Decentralized data processing: personal data stores and the GDPR

Janssen, H.; Cobbe, J.; Norval, C.; Singh, J.

DOI
10.1093/idpl/ipaa016

Publication date
2020

Document Version
Final published version

Published in
International Data Privacy Law

Decentralized data processing: personal data stores and the GDPR

Heleen Janssen*, Jennifer Cobbe**, Chris Norval*** and Jatinder Singh****

Key Points

- When it comes to online services, users have limited control over how their personal data are processed. This is partly due to the nature of the business models of those services, where data are typically stored and aggregated in data centres. This has recently led to the development of technologies aiming at leveraging user control over the processing of their personal data.
- Personal data stores (PDSs) represent a class of these technologies; PDSs provide users with a device, enabling them to capture, aggregate, and manage their personal data. The device provides tools for users to control and monitor access, sharing, and computation over data on their device. The motivation for PDSs are described as (i) to assist users with their confidentiality and privacy concerns, and/or (ii) to provide opportunities for users to transact with or otherwise monetize their data.
- While PDSs potentially might enable some degree of user empowerment, they raise interesting considerations and uncertainties in relation to the responsibilities under the General Data Protection Regulation (GDPR). More specifically, the designations of responsibilities among key parties involved in PDS ecosystems are unclear. Further, the technical architecture of PDSs appears to restrict certain lawful grounds for the processing, while technical means to identify certain special categories of personal data, as proposed by some, may remain theoretical.
- We explore the considerations, uncertainties, and limitations of PDSs with respect to some key obligations under the GDPR. As PDS technologies continue to develop and proliferate, potentially providing an alternative to centralized approaches to data processing, we identify issues that require consideration by regulators, PDS platform providers, and technologists.

Introduction

Online services are driven by data; functionality and value are derived from its processing. However, individuals generally have little visibility—let alone control—over what, how, why, and by whom their data are captured, analysed, transferred, stored, or otherwise used.[1] In response to this, and to the growing public discourse regarding data-related issues, there is considerable focus by the computer science and engineering communities on developing *privacy-enhancing technologies* (PETs), ie technical tools and measures that can assist in addressing privacy concerns.[2] *Personal data stores* (PDSs) are one such technology, which aims to tackle

* Heleen Janssen, Compliant & Accountable Systems Group, Department of Computer Science and Technology, University of Cambridge, Cambridge, UK. E-mail: heleen.janssen@cst.cam.ac.uk
** Jennifer Cobbe, Compliant & Accountable Systems Group, Department of Computer Science and Technology, University of Cambridge, Cambridge, UK. E-mail: jennifer.cobbe@cst.cam.ac.uk
*** Chris Norval, Compliant & Accountable Systems Group, Department of Computer Science and Technology, University of Cambridge, Cambridge, UK. E-mail: chris.norval@cst.cam.ac.uk
**** Jatinder Singh, Compliant & Accountable Systems Group, Department of Computer Science and Technology, University of Cambridge, Cambridge, UK. E-mail: jatinder.singh@cst.cam.ac.uk

1 Peter Tolmie and others, 'This Has to Be the Cats - Personal Data Legibility in Networked Sensing Systems' (Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing, 2016) 490 <https://nottingham-repository.worktribe.com/output/775056> accessed 21 August 2020; Tim Berners Lee, 'One Small Step for the Web . . .' open letter by Tim Berners Lee (23 October 2018) <https://inrupt.com/blog/one-small-step-for-the-web> accessed 21 August 2020.

2 Royal Society, 'Protecting Privacy in Practice: The Current Use, Development and Limits of Privacy Enhancing Technologies in Data Analysis' (London, March 2019) <https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/> accessed 21 August 2020.

issues of privacy and user control by providing an alternative, decentralized approach to data processing. The European Commission in its European strategy for data expressed that such tools have significant potential, and that a supportive environment to foster its development is necessary in order to realize its benefits.[3]

The current dominant data processing model involves the capture or transfer of data from an individual to an online service, which makes use of those data. This often involves some centralized aggregation of data; that is, where an organization works to collect, store, and/or otherwise use the data across a range of individuals. In contrast, PDSs propose a model in which the data and its processing are 'decentralized'. Conceptually, they provide a way for an individual to capture (some of) their data, in doing so putting them in a position to decide the processing or transfers that may occur.[4]

Generally speaking, PDS technology entails equipping an individual (a PDS user) with a *device* dedicated to managing their data (section 'PDSs: an overview elaborates this). A device entails a technical environment that encapsulates the user's data, providing mechanisms that enable that user to monitor, mediate, and/or control:

(a) the data flow in/out of the device; and/or

(b) the computation that occurs over the device's data.

In line with this, PDSs purport to empower users through means that 'put individuals in control of their data'.[5] In essence, the concept is that third parties cannot access or take data off, or perform computation over, the data of the PDS without user agreement or intervention. With regards to compute, it is argued that by effectively allowing 'compute to be brought to the data', the privacy concerns commonly associated with the current centralized approaches to data analytics can be tackled.[6]

In a PDS context, app developers seeking to perform computation over a user's data will not require direct access to that data (ie by having it transferred to them). Rather, the computation can be performed in the user realm (on the user's device), in line with the user's agreement and only over a defined set of their data. The app developer receives the results of the computation—which, as PDS proponents argue, has the potential to be less personal, sensitive, or privacy-invasive, and potentially more valuable for companies.[7]

The technology, though nascent, is growing in prominence. There are various PDS initiatives underway at various stages of maturity, though all appear to be still 'underdevelopment' and evolving. Example deployments currently available include Dataswift/Hub of All Things (hereafter 'Dataswift/HAT'[8]), Mydex, DigiMe, or CitizenMe, while others—such as Databox and Solid (led by Sir Tim Berners-Lee[9])—are at early stages in the development process, and not yet readily available for consumers.

This article explores the nature of this emerging technology and how it relates to issues around data protection as well as indicating some of the assumptions, opportunities, possible issues, and areas for consideration regarding PDSs. We begin, in section 'PDSs: an overview', by introducing the general nature of the technology and the benefits that are claimed. We then consider issues of responsibility in the context of the General Data Protection Regulation (GDPR[10]). We focus on allocations of responsibilities in a PDS ecosystem from the perspective of three key actors: *users* (those who hold a PDS), *PDS platform providers* (those building or operating the PDS infrastructure), and *app developers* (those organizations seeking to leverage the data

3 European Commission, A European Data Strategy (Brussels, 19 February 2020) COM (2020) 66 final, 10.

4 Opinion 9/2016 of the European Data Protection Supervisor (EDPS) on Personal Information Management Systems: Towards more user empowerment in managing and processing personal data (20 October 2016) OJ C463/10.

5 Some PDSs, like the Databox, explicitly aim at user control, privacy, and data protection, see Andy Crabtree and others, 'Building Accountability into the Internet of Things: The IoT Databox Model' (2018) 4 Journal of Reliable Intelligent Environments 39, 51. Other PDSs offer users opportunities to monetize their personal data, see eg Dataswift/Hub of All Things <https://www.hubofallthings.com> and <https://dataswift.io/about>; Solid <https://Solid.inrupt.com>; CitizenMe <https://citizenme.com> accessed 21 August 2020.

6 Crabtree and others, ibid; Aantti Eskola and others, 'MyData, A Nordic Model for Human-centred Personal Data Management' (2014) 3 <http://julkaisut.valtioneuvosto.fi/handle/10024/78439> accessed 21 August 2020.

7 Eskola and others, ibid 3.

8 HAT recently changed its name into Dataswift/HAT, see <https://dataswift.io/about> accessed 21 August 2020. Throughout this article, we refer to Dataswift/HAT. Dataswift Ltd (formerly HAT Data Exchange Ltd) is a commercial enterprise spun out from the university led HAT research project tasked to build the decentralized HAT infrastructure and the governance model. From 2015–18, Dataswift built technology services and set up the legal and economic model (the HATDeX Governance Platform) around the open-sourced HAT to scale the issuance of HAT personal data accounts.

9 <hubofallthings.com>; <mydex.org>; <https://digi.me>; <https://citizenme.com>; <http://iotdatabox.com>; <https://solid.inrupt.com> accessed 21 August 2020. Note that Solid is operational; however, the website informs visitors that the Solid POD is currently designed for use by a highly technical audience of developers.

10 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 personal data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016) OJ L119/1. Art 4(1) GDPR defines personal data.

of PDS users). In the context of PDSs, platforms and app developers are usually the entities developing and providing the apps entailing data transfer and/or computation. In section 'How legal concepts might challenge the design of PDS technology', we explore how the law might challenge the technical architecture, thereby focusing on how the GDPR's lawful grounds for processing in federated systems might function, and whether technical identification of special categories of personal data in PDSs can be achieved. We conclude with some observations.[11]

PDSs: an overview

PDS platforms seek to give users control over the management of their personal data. We consider PDSs as platforms as they provide the means and infrastructure for mediating between users and those seeking to process their data. In essence, PDSs represent a point for user intervention and mediation in digital ecosystems. Several PDSs exist, at various stages of development, with some variance in their approach. As a nascent technology, and as an exploratory paper, we now present some of the common and general properties of PDSs. Naturally, as the offerings and architectures vary, not all properties will apply to all platforms.

PDS architectures: technical aspects, devices, and the role of apps

PDSs aim to provide an alternative to the current, more centralized approaches of data processing—which typically entail data being transferred and collected by app developers, who then perform computation over the data they hold. Current data processing environments mean that users generally have limited visibility over these aspects, let alone control. In contrast, PDSs seek to localize capture and processing of data to the individual.

PDSs involve equipping a user with a *device* that is dedicated to a user's personal data. In essence, it represents a 'container' (in a non-technical sense of the word) within which the user's personal data can be captured, stored, and where computation can occur.[12] This device may entail a physical piece of equipment (eg Databox envisages a physical device designed for a home), and/or a virtual component that is (logically) separated by software,[13] possibly stored in the cloud.[14] Through this device, PDSs generally seek to provide functionality pertaining to one or more of the following[15]:

1. local capture and storage of a user's personal data within their own device—managed by the user;
2. local, on-device computation (analytics) through constrained and managed apps, in line with user preferences;
3. management and control over the transfer of the (raw) data and/or results of analytics or other computation from the device; and
4. a degree of user monitoring, management, and control regarding the above.

Through their functionality, PDSs seek to give individual users *control*, not only by giving more detail regarding the apps they install, but also through mechanisms that allow users to define what apps may do with their data. More user control is envisaged as being facilitated by higher levels of *transparency* over data processing. That is, PDSs generally operate to provide information to users, through various means, about what data the app wants to access and how those data will be processed. At the same time, PDS platforms may also themselves undertake risk assessment processes for the apps they support. We now elaborate on these properties of PDSs.

Data capture and storage

Users generate data in various ways—be it via sensors, online platforms, browsing habits, data entry, typing habits, generated by software or in apps, etc. In a PDS context, data capture and storage happen 'locally', within a user device. PDSs offer users some means to decide and control the capture of data, through technical mechanisms that enable, mediate, and manage the input flows of data. In addition to allowing users to manually input data, some PDSs provide for 'data collectors' or for 'drivers', which once set up can

11 Note, as an exploratory paper, we do not attempt exhaustively consider the range of PDS platforms. Instead, we focus on the common traits and purported functionality of this emerging technology—in light of its current state (and available descriptions). Our aim is to highlight issues and influence the discussion as this technology continues to develop.

12 We use 'device' in a broad conceptual sense, where the data or compute on, in, or within device generally refers to that the device governs, *regardless of the specifics of its underlying technical implementation*. Note, however, that technical architectures or implementations can vary, and there is potential for PDSs where devices govern through interfacing, managing, and mediating other technical components. Such technical specifics will depend on the particulars of the implementation and are beyond the scope of this article.

13 For instance, through the use of virtualized containers—see, for example, Donald Firesmith, 'Virtualization via Containers', Blogpost at Carnegie Mellon University (25 September 2017) <https://insights.sei.cmu.edu/sei_blog/2017/09/virtualization-via-containers.html> accessed 21 August 2020.

14 Some platforms offer users cloud storage of their PDS (HAT), while other platform let the choice for storage to users (eg Digi.me <https://developers.digi.me/private-sharing-overview> accessed 21 August 2020).

15 EDPS (n 4).

automatically feed a PDS with user data originating from various sources, be they sensors, online services, and so on.[16]

Managed, on-device computation, and transfer

Many PDSs seek to allow the computation (including analytics) to be 'brought' to the data. This contrasts with today's common approach where data are transferred to remote, third-party operated servers for computation to occur.

Computation on a PDS device generally occurs through '*apps*'. An app executes on a user's device, working to process the local data stored on that device.[17] Some apps may perform processing and provide the user with functionality without involving any external transfers of data. Such apps might require access to data from the PDS to do the computation locally (on the PDS), but that app does not transfer nor store user data anywhere outside the PDS. That is, the app performs all processing locally, within the particular user's device. Other apps may transfer data from the device to the app developer—if permitted by the user. These transfers can include user data itself (stored or real-time), or the results of computations/analytics that the apps have performed over data.

PDSs generally provide means for constraining apps, particularly regarding the data accessed, aggregated, or released. This means that usually, app developers must specify the types of data they will access, transfer (perhaps raw data, or perhaps the result of computation), and the computation they will undertake.[18] Similarly, users may be enabled to specify their own constraints over an app's behaviour.[19] Users might for instance limit the number of times data are accessed by an app through granular settings (for instance by limiting access to motion data from a bedroom sensor).

These constraints are generally enforced and guaranteed by the PDS's technical operating environment and by legal arrangements between platform providers and app developers. In short, the user must agree to the general nature of the technical functionality provided by the app, as well as the data transfers that the app enables. Legal constraints will generally also apply; that is, platforms often have terms of services (ToS) and conclude contracts with app developers to which an app developer (and the platform) is bound (see section 'Constraints on apps to leverage user control').

PDSs work to mediate between a user and an entity seeking to process user data (ie an app developer). However, PDSs also facilitate large-scale (or collective) personal data processing. That is, computation can be conducted across a population, by operating over a range of PDSs (again, via apps). This might be, for example, to support participative health studies such as those relating to epidemiology,[20] or for a supermarket to monitor product usage within homes, as relevant for advertising, stock management, etc. A key difference between these collective computations in a PDS context, as compared with computations in more centralized data processing environments (ie where the organization holds and analyses the data of individuals), is that such computation is executed on the devices of PDS users, with only the results of that computation transferred to the organization. This contrasts with the centralized data processing models of today, where the data of many users are collected and aggregated by the organization for processing. In this way, (i) the PDS provides a potential intervention point for giving each user some degree of control over the data processing that may occur, and (ii) the party interested in the results of computation does not have access to, or need to manage or be responsible for, the (raw) user data that drives such computation.

Monitoring

Some PDS platforms provide measures, such as logs, audits, and visualizations, to give insight over the happenings within a PDS device, including that of app behaviour. These aim at increasing levels of transparency, by allowing users to (pre)view data processing operations, data transfers, or to inspect historical operations of apps,[21] as well as to review and evaluate their

16 CozyCloud (through its CozyCollect application) or Digi.me provide their users with a catalogue of connectors to retrieve many kinds of personal data, including financial data (eg Paypal or banks), administrative data (eg electricity data or consumption traces, insurance contracts), music (eg Spotify) or fitness data (eg Fitbit), examples borrowed from Nicolas Anciaux and others, 'Personal Data Management Systems: The Security and Functionality Standpoint' (2019) 80 Information Systems 16; Dataswift/HAT for instance collects these data through its 'Data Plugs', 'HAT Glossary of terms used in the HAT ecosystem' <https://www.hatcommunity.org/hat-ecosystem-glossary> accessed 21 August 2020; see tab 'Data Plugs' <https://hatdex.dataswift.io/hatstore2/2018/5/7/spotify-data-plug> accessed 21 August 2020; for Databox, see Crabtree and others (n 5) 44.

17 Note that the terminology varies by platform; not all platforms would describe processing as occurring through apps, though generally there is some conceptually similar construct.

18 Crabtree and others (n 5) 47; Lachlan Urquhart, Tom Lodge and Andy Crabtree, 'Demonstrably Doing Accountability in the Internet of Things' (2019) 27 International Journal of Law and Information Technology 1, 17.

19 Platforms envisage to provide granular controls (eg Crabtree and others (n 5) 42, 49.

20 Anciaux and others (n 16) 21. The developers of Databox envision possibilities for users to share apps with other users within the home and between distributed Databoxes in other homes.

21 Crabtree and others (n 5) 51.

preferences, constraints, etc. PDS proponents argue that such measures are key to user empowerment.[22]

Constraints on apps to leverage user control

PDS platforms generally provide legal and technical mechanisms through which they seek to ensure that app behaviour (including computation and transfers of data) is properly constrained in line with user preferences, and developer and platform requirements. We elaborate some of these aspects below. Note that the mechanisms driving some PDSs—consisting of apps, app stores, etc—appear similar to current mobile app ecosystems, such as that of Google's Play Store or Apple's App Store. However, the vision of PDS is for far more granular and detailed controls, interventions, and mechanisms for transparency.

Legal constraints on apps

PDS technology will commonly be accompanied by legal constraints on app developers, which work to regulate app behaviour. Such constraints are usually written into a platform's ToS. ToS commonly include definitions of terms, how the platform can be used, and how responsibilities and liabilities are envisaged by the platform. Platform's ToS require app developers to have appropriate technical and organizational measures in place to safeguard users against unlawful processing and against accidental loss or destruction of data, while requiring users to, for instance, safely keep their passwords or to regularly update their PDSs to keep their PDS secure.

Contracts between a platform and app developer are commonly based on the platform's ToS and made business-specific. An app developer's technology must be compliant with the constraints determined in the platform's legal documents. Contracts might for instance prescribe that app developers fully comply with platform policies and principles to the processing of user data, or stipulate that the platform can terminate data processing on a specific user's PDS whenever the app developer acts in a non-compliant manner.[23]

Technical constraints on apps

PDSs propose to implement a variety of technical constraints on apps, as follows:

1. Manifests—app data processing specifications: PDS platforms seek to give users control through technical mechanisms that allow them to define what apps may do with their data. To that end, platforms require app developers to put in place some form of 'manifest' (specification or notice) with their apps as a precursor to being able to run on the PDS device.
   Manifests are usually a technical, 'machine-readable' file (though can also include 'human-readable' descriptions) which describe the data sources that the app seeks to access, and how the app will process that data. This includes, for example, details of the data sources and types that the app seeks to access, transfer, and conduct computations on, the app developer's purpose of the processing, the risks attached to the app (see below), or it may list third parties with whom the data are shared by the app developer. Some aspects of the manifest, such as the data sources, will be enforced by the platform—ie an app may only access the data sources that it specifies in its manifest (see below).
   Technical manifests can be presented in various ways, so as to be legible to users. This may include or otherwise be represented (in more a user-accessible manner) in the PDS platform's 'app store' (should one exist for that PDS ecosystem).[24]
2. User-defined data control specifications: Some PDS platforms also provide generic mechanisms for users to specify their data preferences, which apply 'horizontally' across apps. That is, a user can specify, for example, that particular data, such as that from a particular data source (eg sensor), may not ever leave the PDS. Similarly, PDSs also allow various control preferences to apply to specific applications. Mydex, for example, seeks to stimulate the development of apps that let a user choose what information is shared with the app developer, and whether this is automatic or requires your permission each time.[25] Dataswift/HAT operates its Platform Rating Declaration which informs users about the platform's ratings of apps.[26] Databox envisages

22 Ibid 47, 51; Urquhart, Lodge and Crabtree (n 18) 19; Anciaux and others (n 16) 17.

23 Examples borrowed from Mydex <https://dev.mydex.org/terms-for-connections.html> accessed 21 August 2020.

24 For example, for Databox see Crabtree and others (n 5) 44; see for Dataswift/HAT their HATstore <https://hatdex.org/hatstore> accessed 21 August 2020.

25 Users might for instance want to share their address but not their telephone number with an organization, example taken from Mydex <https://pds.mydex.org/what-personal-data-store-0> accessed 21 August 2020.

26 <https://www.hatcommunity.org/hat-dex-rating> accessed 21 August 2020.

stimulating app developers to develop mechanisms that offer users granular data settings.[27]

3. Mechanisms upon installation of apps including risk assessments from app stores: To indicate app behaviour, platforms may (depending on their motivation) perform their own risk ratings and assessments of applications.[28] Some platforms only allow apps to function in PDSs or to appear within their app store after they have been (manually or automatically) screened for risks they might pose.[29] This generally involves an analysis (or parsing) of the app's manifest, and in some cases, potentially an analysis of the app's code.

   These risk ratings are presented to the PDS users looking to install that app.[30] By screening risks, platforms seek to contribute to better-informed user decision-making.[31] In the Databox ecosystem, for instance, it is envisaged that the final risk rating of the app will be determined by factors such as the data they process, what they do with the data, the components they seek to interact with (such as devices in the home), etc.[32] This rating is presented to users prior to the installation of apps on user PDSs, so that they are in a better position to decide whether they are appropriate.[33] Note that risk assessment regimes appear nascent and are mostly still under development.

4. Ensuring preference adherence: Once a user has agreed with an app's manifest and possibly configured various preferences and constraints within, and the app has been installed on the PDS, the app's on-device processing of their data may begin. The device works to ensure technical adherence with the requirements specified by the manifest and by the user.[34]

   PDSs commonly operate to ensure (technical) adherence with the manifest's conditions at run-time (during an app's operation). The platform will use the manifest, user constraints, and possibly other considerations to ensure that apps technically comply (where compliance at a technical level can be ascertained) with the associated data specification and usages at run-time.

   As apps operate on a user's behalf, some PDS platforms envisage enabling users to inspect what has happened with their data—eg why the light was switched on at 5 PM, or why the discount was given on car insurance.[35] Some of this might be apparent from the manifest. However, also envisaged are means that enable the audit, inspection, or visualization of what happens or has happened to data, and the computation that is occurring or has occurred on device. It is thought that this contributes to the transparency of processing.[36]

Purported benefits of PDSs

To convince potential users and app developers to engage with PDSs, platforms claim a wide range of benefits.[37]

Benefits for users

Proponents of PDSs claim various benefits relating to user empowerment, including:

1. granular controls over data processing[38];

27 Users could set their preferred measurement in 'sliding scale' data settings between every minute or every six hours whether a home is occupied for, eg an energy company, see Crabtree and others (n 5) 49.

28 Some PDSs do not undertake any risk rating of apps, see eg <Digi.me https://digi.me> accessed 21 August 2020.

29 Crabtree and others (n 5) 43; in the context of Dataswift/HAT, see <https://hatdex.org/hatstore> accessed 21 August 2020.

30 See for risk rating for instance Dataswift/HAT, HAT Platform Standard Operating Procedures (PSOP) <https://static1.squarespace.com/static/59e5e08bfe54ef97e3d095f0/t/5c2d1ab64d7a9c5a97ad1655/1546459834256/PSOP-2.3.pdf> accessed 21 August 2020; the Databox platform seek to differentiate for instance among three types of risks: 'legal', 'technological', and 'social' risks which sit in their risk spectrum from 0 to 5. 'Social risks' include risks whenever apps access 'sensitive information or produce results that may be deemed sensitive (as articulated by article 9 (1)(a) GDPR)'; 'legal risks include risks associated with the GDPR, particularly those implicated in apps taking data of the PDS, apps that export user data outside the EU or apps that transfer data to other recipients, while 'technical risks' include apps that use devices that have not been validated by the SDK, use unverified code, or physically actuate essential infrastructure or potential dangerous devices in the home, see Crabtree and others (n 5) 47.

31 For instance, the Databox manifests consist of three layers, which provide a user with a short description of the specific purpose of the processing, a condensed description providing the information required by the GDPR and full legal terms and conditions. Dataswift/HAT uses 'Data Debits' as a primary mechanism for app developers to retrieve data from the user PDS. Users can only agree to share data from their PDSs whenever Dataswift/HAT has established the app developer as a 'verified' actor, see HAT PSOP (n 30) 13.

32 Urquhart and others (n 18) 21.

33 Ibid 21. In the HAT, the platform can classify apps as 'low risk', or, where relevant, with 'no risk' rating in the PDS app store, depending on the specifics of the architecture of the app, see HAT Data Exchange Rating Scheme, 1, 2 <https://www.hatcommunity.org/hat-dex-rating> accessed 21 August 2020.

34 Crabtree and others (n 5) 47.

35 Ibid 48, 49.

36 Ibid 50.

37 We perceive a 'user' in GDPR context as a natural person acting in the capacity of a data subject in the sense of art 4(1) GDPR, who is entitled to data subject's rights (this will be further elaborated in section 'The role of users in the PDS ecosystem').

38 Data Protection Working Party Article 29 (hence 'WP29') wrote in 2014 that 'Device manufacturers must provide users with granular choices over data capture', see 'Opinion 8/2014 on Recent Developments on the Internet of Things' (WP 223 of 16 September 2014, 21, 22). WP29 is an independent European body, which contributes to the consistent application of data protection rules throughout the European Union (EU) and promotes cooperation between the EU's data protection authorities. It

2. user consent will become better informed as users have more information over data processing (eg through specifications derived from manifests, risk rankings as well as real-time logs, audits, permanent monitoring, and visualizations[39]);
3. the architecture, including controlled collection, transfer, and on-device processing, means that data (including 'sensitive data') are better protected from access by third parties[40];
4. isolated storage of user data and apps to prevent apps from unobtrusively interacting with data (and other apps) without user agreement/intervention[41];
5. generally incentivizing app developers towards taking more privacy-friendly approaches[42]; and
6. opportunities for users to transact and otherwise monetize their personal data.[43]

Benefits for app developers

A common argument is that if users feel that they better understand how their data are being processed and feel empowered by way of the control mechanisms a PDS platform provides, they may have less 'resistance' against data sharing (cf 'giving up their data for processing'), and harbour a greater 'willingness' to enable the processing of their data by a PDS.[44] This could result in enabling organizations to undertake data processing over data that would otherwise be unavailable.

Similarly, given that a PDS is envisaged to encapsulate a wide range of individual data, PDSs offer app developers' access to a potentially broader range of data *types* (eg enabling the combination of medical data with eating patterns or bank statements with shopping history to analyse health or spending patterns) than they would otherwise have if they attempted to collect the data themselves.

It follows that PDS proponents claim that app developers would benefit by being able to perform computations and analytics on a far richer set of personal data than in more traditional, centralized arrangements. As a result, it is claimed, PDSs can facilitate a broader range of analytics and insights. And while such data and its processing are subject to constraints in a PDS ecosystem (in line with platform restrictions and user demand), it is argued that many companies are interested in analytics across different sets of user data or particular sets across a user population and often show less interest in individuals themselves.[45]

In addition, it is argued that companies would benefit from reduced burdens associated with acquiring and managing the users' personal data.[46] That is, the liability exposure and operational costs of companies are expected to be reduced as companies would no longer have to manage the same degree of personal data on their own infrastructure.[47]

Enabling business models

Most PDS platforms seek to monetize operations in user PDSs; various business models are proposed by platforms, many of which appear under development. Some platforms might require payments per data transaction between users and app developers, where the type of transaction may determine the amount to be paid.[48] Data transactions may for instance include an app developer's access to user data or an app developer's transfer of user data (results of computation or raw data). Other platforms might require app developers to pay an 'entrance fee' to be listed on the platform's app store or otherwise to enter the particular PDS' ecosystem. Entrance fees might also be combined with app developer payment per 'data transaction'.[49] This latter business model might again combine with other, more refined collections of revenue (eg where the platform collects a fee for an app developer's access to user data or for data transfers, and/or for collective computations,

consists of representatives of national Data Protection Authorities and writes opinions interpreting specific elements of EU data protection law. Their Opinions (or Guidelines) are not legally binding; they however do tend to have impact, see Christoper Kuner, *European Data Protection Law: Corporate Compliance and Regulation* (Oxford: OUP 2007) 9. With the entry into force of the GDPR, the European Data Protection Board (EDPB) has replaced WP29. See also Digi.me: 'Only you decide whether to share data with apps for benefits and personal insights' <https://digi.me>; Crabtree and others (n 5) 42; Urquhart, Lodge and Crabtree (n 18) 16.

39 Eskola and others (n 6) 3.

40 Crabtree and others (n 5) 46; Tom Lodge, Andy Crabtree and Anthony Brown, 'Developing GDPR Compliant Apps for the Edge' in *Proceedings of the 13th International Workshop on Data Privacy Management* (Barcelona Springer 2018) 313.

41 Crabtree and others (n 5) 43.

42 Ibid 45.

43 Irene Ng and Hamed Haddadi, 'Decentralised AI Has the Potential to Upend the Online Economy' (*Wired*, 28 December 2018) <https://www.wired.co.uk/article/decentralised-artificial-intelligence> accessed 21 August 2020.

44 For example, for MyData see Eskola and others (n 6) 4; Digi.me <https://digi.me/partners> accessed 21 august 2020) 26; CitizenMe <citizenme.com/public/wp/business/> accessed 21 August 2020.

45 Eskola and others, ibid 4.

46 Ibid 4.

47 Crabtree and others (n 5) 23.

48 CitzenMe rates company queries to users as most expensive; requests for behavioural personal data are less expensively rated, while the distribution of self-supported data is rated as least expensive data transactions, see <https://www.citizenme.com/public/wp/business/pricing/> accessed 21 August 2020.

49 For Mydex, see <https://community.mydex.org/question/how-does-mydex-cic-make-its-money> accessed 21 August 2020.

whereby the platform charges a fee per question per user PDS[50]).

Another approach is for platforms to not require an app developer's payment for entrance fees or payments per transaction, but rather to raise revenue through the sale, renting, or leasing of PDS devices to users, which may perhaps include service or maintenance contracts.

Finally, platforms might exist that instead pursue general public interests (eg charity foundations seeking to assist elderly people, or platforms offering subject-oriented PDSs, eg health purposes), where entrance fees and/or transaction costs may be absent. Such platforms are likely financed by public means (eg by allowing for tax deductions in the case of charity foundations, or subventions in the case of public's health-purpose platforms).

Potential issues for GDPR compliance

We have described how PDSs essentially provide a point of mediation between users (and their data) and organizations (app developers) seeking to use their data. Given that PDS architectures are designed specifically to address issues regarding personal data, the GDPR is relevant and must be complied with by those involved in undertaking that processing whenever it applies. In this section, we consider the roles and responsibilities as part of the PDS ecosystem, with respect to the GDPR. Again, our focus is on the relationships between three key actors in the PDS ecosystem: users, platforms, and app developers.

We note that both the technology and their governance models are still developing.[51] Therefore, we examine the approaches of some current platforms as these represent an important starting point for exploring the effectiveness of PDSs (and their vision) in dealing with data protection concerns. Our aim is to indicate areas for consideration, which, in turn, can influence the way the technology (and associated governance regimes) might evolve.

This section starts with some legal background and context, through the GDPR, opinions by WP29, and relevant case law. We subsequently analyse some currently available terms of service documents from PDS platforms as to how these platforms envisage distributions of responsibilities among users, app developers, and the platforms themselves. We then explore how case law would allocate roles and responsibilities in a hypothetical setting. We conclude that the roles and responsibilities of users, platforms, and app developers in a PDS context are an area requiring further attention.

Setting the scene for responsibilities in PDSs: developments under the GDPR

There are a range of perspectives on legal responsibility and its connection to accountability as it relates to technical environments.[52] Given the relevance of the GDPR to PDSs, our focus is on the notion of 'responsibility' as enshrined in the GDPR. Relevant here is the concept of the *data controller* (or simply 'controller').[53] Article 4(7) GDPR defines data controllers as 'the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data [...]. Article 5(2) GDPR allocates *responsibility* to controllers for compliance with the data protection principles as enshrined in Articles 5(1)(a)–(f) GDPR[54] and with a range of other obligations relating to that responsibility (Table 1).

50 Ibid.

51 Moreover, we do not analyse all initiatives, nor do we purport to be complete in overseeing all aspects of the PDS landscape.

52 Helen Nissenbaum, 'Accountability in a Computerized Society' (1996) 2 Science and Engineering Ethics 25 <https://pdfs.semanticscholar.org/cf32/d610202300ce228eb6a5a0149ea66af1fac2.pdf> accessed 21 August 2020; Mark Bovens, 'Analysing and Assessing Accountability: A Conceptual Framework' (2007) 13 European Law Journal 448; Colin J Bennett, 'The Accountability Approach to Privacy and Data Protection' in Daniel Guagnin and others (eds), *Managing Privacy Through Accountability* (London: Palgrave MacMillan 2012); Joseph Alhadeff, Brendan Van Alsenoy and Jos Dumortier, 'The Accountability Principle in Data Protection Regulation: Origin, Development and Future Directions' in Guagin and others (eds), *Managing Privacy Through Accountability* (London: Palgrave MacMillan 2012) 49; Jatinder Singh and others, 'Responsibility & Machine Learning: Part of a Process' (27 October 2016) <http://dx.doi.org/10.2139/ssrn.2860048> accessed 21 August 2019; Chris Reed, Elizabeth Kennedy and Sara Silva, 'Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning' (2016) Queen Mary School of Law, Legal Studies, Research Paper 243 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2853462> accessed 21 August 2020; Urquhart, Lodge and Crabtree (n 18); René Mahieu, Joris van Hoboken and Hadi Asghari, 'Responsibility for Data Protection in a Networked World: On the Question of the Controller, "Effective and Complete Protection" and Its Application to Data Access Rights in Europe' (2019) 10 JIPITEC 85, para 1; Jatinder Singh and others, 'Accountability in the IoT: Systems, Law, and Ways Forward' (2018) 51 IEEE 54; Lilian Edwards and others, 'Data Subjects as Data Controllers: A Fashion(able) Concept?' (*Internet Policy Review, Internet Journal on Internet Regulation*, 13 June 2019) <https://policyreview.info/articles/news/data-subjects-data-controllers-fashionable-concept/1400> accessed 21 August 2020; Jiahong Chen and others, 'Who Is Responsible for Data Processing in Smart Homes? Reconsidering Joint Controllership and the Household Exemption' (18 November 2019) <https://ssrn.com/abstract=3483511> accessed 21 August 2020.

53 WP29, 'Opinion 3/2010 on the Principle of Accountability' (WP 193 of 13 July 2010) and 'Opinion 1/2010 on the concepts of "controller" and "processor"' (WP 169 of 16 February 2010); EDPS, as of 25 May 2018 the successor of WP29, Guidelines on the concepts of controller, processor, and joint controllership under Regulation (EU) 2018/1725 (7 November 2019).

54 Art 5(1)(a) refers to requirements of lawfulness, fairness, and transparency of processing, art 5(1)(b) to the requirement of purpose limitation, art 5(1)(c) to the requirement of data minimization, art 5(1)(d) to the requirement of data accuracy, art 5(1)(e) to storage limitation, and art 5(1)(f) to the requirement of data integrity and confidentiality.

Table 1. Controller and processor core obligations

GDPR articles	Obligations	Controller	Processor
Article 5	Comply with fundamental data protection principles	+	+ (according to controller instructions)
Article 13 and 14	Comply with information obligations		
Articles 15–22	Data subjects' requests to exercise their rights	+	
Article 25	Implement data protection by design and by default	+	
Article 30	Keep records of processing activities	+	+
Article 32	Implement appropriate technical and organizational measures	+	+
Articles 33 and 34	Comply with the personal data breach notification obligations	+	+ (Article 33 only)
Article 35	Carry out data protection impact assessments prior	+	
Article 36	to risky processing in certain situations	+	
	If necessary, consult with the supervisory authority prior to such processing		
Articles 44–49	Comply with rules regarding transfers of personal data outside the EEA	+	+

Knowing who is (joint) controller or processor is crucial, as these roles relate to responsibilities, obligations, and liabilities under the GDPR. (Joint) controllers and processors can face significant penalties for not complying with the GDPR. However, distinguishing controllers from processors (and circumstances where courts might consider controllership as 'joint') becomes less clear. Recent case law indicates that the technical configurations have significance in determining where GDPR-compliant assignment of (joint) controllership is involved; however, as we will see in the context of assigning responsibilities in PDS ecosystems, these assignments need further clarification.

Controllers and processors

Article 24 GDPR obliges controllers to implement appropriate technical and organizational measures to ensure that their data processing complies with the GDPR. These measures must be designed to meet the GDPR's data protection by design and by default obligations, non-compliance with which can potentially result in fines, orders to cease processing, etc.[55] The *accountability* principle as enshrined in Article 5 GDPR obliges controllers to demonstrate that their processing of personal data accords with the GDPR. Whenever responsibilities are shared among several parties, they operate as '*joint controllers*'.[56] Joint controllers determine their respective responsibilities for compliance with the data protection principles and obligations, in particular in their relations with data subjects. Irrespective of the terms of how responsibilities for compliance are allocated, the data subject may exercise their rights in respect of and against each of the controllers.[57]

A '*data processor*' (processor) is defined as a natural or legal person, public authority, agency, or other body which processes personal data on behalf of the controller.[58] A processor is to implement the instructions given by the controller with regard to the purpose of the processing and the essential elements of the means for that processing.[59] The lawfulness of the processor's data processing activity is determined by the mandate given by the controller.[60] Both the controller and the processor must implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk posed by the processing to the

55 Art 25 GDPR. Note as regards administrative fines, regulators shall, when deciding whether to impose an administrative fine and deciding on its amount, pay due regard to the degree of responsibility of the controller or processor, and pay due regard to the technical and organizational measures implemented by them pursuant to art 25 (art 83 (2)(d) GDPR).

56 Art 26 GDPR.

57 Art 26(3) GDPR.

58 Art 4(8) GDP; see art 28 GDPR for obligations and responsibilities.

59 WP29 1/2010 (n 53) 25.

60 This is eg reflected in art 28(2) that stipulates that a processor shall not engage another processor without prior specific or general written authorizations of the controller.

rights and freedoms of natural persons.[61] Table 1 presents an overview of controller and processor obligations.

Importantly, although claims of compensation can be brought against both the controller and processor, a processor can only be held liable for the damage caused by processing where they did not comply with their obligations under the GDPR, or if they acted in contrary or outside the instructions given by the controller.[62] Whenever data subjects suffer material or non-material damages as a result of non-compliance with the Regulation, they are entitled to receive compensation from the controller or processor.[63]

The basic elements of the GDPR centre around its legal categories of stakeholders: controllers, joint controllers, processors, and data subjects.[64] These categories were designed for an environment of data processing in which there are direct relationships between data subjects and data controllers. This approach was prevalent at the time the Data Protection Directive (DPD) was written,[65] and has largely remained intact in the GDPR. The underlying logic is that the controller is the main 'architect' of an information system, who determines the entire system's operation[66] and can engage with services or systems of other organizations to perform specifically defined tasks. That is, the controller is in 'control' of processing (hence the name) and, therefore, responsible for it.

WP29 interpreted the wording 'determines the purposes and means' in its 2010 Opinion. It envisaged that determining control can be given to a particular actor by three means: (i) by explicit power, as allocated by law; (ii) by implicit power arising from legal practices (for instance by an employer handling employee data), and (iii) by factual influence, which can be triggered by mere fact.[67] Regarding the latter, responsibilities are to be assigned where non-legal facts 'dictate' who the controller is.[68] This might, for instance, apply whenever there is a contract or legal provision, but where the factual influence of an actor in that situation does not accord with the stipulations of the contract or the legal provision. The GDPR stipulates that where a processor determines the means and purposes, they will be considered a controller.[69] As regards 'joint control', WP29 indicated that this might arise in various ways—in evaluating, one must consider whether, at the macro level, the processing operations form 'a set of operations' where joint purposes and means are involved. This may apply where shared infrastructures were established by the parties.[70]

This underpinning logic might not easily apply in relationships between actors in contemporary and emerging technical ecosystems.[71] These ecosystems can involve complex technical supply chains, eg with cloud and other service providers involved.[72] Decisions traditionally reserved for controllers—over which hardware and software shall be used, which data shall be processed, how it should be processed to achieve a controller's purposes, for how long it shall be processed, who shall have access to it, what security measures should be taken, for how long data should be stored—are decisions that are nowadays often considered by processors.

As a result, considerations of factual influence over the data processing might still not result in clear determinations of roles as a controller or a processor.[73] This might for instance be demonstrated with the widespread use of cloud computing service providers, which underpins a great many of online services, and which is frequently relied on by controllers. That is, cloud services provide storage, compute, and data management

61 Arts 28(3)(c) and 32(1) GDPR; once a data breach occurs, a processor is, after becoming aware of a personal data breach, obliged to notify the controller without undue delay. The controller is obliged vis-à-vis the supervisory authority and the data subjects to notify the personal data breach.

62 Further, processors are obliged to make available all information necessary to the controller to demonstrate their compliance with the GDPR, see art 28(3)(h) GDPR. This information should support audits, including inspections by the controller, see art 28(3)(h) GDPR.

63 Should more than one controller or processor be involved in the same processing and be responsible for any damage caused by that processing, each controller or processor shall be held liable for the entire damage to secure effective compensation of data subjects (art 82 GDPR); see for liability of the processor art 28(10) GDPR.

64 Art 2(1) GDPR.

65 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (1995) OJ L281/31.

66 Mahieu, van Hoboken and Hadi Asghari (n 52) para 7.

67 WP29 1/2010 (n 53)10 ff.

68 Ibid 1.

69 Art 28(10) GDPR.

70 WP29 1/2010 (n 53) 20ff.

71 Kuner (n 38) 71; Patrick Van Eeck and Maarten Truyens, 'Privacy and Social networks' (2010) 26 Computer Law & Security Review 535; W Kuan Hon, Christopher Millard and Ian Walden, 'Who Is Responsible for "Personal Data" in Cloud Computing? The Cloud of Unknowing, Part 2' (2011) 2(1) International Data Privacy Law 3; Brandon Van Alsenoy, 'Regulating Data Protection: The Allocation of Responsibility and Risk Among Actors Involved in Personal Data Processing' (2012) 28 Computer Law & Security Review 35; Mahieu, van Hoboken and Hadi Asghari (n 52) para 1.

72 Reed, Kennedy and Silva (n 52) 1; Singh and others (n 52); Jennifer Cobbe, Chris Norval and Jatinder Singh, 'What Lies Beneath: Transparency in Online Service Supply Chains' (2020) 5 Journal of Cyber Policy 1.

73 See for discussion with regard to the responsibilities and roles of cloud service providers WP29 'Opinion 5/2012 on Cloud Computing' (WP 196, 1 July 2012); Hon, Millard and Walden (n 71) 7; Christopher Millard, *Cloud Computing Law* (OUP, Oxford 2013) 89.

services. Traditionally, cloud service providers were perceived as data processors.[74] They may themselves prefer to be qualified as data processors, as being a data processor involves significantly fewer obligations compared to a controller.[75] However, some cloud service providers might, perhaps, given their factual influence over means that were traditionally determined by controllers, be perceived as (joint) controllers. Some providers, through their business models, standard form contracts which often leave little room for negotiation, and technical expertise, can largely shape and influence a controller's data management operations (which, in turn, may also feed into the cloud service's business interests).[76] While such factual influence and power over the means might urge for a shift of roles, data controllers outsourcing data storage to a cloud service provider are—as long as these cloud service providers act within the boundaries of their contracts with controllers—still regarded controllers.[77] This might, however, no longer capture the diversity in business models and the ways in which cloud service providers can shape controller processing operations.[78] As we explore, determining the roles and the corresponding responsibilities across PDS platforms, users, and app developers in PDS ecosystems do not appear straightforward either.

Joint controllers

In recent case law, the Court of Justice of the European Union (CJEU) has considered the assignment of *joint controllership* in complex ecosystems involving multiple actors (in particular, where one is acting as a platform and another is using that platform for their processing). In *Wirtschaftsakademie*, the CJEU held that using another controller's platform for processing does not exempt a controller from responsibilities under the GDPR, but joint responsibility also does not mean that all controllers are *equally* responsible for all of the processing.[79] The level of responsibility will depend on the degree of involvement by the entity in question. That is to say, the assignment of responsibilities for processing is not 'all or nothing' in these contexts—rather, it is more nuanced.

In *FashionID*, the CJEU further clarified this point.[80] This case again recognized that, in these complex environments, processing typically involves various operations performed on personal data by controllers who are involved in different degrees at different stages of the processing.[81] The assignment of controllership for each of those stages would, according to *FashionID*, depend on identifying which entity is responsible for determining the means and purposes of processing in relation to a particular operation (or set of operations) performed at each stage of processing. Similarly, an entity cannot be a joint controller for (sets of) operations performed previously or subsequently to these stages of processing if that entity does not determine any of the purposes and means at these stages.

We now explore the allocations of roles in some existing and developing PDS's ToS in the light of the case law in section 'Role and responsibility allocations as described in PDS platform's ToS', while in section 'Identifying roles and (joint) responsibilities of platforms in a scenario', platform responsibilities will be explored a scenario consisting of three hypothetical platforms.

Role and responsibility allocations as described in PDS platform's ToS

Some PDS platforms inform users and app developers through their ToS about their claimed (or preferred) assignments of roles of (joint) controllers or processors. Further, some platforms that are still underdevelopment have published relevant information on websites or in literature, from which some directions of how these platforms envision distributions of roles and responsibilities can be inferred. We now explore three PDS platforms to give insight into the directions in which PDS platforms are heading. Some common characteristics will be compared with the assignment of roles and responsibilities as per the GDPR and relevant case law.

Allocations in platform's ToS

We first examine the ToS of various PDSs to determine how they envisage the assignment of responsibilities

74 Hon, Millard and Walden (n 71); Millard, ibid 193.

75 See Table 1.

76 Peter Blume, 'Controller and Processor: Is There a Risk of Confusion?' (2013) 3 International Data Privacy Law 142; Edwards and others (n 52) 2; Mahieu, van Hoboken and Hadi Asghari (n 52) para 8.

77 Other relevant factors that may help to identify whether an actor qualifies as processor include (i) a level of prior instructions given by the data controller; (ii) a controller's monitoring of execution of the service; 'constant and careful supervision' to ensure the processor's 'thorough compliance' with instructions and contractual terms indicates the controller is in 'full and sole control'; (iii) the visibility or image (impression) given by controller to data subjects, and expectations of data subjects based on that, and (iv) expertise of the parties—sometimes the provider's 'traditional role and professional expertise' may make it a controller, see WP29 1/2010 (n 53) 33.

78 Edwards and others (n 52).

79 Case C-210/16 of 5 June 2018 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH* [2018] ECJ EU:C:2018:388.

80 Case C-40/17 of 29 July 2019 *FashionID GmbH & Co KG v Verbraucherzentrale NRW eV* (2018) ECLI:EU:C:2018:1039.

81 Ibid, para 72.

under the GDPR. Note that this is not an exhaustive, but rather an exploratory analysis; as a nascent technology, we selected examples of PDS platform's ToS that were readily available at the time or research in order to indicate various considerations, noting that there are other PDSs underdevelopment:

1. Dataswift/HAT: The platform is operated by Dataswift, a commercial organization based in the UK.[82] The Dataswift/HAT platform enables the creation of devices,[83] and users install apps locally on their PDS. Apps appear in the platform's app store after having been rated by Dataswift/HAT.[84] Their rating system tells users where the app stores its data, what conditions are imposed on services delivered by the app, and informs users about the type of data a user will receive if they submit a subject access request to the app developer.[85] The platform operates—given users agree—mediated data transfers. Users are provided with technical mechanisms with which they can permit app developers to do computations over their data, or through which they can permit data transfers. The PDS is stored in a personal container in the Amazon cloud. The PDS is only accessible by the user—and not (directly) by Dataswift/HAT.[86] Security practices of Dataswift/HAT include end-user encryption.[87]
   Users can permit app developers' access to raw user data in return for specific services. The primary purpose of this platform is to create a marketplace, the device providing a mechanism for users to transact their personal data and 'to build them into the personal data value chain'.[88] The device offers users mechanisms to transact and transfer data. The platform's approach has its roots in the so-called 'database rights' as enshrined in EU Directive 96/9/EC on the legal protection of databases, which prompts to harmonize the treatment of databases under copyright law and the *sui generis* right for the creators of databases (which themselves do not qualify for copyright).[89]
   The Dataswift/HAT's ToS describes the platform as a 'controller' and a 'processor' with regard to personal data that is necessary to create the PDS device.[90] Other conditions specifying when, how, and for whose data the platform's roles as controller and processor are envisaged are not defined in its ToS; on the contrary, the platform stipulates that it is neither a data controller nor a data processor under the GDPR for personal data of their users.[91] Perhaps this might be based on the platform's presumption that it has no access to user data in the PDS and that it, therefore, bears no legal responsibility for that data.[92] Dataswift/HAT further stipulates that '[users] are the controller and processor of the data within their [PDS]. They alone have access to the data within the [PDS] and they call on platform services to move, process, transform, exchange or generate new data into the [PDS]'.[93] Further, Dataswift/HAT's ToS explicitly assign responsibility to users for specific aspects of their devices—to protect their PDS against unauthorized access, and for keeping the data in the device protected and

82 Dataswift/HAT comprises various entities (a lab, a regulator, an entity responsible for the technical architecture). It will generate income from, among other, annual fees from each fielded PDS in active use which is charged to the companies that issue the PDSs to users.

83 Dataswift/HAT ToS, s 1.3.1 <https://cdn.dataswift.io/legal/hat-owner-terms-of-service.pdf> accessed 21 August 2020. Users can buy or get their PDSs from Dataswift/HAT, or from Dataswift/HAT trusted companies (eg Fitbit, Tesco, or Facebook) with which a user already has a relationship; see <https://www.dataswift.io/concepts/personal-data-account> accessed 21 August 2020.

84 See for the risk rating scheme <https://www.hatcommunity.org/hat-dex-rating> accessed 21 August 2020.

85 Ibid. Regarding this aspect of Dataswift/HAT risk rating ('the completeness of data contributed by the app') refers to the situation in which a PDS user would submit a subject access request to the app developer. If a subject access request to the app yields the same data as that which is brought into the PDS, the app receives a high rating, particularly if the app also contributes other data such as descriptions and images to make the data more understandable, and if the app contributes data generated by the person on the app back into the PDS.

86 HATDeX Privacy policy 2.0 <https://static1.squarespace.com/static/5a71ebc8b1ffb68777ca627a/t/5c26550670a6ad08a55214ba/1546016006812/HATDeX+privacy+policy+2.1.pdf> accessed 21 August 2020.

87 'HAT Microservers (a Microserver is a PDS) create data debit contracts when granting rights of HAT data to others and data in transit is SSL-encrypted from end to end, in a similar manner to emails', Irene Ng, 'Can You Own Your Personal Data? The HAT (Hub-of-All-Things) Data Ownership Model' (2018) Warwick Manufacturing Group Service Systems Research Group, Working Paper Series (02/18) 9.

88 Crabtree and others (n 5) 52.

89 Dataswift/HAT PSOP under B which explains the ownership rights model (for Dataswift/HAT PSOP, see n 30); for data bases rights, see EU Directive 96/9/EC on the legal protection of databases (27 March 1996) OJ L77, 20. Art 1(2) defines a database as 'a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means'. Computer programmes used to create the database is not included (Recital 23). Copyright protection of software is governed by Directive 91/250/EEC. The platform's approach to user PDSs largely resembles to the databases under the said Directive, which 'by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation' are protected by copyright as collections.

90 Dataswift/HAT ToS (n 83) s 2.4.2.

91 Ibid, s 2.4.2.

92 HATDeX Privacy policy 2.0 (n 86).

93 Also here, conditions specifying towards whom, at what stages, and for what processing users act as a controller or as a processor towards other actors is not fully clarified; see 'Section D Platform and Platform Functions', Dataswift/Hat PSOP (n 30).

accurate through the platform's tools, apps, and other instruments 'when necessary'.[94] These assignments apply from the perspective 'data protection law' which we understand to mean that Dataswift/HAT's assignments are intended to accord with that of the GDPR.[95]

2. Databox: The Databox platform is a UK-based initiative and still under development.[96] The platform designers seek to offer users a physical device, primarily for use in a domestic Internet of Things context, allowing mediated data transfer and local app installation by users.[97] The platform designers envision undertaking risk assessments of apps as part of the platform's Software Development Kit (SDK). That is, the SDK envisages to support app developers not only with the technical tooling to support development but also to assist data protection aspects. This includes certain mechanisms through which the platform seeks to sensitize app developers to potential legal, social, and technical risks that may accompany personal data processing.[98] The designers also intend to undertake a risk assessment, presenting an app's risk rating in the envisaged app store.[99]
   Databox's designers foresee the platform's own role towards users as a 'processor' under the GDPR.[100] The designers assign the role of 'controller' to 'external machines exploited by parties', which we interpret such that Databox claims that the app developers would be the 'controllers' whenever data are transferred, but not in the situation where the data remain in the device. The Databox literature does not clarify who is envisaged to be the controller for user data at rest in their PDS.
3. Mydex: Mydex is a UK-based platform.[101] The platform's purpose is to 'pursue its social mission' to enable users 'to realise the value of their data'.[102] It promotes this mission through commercial activity, as it charges app developers for the delivery of services by the platform (ie for making use of the platform's services, including support fees).[103] App developers pay a transaction fee to PDS users each time they access user data in a user's PDS (eg to process personal data or to deliver services),[104] and the platform collects a percentage of each data transaction.[105]
   The platform presents various relevant documents among which its ToS for users and app developers.[106] The platform provides users with cloud-based devices on which apps can be installed. The data in the device are encrypted by the user before entering the device, such that the platform cannot 'see' the data (in 'a legible form').[107] Users share their data after giving consent to the manifests described by app developers, which must be compliant with the platform's 'Standard Data Sharing Agreement'.[108] Users can, according to the platform's PDS Guide for users, perform on-device computations.[109] Users themselves populate their PDSs or have them populated by organizations of their choice; no specific data capture

94 Dataswift/HAT ToS (n 83): '2.4.3. Responsibilities: It is your responsibility to: 2.4.3.1 protect against unauthorised access to your HAT Microserver; 2.4.3.2. not share, and otherwise ensure no-one else uses your HAT Microserver; 2.4.3.3. not use any account data or account of any other Owner or person than yourself without permission of the Owner or person holding the respective account; 2.4.3.4. keep your data in the HAT Microserver database useful and accurate through available HAT tools, apps and plugs and updating the said tools, apps and plugs when necessary; 2.4.3.5. keep your password or other access information secret (your password and login details are personal to you and should not be given to anyone else or used to provide shared access), and 2.4.3.6. generally, maintain good internet security practices.'

95 Dataswift/HAT ToS, ibid, s 2.4.2.

96 See <https://github.com/me-box/databox/blob/master/documents/what-is-databox.md> accessed 21 August 2020. The platform has not (yet) developed ToS, a privacy policy, or other documents. The findings in this paragraph therefore largely rest on findings in literature (mainly Crabtree and others (n 5) and Urquhart, Lodge and Crabtree (n 18)).

97 Crabtree and others, ibid 53.

98 Ibid 46, 53. The SDK provides developers with supporting manifest construction to assist in meeting the information requirements of the GDPR, enhanced granular choice over data collection, and providing for runtime accountability in surfacing how data flow through an app and how some action or decision is arrived at. The SDK also exploits a risk-based framework to motivate development of GDPR compliant apps providing access to data taken off-the-box.

99 Crabtree and others (n 5) 47.

100 See ibid 43: 'data processors' (external machines exploited by parties, or 'data controllers' in GDPR terminology, who wish to make use of the user's data in some way).

101 Mydex 'Terms for Connections' <https://dev.mydex.org/terms-for-connections.html> accessed 21 August 2020.

102 'Mydex Charter' <https://dev.mydex.org/mydex-charter.html> accessed 21 August 2020.

103 Ibid.

104 Mydex acts as a commercial party towards app developers; these pay a low cost initial fee per service every time they access the Mydex Platform, and then they pay for each access to a user PDS (whether for personal data or for other services), see Mydex ToS for app developers, 'Terms for Connections' (n 101) 7.

105 Mydex Tariff table, see <https://dev.mydex.org/tariff-table.html> accessed 21 august 2020.

106 Most relevant are Mydex ToS for *app developers* 'Terms for Connections' (n 101) 7; ToS for *users* 'Terms for Members' <https://dev.mydex.org/terms-for-members.html> accessed 21 August 2020; Mydex ToS *governing relations between app developer and user* 'Data Sharing Agreement' <https://dev.mydex.org/data-sharing-agreement.html> accessed 21 August 2020; and Mydex privacy policy <https://dev.mydex.org/privacy-policy.html> accessed 21 August 2020.

107 'The Mydex Members PDS Guide' 1 <https://pds.mydex.org/pds-members-guide> accessed 21 August 2020; see also Mydex Member Questions which stipulates: 'Mydex CIC does not have access to this data; Mydex CIC simply provides the vessel in which to hold the data, the pipelines to share it (in a secure, auditable way), and the taps for individuals to turn this data sharing on and off as they please' <https://community.mydex.org/questions/members> accessed 21 August 2020.

108 See Mydex 'Data Sharing Agreement' (n 106); Mydex Member Questions, ibid.

109 Mydex Member Questions, ibid.

mechanisms are provided by the platform itself.

The platform purports to put 'obligations such as transparency regarding use, loss, misuse or breach of data, obligations of audit and verification on all parties connecting into the platform'.[110] It perceives itself as owner and 'controller' of the data architecture and the management process of IT. Mydex' ToS holds app developers responsible for personal data of users that are processed in Mydex' PDSs.[111] The ToS do not specify who is regarded as the data processor. It is unclear from the ToS whether its terminology is intended to align with the terminology of the GDPR.[112]

Considerations on platform allocations of 'controller' and 'data subject'

While operations might differ, platforms' ToS usually attempt to assign the roles and responsibilities of controller, processor, and data subject. As we have conducted a non-exhaustive scan of some available ToS, we restrict ourselves here to highlighting some problematic areas in these ToS in relation to the designations of roles and responsibilities under the GDPR. In the next section we develop a scenario with three imaginary platforms, whereby we aim at elaborating an approach on how responsibilities should be assigned under the GDPR.

1. Having no access to personal data is not a precondition to become a controller: Mydex and Dataswift/HAT stipulate that they have no access to user data, nor can they see that data.[113] They claim that this excludes them from being a controller for user data. The CJEU has, however, determined that having access to or being able to see user data is not a precondition to be assigned the role of controller.[114]
2. There is no consistent approach in designations of roles and responsibilities. Interestingly, none of the platforms assign themselves as a controller for user data (although the Dataswift/HAT's ToS mention that the platform is responsible for the processing of (only) that personal data which is necessary to *create* a PDS). Dataswift/HAT assigns users as data controllers for their own personal data. The Databox platform envisages app developers as controllers.[115] Mydex also holds app developers responsible for the processing of user data.[116]

   Dataswift/HAT assigns PDS users (also) as data processors.[117] Dataswift/HAT does not explain how the roles of controller and processor are performed by users. Some clarification would, however, be useful, as there is some legal uncertainty given these roles cannot be performed simultaneously. Databox assigns the platform itself as a processor for the processing of user data. Mydex' ToS do not specify who is data processor.
3. Allocating responsibilities for platforms in PDS ecosystems is challenging. Given the case law discussed, a platform's decisions over whether (or not) to use a specific technology, how to assemble or integrate various components in that technology, or to involve particular actors, are highly relevant for determining the allocation of responsibilities.[118] Many of those decisions will usually be taken by the platforms as they are designing and developing their systems. Nevertheless, defining the responsibilities and allocating those at these initial stages involved can be challenging. This will be even more so as future uses of that specific technology might evolve. In a chain of actors and activities, platforms themselves might not be in a position to oversee all possible future purposes, contexts, and uses of PDSs—and hence of (future) actors involved.[119] Such unforeseeable developments might make it difficult for platforms to determine, a priori, an allocation of roles and responsibilities that is sustainable and GDPR compliant.

   Moreover, technologies applying in distributed data ecosystems may pose specific challenges, as data processing involves more than a single entity (the earlier mentioned 'architect') that would traditionally be easily identifiable as *the* controller. Rather, these decentralized systems are shaped by the

110 Mydex Charter (n 102).

111 '[App developers] are fully responsible for the use of all Data that is provided to [them], whether that is [an app developer's] own use or use by another person that gains access to the Data through [the app developer] or [the app developer's] systems', Mydex 'Terms for Connections (n 101) 7.

112 Mydex Charter (n 102).

113 See for Mydex their Mydex Members PDS Guide (n 107); see for Dataswift/HAT ToS (n 83).

114 In *Wirtschaftsakademie*, the CJEU has made clear that this is not an argument to escape controller responsibilities; this point was repeated in Case C-25/17 of 10 July 2018 *Tietosuojavaltuutettu v Jehovan Todistajat* ('*Jenovah's Witnesses*') (2018) ECLI:EU:C:2018:551, para 75.

115 Crabtree and others (n 5) 43.

116 See n 106.

117 See section '*Allocations in platform's ToS*'; Dataswift/HAT PSOP (n 30): '[users] are the controller and processor of the data within their [PDS]. They alone have access to the data within the [PDS] and they call on platform services to move, process, transform, exchange or generate new data into the [PDS].'

118 Case C-131/12 of 13 May 2014, *Google Spain v Agencia Espanola de Protección de Datos* (2014) ECLI:EU:C:2014:317, paras 32–41; *Wirtschaftsakademie* (n 79) paras 35–40; *FashionID* (n 80) para 75.

119 Jatinder Singh and others (n 52) 54ff.

collaboration of a multitude of parties, including the key actors considered here—platforms, users, and developers—whose interactions with a particular PDS may differ depending on the specifics of each user device.[120] In that way, PDSs seem to be at odds with the GDPR's initial approach which considered data controllers as single entities who oversee the data processing in a system.

While neither the facts of *Wirtschaftsakademie* nor of *FashionID* are exact analogues for the position found in PDS platforms, the general rule elucidated by the CJEU—that the assignment of controllership at various stages of a processing chain is itself a nuanced, dynamic process that involves identifying who is responsible for determining the means and purposes of the specific operations performed at each of those stages—can be applied across to PDSs.[121] In this context, therefore, it appears that often the likely position in law is that platforms and app developers are joint controllers for processing performed by or on the PDS in the context of providing a particular service to the user. This will be further explored in the scenario in the next section.

In line with *FashionID*, it appears that platforms are controllers for the initial transfer of any personal data out of the PDS ecosystem (eg the results of analytics, or indeed the raw user data); nonetheless, platforms cannot be held responsible for any processing subsequent to that transfer.[122] Note that the specifics of a particular platform or, indeed, of a particular app seem highly relevant to the legal considerations in practice.

4. Strengthening user control does not automatically imply user controllership: PDSs represent federated data processing ecosystems, seeking to give users the means for determining their own purposes for data processing. This might, in the perception of some platforms, challenge the GDPR's approach of perceiving users as mere data subjects, according to which users are only passively involved in processing operations.[123] PDS approaches seeking to bring granular controls to users might indicate new distributions of responsibilities. However, where platforms seek to appoint users as controllers, this might only apply if user activity is not covered by the GDPR's household exemption.

Identifying roles and (joint) responsibilities of platforms in a scenario

Given personal data are processed in PDSs, controllers will be involved. App developers will generally be controllers for the personal data of PDS users, given they process that data for (often) commercial purposes. Platforms will likely act as controllers for that personal data as far as they process data for their own purposes, which includes for usage statistics, or the platform's own direct commercial purposes as well.

We now use a scenario to investigate whether and how controllership may apply to a platform, using a scenario consisting of three hypothetical, yet indicative PDS platforms. Given the limited selection of existing 'real-world' PDS platforms, these exemplar platforms (with strategically selected attributes) were devised to allow a fuller comparative analysis of some key considerations impacting the allocation of roles and responsibilities. PDSs is a nascent technology, but one gaining attention; elaborating the implications of key design decisions platforms might take which impact role and responsibility assignments is particularly relevant, particularly now that few platforms that do exist remain under continued development, and new platforms look likely to appear in future.

Table 2 details the three platforms: *PDS-Cloud*, *PDS-Physical*, and *PDS-Source*. All platforms have the same broad aim, seeking to strengthen user control over third-party (app developer) data processing. PDS-Cloud offers cloud-based user devices, while platform PDS-Physical provides its users with physical devices.[124] Platform PDS-Source does not provide devices to users; instead, users compile and operate the platform's open-access software to produce their own devices, which run on the user's own technical infrastructure. Platforms PDS-Cloud and PDS-Physical envision to monetize some aspects of data processing in user PDSs. PDS-Source does not entail any direct monetization. The details concerning specific activities are elaborated below (Table 2).

120 See in the context of Blockchain Michèle Finck, 'Smart Contracts as a Form of Solely Automated Processing under the GDPR' Max Planck Institute for Innovation and Competition Research Papers 19-01 <ssrn.com/abstract=3311370> accessed 15 March 2020; Finck draws attention to the French Data Protection Authority who recently suggested that smart contract developers (Blockchain represents another example of distributed data processing) can also be deemed data controllers; CNIL, 'Premiers éléments d'analyse de la CNIL' (September 2018) 2 <cnil.fr/sites/default/files/atoms/files/la_blockchain.pdf> accessed 21 August 2020.

121 *FashionID* (n 80) para 72.

122 Ibid, paras 99–101.

123 This can be derived from the fact that some platforms (eg HAT) assign users as controllers with regard to their personal data.

124 See section 'Introduction' about physical and cloud-based devices.

Table 2. An overview of the hypothetical PDS platforms

Platform activity	PDS-Cloud	PDS-Physical	PDS-Source
Platform business models			
1. Platform charges app developers to list their apps in the platform's 'app store'	Yes	Yes	N/A
2. Platform receives revenue from the app developer per data transaction (per transfer of raw or computed from device to app developer)	Yes	No	No
3. Platform receives revenue from app developers whenever app developers charge users for downloading their app via the platform's app store	Yes	Yes	No
4. Platform receives revenue from app developers whenever app developer charges users for in-app purchases	Yes	No	No
Technical architecture			
5. PDS device	Yes (cloud-based)	Yes (physical device)	No (users obtain open-source software which offers them functionality to operate their own PDS)
6. PDS platform offers a technical architecture and provides mechanisms for users to consent and specify preferences for application, installation, and operation of apps	Yes	Yes	Yes
7. Platform offers logging, readable audits, and real-time visualizations regarding the on-device behaviour of apps	Yes	Yes	Yes
8. Platform only lists apps in its app store if they are designed and developed with the platform's technical tools and in accordance with particular standards	Yes—all apps must be developed with the platform's SDK	Yes—all apps must be developed with the platform's SDK	No—there is no app store
9. Platform only allows installation of apps through its app store	Yes—apps must be installed via the platform app store	No—any compatible apps can be installed in the user's PDS	No—any compatible apps can be installed in the user's PDS
ToS and contractual arrangements			
10. ToS	Yes	Yes	Yes
11. Contracts between app developers and platform	Yes; business-specific contracts	No; generic ToS apply	No; generic ToS apply

In the scenario, we generally consider app developers as controllers, given they process personal data of users for defined (and often commercial) purposes. App developers achieve their purposes by processing user data by way of conducting computation, and/or transferring raw data or the results of computation. PDS users are considered data subjects. Our focus in the scenario is on the role of platforms and the nature of their legal responsibilities.

In this subsection, we analyse (i) the (joint) purposes PDS-Cloud, PDS-Physical, and PDS-Source pursue by the processing of personal data of PDS users, and (ii)

through what (joint) means they seek to achieve these purposes. As far as a platform would indeed qualify as a joint controller, we then (iii) consider what the *scope* of the platform's controllership with relation to the app developer. Table 2 presents some characteristics of the imaginative platforms and activities they undertake, potentially relevant to assignments of responsibility.

We now extract the relevant rows from Table 2 and discuss what they mean for the assignment of controllership in terms of the purposes of processing, the means of processing, and the scope of platform responsibilities. To assist the presentation, Tables 3–6 below reflect the rows extracted from Table 2.

Determining the purposes

Determining a purpose for the processing of personal data triggers an actor's qualification as a controller.[125] However, it is not always immediately clear whether an actor processes personal data for a purpose. The CJEU clarified such a situation in *Jehovah's Witnesses*, where the religious Community contested that the collection of personal data by its preachers meant that both the preachers *and* the Community were (joint) data controllers. The Community argued that it was the preachers who paid door-to-door visits (and thereby processed the data), and not the Community itself; it refused to be assigned (joint) controller.

The CJEU, however, held that an entity can still play a part in determining the purposes of processing where they 'exert influence' over that processing for their own purposes, even if that entity has no access to the data.[126] This might be the case where they are coordinating, organizing, and encouraging the processing of personal data in such a way as to benefit them by helping them to achieve one of their own purposes. [127] This can arise even where it is not processing the personal data itself, or where no written guidelines or instructions in relation to the data processing were given.[128]

PDS platforms typically pursue some commercial purpose.[129] In terms of the business models of these three examples, PDS-Cloud and PDS-Physical monetize their services by charging app developers a fee to have their apps included on the platform's app store (point 1 of Table 3).[130] The other commercial activities—for PDS-Cloud, point 2 of Table 3, point 3 of Table 3,[131] and point 4 of Table 3; and for PDS-Physical only point 1 of Table 3—they undertake demonstrates their commercial incentive to encourage the processing of personal data of PDS users. More data transactions, more app downloads, and more in-app purchases generate more revenue for PDS-Cloud and PDS-Physical. Both of these platforms exert influence over the processing for their own commercial purposes; they derive benefit from the data processing by app developers and users using their platform. That processing assists the platforms in achieving their purposes.

PDS-Source does not pursue a direct commercial purpose for the processing of data. Instead, they offer users its open-source software, providing them with a range of tools that the user can assemble (or even extend) as the user desires. Users create their own device, and then work to manage their data, run diagnostics/audits, etc, by leveraging the PDS's functionalities derived from the codebase. Although PDS-Source provides users with their open-source software, it is unlikely that this platform 'exerts influence' over the processing for their own purposes. The platform might still exert some influence over the data processing in the user-built PDS by determining the nature of the tools available to users and thus to some extent the purposes for which they can be used. But that does not necessarily mean that they determine those purposes for the benefit of the platform. Rather, the platform provides tools to users to make use of as they see fit. We, therefore, conclude that PDS-Source probably does not act as a controller for the processing of personal data of the user.

Given that PDS-Cloud and PDS-Physical in the scenario pursue defined purposes for which they process user data, we now explore whether they determine *joint* purposes with the app developer, and where that joint controllership, in the course of the data processing in the PDS, might arise.

Data processing in PDSs consists of discernible 'stages' (such as the capturing of data by the device, computation that occurs on device, and the transfer of data out of device). The data processing in each stage can have more than one purpose.[132] *App developers* might for instance determine some purposes for data

125 See section 'Setting the scene for responsibilities in PDSs: developments under the GDPR'; WP 1/2010 (n 53) 14; *FashionID* (n 80) paras 66 and 68.

126 *Jehovah's Witnesses* (n 114) paras 69 and 75.

127 Ibid, paras 71 and 75.

128 Ibid, para 68.

129 See section 'Purported benefits of PDSs'.

130 See for instance Mydex Tariff table (n 105), according to which app developers have to pay one-time connection fees per system (which could include entrance to the platform's app store, if there is one), as well as annual support fees to ensure sustainability of the platform.

131 An example can be found in the business model of CitizenMe. App developers can request access to user data for which a fixed price per user is determined. The same users receive a fixed amount per request posed from the business. See <https://www.citizenme.com/public/wp/business/pricing> accessed 21 August 2020.

132 For 'stage', see *FashionID* (n 80) para 72.

Table 3. Overview of platform purposes

Platform activity	PDS-Cloud	PDS-Physical	PDS-Source
Platform business models			
1. Platform charges app developers to list their apps in the platform's 'app store'	Yes	Yes	N/A
2. Platform receives revenue from the app developer per data transaction (per transfer of raw or computed from device to app developer)	Yes	No	No
3. Platform receives revenue from app developers whenever app developers charge users for downloading their app via the platform's app store	Yes	Yes	No
4. Platform receives revenue from app developers whenever app developer charges users for in-app purchases	Yes	No	No

processing where data transfers are involved. The data transfer might happen for an *app developer*'s commercial gain, which can also include the development and improvement of the app and the developer's internal processes. At the same time, the *platform* might also have an interest in data transfers, its purpose being to gain revenue (for instance, by way of a 'data transfer charge', which is done, eg by the Mydex platform[133]) and generally to help grow the platform.

In other words, both the platform and the app developer benefit from the same transfer of data. Following *FashionID*, the platform will act in such situations as a joint controller with the app developer for that specific stage of processing of user data and thus towards the user. The platform does not need to determine the exact same purposes as the app developer, their own purpose for the same stage of processing will suffice—which in the example given is their joint commercial purpose. Having discussed the stage of data transfers here as an example, it is important to note that joint controllership might occur at other stages of the data processing in PDSs as well.

It appears from current case law that imprecisely defined purposes, such as those 'commercial' or 'economic', can already qualify as joint purposes.[134] While the GDPR requires controllers to define *specific* purposes,[135] it is not clear how this 'specificity' requirement relates to the CJEU's high-level approach of 'joint' purposes. Considering the purposes of PDS-Cloud and PDS-Physical, it appears that they would, given their commercial activities, qualify as joint controllers with the app developers. In situations where no joint purpose with an app developer can be identified, a platform would, in line with *Google Spain*, rather qualify as an independent controller for the processing of user data.[136] If no platform could be identified as a controller for the processing of user data, which applies to PDS-Source, the app developer will qualify as a sole controller.

Determining the means

PDS platforms may define the means of data processing for their own processing purposes. This includes determining the specific technical functionality that occurs, while they might also seek to influence the method according to which apps can process user data in the platform's PDS ecosystem. Beyond technical specifications, a platform's ToS and business-specific contracts may further assist the platform in achieving their purposes. We now explore how these aspects—be they technical or legal—operate to determine the means through which the platforms seek to achieve their purposes.

1. Technical functionality as a means: Generally, a platform's technical functionality operates to support the capturing and storage of user data in the device and the management of on-device computation and transfer, based on (machine-readable) policies (point 6 of Table 4). For all three platforms, every interaction with user data is logged and audited, so

133 Mydex Tariff table (n 105).

134 Neither *Wirtschaftsakademie* nor *FashionID* indicated that the CJEU requires the jointly determined purposes to be determined at granular levels, see *Wirtschaftsakademie* (n 79) paras 38 and 39 and *FashionID* (n 80) para 80, respectively; in *Wirtschaftsakademie*, the joint purposes of the Wirtschaftsakademie and Facebook were processing of the personal data of visitors for statistical purposes, while the joint purposes of FashionID and Facebook were processing operations performed in the economic (commercial) interests of both FashionID and Facebook.

135 Art 5(1)(b) GDPR.

136 *Google Spain* (n 118) para 40.

Table 4. Overview of platform technical architecture

Platform activity	PDS-Cloud	PDS-Physical	PDS-Source
Technical architecture			
6. PDS platform offers a technical architecture and provides mechanisms for users to consent and specify preferences for application, installation, and operation of apps	Yes	Yes	Yes
7. Platform offers logging, readable audits and real-time visualizations	Yes	Yes	Yes

that users can monitor these interactions and decide whether they want to (dis)engage with an app (point 7 of Table 4). Through the technical functionality provided by their respective platforms, PDS-Cloud and PDS-Physical determine how the processing of personal data of users by app developers will evolve.[137] PDS-Source offers users the source code for a range of components with which users can compile their own PDS. In providing the source, users can extend and customize the functionality provided.

1. In both designing the technology in a particular way, and in encouraging a particular use, PDS-Cloud and PDS-Physical seek to attract and encourage individuals to become PDS users, more app developers to provide services, and hence more data processing in their PDSs. This all serves as means for generating more revenue for the platforms. More specifically, the technical architecture of PDS-Cloud is designed to monetize data transactions between users and app developers as per point 2 of Table 3, to monetize app downloads by users as per point 3 of Table 3, and to monetize in-app purchases under point 4 of Table 3. PDS-Physical's technical architecture will be designed to monetize app downloads as per point 3 of Table 3. Given that PDS-Cloud and PDS-Physical pursue commercial purposes, their systems are designed to facilitate the platform to collect the revenues from the data processing by the app developer during the different stages of data processing.
   PDS-Source also offers the technology to provide particular PDS functionality, including defining and enforcing data control mechanisms, logging, readable audits, and real-time visualizations. However, PDS-Source does not, like the other platforms, determine how the standards in these components are set or used. It is up to the user to create their PDS, which could entail a user customizing their PDS (including by extending the code base) such that it operates in a preferred manner.
2. Platform influence over app development as a means: By instructing app developers as to how apps should be designed to be allowed in the platform's app store, platforms will exert influence over the data processing as envisaged by the app developer. Table 5 illustrates whether PDS-Cloud, PDS-Physical, and PDS-Source envisage such influence.
   PDS-Cloud only lists apps in their app store if they are designed and developed in accordance with the platform's technical standard, including being developed with the platform's SDK (point 8 of Table 5). All their apps are installed in and distributed from the platform's app store.
   PDS-Physical also operates an SDK, thereby also determining the technical design of apps. However, PDS-Physical also allows apps to be 'sideloaded', ie the app can be developed and installed from outside of the platform's standard development distribution processes (ie their app store). Side-loaded apps will still operate within the limits of PDS-Physical's technical architecture, in terms of application, access to data or data flow, but they might not have been vetted or risk assessed by the platform and would give fewer guarantees about conforming to the platform's standards. PDS-Physical, therefore, proposes a hybrid approach.
   PDS-Source does not manage nor maintain any particular app store, nor does it mandate any particular platform tools for app development.
3. ToS and contracts as a means: All platforms set legal conditions in ToS according to which app developers can process user data on PDSs, including PDS-

137 See section 'PDS Architectures: technical aspects, devices and the role of apps'.

Table 5. Platform influence over app development

Platform activity	PDS-Cloud	PDS-Physical	PDS-Source
Technical architecture			
8. Apps must be designed and developed with the platform's technical tools and in accordance with particular standards	Yes—all apps must be developed with the platform's SDK	Yes—all apps must be developed with the platform's SDK	No—there is no app store
9. Platform only allows installation of apps through its app store	Yes—apps must be installed via the platform app store	No—any compatible apps can be installed in the user's PDS	No—any compatible apps can be installed in the user's PDS

Table 6. PDS ToS and contractual arrangements

Platform activity	PDS-Cloud	PDS-Physical	PDS-Source
ToS and contractual arrangements			
10. ToS	Yes	Yes	Yes
11. Contracts between app developers and platform	Yes; business-specific contracts	No; generic ToS apply	No; generic ToS apply

Source (point 9 of Table 6). PDS-Cloud's and PDS-Physical's ToS will, given the platforms' involvement in constraining apps, work to regulate app behaviour, describe how app developers can use the platform, and describe how the platform envisages responsibilities and liabilities. Their ToS stipulate that app developers should, in order to be able to process user data by way of the platform, have appropriate technical and organizational procedures in place to safeguard users against unlawful processing and against accidental loss or destruction of data. As far as PDS-Cloud and PDS-Physical are acting as joint controllers, Article 26(2) GDPR requires them to draw up an agreement with other joint controllers in which they agree on how responsibilities are assigned among the platform and the other joint controllers. This agreement could form part of their ToS for app developers.

PDS-Source's ToS may contain conditions according to which the user can use PDS-Source[138]; however, their ToS will most likely not concern obligations applying to app developers, as it is for users to determine what obligations the app developer should meet.

PDS-Cloud concludes business-specific contracts with app developers (point 11 of Table 6); it contractually obliges app developers to design their apps with the platform's SDK. These legal means provide the conditions according to which developers can process user data and might, therefore, indicate a platform's role as controller.

To determine whether PDS-Cloud and PDS-Physical act as *joint* controllers with app developers, joint *means* of processing must also be identified. As regards PDS-Cloud, their revenue raised from transactions between app developers and users under point 2 of Table 3 likely entails a joint nature of means of PDS-Cloud and commercial app developers. The data transactions enable the app developer to generate user profiles from, for instance, data transfers of raw data or from results of computation. PDS-Cloud 'monetizes' the data transfers by collecting a small fee for every transaction between the app developer and PDS user. User data might, depending on the app developer's request, involve a small or a broad range of personal data. The situation of PDS-Physical entails joint means; this platform collects, as per point 3 in Table 3, a percentage from an app developer's fee which the app developer collects for each app download and installation by a PDS user. Both platform and app developer use these means for economic reasons (to monetize an app

138 And, indeed, much of the focus may be regarding software licencing.

download and installation by users). The type of user data involved might remain restricted to a user's payment data. As regards PDS-Source, the app developer will be the sole controller for data processing appearing between app developers and users of PDS-Source.

Given PDS-Cloud and PDS-Physical pursue commercial purposes, and that they determine the technical and legal means according to which app developers can process user data, we qualify PDS-Cloud and PDS-Physical as a controller in the sense of Article 4(7) GDPR. Though variability in their involvement in the management of user data (hands-on/hands-off approaches) exists, both are to be regarded as controllers. This is different for PDS-Source. This platform does not determine purposes, while its activity implies, at most, a remote influence over the data processing by app developers in user PDSs. If personal data are processed in this PDS, the app developer will be the sole controller for the user's personal data. As regards *joint* controllership, both PDS-Cloud and PDS-Physical appear to act as joint controllers with the app developer regarding the stage in which data are transferred (point 3 of Table 3).

Scope of platform responsibilities

Generally, PDS-Cloud and PDS-Physical act as joint controllers with app developers and, therefore, have specific GDPR obligations.[139] Joint controllers must draw up an arrangement which duly reflects the respective roles and relationships of the platform and the app developers towards PDS users.[140] This GDPR requirement applies in particular in relation to data subject rights and duties to duly inform users as required by Articles 13 and 14 GDPR. The 'essence' of those arrangements must be made available to PDS users and reflects the respective roles and relationships of the joint controllers towards PDS users.[141]

Article 26(2) GDPR is grounded in the approach that joint controllers can agree on how responsibilities should be allocated among them. However, platforms might not be in a position to predict how and for what purposes they might be used by other stakeholders, or be in a position to vet or to intervene when new joint controllers try to get involved in the data processing. This might complicate a platform's compliance with the GDPR, and it might leave PDS users in the dark over how responsibilities are distributed across joint controllers. Generally, data subjects can, independent from how responsibilities are spread (or not spread) across controllers, exercise their data subject access rights against each of the controllers.[142] Applying this to our use case, this means that data subjects can exercise their data subject rights against PDS-Cloud and PDS-Physical, regardless of how responsibilities are distributed among platform and app developers.

The *scope* of responsibilities has not yet been clarified. The CJEU has given some guidance which deserves further elaboration. In *Wirtschaftsakademie*, it held that 'joint responsibility' does not necessarily imply *equal* responsibility of the operators involved in data processing. On the contrary, and as elaborated earlier, these operators may be involved at different stages of that processing of personal data and to different degrees, so that the level of responsibility of each of them must be assessed with regard to all the relevant circumstances of the particular stage.[143] Different degrees of control might give rise to different degrees of responsibility and liability, potentially resulting in joint and several liability.[144]

A demarcation of the scope of responsibilities of joint controllers was given in *Fashion ID*, where the CJEU held that a platform's responsibility would likely be limited to the operation (or set of operations) involving the processing of personal data in respect of which a platform actually determines the purposes and the means, and *not* for any operations performed prior or subsequent to those particular operations in the processing chain.[145] This implies a certain 'temporality' of responsibilities in the CJEU's approach. PDS-Cloud can, in analogy with *FashionID*, for instance not be held jointly responsible for the data processing *beyond* data transfers to the app developer. For PDS-Physical, their joint responsibility would apply to the operations where user data are involved in the platform's collection of revenue from the app developer.[146]

Our analysis of how (joint) controllership and its responsibilities should be assigned to the platforms under prevailing law across PDS-Cloud, PDS-Physical, and PDS-Source has demonstrated that this can be a challenging exercise. However, platforms and others involved in data processing in PDS ecosystems should bear in mind that eventual harms occurring from new

139 See Table 1.

140 Art 26(1) GDPR.

141 Art 26(2) GDPR.

142 Art 26(3) GDPR.

143 *Wirtschaftsakademie* (n 79) para 43; WP29 had predicted that the participation of actors in the determination of purposes and means in the context of *joint* control may take different forms and does not need to be equally shared and that different degrees of joint control may exist; see *Wirtschaftsakademie* (n 79), Opinion of Advocate General Bot of 24 October 2017 ECLI:EU:C:2017:796, paras 75, 76 and 105.

144 Art 82 GDPR; WP29 Opinion 3/2010 (n 53) 24.

145 *FasionID* (n 80) para 85.

146 Ibid para 85.

technology (such as PDSs) can lead to *strict liability*: this applies if a responsible person is held liable, irrespective of any fault on that person's part.[147] 'Strict liability' has developed in response to the recognition of dangerous activities or states of affairs against the consequences of which the responsible person is required to indemnify.[148]

The role of users in the PDS ecosystem

Technical developments continue to blur the traditional border between 'public' and 'private' spheres,[149] with an increasing number of sensors and actuators instrumenting our homes, workplace, and cities. PDSs continue this blurring of the public and private—the processing undertaken through a device may not only be highly personal to the user, but it could also potentially relate to *other* individuals, such as other occupants in a home or members of the general public. In some contexts, it may, therefore, be difficult to discern situations where users will act in a purely domestic manner (where the personal and household exemption likely applies) rather than in a public context (where they act as controllers and, therefore, have particular obligations under data protection law). Questions thus arise as to the legal implications of these technical realities. We now consider the role and responsibilities of users in the PDS ecosystem.

Personal data are regarded as 'social by nature', meaning that personal data often relate to other persons than merely the data subject.[150] To prevent data subjects from becoming controllers for every data processing in purely personal contexts, the GDPR includes a 'personal and household exemption', which precludes the GDPR from applying to data processing whenever that processing happens 'by a natural person in the course of a purely personal or household activity' with no connection to a professional or commercial activity.[151] Such personal or household activities could include correspondence and the holding of addresses, or social networking and online activity undertaken within the context of such activities.[152] In order to address any potential gap in data protection, Recital 18 stipulates that in such cases the GDPR still applies to the 'controllers or processors which provide the means for the processing of personal data for such personal or household activities' (as with the scenarios explored above, which of the platform and app developer will be a controller or a processor and how they relate to each other in this context will depend on the nature of the platform and the circumstances in question).[153] This means that, for example, private users of a social media platform are likely to be covered by the exemption for any processing that they undertake in connection with their personal use of that platform (for instance, uploading photographs of their friends). However, the household exemption would not apply for the platform's processing of its users' data (their posts, behavioural data, and so on) regardless of context in which it is being processed; ie platforms and app developers will still be subject to the GDPR. This exemption is interpreted narrowly, as it entails a general exception from the obligations arising under the fundamental right to data protection.[154]

A second issue arises from the fact that some PDSs, such as Dataswift/HAT, claim that users will act as controllers for the processing of their own personal data. This appears to be problematic, as users will also be the data subjects for that data (ie the identified or identifiable natural person to whom the data relates).[155] Of course, users themselves may indeed have responsibility for some defined aspects—such as taking basic security measures like not sharing passwords with other data subjects or companies. Nevertheless, user responsibilities under the GDPR remain limited whenever they act in their capacity as the data subject. Having users responsible for an undetermined and unpredictable range of actions and decisions—which are in fact taken by app developers or the PDS platform—appears inconsistent with the data protection principle of effective and complete protection of data subjects, or more generally, with the spirit of the GDPR.[156]

We see two questions relevant to PDSs here. The first is the preliminary question of whether the personal and household exemption applies to PDSs used by private individuals. If it does not apply (either to all or some processing), then users will likely be data controllers for the personal data of any of the other individuals that are

147 Reed and others (n 52) 5.

148 Ibid 5.

149 The disappearance of traditional borders has been recognized in fundamental rights context discussions, see eg Bert-Jaap Koops and Merel Prinsen, 'Houses of Glass, Transparent Bodies: How New Technologies Affect Inviolability of the Home and Bodily Integrity in the Dutch Constitution' (2007) 16 Information and Communication Technology Law; <https://ssrn.com/abstract=1350507> accessed 21 August 2020.

150 WP29, Opinion 4/2007 on the concept of personal data (WP 136 of 20 June 2007) 13; Andy Crabtree and Richard Mortier, 'Human Data Interaction: Historical Lessons from Social Studies and CSCW' (2015) in *Proceedings of ECSCW*, Springer, Oslo 1.

151 Art 2(2)(c); see also WP29, Opinion 5/2009 on online social networking (WP163 of 12 June 2009).

152 Recital 18 GDPR.

153 Recital 18 GDPR.

154 Case C-101/01 *Lindqvist* (2003) EU:2003:596; *Google Spain* (n 118) para 38.

155 Edwards and others (n 52).

156 *Google Spain* (n 118) para 32.

processed by their own PDS. The second question arises if indeed the household exemption does not apply: whether, in those circumstances, data subjects can be controllers for the processing of their own personal data by their PDS. We now consider these questions in turn.

Application of the personal and household exemption to PDSs used by private individuals

The two leading cases on the personal and household exemption are *Lindqvist*[157] and *Ryneš*,[158] both of which relate to the exemption as set out in the DPD.[159] According to the CJEU in *Lindqvist*, the exemption includes only 'activities which are carried out in the course of private or family life of individuals'[160] and does not include sharing data online with an indefinite number of people. Moreover, according to *Ryneš*, processing comes under the exemption 'only where it is carried out in the purely personal or household setting of the person processing the data'[161]; accordingly, the exemption applies only where data collection is not 'directed outwards'[162] from that private setting (ie where it does not cover public space or the property of others).

Some have argued that this means that smart home users will struggle to avail of GDPR's personal and household exemption in relation to accessing external services or processing the personal data of visitors to their home.[163] However, taking *Lindqvist* and *Ryneš* together, we argue that determining the application of the exemption will involve a holistic analysis of the processing in question, assessing both (i) the 'setting' of the processing and (ii) the purposes of the processing. Where both the 'setting' and the purpose are determined to be 'personal or household', the exemption will apply—which will often be the case with PDSs.

The 'setting' of the processing is made somewhat more complex by the varying nature of PDSs. We set aside, for now, the issue of from where personal data are collected (although it is important in assessing the setting) and focus first on the location of the PDS and thus of the data's processing subsequent to collection. In some cases, the PDS will be a physical device located in a users' home or on their mobile device; this would likely be in line with the exemption. In other cases, however, it will be a virtual device located in 'the cloud'—ie the processing itself will not be undertaken within the users' home or on their personal property (such as a mobile phone) but will instead involve transferring the data to a third party's system for processing.

In *Lindqvist*, the CJEU considered the maintenance of a personal website and held that the Directive's household exemption could not cover the publishing of personal data on the Internet such that it was made available to an indefinite number of people.[164] However, the transfer of personal data to a closed (access is controlled) device in 'the cloud' would not have that effect. Moreover, unlike the Directive, the GDPR explicitly states that the personal and household exemption includes 'social networking and online activity undertaken within the context of [personal or household] activities'.[165] The GDPR itself thus clearly envisages that the 'personal and household setting' will extend as far as an individual's personal or household social networking and online *activity*,[166] which will, in many cases, necessarily include the transfer of personal data outside the home. Under the current legal framework, obtaining online services (including smart home and home automation services) cannot, therefore, in and of itself take an individual out of the personal and household exemption.

This might appear to conflict with the reasoning adopted by the CJEU in *Ryneš*. However, the CJEU did establish that the setting of the processing was a key consideration in determining whether the exemption applies. The issue in that case was not that processing of personal data was 'directed outwards' in the sense that it would be transferred beyond the home, but that processing was 'directed outwards' in the sense that the claimant was using a video camera to collect (and thus process) personal data of data subjects located in public space outside the claimant's property and of those located on the property of others. That is, the issue in *Ryneš* was not that personal data, once collected, would leave the claimant's home (in that case it did not); the issue was that the claimant was collecting personal data from outside of his personal and household setting. The corollary of *Ryneš* is that the household exemption would have applied had the claimant's camera only monitored his own property, potentially including any visitors to his property. Moreover, *Ryneš*, like *Lindqvist*, involved the DPD, which lacks GDPR's inclusion of explicit reference to 'social networking and online activity'. It is likely, therefore, that, under the GDPR, the

157 *Lindqvist* (n 154).

158 Case C-212/13 *František Ryneš v Úřad pro ochranu osobních údajů* (11 December 2014) EU:C:2014:2428.

159 Art 3(2) DPD.

160 *Lindqvist* (n 154) para 46.

161 *Ryneš* (n 158) para 31.

162 Ibid, para 33.

163 Edwards and others (n 52); Chen and others (n 52).

164 *Lindqvist* (n 154) para 47.

165 Recital 18 GDPR.

166 Ibid.

claimant in *Ryneš* would have been able to avail of the household exemption even where using a smart camera connected to an online service, provided the camera itself was directed only at the claimant's own property.

That *Ryneš* was primarily concerned with data processing that was 'directed outwards' as a result of its collection remains, however, a significant consideration in the context of PDSs. The data processed by PDSs could originate from various sources, both inside and outside the home. It is both possible and likely that personal data relating to others will also be collected and processed at some point. This could be the case with, for example, visitors to a private home, which has smart devices installed (that are not directed beyond the property). However, processing of their data by those smart devices would be well within the personal and household setting so, provided the purposes of the processing are themselves personal (see below), would come within the exemption. Alternatively, personal data relating to others could be processed where data are collected by devices carried with the user as they move about the world beyond their home. On the face of it, this seems to be a clear case of processing being 'directed outside' the personal and household setting. However, in *Linqvuist*, the CJEU remarked that the Directive's household exemption included activities where they 'incidentally concern or may concern the private life of other persons'.[167] While the household exemption should as a general rule be construed narrowly so as to ensure the highest level of protection for data subjects, it is possible that the collection of personal data relating to others by a private individual as they go about their personal lives would be taken to be 'incidental' and thus within the scope of the exemption.

The second issue in determining whether the personal and household exemption applies (the purposes of the processing) will depend largely on what the users are doing. It is difficult to imagine how a home automation system used within a domestic dwelling, for example, could be considered to not involve a personal and household purpose. Running a business from home, on the other hand, would be considered commercial activity[168] and thus would not be personal or household. As such, any processing undertaken in the course of running that business would not be covered by the exemption. However, not all scenarios envisaged by PDSs will be as straightforward, such as where an app returns a financial benefit or some other form of remuneration to users for particular processing. This would arguably involve a purpose that is not purely personal or household, although this would depend on whether it would be considered a transaction engaged in by a user who is still acting in a private, personal capacity, or as a commercial activity in and of itself. This would be further complicated where apps incentivize users to collect data for the financial benefit (eg recording the sound of their immediate environment, which may inadvertently capture the sound of people talking). In that situation, the likelihood of the personal and household exemption applying would be further diminished. In some circumstances—where an app financially incentivizes users to deliberately collect data on other people (such as by taking photographs of them) —the exemption will likely not be available.

Concluding that users of PDSs are, in many cases, covered by GDPR's personal and household exemption does not necessarily mean that there is no data controller for the processing that the PDS undertakes. As discussed above, platforms and app developers will be data controllers (either solely or jointly) for the processing done by a PDS. Indeed, the GDPR itself envisages that even where processing falls under the personal and household exemption, it can still have a data controller: 'this Regulation applies to controllers or processors which provide the means for processing personal data for such personal or household activities'. This supports the conclusion that users of PDSs will often come under the exemption. The reasoning given by the CJEU for a narrow interpretation of the household exemption is to ensure the highest possible level of protection for data subjects,[169] in line with the general purpose of the GDPR of providing for complete and effective protection. The exemption should, therefore, rightly be construed narrowly where its application would mean that the processing falls outside of GDPR's scope. But where there is an alternative controller—such as a PDS platform or app developer—the GDPR will still apply to their processing and thus the need to interpret around the household exemption by construing it narrowly is considerably lessened. Indeed, we note that in both *Lindqvist* and *Ryneš*, unlike in the context of PDSs, there was no other potential data controller for the processing in question. Had the CJEU found that the processing was covered by the Directive's household exemption, then the data subjects in question would have been left without protection by the Directive. However, in most cases, either PDS platforms or app developers (or both) will be data controllers for processing even where the household exemption applies for users. There is, therefore, no need to try to interpret around the household

167 *Lindqvist* (n 154) para 32.
168 Recital 18 GDPR.
169 *Lindqvist* (n 154); *Ryneš* (n 158).

exemption by construing it narrowly so as to exclude personal users of PDSs.

Data subjects as controllers for the processing of their own personal data by their device

Although users will, in most cases, be covered by the household exemption for processing done by or with their PDS, there may be circumstances in which they are not. In that case, platforms or app developers would likely be the controllers for that processing (see section 'Identifying roles and (joint) responsibilities of platforms in a scenario' above). However, some (including the French supervisory authority[170]) have argued that data subjects may in fact in some cases be the data controller for their own personal data.[171] And, as noted previously, some PDS platforms themselves claim that users will act as data controllers.

Although assigning controllership to data subjects may, on the face of it, seem like a possible interpretation given the CJEU's case law around joint controllership, approaching the law purposively leads to a different conclusion. The primary purpose of the GDPR is to provide for the 'complete and effective protection' of data subjects. The GDPR seeks to achieve this, in part, by assigning everyone who falls within the remit of the GDPR to one of three roles—(joint or sole) data controller, data processor, or data subject—and attributing to them a variety of rights, responsibilities, and obligations according to their role. It is upon this division that the rest of GDPR's protective framework hangs. The correct interpretation of the GDPR should, therefore, be the one that provides the highest level of protection to data subjects by maintaining this division of roles so as to ensure that in all cases there is a data controller who is responsible for the processing and is obliged to implement GDPR's compliance requirements, and against whom data subjects can assert their rights. The alternative, making data subjects the controllers for their own processing, would undermine this tripartite division and in doing so go against the purpose of the GDPR, with the effect that it would make data subjects more responsible for their own protection than the GDPR can possibly intend. Arguably, given this, the only interpretation that works purposively is that users are data subjects for their own personal data at all times and some other entity is the data controller (solely or jointly) who has obligations towards that data and those users and against whom users can enforce their rights. In the context of PDSs, in most cases, these data controllers would be platforms, app developers, or both.

How legal concepts might challenge the design of PDS technology

Prior to an app developer's access to or transfer of user data, or on-device conduct of computations, PDS is designed to obtain some form of user agreement to that data processing. That agreement seems usually based on user consent. However, the GDPR contains other legal bases for which app developers might be allowed to process user data. In section 'PDS's technical architecture and lawful grounds for processing', we consider how a PDS's technical design relates to these other grounds for processing. In section 'PDSs technically identifying special categories of personal data', some propositions as to the ability of PDSs to distinguish between special categories of personal data and other kinds of personal data are discussed.

PDS's technical architecture and lawful grounds for processing

Whenever app developers process personal data, the GDPR requires that processing to be predicated on a defined lawful ground as enshrined in Articles 6 and 9 GDPR.[172] Grounds that may be particularly relevant in a PDS context are *consent* (*explicit consent* for special categories of data[173]), *contract*,[174] *legal obligation, public interest*,[175] or *controller legitimate interest*.[176] The GDPR does not favour any one basis over another. It may also be that several and/or different grounds may be appropriate for different aspects of processing.

Each lawful ground for processing has different requirements. For instance, the ground of legitimate interests can only be applied by private controllers if their legitimate interests do not override the fundamental rights of the data subject.[177] Where the lawful ground is public interest, legitimate controller interests,

170 CNIL (n 120).

171 Edwards and others (n 52).

172 Art 5(1)(a) GDPR.

173 Art 6(1)(a) GDPR; explicit consent (art 9(2)(a)).

174 This ground applies where processing is necessary for the performance of a *contract* to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract (art 6(1)(b) GDPR).

175 This ground applies whenever processing is necessary in the performance of a task carried out in the *public interest* or in the exercise of official authority vested in the controller (art 6(1)(e) GDPR), or for reasons of substantial public interest (art 9(2)(g)).

176 This ground applies whenever processing is necessary for the *legitimate interests* of the controller or by a third party (art 6(1)(f) GDPR).

177 Heleen Janssen and others, 'Personal Data Stores and the GDPR's Lawful Grounds for Processing Personal Data' (12 June 2019) <https://zenodo.org/record/3234902#.XQtvrS2Q2YU> accessed 21 August 2020.

or contract, the data processing must be *necessary* for those purposes. This means that these lawful bases can not apply if there are other reasonable and less intrusive ways to achieve the desired outcome—ie if the desired outcome can be accomplished with processing less personal data, by processing personal data in a different way, or even without *any* processing of personal data. The 'most appropriate' basis in a given case will depend on the specific purpose and context for use, the nature of the parties involved (ie whether they are a public or private controller), and their motivations and relationships. Some grounds will be unavailable to commercial actors, while public bodies are likely in many circumstances to be unable to rely on consent as a ground given the power imbalances between individuals and public bodies.[178]

As section 'Personal data stores: an overview' sets out, a key aim of PDS technology is to provide means for users to be better informed, and thereby (in theory) to be better positioned to give consent for processing or to enter into a contract. At the same time, installing apps for that processing, as well as configuring data usage preferences and policies to enable such processing, means that users must be actively involved—ie take active and deliberative steps—in order for processing to occur. In this way, the technical design of PDS platforms, which require active user action for processing to occur, seems to be oriented towards supporting consent[179] and contract-based processing.[180]

What appears less considered in PDS architectures are the grounds that are *not* based on specific user agreement/undertakings. The legitimate controller interests ground, for instance, does not require user agreement before processing occurs. To use this ground in a lawful way, controllers have to be able to demonstrate and to actively undertake balancing of interests with regard to (i) what *legitimate interest* the controller pursues (or the legitimate interest a third party pursues to whom the controller discloses the data); (ii) the *necessity* to process personal data for that legitimate interest (ie why the purpose cannot be achieved by other, less intrusive means); and (iii) why the fundamental rights and freedoms of the data subject do not take precedence over the controller's legitimate interests.[181] Similarly, the user agreement is not required where processing is necessary for the performance of a task carried out in the public interest; ie where the task itself or the controller's authority has a clear foundation in law, which applies whenever a public service would process personal data through an app. This could arise in situations, for example, regarding taxation, or in emergency situations such as a pandemic, where it is imaginable that some Member State law could oblige PDS users to share some of their personal data (eg location data, information about their health status) with health organizations to assist in tracking the spread of a virus.[182]

In contrast, PDS platforms are generally designed technically to predicate processing on a user taking an explicit *action* for the processing to be permitted (eg installing the application, setting preferences, agreeing to terms, etc). Therefore, the key takeaway from this discussion is that currently it is unclear how, in terms of PDSs operate, this reconciles with the lawful processing bases that do *not* require user involvement from the outset. This is an area that has received little consideration, and one requiring further attention.

Note, however, that the technical mechanisms already provided by PDS platforms could perhaps be repurposed to assist these lawful grounds in various ways. Transparency of information, such as the *nature* of the public interests or of the controller's legitimate interests as a basis for processing, could be presented to users through similar means used to provide other information to users, eg through installation processes, online notices, etc.[183]

Another consideration is that user preferences might not align with the controller's legitimate interest or a task in the public interest. A 'mismatch' between user preferences and legitimate controller interests could well occur. As regards public task, this could for instance apply in the earlier mentioned example of a public emergency, where a PDS user would not want to share their location and health-related data with a

178 Recital 43 GDPR.

179 As regards a PDS platform's preferences for consent, see eg Crabtree and others (n 5) 40, 53.

180 As regards a PDS platform's preference for contract, see eg Mydex 'Data Sharing Agreement' (n 106) and Mydex 'Terms for Members' (n 106).

181 See for the obligation to inform data subjects arts 13(1)(d) and 14(2)(b) GDPR; see Case C-13/16 *Rīgas satiksme* (4 May 2017) EU:C:2017:336, para 28; *FashionID* (n 80) para 95; Irene Kamara and Paul de Hert, 'Understanding the Balancing Act Behind the Legitimate Interest of the Controller Ground: A Pragmatic Approach' (August 2018) Brussels Privacy Hub (4) Working Paper 12.

182 But see EDPB Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak (adopted 21 April 2020), 4, which say that 'the use of contract tracing applications should be voluntary and should not rely on tracing individual movements but rather on proximity information regarding users'. Note, however, that the EDPB is an advisory body that issues opinions (art 64(1) GDPR), and that EDPB's view, while authoritative, does not necessarily preclude Member States from enacting legislation that could, for instance, make the use of an app obligatory, or require that certain data from individuals be shared with designated public bodies (though such legislation will be subject to various constraints).

183 See for instance Mydata (n 6) 5.

public organization. The legitimate and public interest grounds entail the right for data subjects to object to such processing; however, the right to object is not necessarily absolute.[184] This means that in practice, it is not necessarily the case that user disagreement requires that the processing ceases. Again, how this works in a PDS context, given the nature of the PDS technical architectures, requires consideration. Another avenue for exploration could be how the PDS's consent management mechanisms might be repurposed to give effect to a data subject's objection where appropriate.[185]

PDSs technically identifying special categories of personal data

Some scholars involved in the development of PDSs have suggested developing technical means to discern between special categories of personal data from personal data.[186] This aims to both inform and motivate app developers to implement technical measures that result in higher levels of data protection for special categories of personal data. Researchers distinguish raw data from computed data, suggesting that raw data might need different level protection, or more user control, compared to computed data.[187] We explore how the suggested approaches relate to legal perspectives on data protection.

Raw versus analysed data

Discussion regarding distinguishing raw data from computed results often occurs in engineering contexts.[188] Raw data, or data originating from its source (in a PDS context, for instance, data that are directly recorded from user activities, captured from sensors, a feed from an online platform, and so forth), are commonly understood by engineers as (personal) data that have not undergone computation.[189] Some imply that computed data or results of computation are data that may be perhaps more *indirectly* linked to users, ie 'less personal', or in some cases, impersonal.[190]

This presumption accords with the suggestion made by the developers of some platforms that users would be more willing to share more indirect data, such as the results from computation.[191] However, although the distinction between raw data and computational results might (in some situations) hold and be relevant from a computational perspective, it appears misleading in a *legal* context. In law, there is no distinction between raw and computed data. Both raw data and the results from computation can contain or otherwise be considered *personal* data.[192] Transferring and sharing results from computation can, therefore, also culminate in data protection risks. The notions of raw and computed data, though perhaps relevant terms in computer science, have no bearing in data protection law—if data of any kind directly or indirectly relate to an identified or identifiable natural person, then, regardless of whether they are 'raw' or 'computed', they will be condidered personal data and data protection law will apply.

Moreover, in a PDS context, it seems unlikely that there could be general, technical means or processes that automatically determine whether computational results entail personal data. Categorization of data as personal data is extremely context-dependent and would be difficult to do in a technical manner, not least because devising a detailed, well-defined specification for such distinctions would be inherently challenging.[193]

Special categories of personal data

Special categories of personal data are personal data that are considered to be particularly sensitive, where its processing creates more significant risks to fundamental rights, including the rights to not be discriminated against, privacy, personal autonomy, human dignity, or freedom of expression.[194] The GDPR, therefore, requires more stringent protections for and management of special categories of personal data.[195] Special categories of personal data include, among other things, data revealing racial or ethnic origin, political opinions, and religious beliefs, as well as genetic data and data concerning health, sexual orientation, and sex life.

Whether personal data is to be regarded as one of the special categories highly depends on the situation in which it is being processed. It has been suggested that PDSs can place extra constraints over the processing of

184 Arts 21(1) and (6) GDPR.

185 Art 21(1) GDPR.

186 With regard to special categories of personal data, see Lodge and others (n 40) 313.

187 Crabtree and others (n 5) 43, 51 and 52.

188 Nick Barrowman, 'Why Data Is Never Raw' (Summer/Fall 2018) The New Atlantis, A Journal of Technology and Society, 133; Urquhart, Lodge and Crabtree (n 18) 25; Crabtree (n 5) 43, 51, and 52.

189 Barrowman, ibid 131.

190 Yves-Alexandre de Montjoye and others, 'openPDS: Protecting the Privacy of Metadata Through SafeAnswers' (9 July 2014) 9 Plos One e98790.

191 Eskola and others (n 6) 8.

192 Barrowman (n 188) 131.

193 Joshua Kroll and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 695.

194 Art 9(1)(a), Recital 75 GDPR.

195 Art 9(1)(a) reads: 'Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited'; see also Recital 51 GDPR.

special categories of personal data.[196] However, this entails practical challenges, given the difficulties in identifying special categories of personal data. While some data will clearly fall within one of the special categories, there will often be situations where seemingly innocuous data can reveal information falling into one of those special categories. This might apply, for instance, when data act as a proxy, eg where a postcode might imply race[197]; or where certain benign data can be combined with other data to become particularly revealing.[198] The text of Article 9 (1)(a) GDPR confirms and encompasses such situations: it classifies data as a special category whenever it *reveals* any of the special categories of personal data.

More broadly, given the wide range of situations where the GDPR might apply, some have argued that in practice there is potential for *all* personal data to, in effect, become one of the special categories of personal data.[199] This has technical relevance, as it may mean that, in practice, a prudent app developer (or indeed, platform) should treat all data as if it concerns one of the special categories. This is particularly pertinent for two reasons. First, the nature of a PDS is to hold and manage a wealth of personal data (much of that with the potential to be sensitive). Secondly, the processing happens on-device (in the realm of the user) without app developers necessarily being in a position to 'know' the nature of all the data and associated contexts for that data and its processing. This is in contrast to processing the data 'in-house', in a company's environment, where the aspects surrounding the data are under their visibility and control.

Processing special categories of personal data is prohibited by Article 9(1) GDPR unless an exemption applies; in commercial contexts, this is generally *explicit consent*. App developers processing special categories of personal data must, therefore, ensure that the explicit consent criteria (or other relevant requirements) are met. As regards explicit consent, the examples given by WP29 might be indicative—to fulfil the obligation of explicit consent, 'a data subject may be able to issue the required statement by filling in an electronic form, by sending an e-mail, by uploading a scanned document carrying the signature of the data subject, or by using an electronic signature'.[200]

PDSs, in being consent-oriented, generally offer means to assist in appropriately obtaining user consent. As part of their user empowerment aims, PDS platforms may claim mechanisms to improve, facilitate, and make consent processes more explicit and informed, requiring proactive steps by a user to enable an app to be installed and operate. PDSs might be better placed than most technological innovations for this, given that PDSs entail onboarding procedures, and generally provide various mechanisms for users to set their preferences, in addition to providing infrastructure for securing data from unauthorized and inappropriate access. Given consent is a key focus of PDS platforms, the mechanisms offered should be designed to facilitate obtaining explicit consent for such processing.

Further, and similar to that described above, means for technically detecting special categories of personal data will inherently be limited—perhaps save some a few, simplistic situations, where data or application is explicitly labelled as being special category, eg a field in a profile describing union membership, or if an app is marked as being for healthcare.

Concluding remarks

The use of personal data to fuel the business models of a few dominant Internet companies has led to the development of PDSs. PDS platforms seek to empower their users to manage their own personal information, with a view to strengthen the user's data protection, or to otherwise enable (eg financial) user benefits from such local processing. PDS platforms provide users with a device within which they themselves capture, aggregate, and control the access to and transfer personal data from the PDS. PDSs are typically conceived of as a decentralized alternative to tackle issues around power, privacy, and data arising from the more centralized approaches.

We explored some challenges regarding PDSs. While PDSs might enable a degree of user empowerment over data processing, their technical architecture raises interesting considerations and uncertainties under the GDPR. Indeed, platforms act as intermediaries in interactions between commercial app developers and users; nevertheless, we saw that platforms themselves can also pursue commercial objectives by processing user data for their own intentions. It may not be clear for those involved in PDS ecosystems how the intentions of platforms as commercial actors and as middlemen should be judged under the GDPR, particularly where platform roles and responsibilities are concerned.

196 Lodge and others (n 39) 313.

197 Anupam Datta and others, 'Proxy Discrimination in Data Driven Systems: Theory and Experiments with Machine Learnt Data' (25 July 2017) <https://arxiv.org/pdf/1707.08120.pdf> accessed 21 August 2020.

198 See, eg Müge Fazlioglu, 'Beyond the Nature of Data: Obstacles to Protecting Sensitive Information in the European Union and the United States' (2019) 46 Fordham Urban Law Journal 271.

199 Ibid 295.

200 WP29 'Guidelines on Consent under Regulation 2016/679' (28 November 2017, WP 259) 18 and 19.

Some early-stage platform ToS we explored do contain assignments of roles and responsibilities, but they seem, particularly from GDPR perspective, deficient. This applies particularly to the assignment of roles in interactions among platform, users, and app developers. Though we recognize that the technology and platforms are nascent, so there are both an opportunity and need for adjustment. This might become urgent, as the European Commission has mentioned decentralized data processing in its EU Data Strategy as one of the ways forward to improve user control and data protection compliance.[201]

In a PDS context, legal uncertainty exists where the dividing lines between the roles lie. Regulators should learn how decentralized ecosystems function, and to help inform and instruct platforms and users about their responsibilities under the GDPR in these federated ecosystems. Platforms must also properly understand and account for the nature of the law, in addition to perhaps developing tools for users in case they act as controllers, so as to help them to comply with the GDPR. This may also stimulate the uptake of PDSs and help consumer trust in using this technology.

Further, we established that legal data protection concepts challenge design choices currently being made in PDS technical architecture. Their architecture generally fails to function in situations where other lawful bases than consent would be more appropriate. This might limit PDS uptake, given that consent can only be used in a limited number of contexts. Platforms should consider extending their architecture to the other lawful bases.

Norms with an open character such as 'special categories of personal data' may entail challenges for (joint) controllers. The application of these norms is context-specific which means that controllers must have proper knowledge of what type of data they process, and what obligations arise from the GDPR. PDSs have the potential to make such concerns more challenging for app developers, where on-device processing may mean the developers have a degree of uncertainty regarding the full context (on device) in which the data processing occurs. Conducting computations on personal data might sooner than later include special categories of personal data; if this type of data is involved, very strict conditions apply to the processing. If results of a computation include personal data or one (or more) of the special categories of personal data, the transfer of these results must be compliant with the GDPR. Guidance by regulators is necessary, especially where decentralized processing of personal data is perceived as a solution to empowering consumers over their personal data.

Another consideration emerging from our exploration of the PDS approach to GDPR compliance is that it has become clear that the law influences the design of technical architecture, while technical architecture in PDSs to enable the processing of personal data also shapes the law. This can for instance be derived from how the CJEU is prepared to involve not only legal but also to consider technical features as relevant aspects in their legal reasoning relating to determining controllership. There are questions that arise regarding the practicalities of the GDPR's assignments of roles and responsibilities in complex technical ecosystems such as PDSs. Our exploration shows that technologists should, in close cooperation with regulators and legal practitioners, foster ways forward in terms of implementing the GDPR's regulatory framework in different technological contexts. This is crucial for ensuring technology is both developed and deployed in an accountable and transparent manner.

Funding

We acknowledge the financial support of the Engineering & Physical Sciences Research Council (EP/P024394/1, EP/R033501/1), University of Cambridge via the Trust & Technology Initiative, and Microsoft via the Microsoft Cloud Computing Research Centre.

doi:10.1093/idpl/ipaa016
Advance Access Publication 28 December 2020

201 European Commission (n 3) 10.