving digital economy. The global Data Protection Market is expected to exceed USD 194.11 Billion by 2026, with Compound Annual Growth (CAGR) at 15.9% in the given forecast period [12].

In response to this parliamentary call for evidence, we leverage on our experience in working with privacy experts, technology developers and business practitioners. **We do so to argue for more appropriate privacy safeguards to be applied in the usage and sharing of individuals' data.** We take an engineering viewpoint, to highlight the weaknesses of current privacy mechanisms, particularly those enshrined in data protection regulations and their pain points. Arguably, it is these pain points that makes regulatory compliance burdensome and not the regulation itself. As such, by addressing the pain points, regulatory compliance becomes less burdensome.

Background

To mitigate the friction that often exist between data use and regulatory compliance, it is our opinion that government policies that seek to maximise the collection, usage and sharing of data across businesses, government, civil society and individuals, should also enable strong policies that support the investigation of new ways of ensuring privacy protection. Without such a dual view, we posit that there is an increased risk of generating an asymmetric data economy with strong acrimony and distrust between businesses and data subjects.

The UK Government's National Data Strategy[1] as stated in its five missions, have proposed the streamlining of its data protection regulations. This is to help innovators and entrepreneurs responsibly use data. In the short term, this appears to be a good incentive to collect and curate data to maximise value across the economy, without undue regulatory uncertainty or risk. But the strategy does little to address the fundamental privacy engineering challenges that the same innovators and entrepreneurs will face when they operate globally – they will still have to satisfy the compliance need of their customers outside of the UK who may have less streamlined data protection regulations.

Fundamentally, a data protection regulation broadly consists of two components. The first is a set of data protection concepts – for example, personal data, special category data, subject, controller, processor, etc. These concepts are the building blocks of the regulation. They embody the attributes of data, the entities that operate on data or are impacted by such operation on data. The second is data protection principles, much like the cement that holds the blocks together. There are 7 principles

[1] https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy

listed in the UK GDPR – these are (1) lawfulness, fairness and transparency; (2) purpose limitation; (3) data minimisation; (4) accuracy; (5) storage limitation; (6) security; and (7) accountability. These principles are essentially an expression of the core values of the data protection regulation and underpin prescriptive obligations necessary to achieve compliance. Compared to other data protection regulations, the differences in concepts and principles across jurisdictions are often subtle and may vary at the level of monitoring and enforcement.

Privacy engineering teams consist of experts with a clear understanding of the regulatory landscape and organisational business goals. They are tasked with facilitating an enabling environment to achieve compliance in a manner that both protects the subject and realises the organisational business objectives. Privacy teams achieve these goals by embedding tools, techniques and privacy governance frameworks into the organisation's business processes. Figure 1 is a mind map illustrating the privacy engineering problem space. An effective privacy solution first requires an understanding of different inputs such as raw data/schemas, the business process and a host of functional/non-functional requirements to be satisfied. There are different target user personas that either contribute to the compliance process or validate that compliance is achieved. These personas may have different

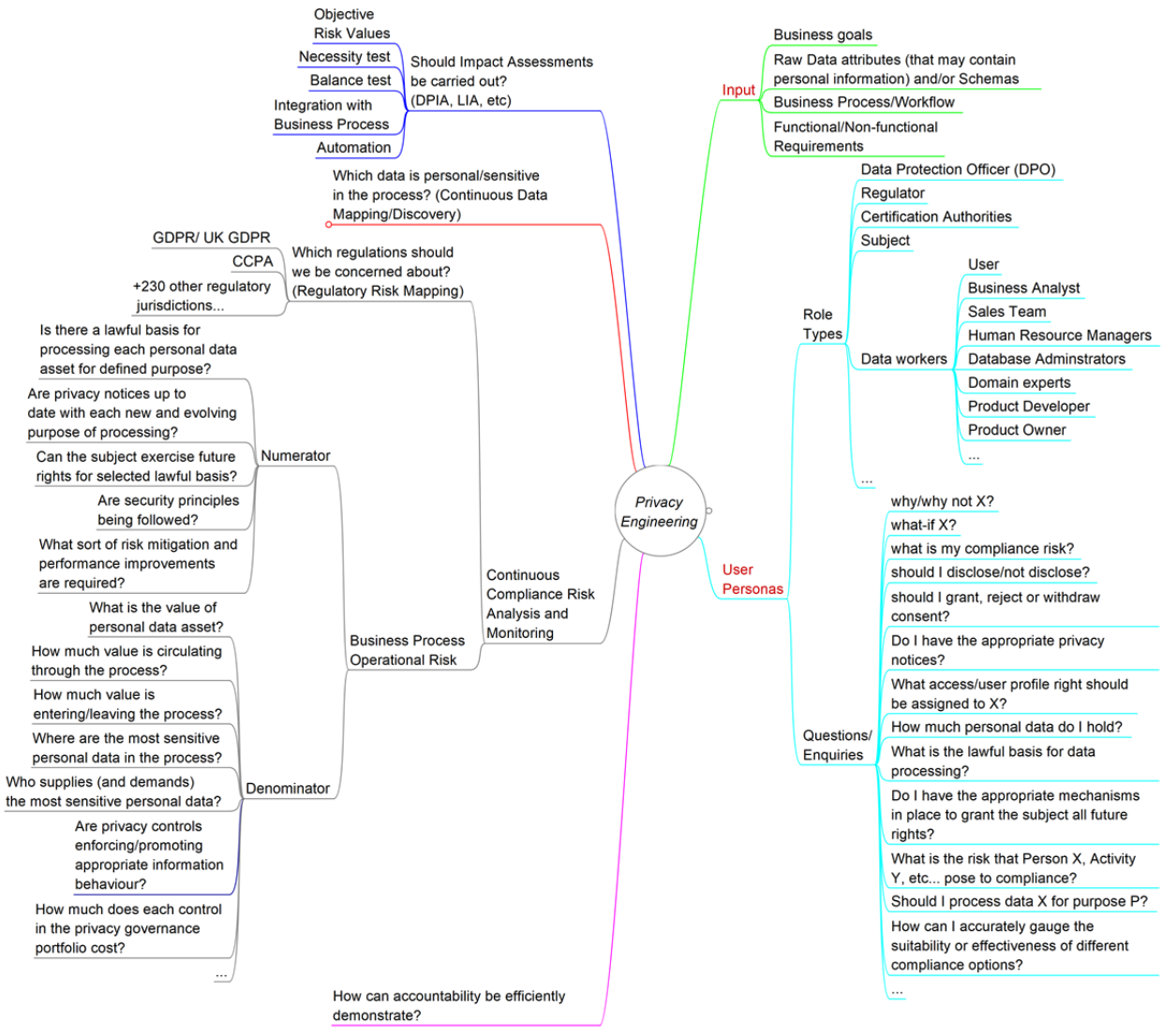

Figure 1: Privacy Engineering Problem Space

Figure 2: Compliance Gap – The misalignment that increases Data Privacy Risk

The state-of-the-art

Existing research primarily mitigates privacy problems by proposing best practices [1]. These practices lack demonstrable pathways to privacy satisfaction, with few accounts on successful implementation due to limited adoption. There are also technology centric solutions [2][3], but studies show that engineers tend to argue in favour of solutions that are match their competencies and experiences, even if they are less optimal [4]. Hence, in situations where a design requirement demands compliance to a privacy regulation, the most that can be expected is limited by the engineers' knowledge and ambitions about privacy [13]. This could explain why technology-oriented solutions such as the Platform for Privacy Preferences Project (P3P) has only gained modest success [5].

Another established privacy engineering approach involves formulating privacy design *strategies* and *patterns* (e.g http://privacypatterns.org). Design strategies express fundamental approaches to achieving privacy goals. Danezis et al. [6] highlighted eight such strategies using terms such as *minimise*, *hide*, *separate, aggregate*, *inform*, *control*, *enforce* and *demonstrate*. Privacy design patterns can then be employed to realize these strategies. But these patterns are very specific and primarily suitable for the last phases of process or system development. At this point, there is little or no room for the engineer to consider design alternatives. Even when they do consider the privacy implications, the software APIs (Application Programming Interfaces) used to implement privacy (such as crypto libraries) can themselves be difficult to use correctly, so leading to privacy breaches [7] or unusable user privacy controls [8].

Finally, it is now common practice to align users' privacy needs and regulatory compliance through privacy notices. An undesirable observation is the inability to use these notices to map back to system behaviour [9], combined with the fact that they have become increasingly hard to interpret, notices are rarely read by users. For example, Google's privacy notice statement has grown from 600 to 4000 words over the past 20 years, and their adjustments in response to GDPR has seen a 30% increase in length (https://policies.google.com/privacy/archive?hl=en-US). The use of policy-based solutions in this way is symptomatic of masking deeply rooted privacy engineering problems with superficial privacy notice statements as solutions.

Although it is easy to see data protection regulations as they apply to the engineering of data intensive systems, they are not necessarily created with such systems in mind. This is because the manner in which existing regulations are written seldom helps abate the problem for the privacy engineer. Often, they are written abstractly to cover a wider audience, making it difficult to measure disclosure risk in a more intuitive way. In other cases, they are formulated as slogans that offer useful explanations of the meaning of privacy, but include little or no information regarding expected systematic and analytic lines of action to achieve such privacy. Arguably, data protection regulations don't submit to straightforward instructions/specifications on privacy that are necessary when building data reliant systems and processes [10].

The privacy engineer needs to ask substantive privacy questions of their users to allow them to implement concrete engineering actions that comply with a regulatory requirement in the design. If regulations provide no guidelines on how such questions may be asked, engineers may find themselves unable to identify and translate privacy requirements of end-users into concrete and verifiable evidence in technology.

Proposed way forward

Irrespective of the regulatory jurisdiction, more robust government policies are required to ease the friction between efficient use of data and compliance. Such policies, while incentivising data collection, use and transfer, should also provoke the investigation of new privacy engineering methods. Particularly, they should incorporate privacy challenges created by the new data dispensation. In conversations with several industry stakeholders, a number of reoccurring pain points were highlighted. These include:

(1) Data intensive organisations generally have very immature data mapping and discovery processes, which cover very small aspects of personal/sensitive data.
(2) In every organisation, although some individuals may care a lot about regulatory compliance, others will not. This creates a weak link.
(3) The majority of IT support, database administrators and software developers lack knowledge of the regulatory implications of their design/development decisions.
(4) A vast majority of COTS (Commercial Off-the-Shelf Software) used in facilitating organisational business processes do not contain necessary data protection features by design.
(5) Product owners/developers/analysts often do not have the required expertise to evaluate regulatory compliance in the systems they build or work with.
(6) There are no privacy risks metrics for benchmarking/measuring organisational business processes.
(7) Data intensive organisations may not have a clear view on how to benchmark the maturity/level of compliance of their processes as the business evolves.
(8) Domain experts often process and/or transfer data without being sure of the consequence on regulatory compliance.
(9) Shared data often exist in an unstructured format and distributed across multiple systems, making deletion management hard.
(10) Organisations have little or no control over how third parties use personal data once transferred.

(11) Developers and database administrators are less likely to think about regulatory risks associated with 3rd party data API integration. Often data is transferred without any form of guarantee.

While this is not an exhaustive list of pain points, it offers a clear indication that existing privacy engineering techniques do not provide the promised regulatory compliance. To address these pain points, we advocate a data use and sharing policy framework that incentivises new privacy engineering solutions with the following six guiding principles at its core:

More focus, less distraction - to automatically translate data protection regulations to fit business day-to-day activities, without the need for data workers to know, hold or work with regulations in depth, giving experts the space to focus on their core activity.

Less complexity, more compliance – using intelligent knowledge-bases to generate specific instructions configured to data intensive processes, cutting through the complexity and sheer volume of content within the regulatory compliance environment.

Intuitive Privacy Helper – providing search functionalities that enable constant access to task specific practical advice, best practice insights, and the knowledge to guide data workers through specific compliance queries at the time of need.

Seamless, just-in-time - by seamlessly translating data protection regulations into business processes, providing data workers and other stakeholders with just-in-time, targeted and actionable regulatory compliance support.

Intelligent Connectors – by leveraging plugin architecture to seamlessly integrates into existing data processes using email, workflow and CRM connectors – providing data workers guidance and insights on the regulatory implications of their day-to-day activities.

Monitoring Compliance – to deal with monitoring, analysing and reporting increasing number of regulations and activities -providing data workers and other stakeholders the feedback necessary to measure compliance levels and demonstrate accountability to the regulator and executives.

These guiding principles offer the benefit of unabated growth in the collection, use and sharing of data, with resulting regulatory compliance concerns continually addressed timely, seamlessly and effortlessly.

January 2022

References

[1] Cavoukian, Ann. "Privacy by design: leadership, methods, and results." *European Data Protection: Coming of Age*. Springer, Dordrecht, 2013. 175-202.

[2] Cranor, L., & Langheinrich, M. (2004). The Platform for Privacy Preferences 1.1 Specification W3C Working Draft.

[3] Inglis, Peter, and Inah Omoronyia. "Analysing Privacy Conflicts in Web-Based Systems." *2021 IEEE 29th International Requirements Engineering*

Conference (RE). IEEE, 2021.

[4] Wiese Schartum, D. (2016). Making privacy by design operative. *International Journal of Law and Information Technology*, *24*(2), 151-175.

[5] Reay, I., Dick, S., & Miller, J. (2009). A large-scale empirical study of P3P privacy policies: Stated actions vs. legal obligations. *ACM Transactions on the Web (TWEB)*, *3*(2), 1-34.

[6] Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J. H., Metayer, D. L., Tirtea, R., & Schiffner, S. (2015). Privacy and data protection by design-from policy to engineering. *arXiv preprint arXiv:1501.03726*.

[7] Patnaik, N., Hallett, J., & Rashid, A. (2019). Usability smells: An analysis of developers' struggle with crypto libraries. In *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)* (pp. 245-257).

[8] Ramokapane, K. M., Rashid, A., & Such, J. M. (2017). "I feel stupid I can't delete...": A Study of Users' Cloud Deletion Practices and Coping Strategies. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)* (pp. 241-256).

[9] Anthonysamy, P., Greenwood, P., & Rashid, A. (2013). Social networking privacy: Understanding the disconnect from policy to controls. *Computer*, *46*(6), 60-67.

[10] Milyaeva, S., & Neyland, D. (2019). Brexit and the Failure of Pre-emptive Reconciliation: A study of the General Data Protection Regulation. *The Sociological Review Magazine*.

[11] Culture Horizon Survey - https://www.privacyculture.com/survey

[12] Data Protection Market Research Report - https://www.marketresearchengine.com/data-protection-market

[13] Omoronyia, I., Etuk, U., & Inglis, P. (2019). A privacy awareness system for software design. *International Journal of Software Engineering and Knowledge Engineering*, *29*(10), 1557-1604.