Computer Law & Security Review 51 (2023) 105873

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

journal homepage: www.elsevier.com/locate/clsr

Using the blockchain to enable transparent and auditable processing of personal data in cloud- based services: Lessons from the Privacy-Aware Cloud Ecosystems (PACE) project

Jose Tomas Llanos [a,*], Madeline Carr [a], Omer Rana [b]

[a] Department of Computer Science, University College London, London, UK
[b] School of Computer Science, Cardiff University, Cardiff, UK

ARTICLE INFO

Keywords:
Data protection
Privacy
GDPR
Blockchain, PET
Data protection-by-design

ABSTRACT

The architecture of cloud-based services is typically opaque and intricate. As a result, data subjects cannot exercise adequate control over their personal data, and overwhelmed data protection authorities must spend their limited resources in costly forensic efforts to ascertain instances of non-compliance. To address these data protection challenges, a group of computer scientists and socio-legal scholars joined forces in the Privacy-Aware Cloud Ecosystems (PACE) project to design a blockchain-based privacy-enhancing technology (PET). This article presents the fruits of this collaboration, highlighting the capabilities and limits of our PET, as well as the challenges we encountered during our interdisciplinary endeavour. In particular, we explore the barriers to interdisciplinary collaboration between law and computer science that we faced, and how these two fields' different expectations as to what technology can do for data protection law compliance had an impact on the project's development and outcome. We also explore the overstated promises of techno-regulation, and the practical and legal challenges that militate against the implementation of our PET: most industry players have no incentive to deploy it, the transaction costs of running it make it prohibitively expensive, and there are significant clashes between the blockchain's decentralised architecture and GDPR's requirements that hinder its deployability. We share the insights and lessons we learned from our efforts to overcome these challenges, hoping to inform other interdisciplinary projects that are increasingly important to shape a data ecosystem that promotes the protection of our personal data.

1. Introduction

Data-driven technological advances are at the centre of the digital transformation our economies and societies have experienced in recent times, bringing about significant benefits in the form of efficiency and innovation. Take the example of the cloud. On account of its lower storage costs, elastic, on-demand service provisioning, enhanced interoperability and insights derived from machine learning, cloud computing has quickly come to dominate online service delivery[1]. Companies across industry segments increasingly rely on cloud vendors' servers and infrastructure to host and operate their websites and mobile apps, whilst cloud platform services are gradually becoming developers' preferred choice to create and deploy middleware and other customised solutions. As a result, data, including personal data, continues to migrate to the cloud, a trend that is unlikely to be reversed in the foreseeable future[2].

On the flipside, the growing amounts of data stored in the cloud, coupled with the complexity of cloud-based services, or ecosystems,

* Corresponding author.
E-mail address: j.llanos@ucl.ac.uk (J.T. Llanos).

[1] This is so much so that according to the consultancy firm Gartner, cloud-native platforms will serve as the foundation for more than 95% of the new digital initiatives by 2025, up from less than 40% in 2021. See Michael Cooney, 'Gartner: Top Strategic Technology Trends for 2022' (*Network World*, 18 October 2021) <https://www.networkworld.com/article/3636972/gartner-top-strategic-technology-trends-for-2022.html> last accessed 09 August 2023.

[2] Accenture observes that the global cloud services industry has grown in over 380% between 2010 and 2020, that it is 'inevitable that more data would invite the need for more data storage', that the COVID-19 pandemic has increased a focus on cloud capabilities, and that around 50% of all corporate data is stored in the cloud as of 2020, See '25 cloud trends for 2021 and beyond' (*Accenture*, 04 March 2021) https://www.accenture.com/nl-en/blogs/insights/cloud-trends last accessed 09 August 2023.

https://doi.org/10.1016/j.clsr.2023.105873

Available online 19 October 2023

raise significant data privacy concerns. Cloud-based services are typically 'layered', involving a chain of cloud service providers and other components[3]. For example, an end-user content-streaming cloud-based application may run on top of a cloud platform which is in turn hosted on a cloud infrastructure[4], with the application itself being a mash-up of other services running on different cloud-based infrastructures. Every component in an ecosystem of this kind processes personal data for multiple purposes, yet individuals are seldom aware of cloud ecosystems' highly intricate and layered architecture. This raises a problem of transparency and accountability. Individuals interact only with a Web interface rather than the larger, composite ecosystem, entrusting their personal data and identity to the consumer-facing component without realising that the cloud-based application may share their data with several back-end services (e.g. providers of cloud-hosted analytics and online advertising). In this opaque context, it is hard for data subjects to exert any control over their personal data[5] (i.e. to exercise 'individual control') - one of the main concerns of EU data protection law[6].

As highlighted by the European Parliament report on blockchain, this technology has the potential to promote transparency, accountability and control over personal data[7]. Thus, solutions building on blockchain can be in theory leveraged to enable the emergence of trustworthy cloud ecosystems. In furtherance of this vision, a group of computer scientists, social scientists and legal scholars in the Privacy-Aware Cloud Ecosystems (PACE) project is elaborating a technological stack designed to enhance transparency and facilitate compliance with the EU General Data Protection Regulation (GDPR)[8] in multi-layered applications hosted over the cloud (the PACE Tool). This stack is also intended to give end-users some degree of control over their personal data. The PACE Tool relies on virtual containers to monitor and log data flows within a cloud-based service, and on the immutability feature of blockchain technology to create a reliable audit trail for the verification of compliance with GDPR requirements.

The PACE project has two overarching goals. First, to develop a blockchain-based automated system for enforcing and auditing compliance with data protection rules. Second, to critically evaluate the practicalities of enforcing the GDPR through blockchain-based solutions, and thus be able to determine whether the blockchain lives up to its promises. These two goals are vitally important because, firstly, on account of the current scale of deployment of data-driven systems and the growing amounts of data being produced, the automated enforcement of data protection rules could improve the overall levels of GDPR compliance and bring about substantial time and cost savings for data protection authorities (DPAs). And secondly, because although the blockchain carries the promise of affording individual control, transparency and accountability without a 'trusted intermediary', we are not aware of any implementation showcasing the successful achievement of these goals. By having legal scholars and computer scientists collaborate so closely on the development of the PACE Tool, we have been able to comprehensively test and evaluate the extent to which GDPR enforcement can be automated through blockchain technology, and thereby distinguish hype from reality. Unfortunately, we have significantly more challenges than successes to report.

In particular, first, we have found that there are substantial barriers to effective collaboration amongst researchers having largely dissimilar backgrounds such as law and computer science. Different ways of reasoning and understandings of the same concepts – such as data protection-by-design – as well as different expectations as to what technology can do for data protection law compliance make communication between the two fields difficult. As a result, work tend to occur in siloes, without input from the other side - a trend liable to result in undesirable outcomes. Second, although the "code is law" idea is certainly appealing to assist under-resourced DPAs and tackle long-standing data protection law enforcement challenges, automating the application and verification of compliance with data protection rules requires encoding them in a manner that accurately represent their meaning and scope, which is highly difficult due to their open-textured nature and flexibility. This challenge has meant that the automated GDPR enforcement goal of the PACE project was unrealistic – and by extension, based on our experience, the blockchain's promises relating to GDPR enforcement are overstated. Ultimately, automating legal provisions is only feasible insofar as they are simple and of straightforward application, which tends not to be the case of most substantive data protection rules the application of which typically involves a balancing exercise. As a consequence, we were forced mid-project to make substantial changes to the PACE Tool's design and objectives, switching away from our original goal of building a tool capable of *hardcoding* the application of legal bases onto efforts to build a tool capable of *guiding* controllers in the correct application of legal bases instead.

Third, more broadly yet not less importantly, there are significant practical and legal challenges that militate against the implementation of the PACE Tool. From a practical perspective, researchers can continue devoting substantial efforts to devise solutions to address the threats and harms to our privacy and associated fundamental rights and freedoms arising from the ubiquitous data-driven technologies deployed in the digital economy; however, the fact remains that the digital economy is surveillance-based, data-hungry and profit-driven, and consequently industry players have little to no incentive to implement any of such solutions, including our PACE Tool. Without any concrete business case for privacy, any privacy-driven initiative must be introduced top down by regulators and forced upon industry players to stand a chance of success. Further, although Turing-complete blockchain networks such as Ethereum can support highly advanced, smart contract-based applications, some of these applications – like our PACE Tool – can prove highly computationally intensive and thus prohibitively expensive to deploy. On the other hand, from a legal perspective, there are important clashes between permissionless blockchains' decentralised architecture and GDPR requirements that are premised on centralised data processing assumptions. As a result of these clashes, we were confronted with a binary choice with no satisfactory outcome: either to deploy a GDPR-non-compliant PET where controllership cannot be determined, or to choose a blockchain architecture that compromises the PET's security and integrity assurances.

Overall, as anticipated above, there are many pressing challenges that hinder the PACE Tool's deployability, scalability, and widespread adoption, yet our interdisciplinary work has not been in vain. The PACE Tool still promotes important objectives of EU data protection law such as transparency, accountability and individual control – albeit in a way and to an extent other than what we originally conceived. Further, most

[3] The average online publisher is embedded with a set of third-party components that include user analytics, UX capture, advertisement, authentication, captcha, performance and cybersecurity, maps and location, search, sales and customer relation management, payment, shipping, reviews, sharing and social media functionality, comment boxes and more. See Seda Gurses and Joris Van Hoboken, 'Privacy after the Agile Turn', *Evan Selinger, Jules Polonetsky and Omer Tene (eds), The Cambridge Handbook of Consumer Privacy* (Cambridge University Press 2018) 587.

[4] This corresponds to the three main types of cloud provisioning models, i.e. *Software-as-a-Service* (SaaS), *Platform-as-a-Service* (PaaS) *and Infrastructure-as-a-Service* (IaaS).

[5] See Article 29 Working Party, 'Opinion 05/2012 on Cloud Computing' (WP 196, 1 July 2012) 5.

[6] See e.g. GDPR, recital 7: 'Natural persons should have control of their own personal data'.

[7] European Parliament, 'Report on Blockchain: A Forward-Looking Trade Policy (AB-0407/2018)' (2018) para 14 <https://www.europarl.europa.eu/doceo/document/A-8-2018-0407_EN.pdf> last accessed 09 August 2023.

[8] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

crucially, we have learned important lessons that we feel compelled to share to warn other privacy-oriented researchers about the overstated power of technological solutions to tackle complex socio-economic problems, and also about some practical implementation challenges they are bound to encounter.

The goal of this article is to introduce the technology we are developing to enable trustworthy cloud-based websites and applications, and to share the main challenges and lessons of our interdisciplinary effort to automate the enforcement of data protection rules. To this end, we proceed as follows. Section 2 presents a real life-inspired example of a multi-layered cloud-based online pharmacy. This example serves to highlight the main challenges arising from cloud ecosystems' complexity from the perspective of data subjects and DPAs. In addition, it is intended to facilitate the reader's understanding of the PACE Tool's design and functionalities, as explained in the following sections. Moving forward, Section 3 first sets out an overview of the two technologies that form the backbone of the PACE Tool - a container-based monitoring system and a blockchain – and then explores the PACE Tool's architecture and functionalities. This is followed by Section 4, where we discuss the challenges that we have faced in the development and implementation of the PACE Tool. After discussing these challenges, we explore the approaches we have followed to overcome them, and the insights and lessons we have learned from our interdisciplinary work. Finally, Section 5 wraps up the discussion with some conclusions.

2. Data protection issues in the cloud: a cloud-based online pharmacy

To facilitate the understanding of both the issues arising from multi-layered cloud-based services and the PACE Tool's operation and functionalities, let us consider a concrete though fictitious example of a cloud-based online pharmacy, in which we trace a transaction and its associated data processing operations.

When a user visits the online pharmacy website to place an order, there are multiple transfers of personal data to the online pharmacy's different components, which are remarkably difficult to ascertain based on the information provided in the online pharmacy's privacy policy, and impossible to see in the literal sense (see Fig. 1 below).

The pharmacy requests *inter alia* the user's name, address, date of birth, an electronic version of the prescription and payment details. The pharmacy uses a non-EU-based IaaS vendor (Cloud4U) to host and operate its website and mobile app. Thus, the aforesaid data is transferred to Cloud4U's servers, which are physically located throughout a non-EU based country. The pharmacy also subcontracts payment and shipping service providers to handle the payment and delivery of orders, and consequently transfers to them the personal data required for these purposes. In addition, the pharmacy's website and mobile app are embedded with so-called 'social plugins' (a 'Like' button and a 'Share' button) from a leading social network (Friendface), which collects highly granular personal data for multiple purposes. Further, the pharmacy uses the real-time bidding (RTB) system of an online advertiser and intermediary (Froogle) to sell advertising inventory space and thus derive another revenue stream – which involves the placement and synching of RTB cookies on users' devices to broadcast highly granular personal data to hundreds of companies in the ad tech chain. Lastly, the pharmacy uses the tools of Fluffy Analytics, which collects personal data for fraud detection, security, business intelligence and service improvement purposes.

Personal data processing operations must serve a 'specified, explicit and legitimate' purpose that has to be informed to the data subject prior to the processing, and also be duly legitimised by a legal ground[9]. In practice, in an opaque setting as above, data subjects typically cannot be aware of what data relating to them is processed by which entity, and for what purposes. Currently, we must tick a box before using a service, a type of interaction that purportedly represents that we have read, understood, and fully agreed with the service operator's privacy policy. This contractual document is typically long, complex, vague, and confusing, and thus fails to accurately depict the actual practices of the service provider, such as how many entities will have access to the user's data, where those entities are located, or any unexpected uses of the data. Given time constraints and a seemingly endless amount of consent requests we are confronted with on a daily basis, we seldom read these documents and just proceed to tick the box to use the service[10]. However, even if we had all the time in the world to read them – and arguably also a law degree - we would still be in the dark as to what happens to our data. After initial disclosure, there is no way to know with certainty whether our personal data is processed in accordance with our privacy preferences and in compliance with the applicable law.

If an interface forces users to read an excessively long and complex text, fully agree with it without room for granularity, and also without any means to help users to understand its content, the consent obtained from that interface cannot be informed. Rather, the consent represented by the box-ticking action is a veneer of choice, as any sort of control purposedly involved in this exercise is illusory: we increasingly agree to whatever terms are presented to us and perceive not having control over our personal data and identity as an inevitable outcome of present-day life. Indeed, empirical research has found that a significant portion of us feels resigned to this lack of control[11].

In turn, Data Protection Authorities (DPAs) are entrusted with the task of policing all the data processing operations performed by their countries' data controllers, as well as those concerning their countries' residents irrespective of the relevant controller's place of establishment. This is an overwhelming undertaking, not least on account of DPAs' infamously known limited budgets, staffing and resources[12], which prevents them from completing investigations within a reasonable timeframe and makes proactive investigations or the expansion in the scope of complaints less likely[13]. As a result, many infringements are likely to escape scrutiny, particularly when they do not attract media coverage yet involve substantive forensic efforts to ascertain how personal data has been actually processed – such as the case of the online pharmacy outlined above.

Against this background, the PACE Tool was theoretically conceived to make improvements in individual control, transparency and accountability through three main mechanisms: making the purposes of processing, the personal data processed for the fulfilment of each purpose, and the legal ground based on which each processing operation is carried out clear and visible. In this way, data subjects may give or deny their consent to each processing operation, or exercise their right to object to processing, as applicable;

[9] GDPR, Articles 5(1)(b), 6 and 13(1)

[10] Caroline Cakebread, 'You're Not Alone, No One Reads Terms of Service Agreements' (*Business Insider*, 15 November 2017) <https://www.businessinsider.com/deloitte-study-91-percent-agree-terms-of-service-without-reading-2017-11> last accessed 09 August 2023.

[11] Joseph Turow, Michael Hennessy and Nora A Draper, 'The Tradeoff Fallacy: How Marketers Are Misrepresenting American Consumers and Opening Them up to Exploitation' (2015) A Report from the Annenberg School for Communication, University of Pennsylvania.

[12] Johnny Ryan and Alan Toner, 'Europe's Governments Are Failing the GDPR - Brave's 2020 Report on the Enforcement Capacity of Data Protection Authorities' (2020) 7-10; see also European Data Protection Board, 'First Overview on the Implementation of the GDPR and the Roles and Means of the National Supervisory Authorities' (Report presented to the European Parliament's Civil Liberties, Justice and Home Affairs Committee (LIBE), 26 February 2019) 7.

[13] Ido Sivan-Sevilla, 'Varieties of Enforcement Strategies Post-GDPR: A Fuzzy-Set Qualitative Comparative Analysis (FsQCA) across Data Protection Authorities' (2022) Journal of European Public Policy 1, 8.

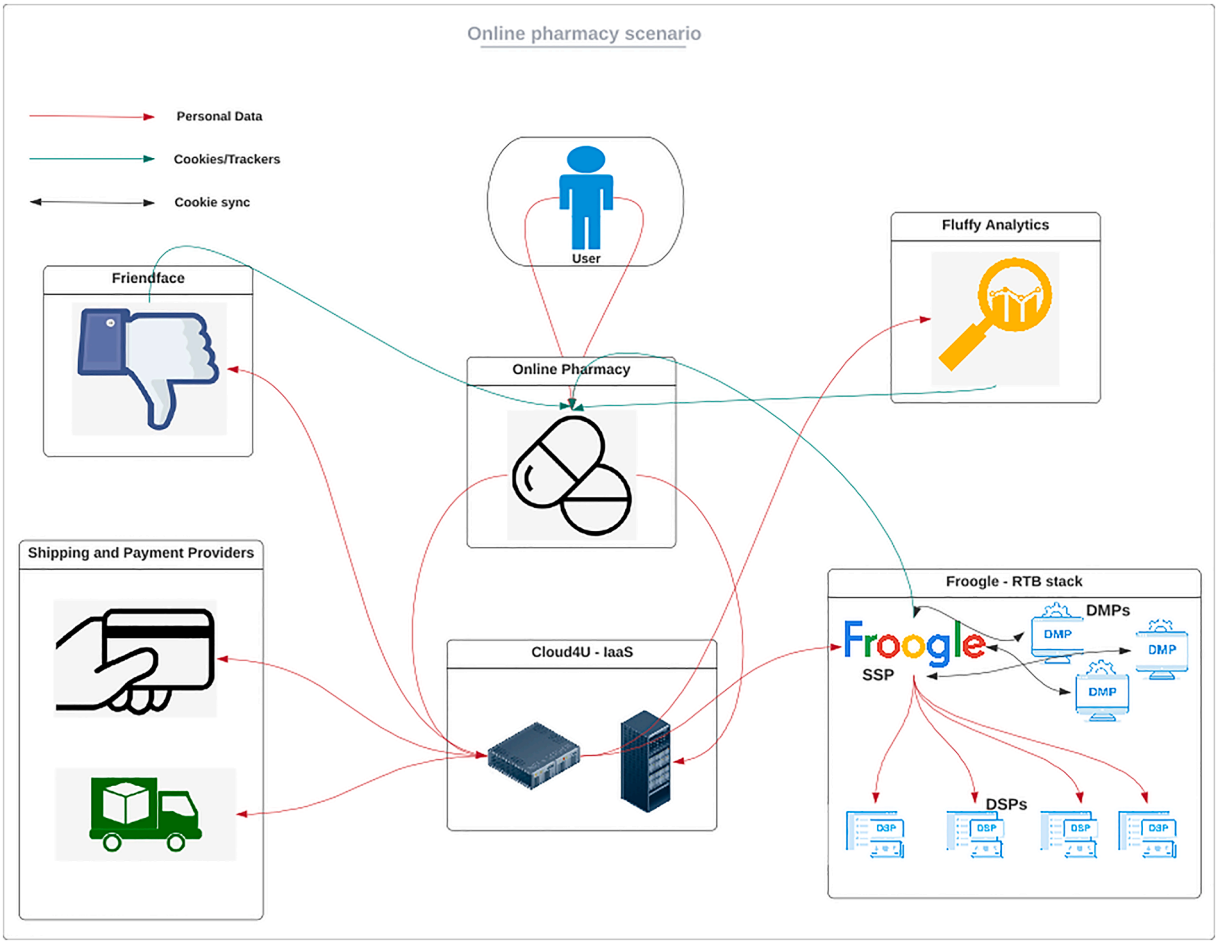

Fig. 1. Cloud-based Online Pharmacy Scenario

- monitoring the different data processing operations that take place within a cloud ecosystem and recording them in a reliable and tamper-proof fashion; and
- automating the verification of compliance with the GDPR by the relevant cloud ecosystem's components.

Two main technologies were chosen to design, build and implement the mechanisms above: virtual containers and the blockchain. It is important to note at this point that, as anticipated in the introduction, we faced certain challenges which meant that not all of the above-mentioned mechanisms proved practicable. In particular, after the realisation that automating legal provisions is only feasible for simple rules, we were forced to reconsider the PACE Tool's original design, moving away from our original hardcoding ideal onto efforts to guide controllers in the correct application of legal bases, also providing functionalities that allow for reliable audits of data processing operations and enhance individual control to an extent feasible[14].

3. The backbone of the PACE Tool: containers and blockchain

3.1. Containers

Generally speaking, a container is a mechanism to perform virtualisation. Virtualisation is a process whereby software is used to create an abstraction layer over computer hardware that allows the hardware elements of a single computer to be divided into multiple virtual computers[15] or 'virtual machines' (VMs) – i.e. several emulations of a physical computer. Virtualisation is a critical part of system optimisation efforts which brings about substantial benefits such as reduction and simplification of server infrastructure, enhanced reliability (by e.g. isolating software faults) and higher security (by e.g. containing digital attacks through fault isolation)[16].

Containers are a lighter-weight, more agile way than VMs of handling virtualisation[17]. They contain everything needed to run a single application or microservice, including all the code, its dependencies and even the operating system itself. This enables applications to run almost anywhere - a desktop computer, a traditional IT

[14] See sections 4.1 and 4.2 below.

[15] On virtualisation see generally Davit Williams, 'An Introduction to Virtualization', *Virtualization for security: including sandboxing, disaster recovery, high availability, forensic analysis, and honeypotting, edited by John Hopes* (Syngress 2009).

[16] ibid 13–14.

[17] For a more in-depth explanation on the differences between containers and VMs see Alexander Kropp and Roberto Torre, 'Docker: Containerize Your Application', *Computing in Communication Networks* (Elsevier 2020) 232–233.

infrastructure, or the cloud. Containers are faster in terms of resource provisioning, more efficient, and produce less overhead as compared to VMs[18]. Also, containers are portable, so applications running in containers can be easily migrated onto different platforms or environments.

On account of their features, containers are a perfect match for complex multi-cloud environments. In cloud-based architectures, containers are normally used to monitor and track performance and system vulnerabilities, and store this information for future verification if required[19]. For the PACE Tool, we leveraged containers' features to monitor and record the data processing operations that are triggered when a user interacts with a cloud-based service (see more details in Section 3.3 below).

3.2. Blockchain

Broadly speaking, a blockchain is an append-only database (or *ledger*) composed of sets (*blocks*) of cryptographically signed transactions that are stored on, shared, and synchronised amongst multiple network participants (*nodes*) based on a consensus algorithm. At the most fundamental level, blockchains give users confidence that stored information (for example, an account balance or property certificate) has not been *tampered with*, thus ensuring a 'single truth' across different participants who may or may not trust each other. Thus, it is commonly said that blockchains are *immutable*; however, in reality they can be modified, although it is very hard to do so - especially the so-called *public* and *permissionless* blockchains[20].

Blockchains rely heavily on 'hash values' and 'references'. Hashing is the process of putting data of arbitrary size (i.e. any data input, such as a video, an image or text) through a mathematical algorithm (the cryptographic hash function). The output of this process (the hash value) is a bit string of a fixed size that is unique to the input data. Hash functions are designed to be one-way and collision resistant, that is, it is computationally infeasible (i.e. practically impossible) to find both any input that maps to any pre-specified output and two or more inputs producing the same hash value.[21] If the original input is altered in the slightest (even one character), the hash function renders a totally different hash value.[22] Therefore, insofar as the hash value remains unaltered, external observers can be certain that the input data has not been changed.[23] In a blockchain, every block has a unique hash value which results from a combination of the block's transactions and the hash value of the previous block,[24] thus creating a 'block chain' that goes back to the genesis (the first) block. This chain is tamper-evident.[25] Given that any alternation of any transaction included in a dataset will invariably change said dataset's hash value and each dataset's hash value is partly built upon the previous dataset's hash value, any such alteration will inevitably disrupt the link between the altered dataset and the following ones.

Blockchains run on multiple nodes comprising a dense peer-to-peer (P2P) network, so there is no central point of failure or attack at the hardware level.[26] Each node holds a copy of the ledger and is able to generate, digitally sign and validate transactions – i.e. verify that the digital signature is correct and that there are no conflicts with previous transactions. Verification is based on asymmetric-key cryptography, which uses two mathematically related keys to encrypt and decrypt data. Data encrypted (i.e. the cypher text) with one of these keys can only be decrypted with the other key, and vice versa.[27] Every blockchain user has a pair – commonly known as public and private keys. Public keys are used to derive user addresses (or accounts), and serve as the user's public identity on the blockchain.[28] Private keys, conversely, are used to authorise (sign) and validate transactions. Whomsoever is transferring a data item must prove that they do intend to complete such transfer, and those verifying the transactions must be able to corroborate that intent. To this end, the transferor has to digitally sign the transaction, which involves encrypting the transaction data with their private key.[29] If the transferor's public key effectively decrypts the data, this proves that the transferor holds the private key,[30] thus confirming the transaction's authenticity. All other nodes on the blockchain can verify the transaction by using the transferor's public key.[31]

Nodes forward on verified transactions to their peers, and at periodic intervals special nodes - 'miners' – assemble candidate blocks by grouping together a set of verified yet unconfirmed transactions.[32] Upon assembling a new block, the miner broadcasts it to the blockchain network so the other nodes proceed to validate it, that is, they verify that the block meets the consensus protocol's specifications.[33] Blocks are accepted only if they contain valid transactions which do not conflict with each other or with those within previous blocks.

Every blockchain employs a type of strategy (the consensus protocol) to ensure that no malicious individual or small group of nodes can take control over the network and manipulate the ledger. Public and permissionless blockchains rely on 'proof-of-work' (PoW), which involves a competition to solve a mathematically difficult puzzle[34]. The winner gets to generate a new block and claim a reward – newly minted coins. The puzzle can be solved only by trial and error, which consumes a lot of computational power, time, and electricity[35]. Thus, nodes with greater computational power and incurring higher electricity costs are more likely to solve the puzzle first. The economic incentive to mine new blocks and the costly nature of such activity ensures the blockchain's security. Making any alteration at any point in the blockchain requires that all hash values from that point onwards be recalculated,[36] and a

[18] Masoud Barati and others, 'Privacy-Aware Cloud Auditing for GDPR Compliance Verification in Online Healthcare' (2021) 18 IEEE Transactions on Industrial Informatics.

[19] ibid.

[20] In public and permissionless blockchains, anyone can join the network, run a node and mine new blocks.

[21] Dylan Yaga and others, 'Blockchain Technology Overview' (2018) Draft NISTIR 8202 12.

[22] Jean Bacon and others, 'Blockchain Demystified: A Technical and Legal Introduction to Distributed and Centralized Ledgers' (2018) 25 Rich. JL & Tech. 1, 9–10.

[23] ibid 10.

[24] ibid 11.

[25] ibid 12.

[26] Michèle Finck, *Blockchain Regulation and Governance in Europe* (Cambridge University Press 2019) 7.

[27] Daniel Drescher, *Blockchain Basics* (Springer 2017) 96.

[28] Yaga and others (n 21) 14.

[29] Bacon and others (n 22) 15.

[30] ibid.

[31] Drescher (n 27) 100.

[32] Yaga and others (n 21) 24.

[33] Finck (n 26) 20.

[34] To dispense with highly intensive computations an attain energy efficiency, alternative consensus protocols have been put forward, the most salient of which being the Proof-of-Stake (PoS). PoS is a way to prove that validators have put something of value into the blockchain network that can be destroyed if they act in a dishonest way. Validators typically stake capital in the form of cryptocurrency, and are then responsible for checking that new blocks propagated over the network are valid, occasionally creating and propagating new blocks themselves. For a more detailed explanation of PoS, see 'Proof-of-Stake (POS)', <https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/> last accessed 9 August 2023. For a comprehensive overview of consensus protocols, see Shehar Bano and others, 'SoK: Consensus in the Age of Blockchains', AFT '19: Proceedings of the 1st ACM Conference on Advances in Financial Technologies, October 2019

[35] Drescher (n 27) 91.

[36] Daniel Conte de Leon and others, 'Blockchain: Properties and Misconceptions' (2017) 11 Asia Pacific Journal of Innovation and Entrepreneurship 286, 290.

malicious node would need to be in control of the majority of the network's hashing power (a so-called 51% attack[37]) to steadily solve PoW puzzles first and thereby be able to 're-write' the blockchain[38]. This is, however, a prohibitively expensive strategy, only bound to become more expensive the more blocks are added to the blockchain.

In *private* and *permissioned* blockchains, conversely, access permissions are more tightly controlled, although they still retain many of the authenticity verification mechanisms and the distributed architecture of public blockchains[39]. Two types can be distinguished. First, consortium blockchains, where the ability to verify transactions and add blocks is restricted to a pre-selected set of nodes[40], and the right to read the blockchain may be public or restricted to the participants. Second, fully private blockchains, where only one central organisation has the power to add new blocks, and read permissions may be public or restricted to an arbitrary extent[41]. Compared to public blockchains, changing the rules of the blockchain, reverting transactions or modifying balances is significantly easier: the consortium or company running a private blockchain does not have to invest in computational power to this end. Rather, a majority of participants need to simply agree on the terms of the change, and then 'allow the chain to continue as if nothing happened'[42]. Thus, 'immutability' in private blockchains is not grounded in PoW puzzles, but in the good behaviour of a majority of predefined validator nodes, backed by contracts and potentially adjudication in legal proceedings[43].

After their appearance as the underlying technology of Bitcoin, blockchains soon became a general-purpose technology, enabling a wide range of applications. For example, the terms of a contract can be encoded into the blockchain's operations, and their execution takes place automatically upon fulfilment of pre-defined conditions without reliance on third parties to enforce the transaction (a so-called 'smart contract)[44]. Since smart contracts are run on a blockchain network, they have certain distinguishing features as compared to other types of software. Firstly, the program itself is recorded on the blockchain, so it benefits from the blockchain's characteristic tamper-proof nature and censorship resistance.[45] Once the smart contract is recorded as a transaction on the blockchain, it cannot be reversed. Secondly, and most importantly, the program is executed by the blockchain, so it will always execute as programmed[46]. Put in other words, as contract performance is 'hardcoded', contractual breaches are impossible[47] - although from a coding perspective only[48].

As seen in Section 3.3 below, smart contracts are relied upon for both producing the audit trail of data processing operations and verifying GDPR compliance by a cloud ecosystem's different components.

3.3. Overview of the PACE Tool

In what follows, we provide a simplified explanation of the PACE Tool's architecture and functionalities.

3.3.1. Recording users' privacy preferences

When a user installs the PACE Tool in the device of her choice, she gains access to a *privacy manager* interface, where she can see each purpose of processing pursued by each component of the cloud-based service, along with the applicable legal ground and the categories of data the processing of which is intended. Here, individuals can give/deny their consent with granularity, or alternatively exercise their right to object to processing, as applicable (see Fig. 2).

The setting of privacy preferences in the *privacy manager* interface depicted in Fig. 2 involves a smart contract-based *ratification* phase between the main controller and the data subject before service delivery and any data processing. A sequence diagram representing the protocol of this phase is illustrated in Fig. 3. In particular, the cloud-based service operator deploys a smart contract called *privacy preferences*, and activates a function called *purposes* in order to send data processing purposes-relevant information[49] into the Ethereum blockchain as *privacy-preference logs*. This data processing purposes-relevant information determines what options the user has on her *privacy* manager. The data subject is then provided with the deployment address of the smart contract, whereupon she can activate the function *vote*; in this way, she is able to retrieve and observe the purposes of data processing (which are shown in the manner depicted in Fig. 2), and on this basis 'vote' on them – i.e. give/deny consent or object/not object to processing. The outcome of this decision is stored in the smart contract (see Figs. 4 a and b below) and then recorded on the blockchain as *privacy-preference logs* after validation by trusted nodes. This enables future automated verification of whether users' privacy preferences were respected or overridden[50].

3.3.2. Monitoring system

After the ratification phase, the container-based monitoring system is activated[51]. This system tracks the different instances of data processing underpinning the cloud components' operations, and records them on the blockchain. Containers are hosted on the servers of each component of the cloud-based service – i.e. one container per component.

[37] Finck (n 26) 21.

[38] However, it is not strictly necessary to hold 51% of the hashing power to successfully re-write a blockchain, as the attack's likelihood of success also hinges on the number of blocks in the blockchain to be re-written and the number of confirmations of the last valid transaction by validating nodes. For a detailed explanation of the likelihood of success of hashrate-based attacks, see Meni Rosenfeld, 'Analysis of hashrate-based double-spending' (2014), arXiv preprint arXiv:1402.2009

[39] Vitalik Buterin, 'On Public and Private Blockchains' (*Ethereum Foundation Blog*, 6 August 2015) <https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/> last accessed 09 August 2023.

[40] For example, in a blockchain composed of 15 financial institutions, each institution operates a node, and 10 of them must sign every block for the block to be valid. ibid.

[41] ibid.

[42] Gideon Greenspan, 'The Blockchain Immutability Myth' (*CoinDesk*, 9 May 2017) <https://www.coindesk.com/blockchain-immutability-myth> last accessed 09 August 2023. For example, in the consortium blockchain composed of 15 financial institutions above, the 10 nodes required to add a new block could in turn have the power to replace an old block.

[43] ibid.

[44] For example, a lender and a borrower can program a smart contract under which collateral kept on the blockchain (e.g. a certain amount of cryptocurrency) is transferred automatically to the lender if the borrower does not satisfy her payment obligation by certain date.

[45] Josh Stark, 'Making Sense of Blockchain Smart Contracts' (*CoinDesk*, 4 June 2016) <https://www.coindesk.com/making-sense-smart-contracts> last accessed 09 August 2023.

[46] ibid.

[47] Kevin Werbach and Nicolas Cornell, 'Contracts Ex Machina' (2017) 67 Duke LJ 313, 332.

[48] Contractual breaches are impossible in the sense that a smart contract will not do something which it is not supposed to do, technically, a 'breach'. In reality, given that automating complex provisions is largely unfeasible – see sections 4.1 and 4.2 below – contractual obligations may not be coded in an accurate and comprehensive way. Under these circumstances, the performance of an obligation via a poorly coded smart contract which does not quite capture a legal position may well amount to a contractual breach.

[49] This information includes: the cloud components' identity (*p*), the type of data processing operation each component intends to execute (*po*), the types of personal data items involved in each processing operation (*pd*), and the purposes of processing (*pur*).

[50] See subsection *Automated verification of GDPR compliance* below.

[51] For details of the monitoring system see Gagangeet Singh Aujla and others, 'COM-PACE: Compliance-Aware Cloud Application Engineering Using Blockchain' (2020) 24 IEEE Internet Computing 45.

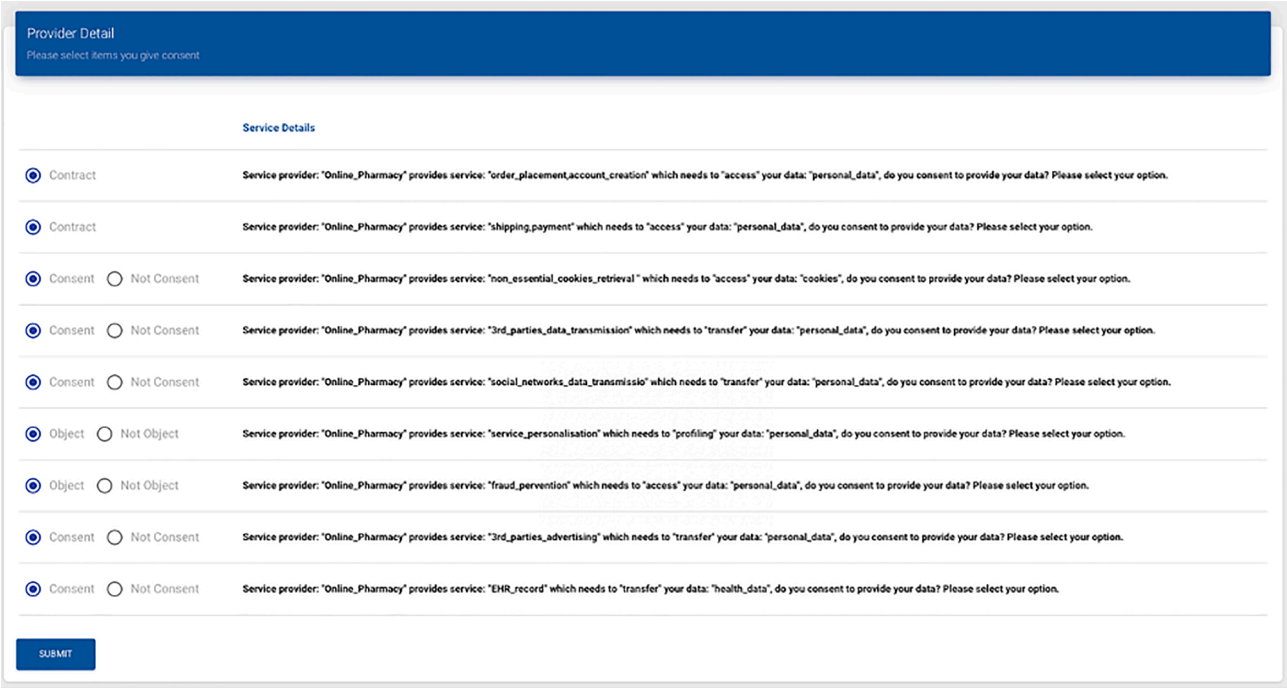

Fig. 2. PACE Tool's *privacy manager* interface

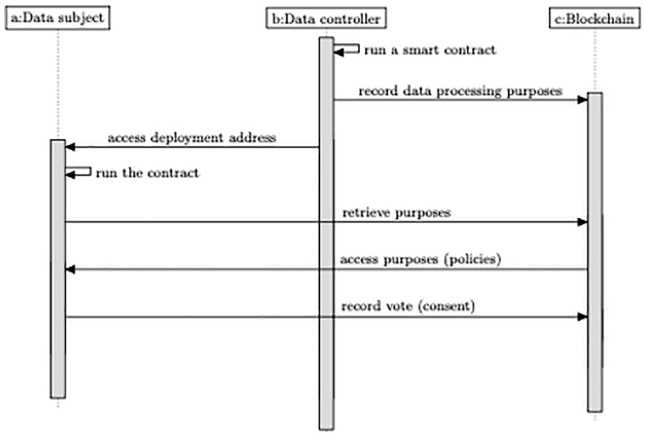

Fig. 3. Protocol for the ratification phase

Each container contains a lightweight software called *GDPR-Agent* which captures the 'events' generated by the relevant component – that is, statistics and details of data processing operations. These events are then sent to a collection engine and thereafter to a filtering engine, both of which are hosted on the *GDPR-Manager* - another lightweight software hosted on the cloud-based service operator's server that is in charge of managing all containers' *GDPR-Agents*. The *GDPR-Manager* filters out GDPR-specific metrics from the data collected by the *GDPR-Agents* and sends them to the Ethereum Blockchain as *container-logs* – i.e. they are added as a transaction and ultimately as a block[52]. The monitoring system is illustrated in Fig. 5 below.

3.3.3. Automated verification of GDPR compliance

Once the blocks containing the *privacy-preference logs* and *container logs* are added to the blockchain, anybody with the required credentials (e.g. the data subject, the controller or a DPA) can deploy smart contracts (called *verification*) to verify GDPR compliance by the cloud ecosystem's different components.

For example, one of the *verification* smart contract's functions is *privacy preferences*. When the smart contract is deployed, trusted nodes can retrieve the *privacy-preference log*'s content, which include: the cloud components' address (*p*), the type of data processing operation each component intends to execute (*po*), the types of personal data items involved in each processing operation (*pd)*, the purposes of processing (*pur*), and the data subject's 'vote' on these purposes (consent/denial of consent, objection or no objection to processing, represented by *pref)*. Based on this information, a violation of an individuals' privacy preferences – and by extension of the GDPR – is flagged if: a component (*p)* executes a data processing operation (*po)* and/or processes personal data other than (*pd*) in contravention with (*pref*)[53].

The details of the other smart contract functions to determine GDPR compliance are explored in Section 4.2.

4. The PACE project: Challenges, lessons and insights

Developing the PACE Tool has proved highly challenging. As anticipated in the introduction, we have found barriers to effective collaboration amongst researchers of vastly different fields such as computer science and legal studies, and struggled with the different expectations between these fields as to what blockchain technology can do for data protection law compliance. After coming to terms with the infeasibility of hardcoding substantive data protection rules, we were forced to reconsider the PACE Tool's original design, replacing our hardcoding ideal with efforts to guide controllers in the correct application of legal bases, facilitate reliable audits of data processing operations, and enhance individual control to a practicable extent. Further, when we tested and implemented the PACE Tool, we found practical challenges. Most actors in the data-driven economy have no incentive to install a container on their servers to have their data processing operations monitored, as this threatens their profitability. Consequently, the PACE

[52] The content of *container-logs* is detailed in section 4.2 below.

[53] This could be the case, for example, if a component processes personal data in spite of the data subject's denied consent.

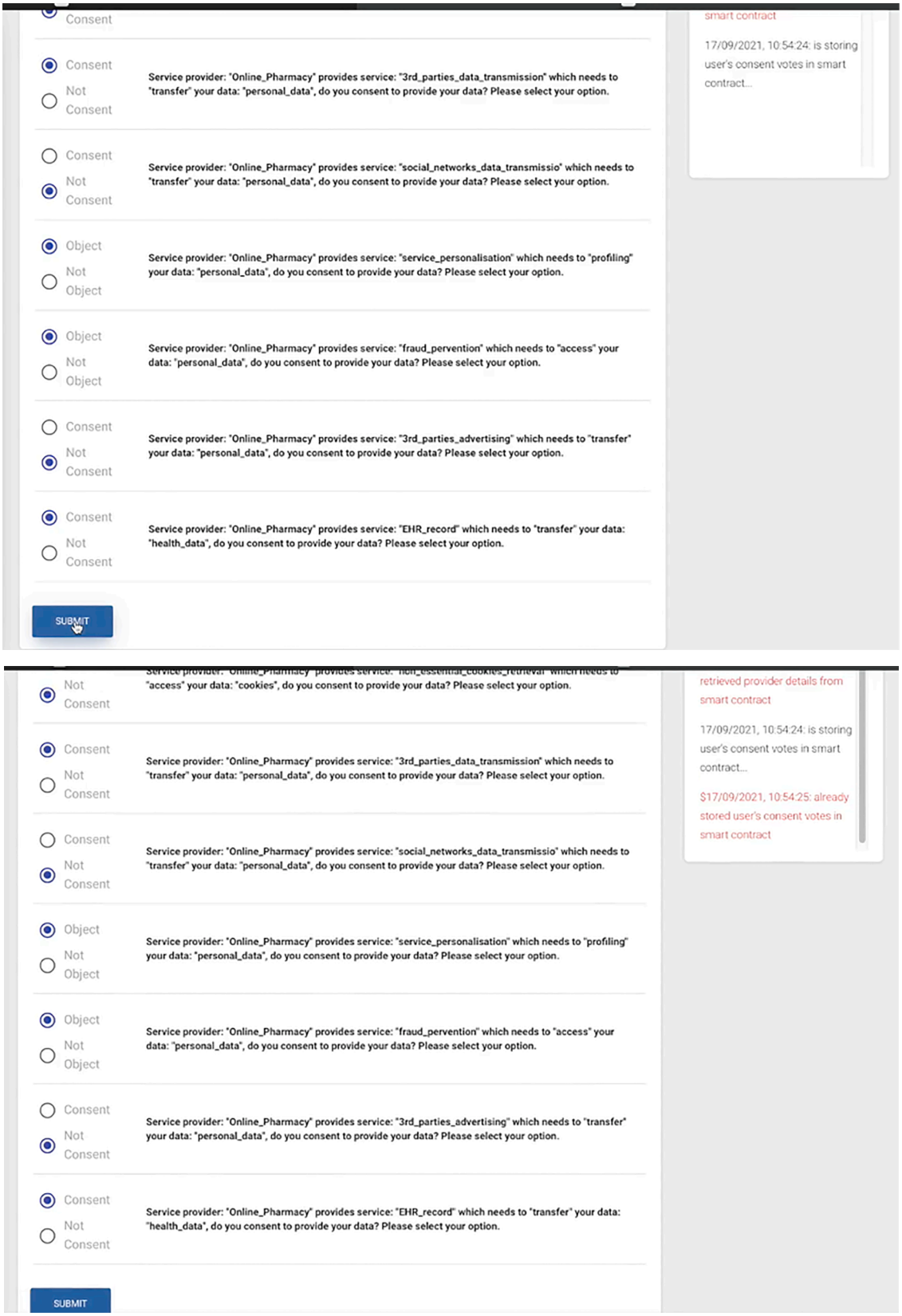

Fig. 4. a and b: Recording of privacy preferences on the blockchain. a. Upon clicking submit, privacy preferences are stored in the smart contract, as seen in the right-hand side of the screen, and later recorded on the blockchain after validation by trusted nodes. b. Confirmation of recording of privacy preferences on the right-hand side of the screen.

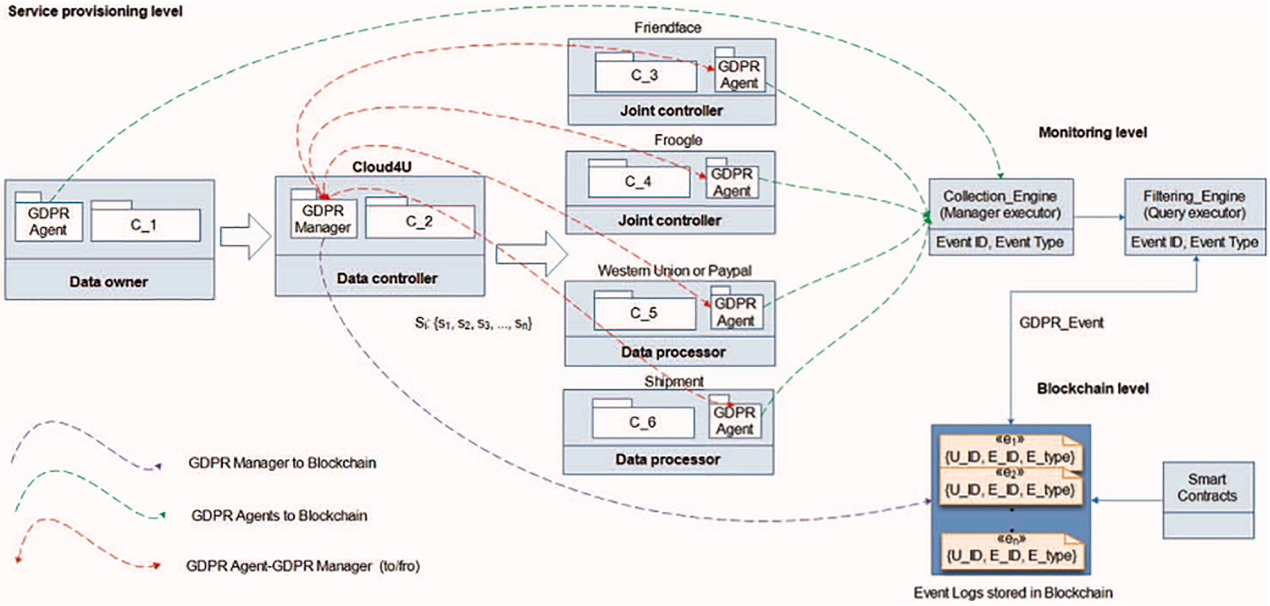

Fig. 5. Operation of the Monitoring System

Tool is unlikely to be deployed in many scenarios for which it was originally conceived - such as the online pharmacy scenario explored in Section 2. In addition, running the PACE Tool's smart contracts in the Ethereum network is computationally intensive and consequently prohibitively expensive, which further makes the adoption of the PACE Tool unlikely. And there are legal challenges as well. To be legally compliant, the PACE Tool must accommodate to the GDPR's requirements, which in practice means switching to a private blockchain and thus sacrificing what is perhaps the tool's greatest advantage: the ability to produce tamper-proof records.

In the following subsections, we explore these challenges, what was our approach to overcome them, and the lessons and insights we have learned from our interdisciplinary endeavour.

4.1. Challenge 1: Barriers to inter-disciplinary collaboration (illustrated with efforts to encoding legal bases under the GDPR)

4.1.1. The challenge

The PACE project team is composed mostly of computer scientists and software engineers, with only a few members having a socio-legal orientation. At the outset of the project, the computer scientist wing had fully embraced the *data-protection-by-design* (DPbD) construct, in the sense of having data protection requirements embedded in the design of data processing systems. This is consistent with the 'code is law'[54] or techno-regulation notion, according to which technology can be intentionally deployed to influence how people behave more effectively than through legislative or contractual measures: legal norms can be 'hard-wired' or 'hardcoded' and automated *ex-ante*, leaving little to no room for noncompliance[55]. Thus, computer scientists in the PACE project conceived a design for a system where the correct application of legal bases and other substantive provisions such as the data quality principles would be automated, and any potential GDPR breach could be detected by trusted nodes after deployment of GDPR compliance verification smart contracts. This approach hinges on the accurate translation into code of highly contextual and interpretable rules (such as Articles 5 and 6 of the GDPR). However, socio-legal scholars have a different understanding of DPbD: due to their flexibility, 'encoding' GDPR rules in this way is hardly always practicable.

Contrary to machine-readable instructions that are concise, typically involving binary 'if/then' type of language and therefore rigid, legal rules tend to be 'open-textured'[56], flexible and subject to interpretation. Their meaning 'is not encapsulated in the words, but reveals itself in the way the rule is used, followed, interpreted, enforced and so on'[57]. Thus, the meaning of terms like 'fairness' or 'reasonable care' will vary depending on the context within which they are implemented and the views of those implementing them, and may still remain imprecise after interpretation. For example, whether someone employed 'reasonable care' depends on many factors, and the outcome of the weighing may range from 'naught to full'[58]. Norms that involve a 'balancing exercise' between competing interests tend to be particularly abstract and require contextual and expert knowledge for their correct application in a given situation. Moreover, what a rule means depends on a number of linguistic and social conventions, which are sometimes fuzzy and susceptible to change[59]. Further, there is a plethora of sources of interpretation of legal norms, including case law, literature, guidance by regulators and customary law, and only the highest court of the relevant jurisdiction is called upon to issue a final authoritative interpretation that trumps any others[60]. The foregoing factors make it significantly harder to hardcode *ex-ante* all the specific scenarios where behaviour is either allowed or

[54] Lawrence Lessig, *Code: And Other Laws of Cyberspace* (Basic Books, New York 1999).

[55] i.e. they can achieve 'digital preemption'. See Danny Rosenthal, 'Assessing Digital Preemption (and the Future of Law Enforcement?)' (2011) 14 New Criminal Law Review 576.

[56] Roger Brownsword, 'So What Does the World Need Now? Reflections on Regulating Technologies', *in R. Brownsword and K. Yeung (eds.), Regulating technologies: Legal futures, regulatory frames and technological fixes* (Hart Publishing 2008) 43.

[57] Erik Claes, Wouter Devroe, and Bert Keirsbilck, 'The Limits of the Law (Introduction)', in *Erik Claes, Wouter Devroe, and Bert Keirsbilck (eds.), Facing the Limits of the Law* (Springer, Berlin 2009) 14.

[58] Sandra Olislaegers, 'Early Lessons Learned in the ENDORSE Project: Legal Challenges and Possibilities in Developing Data Protection Compliance Software', *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life* (Springer 2011) 79.

[59] Brownsword (n 56) 44.

[60] Olislaegers (n 58) 79.

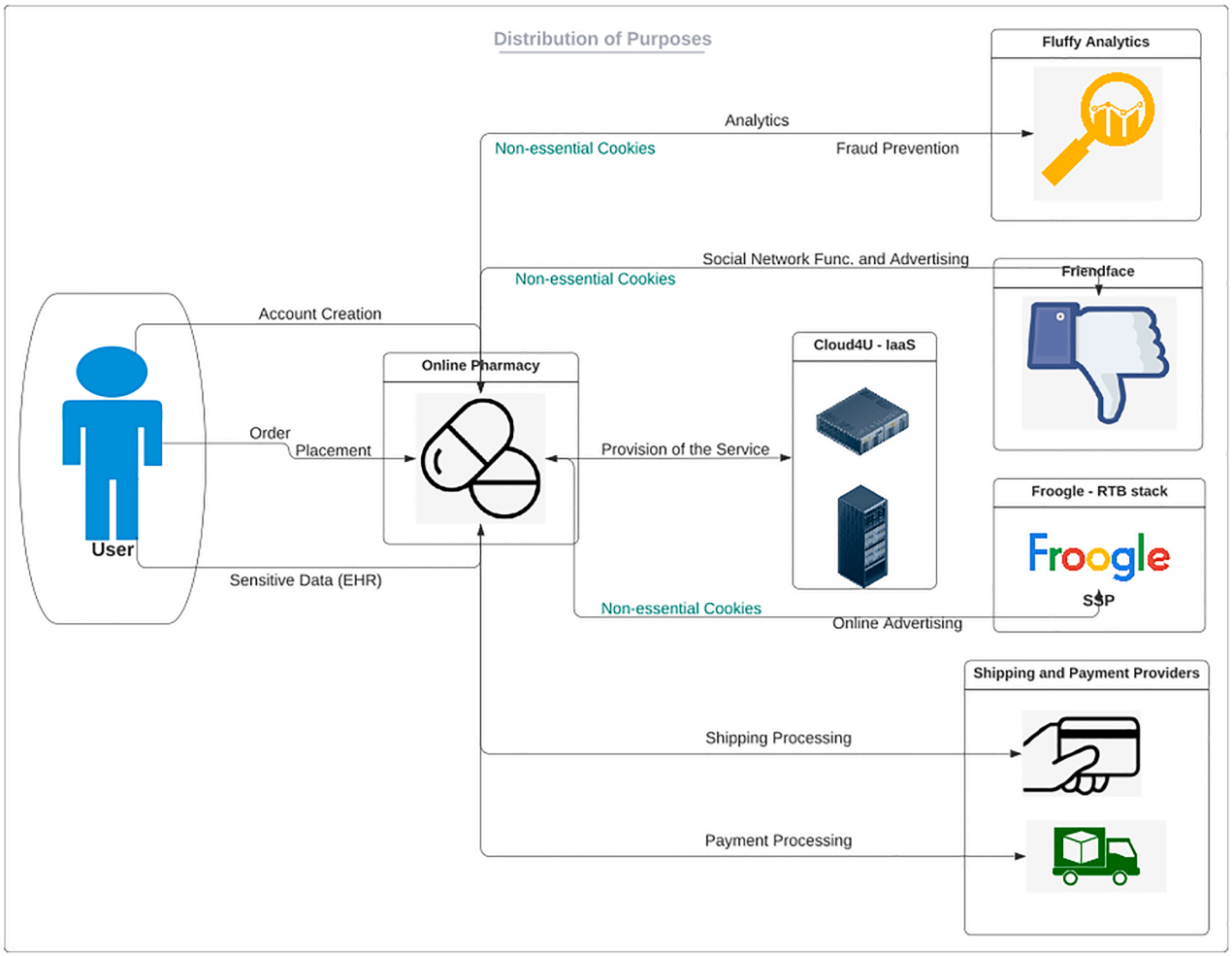

Fig. 6. Distribution of Purposes

prohibited by a given open rule than determining this *ex-post,* typically in legal proceedings[61]. This is particularly the case of data protection law, which is rife with open-textured norms[62]. A prime example of these norms are the lawful bases for processing.

Consider the concept of 'necessity', which is paramount for the application of many GDPR provisions, including the legal bases other than consent. This concept has 'its own independent meaning' in EU law[63], it being the second prong of the proportionality principle. The 'necessity' prong asks: "is the measure concerned necessary (indispensable) to realising the goals it is aimed at meeting?" [64] Thus, when applying the basis set out in Art. 6(1)(b), the necessity assessment involves asking 'is the processing of personal data necessary for the proper performance of the contract at hand? The processing of personal data to perform a contract is not necessary unless such processing is of the essence and unavailable to complete the transaction[65]. It follows that the processing of personal data that is useful or facilitates the performance of a contract, or which renders such performance more profitable for the data controller, is not necessary. As the A29WP explains, the exact rationale of the contract must be determined - i.e. its substance and fundamental objective – 'as it is against this that it will be tested whether the data processing is necessary for its performance.' [66] This is a controller-specific assessment: the contract at hand will vary depending on the services controllers provide, and whilst processing certain personal data may be necessary for the performance of one contract, it will not be necessary for the performance of others. Translating all the contextual specificities and subtleties of diverse cloud-based services into executable smart contracts is not feasible, and even if it were, there would still be likely substantial room for disagreement amongst controllers, data subjects and DPAs as to whether certain forms of processing concerning specific elements of personal data are in fact 'necessary'.

Similar considerations apply to the automation of the 'legitimate interests' basis, which in addition to the necessity assessment it involves a balancing exercise: the relevant interests of the controller or third parties must be balanced against the interests or fundamental rights and freedoms of the data subject[67]. This entails, on one hand, looking at the nature and source of the legitimate interests, and on the other hand,

[61] Bert-Jaap Koops, 'The (in) Flexibility of Techno-Regulation and the Case of Purpose-Binding' (2011) 5 Legisprudence 171, 176.

[62] Olislaegers (n 58) 79.

[63] Case C-524/06, *Heinz Huber v Bundesrepublik Deutschland* [2008] ECLI:EU:C:2008:724 [52].

[64] Lee A Bygrave and Dag Wiese Schartum, 'Consent, Proportionality and Collective Power', in Serge Gutwirth, Yves Poullet, Paul De Hert, Cécile de Terwangne and Sjaak Nouwt (eds), *Reinventing Data Protection?* (Springer 2009).

[65] Christopher Kuner, *European Data Protection Law: Corporate Regulation and Compliance* (Oxford University Press 2007) 234–235.

[66] Article 29 Data Protection Working Party, 'Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC 844/14/EN WP 217' (2014) 16–17.

[67] GDPR, Art. 6(1)(f)

looking at the impact on the rights of the data subject[68]. If the outcome of this assessment is ambiguous, it is necessary to consider whether there are any safeguards intended to protect the data subject[69]. The multiplicity of elements that must be weighed and assessed in the balancing exercise entails a degree of flexibility which sits at odds with the rather rigid nature of technology-embedded rules, and as a consequence, the application of this basis cannot be accurately translated into code.

Ultimately, automating legal provisions is only feasible for simple rules, which are strongly specified and literally applied, as they have low representational complexity and therefore are best-suited to be represented computationally[70].

Unfortunately, work in siloes at the design stage of the PACE Tool meant that the legal side's input was not taken on board. As a result, substantial additional time and effort had to be devoted to change the PACE Tool's initial design.

4.1.2. Our approach to this challenge

Instead of *hardcoding* the application of legal bases, the PACE Tool was re-designed to *guide* controllers in the correct application of legal bases.

To this end, we focused on the data processing purposes which we found to be the most common, and consequently, which users should regularly be expressly asked to consent to. As one of the reviewers of this article rightly pointed out, attempting to condense long and complex 'walls of text' – i.e. privacy policies – into a comprehensive list of purposes is unrealistic, and at any rate does little to render users' consent duly 'informed'. Conversely, a focus on the most common purposes gives users some degree of choice on data processing operations they are routinely subjected to, with which we promote individual control to the greatest extent we found practicable.

Thus, a list of processing purposes was prepared and built-in into the PACE Tool. The purposes were paired with their relevant legal bases based on abstract thinking - i.e. without specifying the elements of data that may be required for their fulfilment. This list serves as a template intended to guide cloud-based service operators (which we refer to as '*main* controllers', such as the online pharmacy operator) in the definition, implementation and enforcement of their GDPR-compliant privacy policies (see Table 1 below). Our focus on main controllers is justified by the fact that these entities make it possible for third-party providers – which may be ultimately deemed joint controllers or even sole controllers depending on the processing purpose at hand - to access personal data of their website/app's users, and that such possibility is dependant on main controllers' design of their cloud-based ecosystems. For example, after *FashionID*, it is clear that the relationship between a website and a social network embedding plugins into that website is that of a joint controller in respect of the collection and disclosure by transmission to the social network of the website users' personal data[71]. This is because the website cannot determine the purposes and means of subsequent operations involving the processing of personal data carried out by the social network after the transmission of that data to this entity[72]. However, without the website authorising the social network to embed its plugins, the transmission of its users' personal data to the social network would not occur. It is the website's design decision that enables such transmission, along with all the privacy risks it involves.

Within the PACE Tool, main controllers are expected to designate which elements of data are required for the fulfilment of each purpose, which presupposes coordination and agreement with the other controllers, joint controllers and processors that comprise the relevant cloud-based service. Main controllers then must assess whether the personal data required by each component of the cloud-based service is 'necessary' to attain the purpose[73] which that[74] data relates[75]. The list of[76] purposes that[77] are relevant[78] to the online[79] pharmacy[80] scenario[81] discussed in Section 2 is presented below:

List of purposes and legal bases

	Purpose	Legal basis
1	Order placement / conclusion of a transaction / provision of a service	Art. 6(1)(b) (contract)
2	Shipping processing	Art. 6(1)(b) (contract)
3	Payment processing	Art. 6(1)(b) (contract)
4	Account creation	Art. 6(1)(b) (contract)
5	Placement of non-essential cookies/ trackers / retrieval of data from non-essential cookies/trackers	Art. 6(1)(a) (consent) / Art. 5 (3) e-Privacy Directive (consent)
6	Online advertising	Art. 6(1)(a) (consent)
7	Collection and transmission of data to social networks	Art. 6(1)(a) (consent)
8	Social network functionalities and advertising	Art. 6(1)(a) (consent)
9	Analytics / Service personalisation	Art. 6(1)(f) (legitimate interests)
10	Fraud prevention / ensuring network and information security	Art. 6(1)(f) (legitimate interests)
11	Sensitive (special categories of) personal data*	Art. 9(2)(a) (consent)
12	International transfers*	Arts. 45, 46, 47 and 49

*When special categories of personal data or international transfers are involved in any of the purposes set out in 1-10, additional bases in Arts 9(2)(a), 45, 46, 47 and 49 are required.

Once the main controller concludes the final list of purposes paired with their applicable legal basis and the elements of data required for their fulfilment, it then must proceed to allocate each purpose to each component of the cloud-based service. At this stage, the PACE Tool is ready to be deployed. Continuing with the online pharmacy example, this distribution would look as seen in Fig. 6.

On the user side, after deployment of the PACE Tool, the list of purposes determines users' available options on their *privacy manager* interface (see Fig. 2 above), where they can see each purpose of processing pursued by each component of the cloud-based service, along with the applicable legal ground and the categories of data involved. On this interface, individuals have the ability to give/deny their consent to each processing purpose, or alternatively exercise their right to object to processing, as applicable.

4.2. Lessons and insights

Even though there is wide agreement in that technical, legal and

[68] Article 29 Data Protection Working Party (n 66) 33.
[69] ibid 42.
[70] Koops (n 61) 193.
[71] Case C-40/17, *Fashion ID GmbH & CoKG v Verbraucherzentrale NRW eV* (Fashion ID) [2019], ECLI:EU:C:2019:629 [76].
[72] ibid.
[73] See EDPB, 'Guidelines 2/2019 on the Processing of Personal Data under Article 6 (1)(b) GDPR in the Context of the Provision of Online Services to Data Subjects' (2019) 3.
[74] Shipping processing, and the processing of data to this end, are steps necessary for the conclusion of a contract.
[75] Same as shipping processing above.
[76] Same as order placement above.
[77] See ICO, 'Update Report into Adtech and Real Time Bidding' (2019) 17–18; See also Article 29 Data Protection Working Party (n 66) 18.
[78] Social plug-ins embedded in websites and apps such as the 'Like' button entail the placement and retrieval of information stored in a user's terminal equipment, for which prior consent is required.
[79] Same as online advertising above.
[80] See Article 29 Data Protection Working Party (n 66) 25–26; see also Article 29 Working Party, 'Opinion 04/2012 on Cookie Consent Exemption' (2012) WP 194 11.
[81] Recital 47 GDPR mentions these purposes as examples of legitimate interests.

other stakeholders must work together to devise data-driven technologies that take privacy into account from the start[82], there is no obvious effective method to put inter-disciplinary collaboration into practice. Ideas for privacy-enhancing technologies (PETs) and DPbD approaches, methodologies and tools do not come into existence just by putting together a number of computer scientists, software engineers and privacy lawyers in the same room. Deep-rooted convictions of a project's leading field may steer the project in the wrong direction if the input of the other fields involved is not taken on board from the outset. Based on the 'code is law' ideal, software engineers – including members of the PACE team during the course of the project - have devoted significant time and effort to devise solutions capable of automating GDPR compliance[83]. However, without the requisite expert legal knowledge, they have grounded their work in either substantial legal misconceptions or mistaken interpretations of this Regulation[84]. As a result, the actual value and impact of their designs on GDPR compliance in particular and the protection of privacy and personal data in general are limited.

Avoiding work in siloes should be thus a guiding principle in interdisciplinary projects, especially during the design stage of a PET. In the PACE project, valuable time, energy and resources could have been saved had we reached from the onset a common understanding on how to attain the project's goals and build the PACE's tools core mechanisms. This is easier said than done. Oftentimes, legal scholars and computer scientists felt like we were speaking two different languages. To some extent, we were. Legal scholars are typically familiar – and even comfortable – with the highly contextual assessments that must be conducted to determine whether a specific use of technology has a negative impact on privacy and data protection, and with the fact that the outcome of such assessments is commonly up for debate and subject to different interpretations, oftentimes leading to disagreement and dispute. Computer scientists and engineers, conversely, tend to struggle with the lack of definition, clarity and conclusiveness that is inherent to the legal field, as these traits are completely alien to their field of expertise. To put it bluntly, programming instructions follow an 'if/then = yes or no' pattern, as opposed to 'if/then = perhaps, depending on whether X, Y or Z, or a combination of the three, takes place'.

Awareness of the abovementioned different way of reasoning in particular, and of how difficult communication between the technical and legal sides can be more generally, is the first step to avoid a silo-based type of interdisciplinary collaboration. In the PACE project, clashes between different ways of reasoning led to frustration: the legal input was seen as too confusing and indeterminate, and ultimately as a barrier to the automation ideal. Work in siloes naturally ensued. Awareness is logically not enough; effective measures to foster interdisciplinary collaboration must be implemented. Jointly agreeing on a project blueprint setting clearly defined interdisciplinary deliverables based on the input of all fields involved, team building and gamification activities[85], as well as periodic reviews of project milestones having an interdisciplinary component, are three examples of such measures that we tried.

On the other hand, although some were optimistic about smart contracts built on blockchain technologies potentially becoming 'the most important example yet of "self-executing, customised rules"' that could serve as a substitute for law[86], the difficulties in encoding flexible data protection rules suggest that the promises of techno-regulation will remain unfulfilled for the time being. However, this is not to say that technological approaches to the enforcement of data protection law are infeasible. First and foremost, technological tools can provide data controllers with guidance in applying the law correctly[87]. This is the approach we ultimately followed. In the re-designed PACE Tool, the correct application of legal bases is not automated via software performing a legal analysis (as originally intended), but instead is made by main controllers at the pre-deployment stage of the PACE Tool, guided by the built-in list of purposes. In short, the PACE Tool serves as a 'choice architecture'[88] that nudges main controllers into applying the correct legal bases. Thus, the highly open-textured provisions of Article 6 of the GDPR are not hardcoded; instead, they are applied by a technological tool via an interactive interface that features data protection information and insights on the basis of which main controllers can structure a legally compliant cloud-based ecosystem.

Second, technological tools can also be leveraged to give individuals control over their personal data. A data subject normally has no choice but to trust that the data controller has technical means in place that honour her privacy preferences, and that these are not bypassed[89]. Conversely, the PACE Tool allows data subjects to check by themselves whether or not their privacy preferences are respected. In concrete, the distribution of purposes made by main controllers is replicated in the *privacy manager* interface, which provides individuals with the ability to give and withdraw consent through the same action, and also to exercise their right to object to processing through an opt-out option when 'legitimate interests' is the legal basis relied upon. Users' privacy preferences are recorded on the blockchain in a tamper-proof fashion, and users can later avail themselves of the verification smart contract's *privacy-preference* function to confirm whether such preferences are respected or bypassed, and take action accordingly – such as changing service providers or filing a complaint with the DPA in the event of a breach[90].

4.3. Challenge 2: Hardcoding GDPR Rules (substantive provisions) for automated compliance verification

4.3.1. The challenge

As seen in the preceding subsection, encoding the GDPR is a daunting exercise, as many of its core provisions – such as the legal grounds for processing - feature terms that are either interpretable or involve a balancing exercise that is highly context-dependent. Other substantive provisions of the GDPR, such as the data quality principles set out in Article 5, are no exception[91].

[82] Pagona Tsormpatzoudi, Bettina Berendt and Fanny Coudert, 'Privacy by Design: From Research and Policy to Practice–the Challenge of Multi-Disciplinarity', *Annual Privacy Forum* (Springer 2015) 200.

[83] See footnote 84.

[84] See e.g. interpretations of legal bases, exceptions to article 9 and discussion on anonymisation in Tom Kittmann, Jens Lambrecht and Christian Horn, 'A Privacy-Aware Distributed Software Architecture for Automation Services in Compliance with GDPR', *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)* (IEEE 2018) 1069; interpretations of data minimisation, data security and data transfer in Masoud Barati and others, 'GDPR Compliance Verification in Internet of Things' (2020) 8 IEEE Access 119697, 119702–119703; the mechanisms that 'fully comply with the GDPR' in Nguyen Binh Truong and others, 'GDPR-Compliant Personal Data Management: A Blockchain-Based Solution' [2019] arXiv e-prints arXiv, 11; and the 'Integrated Knowledge Graph' to automate GDPR compliance in Lavanya Elluri, Ankur Nagar and Karuna Pande Joshi, 'An Integrated Knowledge Graph to Automate Gdpr and Pci Dss Compliance', *2018 IEEE International Conference on Big Data (Big Data)* (IEEE 2018).

[85] See generally Gloria Piedad Gasca-Hurtado and others, 'Gamification Proposal for Defect Tracking in Software Development Process', *European conference on software process improvement* (Springer 2016).

[86] Christopher Millard, 'Blockchain and Law: Incompatible Codes?'(2018) 34 Computer and Law Security Review 846

[87] Olislaegers (n 58) 80.

[88] Richard H Thaler and Cass R Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness.* (Penguin 2009) 6.

[89] Christophe Lazaro and Daniel Le Metayer, 'Control over Personal Data: True Remedy or Fairy Tale' (2015) 12 SCRIPTed 25.

[90] There are, however, important practical and legal challenges that curb the adoption of the PACE Tool, see sections 4.3 and 4.4 below.

[91] Koops (n 61) 175–176.

4.3.2. Our approach to this challenge

One of the main goals of the PACE Project was to develop an automated system for auditing compliance with data protection rules, so abandoning the automation endeavour altogether was not an option. Therefore, we were forced to find a compromise. We acknowledged that translating most GDPR substantive provisions into machine-readable instructions in an accurate fashion – that is, contemplating all potential interpretations and contextual scenarios – is close to impossible. However, it is nevertheless possible to translate *some* provisions into code by attempting to replicate the meaning of the relevant provision to the greatest extent possible.

For example, according to the 'data minimisation' principle, the personal data being processed must be limited to what is necessary in relation to the purposes for which it is processed (Art. 5(1)(c) of the GDPR). This principle is very difficult to accurately convert into code, as determining what data is 'necessary' depends on the purpose at hand, which will vary depending on the specific task a component is supposed to execute. However, we can determine necessity in broad terms *by proxy*, relying on the labels assigned to the different pieces of information included in the *container logs*.

The *container logs* include (i) the relevant cloud component's address (*p*), (ii) the data processing operations performed by the component (*Ap*) (which include the relevant processing purposes authorised by the user (*Apur*)), (iii) the types of personal data processed by the component (*Dp*) (e.g. name, home address, location), (iv) the types of personal data collected from the user (*Dcp*), (v) any security measures implemented in the data processing operations (*Eap*) (e.g. encryption or pseudonymisation), (vi) the physical location of the provider (*locp*), and (vii) the period of time claimed by the component for storing personal data (*tp*). Thus:

Data minimisation verification: if a component *p* collects different types of data (*Dcp*) but only uses a subset of it (*Dp)* for the processing it is expected to perform, then a potential violation of this principle can be flagged.

Other GDPR requirements can be represented in this way.

Data security: this principle requires that appropriate technical or organisational measures are implemented when processing personal data to protect the data against accidental, unauthorised or unlawful access, use, modification, disclosure, loss, destruction or damage (Arts. 5 (1)(f) and 32(1) of the GDPR). These measures may include, for example, pseudonymising and encrypting personal data.

Data security verification: A component *p* executing a set of operations on personal data (*Ap*) can be flagged as a potential violator if there is an operation (*ap*) in which personal data is not encrypted or pseudonymised (*Eap*: false).

Transfers of personal data to a non-EU country: transfers of this type may take place on the basis of an adequacy decision by the European Commission, or in lieu thereof, where the controller or processor provides appropriate safeguards (Arts. 45 of the GDPR). These appropriate safeguards can take the form of, for example, Binding Corporate Rules (BCR), or adherence to codes of conduct or certification mechanisms (Arts. 46 and 47 of the GDPR).

International transfer verification: A potential violation may be flagged if personal data is transferred to a component (*p)* in a country (*locp)* which has no adequacy decision with the European Commission, and if other appropriate safeguards (which are globally subsumed within the concept *BCR*) enabling the transfer have not been implemented[92].

Storage limitation: according to this principle, personal data may not be kept for longer than necessary for the purposes for which it is processed (GDPR, Art. 5(1)(e)). Service providers must state their retention periods in their privacy policies (Art. 13(2)(a)).

Storage limitation verification: A potential violation may be flagged if a component (*p)* retains personal data for a period (*ts*) longer than that stated in its privacy policy *(tp*).

Purpose limitation: according to this principle, personal data may only be processed for specified, explicit and legitimate purposes, and not further processed in a manner that is incompatible with said purposes (Art. 5(1)(b)).

Whilst determining 'compatibility' is a highly context-dependent assessment, it can be determined by proxy, albeit admittedly with less-than-ideal accuracy[93]. As noted above, the processing purposes are sent to the blockchain during the ratification phase[94], and the monitoring system tracks the data processing operations performed by the components of the cloud-based service. Thus:

Purpose limitation verification: A potential violation may be flagged if a component (*p)* carries out data processing operations (*Ap*) for purposes other (*Opur*) than those disclosed to the user in the *privacy manager* interface *(Dpur*).

To verify compliance with the abovementioned GDPR provisions, a *verification* smart contract is deployed. The smart contract has different functions, which correspond to the requirements outlined above. Upon deployment, trusted nodes in the Ethereum blockchain run a transaction to retrieve the information contained in the *container-logs,* and then flag any observed GDPR violations in an automated way. The results of the verification are recorded on the blockchain, and can be consulted for auditing purposes.

4.3.3. Lessons and insights

DPAs have been historically under-staffed and under-resourced[95]. Against this background, techno-regulatory approaches to the enforcement of data protection law are all the more appealing. A GDPR violation detected and flagged after deployment of a smart contract could eliminate the need for conducting an investigation altogether, or at least it could make it significantly shorter. Thus, automated tools for GDPR compliance verification could relieve DPAs from their budgetary and staffing constraints, this being one of the PACE project's underlying motivations. However, due to the uncertainty that arises from the indeterminacy of data protection rules, the PACE Tool cannot be relied upon to establish a GDPR violation without proper human intervention and expert knowledge. This is because, if violations flagged with the aid of the PACE tool had authoritative power – i.e. were deemed conclusively determined by a DPA – a number of issues would arise.

First, the regulatory response arising from the use of the PACE Tool would not necessarily align with the relevant DPA's underlying policy objectives. As Brownsword observes, we are not able to anticipate or foresee the full set of scenarios to which a rule with indeterminate terms – e.g. necessary – applies[96]. This challenge can be addressed by equipping the automated system with a default rule, which essentially entails a simplification exercise: once the default is implemented, 'the system knows what to do even if the scenario is not specifically anticipated'[97]. This is what we did when encoding the purpose limitation principle. Confronted with the impossibility to anticipate every scenario in which further processing is compatible with the purposes of the original one, we instructed the system to reach finding of incompatibility – and therefore a violation of this principle - when a component of the cloud ecosystem processes personal data for purposes different than those originally informed to the data subject. This default rule effectively prevents mission creep, and as such is in line with the goal of protecting

[92] To determine the existence of appropriate safeguards, main controllers must ascertain and record this fact based on the contractual documentation they have in place with the components of its cloud-based service.

[93] See paragraphs containing footnotes 96 and 97 in section 4.2, *Lessons and Insights.*

[94] See section 3.3 and Figure 3.

[95] See text accompanying footnotes 12 and 13 above.

[96] Brownsword (n 56) 44.

[97] ibid.

individuals with regard to the processing of personal data. On the flipside, it removes the possibility of further processing altogether, and as such is not aligned with the goal of ensuring the free flow of personal data between Member States.

Second, largely as a consequence of the above, there would be likely an over-inclusiveness tendency resulting in a high number of false positives. For example, violations of Articles 5(1)(e) and 32 would be found whenever a component processes personal data without encrypting or pseudonymising it. The pseudonymisation and encryption of personal data, however, are only two technical measures out of many that controllers and processors can implement to comply with these provisions. Moreover, organisational measures ensuring 'a level of security appropriate to the risk' would be completely ignored, and violations would be found in many scenarios where 'the nature, scope, context and purposes of processing', as well as the risks involved, do not warrant the pseudonymisation or encryption of personal data.

And third, the vital role of the judicious exercise of discretion by data protection watchdogs would be dramatically reduced – this effect stemming more generally from the techno-regulation idea itself. As a result, data protection rules would be applied in a rigid fashion, even in scenarios where DPAs believe their strict application would be counterproductive[98].

In the light of the above, instead of being used to automate law enforcement in a way that dispenses with proper human input, technologies such as the PACE Tool should be used only to assist the effective enforcement of data protection law. This could be done, for example, by deploying it to identify potential instances of noncompliance, which can be subsequently investigated in more detail to determine whether noncompliance actually took place. In fact, by providing a tamper-proof 'single truth' of all data processing operations arising from the interaction of a data subject with a cloud-based composite service, the PACE Tool's blockchain-based architecture can facilitate investigations into data protection law breaches, thereby fostering accountability. As Lazaro and Metayer observe, accountability depends on the extent to which its main piece of evidence – the execution logs of the relevant system – meet certain requirements. First, they must include sufficient information to determine compliance or detect non-compliance; second, they must depict the actual behaviour of the system, in such a way that hiding operations or providing false evidence is highly difficult; and third, their security and integrity must be guaranteed – i.e. it must be impossible to modify them and no non-authorised users may be able to read their content[99]. As the PACE Tool's *container-logs* include detailed information on the cloud-based service's data processing operations and are recorded on the blockchain, they seem to meet these criteria. These logs can be queried at any point to assert lawful processing, either via the automated GDPR compliance verification functionality or manually (see Fig. 7).

4.4. Challenge 3: Lack of incentive to deploy the PACE Tool

4.4.1. The challenge

The predominantly surveillance-based business model of the Web 2.0 has proved remarkably profitable, and as shown by the failed 'Do Not Track' initiative, corporations go great lengths to defend it[100]. Technological solutions such as the PACE Tool, which seek to ensure observance of the limitations on the collection and processing of personal data imposed by data protection law, undermine the significant leeway industry players have had thus far to access, use and experiment with personal data - and by extension threaten the profitability of their business model. Accordingly, they have little to no incentive to deploy the PACE Tool, and consequently the PACE Tool's scalability is inherently limited.

Further militating against the widespread adoption of the PACE Tool are its high transaction costs. Deploying and running the PACE Tool's smart contracts in the Ethereum network consume a fair amount of *gas*, a unit that measures the computational effort required to execute transactions by a miner[101]. Gas units are expressed in *wei,* which is the smallest unit of the Ethereum network's cryptocurrency *ether*[102].

To test the smart contracts, we deployed them on Ropsten, a public blockchain test network, contemplating three different Service Packages (SPs) for the online pharmacy scenario outlined in Section 2:

- Service Package 1 (SP1) involved two cloud components performing 9 operations on personal data;
- Service Package 2 (SP2) involved four cloud components performing 16 operations on personal data; and
- Service Package 3 (SP3) involved six cloud components performing 23 operations on personal data.

The smart contracts *container-log* and *verification* were executed five times to calculate the average results. As seen in Table 1 below, the experimental results show that a higher number of operations and components involved entails a sharp increase in gas consumption. Moreover, the amount of transaction costs hinges on the complexity of the verification at hand. In particular, the verification of compliance with data security requirements is the less costly, as it only assesses the implementation of encryption and pseudonymisation, and consequently its complexity is comparatively lower. Conversely, the verification of compliance with data minimisation is the most expensive, as it requires checking the data processing operations involved as well as the types of personal data collected and actually processed by the different cloud components.

Considering the complexity of the average cloud-based service[103] and the high number of these services with which individuals interact on a daily basis, the cost of using the PACE Tool's GDPR compliance verification functionality would be prohibitive. As of the time of testing, the average cost for running the *verification* smart contract's four functions included in Table 1 on a single occasion under the SP3 scenario was USD 115.17[104].

4.4.2. Our approach to this challenge

We held a workshop at UCL Computer Science with the participation of civil society organisations, data protection law scholars, computer scientists, software engineers, UK regulators, and industry players[105]. After exploring the PACE Tool's architecture and functionalities, we asked participants to identify the potential of the PACE Tool to foster individual control, transparency and accountability, as well as any challenges capable of undermining such potential. A discussion on what

[98] See Karen Yeung, 'Towards an Understanding of Regulation by Design', *in R. Brownsword and K. Yeung (eds.), Regulating technologies: Legal futures, regulatory frames and technological fixes* (Hart Publishing 2008) 93.

[99] Lazaro and Metayer (n 89) 28.

[100] See Lee A Bygrave, 'Hardwiring Privacy', *in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2016) 764 and references cited therein.

[101] 'Transactions' (*ethereum.org*) <https://ethereum.org/en/developers/docs/transactions/> last accessed 09 August 2023.

[102] 1 ether = 10x18 wei

[103] See footnote 3 above.

[104] More details on this testing can be found in Barati and others (n 18).

[105] Participants were selected on the basis of contacts of the PACE Project's team members.

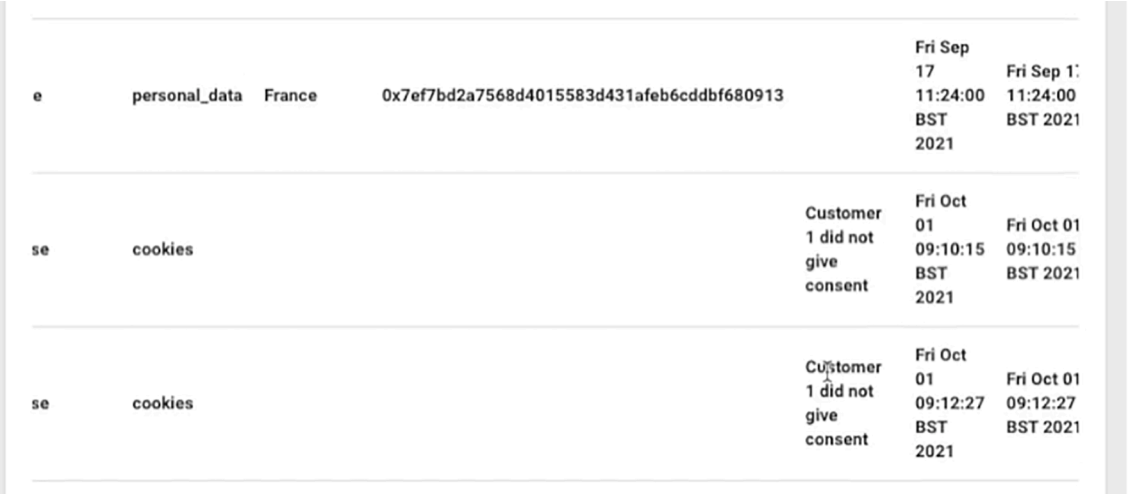

Fig. 7. Blockchain logs
A hypothetical DPA consulting blockchain logs to ascertain users' privacy preferences and identify data transfers.

Table 1
Transaction costs of smart contracts deployment

	SP1	SP2	SP3
Number of components	2	4	6
Number of data processing operations	9	16	23
Container-log smart contract (wei)	1562478	2782774	3882652
Verification smart contract, i.e.:			
Data security (wei)	297628	743436	1401864
Data minimisation (wei)	905648	1582621	2305178
International transfer (wei)	323501	1112821	1803427
Data storage (wei)	304562	762341	1522370

could be done to overcome such challenges then followed.

There was agreement in the workshop in that incentivising data-driven companies to adopt a technology such as the PACE Tool is beyond our capabilities as researchers. This is because, in the data-driven economy, there is still no compelling business case for PET adoption: freedom to experiment with data is more profitable that implementing restrictions to do so. Industry players consistently held the view that it made no sense from a business perspective to implement a technology intended to constantly generate tamper-proof records of any potential wrongdoing on their part, as under the status quo, DPAs' powers are perceived as limited, the threat of a fine distant, and privacy has yet to consolidate as an added value that can be profitably exploited. Conversely, collecting and processing personal data is a well-tested and successful business proposition. Civil society representatives shared the same view, noting that for most data-driven firms, 'business as usual is good business'.

In turn, participants in the PACE workshop shared the view that, if individuals were to bear the costs of deploying the smart contracts, most would be deterred from adopting the PACE Tool in the first place. Moreover, given the costs involved, those who decided to try it out would likely soon stop using the automated GDPR compliance verification functionality altogether. On their part, industry players tended to agree that, if they had to bear the deployment costs, they would either refrain from using the PACE Tool or pass on to consumers the costs they would incur. If the last option were chosen, their offering could over time become more expensive and consequently less competitive, which would serve as an additional motivation to abandon the use of the tool out of fear of consumer switching.

A potential solution to the high transaction costs is to switch to a blockchain that operates on a consensus protocol other than PoW and consequently requires lower computational effort to run transactions. Subject to funding, future versions of the PACE Tool will try this alternative, although this would entail compromising the tool's security and integrity[106].

4.4.3. *Lessons and insights*

The aforementioned lack of incentive is not exclusive to the PACE Tool, but instead affects PETs more generally. Whilst legal and regulatory pressure regarding data protection is a factor capable of having a positive impact on the PET adoption process[107], this pressure hinges on the extent to which data protection rules incentivising PET adoption are effective and enforceable. Unfortunately, legislative support for widespread deployment of PETs is rather shaky.

Article 25 of the GDPR enshrines controllers' obligation to observe data protection-by-design (DPbD), a notion intended to ensure that privacy-related requirements be duly accounted for in data processing systems' design and subsequent development, in order to improve such requirements' traction[108]. One way to realise DPbD is through the deployment of PETs. However, there are significant challenges impeding the effective application of this provision.

First, Article 25 contains a number of factors that must be weighed to decide what DPbD measures may be implemented, including 'the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing.' It is arguably easy to get lost in this sentence. Moreover, balancing these factors is bound to be a daunting task, not least given that 'there is no further explanation on how to interpret and prioritise them in relation to one another'[109]. Second, there are few compelling reasons to observe DPbD other than the risk of incurring sanctions[110],

[106] See section 4.4 below.

[107] John J Borking, 'Why Adopting Privacy Enhancing Technologies (Pets) Takes so Much Time', *Computers, privacy and data protection: an element of choice* (Springer 2011) 322.

[108] Lee A Bygrave, 'Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements' (2017) 4 Oslo Law Review 105, 106–107.

[109] Tsormpatzoudi, Berendt and Coudert (n 82) 204.

[110] Whilst there are some incentives in, for example, Article 83(2)(d) and Article 6(4)(e), these "are few in number and stunted by obtuse phrasing". See Bygrave (n 100) 771.

and the fact that DPAs are notoriously under-resourced and understaffed make the imposition of sanctions an unlikely scenario. And third, since DPbD obligations are imposed mainly on controllers only, Article 25 presupposes a market in which controllers demand PETs and DPbD products and services or otherwise fuel their production and availability[111]. However, due to winner-takes-all dynamics, traditional and data-driven network effects, and the overwhelming market power a handful of tech firms has managed to amass, such market hardly exists in reality.

In the light of the above, improvements on different fronts must be attained for PETs like the PACE Tool to be widely adopted and thereby have a meaningful positive impact on the levels of data protection individuals can currently enjoy. In particular, we must move away from the formulation of DPbD principles as slogans that 'are almost totally silent regarding line of action'[112], and instead come up with clear guidance on how to translate said principles into engineering methodologies and practices. The work of some researchers in this regard is noteworthy. For example, Hoepmann has proposed eight privacy design *strategies*, which are abstractions derived from the GDPR's core principles and requirements that achieve (some level of) data protection as their goal: minimise, hide, separate, aggregate, inform, control, enforce and demonstrate[113]. These strategies are realised through privacy design *patterns*, which are available reusable software solutions that implement the strategies in concrete terms. For instance, the strategy *inform* can be achieved by using patterns such as the *data breach notifications*[114] or the *multi-layered presentation approach*[115]. Others like Perera *et al.* have proposed a number of privacy *guidelines* to be applied in Internet of Things (IoT) application design processes[116]. Nonetheless, although the availability of a catalogue of privacy design strategies, patterns or guidelines is a positive development, the question of how they can be put to use in practice remains open[117].

Also, the benefits arising from the implementation of PETs and DPbD measures should be both made more explicit and larger in number. As Bygrave argues, 'greater consideration ought to be given to how to best craft the carrots that can ensure the [DPbD] goals become more simply than aspirational'[118]. For example, future revisions of the GDPR could establish a rebuttable presumption of no fault in data protection investigations in favour of controllers that employ certified PETs. This would require, however, an expansion of the scope of the certification mechanism contemplated in Article 42 in relation to Article 25(3) of the GDPR, which is currently limited to 'processing operations', as opposed to certification of a technology or IT system as a whole.

And lastly, most importantly, regulators and lawmakers must come to terms with the fact that the privacy and data protection crisis we are experiencing is intrinsically connected to, and fuelled by, the problem of lack of healthy privacy-driven competition in digital markets. Decisive action to correct these seemingly different regulatory failures in a holistic way must be taken. For example, the fact that the behavioural advertising industry is notoriously privacy-invasive is well-documented[119]. Thus, efforts should be deployed to limit in a meaningful way what the actors in the ad tech value chain can do with our personal data, in such a way that their privacy-intrusive practices are made too risky and potentially costly, thus forcing them to consider alternative, more privacy-friendly business practices. This could be achieved based on more robust data protection law enforcement in combination with a sector-specific regulation on online advertising[120]. Unfortunately, recent legislative initiatives completely overlook the negative impact on privacy arising from tracking-based business practices, and even seek to foster their growth. For example, the UK Data Reform Bill is intended to 'reduce burdens of businesses' by *inter alia* cutting 'down on 'user consent' pop-ups and banners - the irritating boxes users currently see on every website - when browsing the internet', switching to an opt-out mechanism via automated tools the effectiveness and availability of which is anything but confirmed[121]. This has the potential of normalising even further pervasive tracking, not least on account of the so-called default setting bias. Similarly, the Digital Markets Act seeks to promote contestability in the ad tech value chain, yet contains little 'tackling head on the surveillance-based core characterizing several of the gatekeepers' business model, with their negative impact on consumers and the society as a whole'[122].

As for the PACE Tool's high transaction costs, let us remember that the PoW protocol is what ensures the blockchain's security and integrity, yet it consumes a lot of computational resources, thereby making mining highly costly[123]. Thus, there is currently an unavoidable trade-off between blockchain's technical assurances and economic considerations. To reduce overall levels of energy consumption, alternative consensus protocols have been put forward, chief amongst which being the 'proof-of-stake' (PoS), which is expected to reduce Ethereum's energy consumption by ~99.95%[124]. However, factors such as 'weak subjectivity' and 'costless simulation' make PoS-based blockchains highly vulnerable to 'alternative history attacks' that are unfeasible in PoW-based ones, largely due to the required computational effort for generating previous blocks and outpacing the main chain[125].

In short, whilst the blockchain offers valuable assurances in terms of security and integrity, its current high demands of computational power dramatically curb the scalability of the PACE Tool, which involves the analysis of an astonishingly high number of data processing operations performed by a multiplicity of actors, and consequently is bound to be

[111] Bygrave (n 108) 119.

[112] Dag Wiese Schartum, 'Making Privacy by Design Operative' (2016) 24 International Journal of Law and Information Technology 151, 157.

[113] Jaap-Henk Hoepman, 'Privacy Design Strategies', *IFIP International Information Security Conference* (Springer 2014).

[114] ibid 455.

[115] Article 29 Working Party, 'Guidelines on Transparency under Regulation 2016/679' (2018) 25.

[116] Charith Perera and others, 'Designing Privacy-Aware Internet of Things Applications' (2020) 512 Information Sciences 238.

[117] Seda Gürses, Carmela Troncoso and Claudia Diaz, 'Engineering Privacy by Design Reloaded', *Amsterdam Privacy Conference* (2015) 2.

[118] Bygrave (n 100) 771.

[119] See generally ICO (n 77).

[120] See e.g. Datenethikkommission, 'Opinion of the Data Ethics Commission' (2019) 98.

[121] DCMS, 'New Data Laws to Boost British Business, Protect Consumers and Seize the Benefits of Brexit' (*GOV.UK*, 17 June 2022) <https://www.gov.uk/government/news/new-data-laws-to-boost-british-business-protect-consumers-and-seize-the-benefits-of-brexit> last accessed 09 August 2023.

[122] Simonetta Vezzoso, 'The Dawn of Pro-Competition Data Regulation for Gatekeepers in the EU' (2021) 17 European Competition Journal 402.

[123] See text accompanying footnotes 35 to 37 above.

[124] 'The Merge' (*ethereum.org*, 30 June 2022) <https://ethereum.org/en/upgrades/merge/> last accessed 09 August 2023. See also footnote 34.

[125] Evangelos Deirmentzoglou, Georgios Papakyriakopoulos and Constantinos Patsakis, 'A Survey on Long-Range Attacks for Proof of Stake Protocols' (2019) 7 IEEE Access 28712, 28712–28713.

excessively costly. Ultimately, insofar as the transaction costs of Turing-complete blockchains are not significantly reduced in a way that does not compromise their integrity and security assurances, the PACE Tool is unlikely to move past the proof-of-concept stage.

4.5. Challenge 4: Finding a GDPR-compliant blockchain that provides the assurances of a public permissionless blockchain

4.5.1. The challenge

If a regulation is too technology-specific, it can quickly become outdated and as a result will have to be adapted sooner than later. A technologically neutral regulation avoids this trap, 'unless a technological advance is so disruptive that it effectively overturns the fundamental assumptions on which that regulation is based'[126]. This is the case with the GDPR and the blockchain. The GDPR was conceived for a setting in which data is collected, stored and processed in a centralised fashion, yet blockchains decentralise these processes[127]. Thus, '[w]hen it comes to data privacy and your personal data, the [blockchain] represents the proverbial round peg that does not fit squarely within the four corners of the law'[128]. Unfortunately, the conflicts that arise from the GDPR's and the blockchain's opposite mindsets hinder software engineers' ability to rely on the blockchain's decentralised features to devise innovative privacy-preserving effective solutions.

In order to minimise privacy risks, we ruled out recording on the Ethereum blockchain the personal data users provide to the different components of a cloud-based service upon interaction with it, such as name, address, and payment details. However, the fact remains that users' public keys constitute personal data, and therefore the whole GDPR edifice applies to the PACE Tool. One of the first steps in an analysis under the GDPR is determining controllership. The identification of controllers is straightforward in scenarios where it is possible to find a central entity that determines 'the purposes and the means of the processing of personal data'[129], yet highly difficult where such determination is distributed amongst multiple actors, as is the case in public and permissionless blockchains such as Ethereum. By choosing the relevant software and its embedded protocols, nodes and miners have significant control over the *means* of processing (in the PACE Tool's case, the Ethereum blockchain), yet they hardly determine the *purposes*. Accordingly, they may be considered processors[130], but the question as to on behalf of whom they are processing personal data remains, as well as how they are supposed to formalise their relationship in the form of a contract with an unknown controller.

The Court of Justice of the European Union (CJEU) consistently espouses a broad interpretation of the concept of controller and joint controller 'in the interest of the effective protection of the right to privacy'[131]. However, the absence of a single entity - or even a group of entities - in control of the data flows within a permissionless blockchain means that there is no controller in the traditional sense, as the application of the 'purposes and means' test results in the conclusion that the users of the blockchain determine the purpose (i.e. recording a given transaction onto the blockchain) and the means of processing (that is, the choice of the blockchain in question to execute a transaction)[132]. Thus, users are the controllers of personal data relating to both others (e.g. the counterparty of a transaction) and themselves, a conclusion of little value for the proper allocation of responsibilities under the GDPR, and by extension for the 'effective protection of the right to privacy'. Conversely, the determination of controllership in permissioned blockchains is significantly simpler, as it is always possible to identify a group (consortium blockchains) or a single entity (fully private blockchains) determining the purposes[133] and the means[134] of the processing of personal data. Accordingly, DPAs are already advising companies considering the use of blockchain technology to choose permissioned blockchains, as controllers can be determined with relative ease[135].

Choosing a consortium or fully private blockchain over a permissionless one is, however, a counter-productive decision for the type of data protection improvements the PACE Tool seeks to elicit. In a permissionless blockchain, the PoW protocol ensures that no individual or group of nodes controlled by a data controller or processor be able to tamper with the ledger to conceal data protection violations on their part. This assurance is essential for reliable individual control, proper transparency, and robust accountability: if the ledger can be amended, there is no assurance that individuals' privacy preferences have been respected, the execution logs cannot be trusted as a faithful depiction of the cloud-based service's operation, and controllers/processors can manage to remain unaccountable for potential violations. In private and permissioned systems, conversely, it is a lot easier for participants to collude to re-write the ledger, as only a few parties need to agree on the terms of the intended modification[136]. It is easy to imagine a consortium or single entity in charge of operating a private blockchain within which controllers and processors soon absorb most decision-making powers and individuals become largely under-represented – a perfect scenario for abuse of the system. However, to be GDPR-compliant, a blockchain-based PET such as the PACE Tool must necessarily use a private blockchain. Therefore, in the context of PACE Tool's design, the need to identify a controller that stems from the GDPR's centralised assumptions results in a lower level of control, transparency and accountability than that which can be achieved by relying on permissionless blockchains where there is no *per se* controller.

4.5.2. Our approach to this challenge

This challenge is the flipside of the trade-off discussed above in connection with the PACE Tool's high transaction costs and the choice between PoW-based and non-PoW-based blockchains. The security and integrity assurances provided by public blockchains' PoW protocol are indispensable to ensure that the records of data processing operations

[126] Chris Reed, 'Taking Sides on Technology Neutrality' (2007) 4 SCRIPTed 263, 275.

[127] Michèle Finck, 'Blockchains and Data Protection in the European Union' (2018) 4 Eur. Data Prot. L. Rev. 17, 17.

[128] Tom Kulik, 'Why Blockchain And The GDPR Collide Over Your Personal Data' (*Above the Law*, 8 October 2018) <https://abovethelaw.com/2018/10/why-blockchain-and-the-gdpr-collide-over-your-personal-data/> last accessed 09 August 2023.

[129] GDPR, Art. 4(7).

[130] CNIL, 'Blockchain - Solutions for a Responsible Use of the Blockchain in the Context of Personal Data' (2018) 2–3.

[131] Case C-210/16 *Wirtschafsakademie Schleswig-Holstein* [2018] ECLI:EU:C:2018:388, para 23; Case C-131/12 *Google Spain* [2014] ECLI:EU:C:2014:317, para 34.; Case C-25/17 *Jehovan todistajat* [2018] ECLI:EU:C:2018:551, para. 66

[132] Bacon and others (n 22) 64; European Parliament (n 7); CNIL (n 130) 1.

[133] For example, a government agency implementing a land registry blockchain-based platform.

[134] The means will always be the choice of blockchain.

[135] CNIL (n 130) 2, 5.

[136] Finck (n 26) 15.

performed within a cloud ecosystem remain unaltered; switching to a private blockchain would jeopardise the integrity of such records, yet it would allow for the determination of a controller and thereby make the PACE Tool GDPR compliant. Without authoritative guidance – i.e. a CJEU judgment – as to how controllership is to be determined in public blockchains, this trade-off is bound to result in a binary outcome: either a non-GDPR-compliant PACE Tool with a reliable ledger, or a compliant PACE Tool of dubious reliability.

4.5.3. Lessons and insights

Whilst there are significant challenges impeding the fulfilment of the data protection-relevant promises the blockchain carries, the fact that the GDPR may be steering innovation away from public blockchains and towards private ones deserves close attention. More specifically, a strict application of its provisions that are based on centralised assumptions risks curtailing the ability to experiment and innovate with public permissionless blockchains, thus calling into question its technological neutrality[137].

Therefore, as Tatar *et al.* suggest, '[p]utting an emphasis on what [the GDPR and the blockchain] are trying to achieve [may] be the right starting point for accommodating the technology and the GDPR'[138]. Put in other words, a teleological interpretation of the GDPR may be in order when a technology clashes with it on a micro-level, i.e. at the level of concepts and assumptions, but aligns with it on a macro-level, that is, at the level of objectives. Following this line of reasoning, when permissionless blockchains are the backbone of technological solutions which promote important objectives of EU data protection law, concepts such as controllership could perhaps be adapted or reinterpreted. In particular, as long as the GDPR is not amended to account for the dynamics of decentralised data processing, *ad-hoc* interpretations could be relied upon to circumvent the difficulties in determining controllership. There is no point in holding that either nodes, miners or users of permissionless blockchains are controllers or joint controllers, as neither of them will be able to comply with the obligations the GDPR imposes on controllers due to their lack of control over the relevant data and the blockchain's operations. Yet, the blockchain can promote some of the outcomes that traditional controllers are called upon to ensure, such as individual control, transparency and accountability. Thus, for example, the blockchain could be considered as the underlying protocol on which the PACE Tool is run – much like the Internet's classic TCP/IP – and the entities that deploy the PACE Tool (such as the online pharmacy) could be deemed controllers of their users' personal data that is stored on the blockchain (the cryptographic keys). Teleological interpretations of the GDPR like this one could ensure that this Regulation is applied in a way that does not suffocate permissionless blockchains' potential to promote data protection, with some of its objectives being achieved through means other than those originally contemplated in it.

5. Conclusions

This article has presented the main outcome of the PACE Project, a blockchain-based PET intended to enable trustworthy cloud-based websites and applications. In doing so, we explored the lessons, challenges and insights derived from our interdisciplinary effort. In particular, we focused on how different ways of reasoning and understandings of concepts such as DPbD can make communication between the fields of law and computer science difficult, ultimately resulting in work in siloes and suboptimal decisions that have to be revised at a later point in time. We also showed that the promises of techno-regulation are overstated, and that hardcoding legal provisions is only feasible for simple rules of low representational complexity. Given this realisation, we had to change the design of the PACE Tool, creating instead a tool intended to guide controllers in the correct application of data protection rules, to facilitate audits of data processing operations, and to promote individual control in a realistic way. We showed that factors such as a shaky legislative support for DPbD and the absence of privacy-based competition entail that many actors in the digital economy have no real incentive to deploy PETs. Also, deploying smart contracts can be computationally intensive and therefore prohibitively expensive. These considerations, coupled with the clashes between the GDPR's centralised tenets and public blockchains' decentralised features, dramatically reduce the PACE Tool's likelihood of adoption. We wanted to share the aforementioned challenges and disseminate the insights and lessons we learned from our efforts to overcoming them, hoping to inform other interdisciplinary projects that are increasingly important to shape a data ecosystem that respects our privacy and promotes the protection of our personal data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This article is supported by the Engineering and Physical Sciences Research Council (EPSRC) funded project PACE: Privacy-Aware Cloud Ecosystems (EP/R033293/1, EP/R033439/1).

[137] See generally Mireille Hildebrandt and Laura Tielemans, 'Data Protection by Design and Technology Neutral Law' (2013) 29 Computer Law & Security Review 509.

[138] Unal Tatar, Yasir Gokce and Brian Nussbaum, 'Law versus technology: Blockchain, GDPR, and tough tradeoffs' (2020) Computer Law and Security Review 38, 6