Privacy and Security in Personal Data Clouds

FINAL REPORT

PUBLIC

NOVEMBER 2016

www.enisa.europa.eu

European Union Agency For Network And Information Security

About ENISA

The European Union Agency for Network and Information Security (ENISA) is a centre of network and information security expertise for the EU, its member states, the private sector and Europe's citizens. ENISA works with these groups to develop advice and recommendations on good practice in information security. It assists EU member states in implementing relevant EU legislation and works to improve the resilience of Europe's critical information infrastructure and networks. ENISA seeks to enhance existing expertise in EU member states by supporting the development of cross-border communities committed to improving network and information security throughout the EU. More information about ENISA and its work can be found at www.enisa.europa.eu.

Contact

For contacting the authors please use [isdp@enisa.europa.eu.]
For media enquiries about this paper, please use press@enisa.europa.eu.

Acknowledgements

We would like to thank the following persons for their valuable insight during the data collection phase of this study: Yves-Alexandre de Montjoye (openPDS/SafeAnswers), David Alexander (Mydex CIC), Leslie Liu (Google), Mary Ruddy (Personal Data Ecosystems Consortium).

Legal notice

Notice must be taken that this publication represents the views and interpretations of the authors and editors, unless stated otherwise. This publication should not be construed to be a legal action of ENISA or the ENISA bodies unless adopted pursuant to the Regulation (EU) No 526/2013. This publication does not necessarily represent state-of the-art and ENISA may update it from time to time.

Third-party sources are quoted as appropriate. ENISA is not responsible for the content of the external sources including external websites referenced in this publication.

This publication is intended for information purposes only. It must be accessible free of charge. Neither ENISA nor any person acting on its behalf is responsible for the use that might be made of the information contained in this publication.

Copyright Notice

ISBN: 978-92-9204-182-3, DOI: 10.2824/24216

Table of Contents

Executive Summary

For the purposes of this study, Personal Data Clouds ("PDCs") are defined as technological solutions aiming to provide to end-users the typical data collection and storage capabilities of data management systems but also, to help end-users regain control over their data. Accordingly, PDCs are ideally embedded by privacy-enhancing elements allowing individuals to determine on their own how they want their data to be managed in and outside of the solution and with whom they should be shared.

The main objective of the study is to identify the different architectures and components of PDCs and discuss their privacy and security challenges. Based on an empirical analysis of various applications that fall under, or are close to, the definition of PDCs, the study presents a "state of the art" analysis of the security and privacy features of PDCs. It assesses to what extent current PDC solutions, either available on the market or in a research and development phase, are supported by functionalities that enhance the level of security and privacy they offer to their users by enabling the latter to take decisions over their data and, ideally, enforce them (user centric model). Given that mobile health applications have been gaining considerable attention nowadays, especially through the data storage and communication capabilities of wearables, the study identifies in particular privacy-enhancing features already adopted by certain PDCs in the health sector.

Although PDCs represent a relatively new concept with which system designers and end-users are not fully familiar, the study has identified two key characteristics to distinguish PDCs from other system categories: user-centricity, meaning the ability of the tool to place the user in the centre of data management, and privacy enhancing technologies, setting up PDCs in a way that users' privacy will prevail over default functions that may put the protection of users' personal data at risk. These key attributes can be concretely expressed through specific features of the PDCs architecture such as:

- Data Management (the ability of users to store, access, and share data within the PDC);
- Privacy by design (incorporation of privacy protections into the design and development of the system);
- Definition of user-centric preference (the ability of users to define their own privacy preferences);
- Privilege and access management (the ability of users to define which parties can access their data);
- Privacy by default (the design of the tool builds from the outset upon security and privacy settings);
- Data deletion (the ability of users to determine the erasure of their data and the conditions of deletion);
- Data Portability (the possibility of retrieving data stored in the PDC and transferring data between PDCs);
- Security (technical measures and controls to ensure the confidentiality, integrity and availability of personal data stored and managed on the PDC – with particular focus on data encryption mechanisms);
- Traceability (having detailed logs of actions performed by users or any third parties).

The state of the art analysis of the present study revealed that, although some of the above features are considered by certain PDCs, there is still room for improvement. Some of the key findings were the following:

- Privacy by design: overall, privacy cannot yet be said to be taken into account from the outset of the development of a product for most identified PDCs, since the majority of these did not feature built-in policies of data minimization, anonymization, or pseudonymisation;
- Definition of user-centric preferences: the majority of the identified PDC solutions offered a simple binary consent system ("allow/deny") for access control;
- Privacy by default: although simple "do not share" profiles are enabled by default in most identified PDCs, the applications still do not feature, on a general level, tools to ensure that personal data is only processed to the extent necessary and stored for no longer than what it is needed;

- Data Portability: this feature was found to be sorely lacking in most of the identified PDCs. The lack of standards in this field prevent the data from being exported to another PDC developer;
- Security: while encryption was found to be widespread amongst the identified PDCs, no identified solution employed client side encryption or layered encryption. The lack of cryptographically enforced preferences was also verified for all the identified applications.

Based on these findings, it is possible to highlight certain privacy and security challenges in the future development of this field. With regard to privacy and user-centricity, a fundamental challenge is the possibility that systems that depend on users setting a large number of options by themselves may not be adopted by the general public due to the difficulty in their use. This reflects a tension between granularity and usability, both of which need to be taken into account during the design of the PDC.

From the point of view of security, the main challenges are related to the lack of adoption of client side encryption. The limits of cryptography (especially in the absence of client side encryption) should remain an open point for future development in this field, as they would enable not only stronger security levels but also the possibility of allowing PDCs to complete "link contracts" or, more generally, act as vectors for technically enforceable user preferences. On a related challenge, stronger authentication measures are also recommended, as well as more transparent procedures for dealing with data breaches and other incidents.

Following the aforementioned analysis, the studies draws a number of conclusions and subsequent recommendations for the further use of PDCs as privacy enhancing technologies:

- **PDCs and Information Management Tools**

A PDC hinges on the control granted to the user while authorizing what data to share, with whom and when. A backdrop layer of privacy-enhancing technologies is thus central to PDCs.

The research community and the developers of PDCs should continue to implement privacy-enhancing technologies in these solutions, taking into consideration that comprehensive information management tools can be combined with proper data protection mechanisms. Policy makers and regulators at national and EU levels should promote the use of PDCs as privacy enhancing technologies that can put users in control over their personal data.

- **Degree of User Control**

The study has identified a tension between granularity and usability, partly expressed in the limitations of consent as an informed basis for processing of personal data. For the most part, PDC tools still rely on the traditional and limited consent-based model, fostering binary (“allow/deny”) systems that do not easily allow to manage large quantities of data. On the other hand, full granularity of choice, for each data set, authorised party, and purpose, may engender “consent fatigue” and alienate users.

The research community and the developers of PDCs must take into account the need to offer solutions that combine a robust framework for managing personal preferences and an easy to use interface or mechanism. The European Commission should promote research and development in the field of ‘usable privacy’, especially in the context of personal information management systems, such as PDCs.

- **Enforceability of Rights**

There is still a lack of users’ rights management mechanisms in the PDC market. This represents a potential hindrance to the widespread adoption of PDCs because users have no way to ensure or enforce (in the absence of legal or contractual arrangements) that third party applications or providers will not process their personal data for other purposes.

As a key element in restoring user trust, PDC developers and the research community should place priority in implementing systems that allow users to enforce their personal choices within the PDC through the use of technical means. The European Commission and Data Protection Authorities should raise awareness of the existence and advantages of such mechanisms, as a means to facilitate the adoption technologies that are still not well understood by the general public.

- **Lack of Standards**

Without unifying standards to allow for export of PDC databases, data portability between providers will be achieved only with great difficulty.

The European Commission, Data Protection Authorities and security-focused international bodies should promote the use of standards in the fields of encryption and data management. Standards-setting bodies may play a key role in the development of new technical specifications that will promote interoperability of PDCs with other solutions they have to communicate with, or between PDCs themselves for the implementation of data portability. The research community and PDC developers should also strive to collectively work on the elaboration of widely-recognised standards, and to implement those, enabling users to transfer their information between different providers.

- **Limits of Cryptography**

Encryption alone cannot protect data inference. Other privacy-preserving computations, such as Oblivious RAM or secure multi-party computation should also be considered as means to enhance the protection of personal data.

PDC developers should not rely only on commonly used cryptographic protocols, but actively roll out more advanced forms of encryption such as client side encryption. The research community should continue developing privacy-preserving computations and relevant key management and infrastructure processes with a view to making them functionally and commercially viable.

- **Variable Level of Security**

Secure coding, regular security audits and penetration testing should be considered as a mainstay of any PDC development and distribution cycle. More efforts should be undertaken to encourage adoption of client side encryption.

PDC developers should combine robust code writing with standard operating procedures that ensure a high level of security. Data Protection Authorities and security-forced international bodies can provide the incentives to foster active, regular security monitoring of systems and procedures for the processing of personal data.

1. Introduction

1.1 Background of the Study

Data is often presented as "the new oil" of our (digital) world, a key asset with both economic and social value [1]. The term "big data" is used to describe the massive processing of high volumes of data produced very quickly by various sources. Despite the opportunities, innovation and growth arising from this omnipresence of data, users find it all the more difficult to have their privacy boundaries clearly delineated and respected in the era of big data. Increasingly, the same sets of personal data are collected by different service providers, each with their purpose and specific approach. Data subjects seeking to access their own data must acquiesce to the terms imposed by these providers, which dramatically decreases the effective control that users have over their personal data [2] [3]. The big data landscape also poses other challenges, such as the increased frequency of personal data breaches, the deployment of wide scale electronic surveillance systems, or the profiling of users without their knowledge or consent.

The challenges outlined above, coupled with revelations on the existence of massive governmental surveillance programs [4], have weakened the confidence of users in certain aspects of the digital economy [5]. To this end, the European Commission ("EC"), in an effort to restore users' trust, has called for a more balanced relation between the interests of data controllers and the rights of end-users regarding the processing of personal data [6]. The General Data Protection Regulation ("GDPR") [7], which will replace Directive 95/46/EC [8] and is set to apply from 25 May 2018, is part of this reform.

Along the same line, ENISA, in its study "Privacy by design in big data", underlined how crucial it is to shift the discussion from "big data versus privacy" to "big data with privacy" [9]. The purpose of the study was to explore ways to address privacy issues with the opportunities afforded by technology, focusing on strategies involving privacy by design and privacy enhancing technologies ("PETs") in big data analytics. Key areas to explore, according to ENISA, are solutions that can increase transparency and control for end users over their personal data.

Personal Data Clouds ("PDCs") is one such privacy enhancing solution that has recently garnered considerable attention[1]. PDCs are data management and sharing systems designed to empower individuals and help them regain control over their data. The term "PDC" is sometimes used interchangeably with "Personal Data Vaults", "Personal Data Stores", or "Personal Data Services". The origin of PDCs can be traced to the development of personal information management systems, which in turn arose out of the need to achieve better management of dispersed data. These systems, however, were not primarily concerned with the preservation of privacy, which is why the idea of PDCs is not only to enable individuals to collect, store, manage, use, and share their personal data, but to do so according to their own levels of privacy comfort, trust and needs. These developments have been bolstered by the appearance of technical standards, such as XDI ("eXtensible Data Interchange"), a semantic data interchange format and protocol under development by the OASIS XDI Technical Committee.[2]

[1] The term 'cloud' in PDCs is used to refer to the processing of different types of personal data collected from various sources and should not be confused with the use of cloud computing services. Still, it is important to note that many PDC solutions are indeed based on cloud services for the storage of their users' data (although local storage is in many cases also supported).

[2] For more information on XDI, OASIS XRI Data Interchange (XDI) TC, see https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xdi

According to a relevant study commissioned by the EC, PDCs could enable a paradigm shift in the way data can be managed online because their underlying model is user-centric [10]. This model is opposed to the traditional distributed model where data is broken down in silos controlled by third parties. One of the privacy-enhancing features of PDCs is their potential of providing granular control to users. Granular control means that there are many levels of permissions at the disposal of the users, which allow them to restrict specific actions while authorizing others. For these reasons, the European Data Protection Supervisor ("EDPS") has also specifically referred to, and encouraged, the use of PDCs as user-centric, safe and secure places to store and possibly trade personal data [11].

Nevertheless, putting users fully in charge of their data in an online environment presents various risks that, if not appropriately handled, could undermine the exact purpose of PDCs, compromising user privacy and personal data. Certain topics should therefore be borne in mind when considering the use of PDCs:

i. Firstly, if privacy policies and relevant user preferences are not adequately embedded in PDC solutions (e.g., if the options provided to the users are not flexible and/or editable enough), the users might give away more personal data than they would otherwise;
ii. Moreover, even if the users have full flexibility and control in choosing their privacy preferences, the whole system may fail if these policies are not adequately and persistently enforced and monitored across different service and applications providers (with the danger of giving the users the false perception of data protection);
iii. In addition, to reap fully the benefits of PDCs, strong security measures are needed, such as access control policies and secure data storage in cloud environments;
iv. Interoperability is another serious concern, both with regard to the interaction with other services and applications, as well as to data portability across different platforms. It is thus necessary to adopt appropriate security protocols and standards for data exchange and communication in a quite heterogeneous environment and taking into account sector specific needs (e.g. in health data communication).

A final aspect of PDCs relates to the very relationship between the individuals and their personal data. In order to assume control, users may need to be actively involved in the management of their data. Therefore, we must account for the possibility that too much granularity in information management (for example, in access control settings) may, in the end, overburden the users and alienate them from carrying out adequate management of their data. Therefore, striking a balance between granularity and usability may be of utmost importance in that respect.

1.2 Study Objectives

The main goal of this study is to explore the different architectures and components of PDCs, with an emphasis on their privacy enhancing features. Within this scope, the particular objectives of the study are as follows:

- To conduct a state-of-the-art analysis on the PDC landscape.
- To define the privacy and security challenges of PDCs, along with potential mitigation measures.
- To draw relevant conclusions and recommendations with regard to the enhancement and wider adoption of PDCs as privacy enhancing technologies.

In addition, attention is especially given to the mobile health sector, as this will likely be one of the areas where PDCs will have a significant impact. Specific challenges from this sector are therefore considered in connection with specific use-case scenarios.

1.3 Methodology

In order to approach the topic of this study, the following methodology was applied.

Firstly, information on the concept of PDCs was gathered through desktop research, including the review of specialised documentation and a variety of sources in associated fields [12] [13] [14]. Since specific studies on the topic of PDCs are still scarce, this report builds also on the prior work and investigation by ENISA, in particular its study on Privacy by Design in big data.[3] The subject of PDCs has also garnered the attention of the EC, which has produced relevant communications on the roadmap towards a data-driven economy and the potential of cloud computing in Europe [15] as a part of the European Cloud Computing Strategy 2012.[4] Previous studies on this topic were also considered, and in particular the 2015 EC study on Personal Data Stores conducted by the University of Cambridge Judge Business School.[5]

Secondly, a list of potential PDC solutions was identified on the basis of our understanding of the concept. This list was then clustered around four representative categories: relevant applications were considered due to their information management capabilities, market presence, the extent of innovation, the transparency of information provided by developers, and the connectivity of the applications with other platforms and devices. Attention was given to documentation provided by the developers of such solutions, including terms and conditions, privacy statements, and other guidance materials. These elements and the functionality of the PDCs were subsequently discussed in detail with developers, designers, or vendors of the PDCs. Wherever applicable, working versions of such solutions were used.

Finally, a state of the art review was performed on the identified solutions with a view to assessing the current state of implementation of the characterising features noted above. Based on this investigation, existing and potential privacy and security challenges could be highlighted. Recommendations and possible mitigating measures are also proposed.

1.4 Structure

Following the methodology indicated above, this report adopts a threefold structure:

- In the first part (Chapter 2), an approach is made to the concept of PDCs. Drawing from three main categories (information management capabilities, user-centricity, and implementation of privacy enhancing technologies) the report explores the characterising features of PDCs.
- In the second part (Chapter 3) a review is made of tools and applications that were identified on the basis of their approximation to the adopted concept of PDCs. The review aims at analysing the current state-of-the-art in PDCs with regard to the previously defined characteristics.
- In the third part (Chapter 4) the main security and privacy challenges involved in the use of PDCs are highlighted. Potential mitigation measures, both technical and organisational, are considered, as well as open issues for future work and research in the field. Where applicable, risks and mitigation measures are specified with relation to the mobile health scenarios, with due consideration to interoperability and communication with mobile health devices and relevant cloud-based solutions.

[3] See [9]. For the latest publications by ENISA, see https://www.enisa.europa.eu/

[4] For more information on the European Cloud Computing Strategy, see https://ec.europa.eu/digital-single-market/en/european-cloud-computing-strategy

[5] See [10].

Based on the aforementioned analysis, the report finally draws a number of conclusions and recommendations with regard to the future use of PDCs as a key privacy enhancing technology.

The target audience of this study are public authorities at the national and EU level, in their capacity both as regulators and policy makers in the area of privacy and data protection. The study may also be of interest for developers, vendors, and distributors in the field of information management solutions, as well as Research & Development communities seeking to implement privacy by design into their current and future projects. Any actors wishing to perform privacy impact assessments in similar tools may also draw insight from this report.

2. Understanding Personal Data Clouds

2.1 Personal Data Clouds in Context

In this section we focus on the concept and the characteristics of PDCs. While the goal of this study is not to analyse the historical origin of PDCs, the analysis of different solutions reveals that there are several broad categories from which these applications may have developed. The most relevant is the category related to Personal Information Management Systems ("PIMS"). Solutions included in this cluster tended to feature user-centric approaches and also, occasionally, privacy-enhancing technologies. Our understanding of the subject matter can therefore be significantly enriched if we first conduct a brief incursion into the concept of PIMS before addressing the notion and possible definition of PDCs.

2.2 Personal Information Management Systems (PIMS)

Today online and mobile users' data are typically stored on several devices (desktop, mobile devices) and/or systems of different service providers, sometimes even without the knowledge of the user. Personal data are thus fragmented and dispersed in various locations, making it difficult for individuals to have a full overview and control over the data they share. This diffusion is heightened by the increasing number of sources of personal data, which may range from data provided by the users themselves (e.g. pictures, tweets, posts), data co-produced with other users (e.g. as part of online social interactions, such as a user tagging a photograph or another user), data produced by organisations about users (e.g. billing, location data, traffic data), data captured by software, hardware sensors, or devices (e.g. web navigation, smartphone use, geo-localisation, surveillance cameras, and medical or fitness devices).

PIMS emerged as a response to this increasing dispersion of personal data, aiming to put users back in control over this information. PIMS have been described as the "digital home" of users [16]. They are software-based solutions which allow users to centralize their information and attain a global perspective on their 'digital lives'. For this purpose, PIMS run a "use software" with the user's data on specific servers, which can either be physical or virtual machines, owned by the user or hosted on another server. In general, the server is in the cloud and can be reached from any location or connection point.

PIMS embody the notion that the best response to dispersion of information may lie in better centralization of data.[6] However, since PIMS emerged primarily as a reaction against the increasingly fractured information landscape, their main concern and purpose was not to address privacy challenges related to the use of big data, but rather how to implement adequate information management capabilities.

In the next section, we will see that the PDCs and PIMS are closely related and might even overlap. However, PDCs and PIMS belong to autonomous categories. A PIMS' chief concern is the centralization and organisation of data around a structured index, where information sharing mechanisms can take an important role. A PDC will fulfil a similar role, but the element of user consent and the presence of privacy-enhancing technologies are more pronounced. Ultimately, PDCs have the aptitude to become hubs of personal data with which other

[6] An alternative view could consist, for example, in allowing users to transfer all their information to a single entity (such as a search engine provider, or a major social media platform). This entity would subsequently provide users with global search tools, data integration capabilities and synchronization options. Delegating large sets of data to such an entity, however, could lead users to a situation where they would be increasingly dependent on this entity for the overall use and further processing of their data.

applications interface. The relation between the two categories may be likened to a set of partially overlapping, concentric circles, as illustrated below:

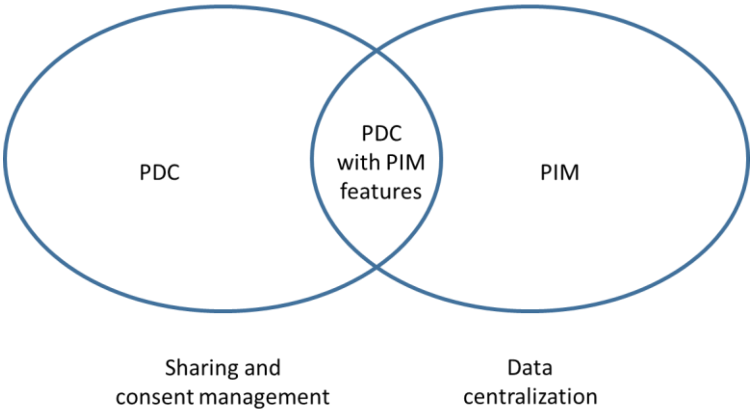

Figure 1: Relation between PDCs and PIMS

If PDCs and PIMS present some differences on a conceptual level, the two are sometimes difficult to distinguish in practice. PDCs are often interweaved with PIMS-like functionalities. This is especially evident in the way that personal data is structured in a PDC to allow for better access and more efficient sharing. However, an integral part of PDCs is the implementation of privacy-enhancing technologies or mechanisms to ensure that users control various aspects related to their own information. Such concerns are secondary in a PIMS, which primarily seek to offer a workable framework for the treatment of data.

2.3 Regulatory Framework

Regardless of its characteristics, which will be analysed below, a PDC is almost by definition implicated in the topics of personal data protection and cyber security. On the one hand, a PDC will invariably serve as an index or storage for at least a limited set of personal data. On the other hand, PDCs are cloud-heavy applications interfacing with a myriad of devices and actors, bringing to the foreground the issue of security. Therefore, when discussing the aspects and challenges of PDCs, it is advisable to keep in mind the regulatory framework applicable to both the fundamental rights to privacy and the protection of personal data, and cyber security.

The centrepieces of such frameworks are the following:

i. The EU Charter of Fundamental Rights, establishing the fundamental right to privacy (Article 7) and the right to protection of personal data (Article 8) [17];
ii. Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data, setting out the main principles and rules applicable to the processing of personal data;[7]
iii. General Data Protection Regulation, which will replace Directive 95/46/EC and will apply from 25 May 2018, carrying out a comprehensive reform in personal data protection with the addition of new principles, such as accountability and concepts such as Privacy by Design and by Default;[8]

[7] See [8] below.
[8] See [7] below.

iv. The guidelines and interpretation issued by the Article 29 Data Protection Working Party, notably in relation to the definition of data controller and processor [18], consent,[9] purpose limitation [19], big data [20], cloud computing [21] and apps on smart devices [22];

v. Security standards acknowledged and supported by international, national, and industry-based forums completing or implementing the regulatory requirements. The three formally-recognised standards-setting bodies in the EU in this field are the European Telecommunications Standards Institute (ETSI),[10] the European Committee for Standardization (CEN),[11] and the European Committee for Electrotechnical Standardization (CENELEC).[12] ENISA supports the ETSI CEN-CENELEC cyber security coordination group and works with Member States to define common European ICT security certification frameworks.[13]

2.4 Concept and Definition of Personal Data Clouds

As a relatively new and unexplored concept, there is yet no common definition of PDCs. Approximations to the concept often fasten on its core functionality, such as the possibility of allowing a user to store, access, and share personal data [23]. However, simple aggregation of data in a structured manner is the defining feature of PIMS in general, and does not take into consideration the nature of the data itself or the role of the user in the management of such a data ecosystem.

Another way of looking at PDCs includes “user-centricity” [24] as a defining trait.[14] This additional layer illustrates the need to go beyond a purely descriptive model of PDCs, acknowledging the user as a central piece in the architecture and design of the solution.[15] If PDCs are to empower users and help them regain control over their data, then they must enable them to control access on a case-by-case basis to their information. In practical terms, user-centricity entails the choice of which data or sets of data can be accessed, by whom, and for which purpose. Logically, user-centricity also implies the possibility of withdrawing consent at any time. This trait exemplifies how PDCs shift control over data from service providers back to users through a framework for a customized and granular control.

While user-centricity illustrates the specificity of PDCs, it does not sufficiently distinguish them from other information systems. Information management tools that do not encompass personal data can also be designed in a user-centric manner, and advanced access rights management systems is a feature that can be also be found in PIMS in general. A third layer is called upon: that of privacy enhancing technologies (“PETs”), required to address the protection of privacy and the right to data protection. This additional feature addresses the specificity of PDCs as tools for the management of personal information, which in fact can classify PDCs per se as PETs offering increased user transparency and control in online environments.

[9] See [28] below.

[10] For more information, see http://www.etsi.org/

[11] See also http://www.cencenelec.eu/Pages/default.aspx

[12] For more information, see https://www.cenelec.eu/. In 2010, CEN and CENELEC further consolidated their cooperation through the creation of a common CEN-CENELEC Management Centre in Brussels (http://www.cencenelec.eu/aboutus/MgtCentre/Pages/default.aspx).

[13] For more information, see https://www.enisa.europa.eu/topics/standards.

[14] See [10], although with a greater focus on the ability to grant and withdraw consent to third parties for access to personal data.

[15] See [24] for a different notion of user-centricity. For the World Economic Forum, user-centricity means involving “stakeholders in the co-creation of valuable services and experiences”. In this report, we prefer to use the term in the sense that solutions are designed around the user, or that “control” over personal data is effective. For example, it is mentioned that users should have “at least” a copy of their data. A true user-centric model, on the other hand, goes far beyond this residual right and ensures increased, effective control over personal information.

PETs have been defined as a coherent set of information and communication technology measures that protects privacy by eliminating or reducing personal data or by preventing unnecessary or undesired processing of personal data, all without losing the functionality of the data system [25]. The notion of PETs is not limited to specific technologies or features such as data minimisation or encryption,[16] but can encompass all kinds of technologies that support privacy or data protection features or consider protection goals for privacy engineering.

The combination of these three layers (information management, user-centricity, and implementation of PETs) is the preferred way of approaching the definition of a PDC. When such layers are present, a PDC can enable more targeted and personalised services, made possible by the integration of third-party providers into an ecosystem and the use of clearer, more accurate, user-curated services. Therefore, users can finally benefit from a customised level of service unique to their needs and preferences.

2.5 Characteristics

The layers identified above can be broken down into different characteristics, which will be used to assess existing PDC architectures and to identify potential privacy and security challenges. The following section contains a more detailed description of the main characteristics of a PDC.

2.5.1 Data Management

This characteristic refers to the core information management functionality of PDCs, closely linked to their PIMs roots. As information management tools, a PDC is predicated around the collection, structuring, and management of data. Generally, users either input their own data, gradually building a personal database, or else they make use of additional applications of devices, which gather and upload data in accordance with a group of pre-defined settings. In some cases data might be inserted by authorised third parties (e.g. a medical practitioner that has been authorised by the user).

It is interesting to note that the choice to accept user inputs may have repercussions on the level of the accuracy of the data (e.g. a user inserting data in his/her medical file). PDCs that rely solely on the user correctly inputting and managing the corresponding information may, over time, accumulate inaccurate data. In this context, the ability of a PDC to accept inputs from third parties should not be conceived as a mere afterthought. On the other hand, controls for ensuring the accuracy of data inserted by third parties should also be in place, especially when this is automatically done by machines/sensors. Moreover, in certain cases it will be necessary that the user is the one inserting the information (e.g. due to the fact that he/she is the only one in possession of such information). Therefore, a balance (between the user and other involved parties) should be in place with regard to the burden, as well as the responsibility of maintaining the accuracy of personal data in the PDC.

Interconnectivity plays a large role in a PDC ecosystem and the way in which it acts as an information management tool. It is understood as the ability of a PDC to accept inputs from different applications and devices and export data for immediate use in other platforms. This is the defining factor in whether websites, computer software, mobile apps, or wearable devices have the ability to connect with the PDC, storing and accessing information.

In this context, an important mechanism for ensuring accuracy of personal data is synchronisation, the process of coordinating or maintaining the consistency and uniformity between two or more devices or processes in time. If adequately implemented, synchronisation enables data storage devices and applications to have

[16] See [25], p. 13.

exactly the same information at any given time. Therefore, for all files and folders marked for synchronisation, the content will be identical across all linked devices.

2.5.2 Privacy by Design

"Privacy by design" refers to the incorporation of privacy protection into the design and development of a system [26]. It has been defined as a holistic concept that may be applied to operations throughout an organisation, end-to-end, including its information technology, business practices, processes, physical design and networked infrastructure [27]. Privacy by design acknowledges that neither the legal and business framework nor the simple accumulation of PETs on an already designed tool are sufficient to safeguard the rights of data subjects up to the desirable level of protection. Privacy should instead be embedded into the design, operation, and management of information communication technologies and systems, across the entire information life cycle.

The GDPR for the first time introduces a legal obligation on data protection by design for data controllers. In this context, data protection by design is referred as the implementation of "*appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner*" and the integration of "*the necessary safeguards into the processing in order to meet the requirements of [the GDPR] and protect the rights of data subjects*".[17]

As defined in previous ENISA's work,[18] privacy and data protection by design can be broken down in practice into specific design strategies (Table 1). These strategies can be attained through the use of particular technologies, such as authentication, attribute based credentials, secure private communications, anonymity and pseudonymity, statistical disclosure control, privacy preserving computations, and others.

	PRIVACY BY DESIGN STRATEGY	DESCRIPTION
1	Minimize	The amount of personal data should be restricted to the minimal amount (data minimization).
2	Hide	Personal data and their interrelations should be hidden from plain view.
3	Separate	Personal data should be processed in a distributed fashion, in separate compartments whenever possible.
4	Aggregate	Personal data should be processed at the highest level of aggregation and with the least possible detail in which it is (still) useful.
5	Inform	Data subjects should be adequately informed whenever processed (transparency).
6	Control	Data subjects should be provided agency over the processing of their personal data.
7	Enforce	A privacy policy compatible with legal requirements should be in place and should be enforced.
8	Demonstrate	Data controllers must be able to demonstrate compliance with privacy policy into force and any applicable legal requirements.

Table 1 – Privacy by design strategies[19]

[17] See GDPR, Article 25(1).
[18] See [9] for more information.
[19] See [9] p. 22.

In the course of this study, we consider privacy by design with regard to the aforementioned design strategies and, in this respect, it is expected that different measures and different levels of implementation might be available among existing PDCs. Having said that, it is also important to note that many of the PDC characteristics (Table 1) are in fact touching upon different elements of privacy by design (e.g. on user choice and control, consent mechanism, data deletion, etc.).

2.5.3 Definition of User-Centric Preferences

This characteristic refers to the ability of users to define their own preferences with regard to the processing of their personal data. As such it is directly linked to the existence of choice from the part of the users or otherwise the gamut of options available to the users regarding the overall terms of processing of their personal data.

In that sense, user-centricity is primarily and most importantly formulated around the notion of consent,[20] providing the users with the possibility to clearly define if and how the processing of their data will be done [28]. This is in fact the main privacy enhancing feature of PDCs, as they could ideally provide a platform where users can be the actual decision makers regarding the processing of their personal data[21]. Moreover, user-centricity involves the broader possibility of adjusting the PDC's settings, in order to define specific aspects of the processing of personal data, e.g. if data may be processed for other purposes, or whether or not certain security measures should apply (such as end-to-end encryption or two-step authentication mechanisms).

Having said that, it is important to note that user-centricity is also very closely linked to usability. Indeed, the expression of users' preferences should be performed in a way that is easy and practical for the users (in contrast with the existing often impractical ways to obtain consent[22] [29] [30]). Striking the right balance between user control and usability should in fact be the strongest dimension and differentiating element of PDCs that could boost them as a privacy enhancing technology.

2.5.4 Privilege and access control management

This characteristic refers to the ability of users to, one the one hand, define who has access to the information stored in the PDC and, on the other hand, to further modulate access to their data by sharing only specific parts of their personal data sets (granularity)[23].

Although clearly linked to user-centricity, we have included this point as a particular characteristic of PDCs, due to the fact that third party access is core in the very role and implementation of PDCs, and thus requires special attention.

In order for privilege and access control management to be efficient, the PDC should ideally not only allow the user to set his/her preferences, but also to maintain a level of control on the data when it leaves the application. This can be achieved by technically enforcing the desired privacy and security policies, in a way that the users' preferences practically 'travel' together with the personal data, as integral part of their

[20] For an analysis on the notion of consent as deriving from Article 2(h) of Directive 95/46/EC ("any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed") see [28].

[21] When the legal basis of the processing is based on consent.

[22] Long, difficult to read terms and conditions and ubiquitous "cookie banners" can prompt users to thoughtlessly accept all notices presented to them, regardless of their content. The term "consent-fatigue" is used to describe the feeling of alienation experienced by users who are so often prompted for consent that they cease to consider the personal data implications of each request.

[23] This feature is commonly found even in non-PDC solutions, as in the ubiquitous case of a collaborative shared folder or drive in which access is granted to specific group members, and only in relation to certain files or folders.

processing conditions. Such a feature could be implemented with the use of encryption technologies, although its level of implementation in practice is currently very low (see also Sections 3.2.3 and 3.2.7 below).

2.5.5 Privacy by Default

Privacy by default refers to a pre-defined behaviour and configuration of the PDC to provide the highest level of privacy and security protection. Conceptually, it can be distinguished from the offering of user-centric preferences because it refers to a moment where such preferences cannot yet be taken into account. Therefore, privacy by default is a pre-emptive, turnkey form of protection of personal data implemented by the PDC developer.

The GDPR for the first time introduces an obligation for data protection by default to data controllers. In this context, data protection by default is described as the implementation of "*appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons*."[24]

Taking into account their overall privacy enhancing role, PDCs are expected to ensure that only personal data that are really necessary for each specific processing purpose are stored and used, a limitation that extends to the amount of data collected, the extent of the processing, the period of their storage, and their accessibility. A crucial feature in that sense is a default instruction to "not share", a point also relevant to the granularity aspect addressed under privilege and access control management. This assumes that the PDC offers a default protective setting that can be changed by the users according to their intended purposes and data recipients. However, if the users do not perform any active choices, their privacy and personal data are still protected.

2.5.6 Data Deletion

This characteristic refers to the ability of users to freely and permanently erase their personal data stored in the PDC at any time. Given the overarching goal of PDCs, this feature should be a core component of the solution, as the act of deleting data is an integral part of information management.

The right to deletion or erasure figures prominently in the GDPR, setting out the obligation, for the data controller, to erase personal data without undue delay upon request of the data subject. The data subjects can substantiate their requests with one of the grounds foreseen in GDPR, such as when the data is no longer necessary for the initial purpose of processing or when a data subject wishes to withdraw consent[25]. Owing to the user-centric nature of PDCs, these criteria are defined by the users themselves. For example, a user may decide that it no longer serves a purpose to store information about a former bank account number: in such a case, the purpose for which the data was initially uploaded and subsequently processed ceases to exist and, thus, the user may ask for deletion.

Having said that, it is important to note that the effectiveness of this category depends on two key parameters. The first parameter is the enforcement of user choices, which (as in the case of privilege and access control management) should ideally be technically enforced, e.g. with the use of encryption. The second parameter

[24] See GDPR, Article 25(2).
[25] See GDPR Article 17.

is the permanent character of deletion, which is directly linked to technical measures applied for secure data deletion.

2.5.7 Data Portability

This characteristic refers to the possibility of users retrieving the personal data that they have stored on the PDC in a structured, commonly used, and machine readable format. In addition, data portability refers to the transfer of data between PDCs at the request of the users.

The above definition follows the relevant legal obligation introduced in GDPR, according to which a data subject has the right to receive "*the personal data concerning him or her (...) in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance*". In the GDPR, this right is limited to cases where processing is based on consent or a contract and carried out by automated means[26].

Data portability often involves storing the information in an open or commonly-read format, but more specific formats can also be used depending on the type of data. For example, certain standards are used by the health industry, such as Continuity of Care Record ("CCR")[27] and Continuity of Care Document ("CCD").[28] While only applicable to certain data, these standards contribute towards data portability since they are designed to facilitate the transfer of information between entities.[29]

2.5.8 Security

This characteristic refers to the implementation of the three pillars of data security: confidentiality, integrity, and availability of personal data (commonly referred to as "CIA") [31]. Confidentiality refers to data that is protected against unauthorised access, while integrity refers to data that has not been altered with reference to an initial state. Availability, on the other hand, refers to the possibility that authorised parties will be able to access the information when needed.

Security is central in GDPR, which imposes a legal obligation to the data controller and processor to '*implement appropriate technical and organisational measures, to ensure a level of security appropriate to the risk, including inter alia, as appropriate: (a) the pseudonymisation and encryption of personal data; (b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of systems and services processing personal data; (c) the ability to restore the availability and access to data in a timely manner in the event of a physical or technical incident; (d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.*[30]

Security is a holistic category in so far as its effectiveness hinges upon the combination of various technical and organisational measures. Therefore, to assess the overall level of security in PDCs, we need to take into account different measures in place, such as authentication mechanisms, access control policies, network security measures, etc.

[26] See Article 20 of the GDPR.

[27] For more information, see http://www.astm.org/Standards/E2369.htm.

[28] For more information on this standard, see http://cdatools.org/infocenter/index.jsp?topic=%2Forg.openhealthtools.mdht.uml.cda.hitsp.doc%2Fintroduction%2FOverview.html

[29] However, certain standards have emerged for specific aspects of an application (such as "UMA" for authentication), http://kantarainitiative.org/confluence/display/uma/Home?src=contextnavchildmode

[30] See Article 30 of the GDPR.

2.5.8.1 Encryption

Encryption merits a separate mention in the course of this study, since it is an essential security measure to implement secrecy and integrity for electronic communication, the equivalent for electronic communications of the letter cover, seal, and rubber stamp in the brick and mortar world [32].

Broadly speaking, depending on the threat model, encryption can be implemented in the following manner:[31]

- **At client side level:** Data is encrypted on the client side (user), and the PDC provider never has access to clear-text data (it only sees encoded data). This kind of implementation not only gives back control to the user but can also be used for secure deletion purposes (e.g., the user can securely delete his/her encryption key thus rendering the data impossible to use from the server side) or enforcing access control systems (e.g., the access control system can rely on a software access list and encryption which would reduce the risk of unauthorised data disclosure).
- **At the server side level:** Data can be encrypted either at the application level or database level, in order to reduce the risk of data leakage or enforce an access control mechanism. It has to be noted that as long as the encryption and decryption process takes place at the server's side, clear text data and encryption keys can be accessed by the PDC server's administrators.
- **At link level:** the communication between the client and the server is encoded and authenticated. The aim is to mitigate the risk of an attacker eavesdropping on the connection or trying to alter the information (for instance to mitigate the risk of an interception when connecting to the PDC service in an airport).

In addition to 'traditional' mechanisms for data encryption, other privacy preserving techniques can be also envisaged, especially in the field of encrypted search[32], although these are in very primitive stages with regard to practical implementation (as shown for example in [33]). This is also the case of homomorphic encryption that could potentially support privacy-preserving computations but is still in research phase.

In the area of PDCs, encryption, on top of providing data security, can also be an integral tool to guarantee technical enforcement of privacy preferences and relevant user controls (i.e. to achieve effectiveness of privacy settings through objective, machine-readable actions). For example, contrary to legally-based enforcement measures, which are dependent on human, voluntary observance, a well-implemented encryption system cannot be bypassed even if an administrator or a PDC provider decides not to respect its legal obligations.

Cryptographic enforcement of privacy policies and preferences can be envisaged in combination with the use of "Link Contracts" [34] [35] or "sticky policies" [36] [37] supported by certain protocols such as XDI [38]. Such implementations can ensure that any third party accessing PDC data must comply with the user's preferences, which are embedded in the data and thus technically protected against any attempt to override them. For example, as a means to palliate the risk of a bug on the consent sharing mechanism, a second layer of control on the data can be set using encryption. The data stored in the PDC would thus be encrypted with a specific key for each authorized third party, on the client side (this approach is an extension of the multiple layered encryption based on data classification).

[31] See ENISA's report on algorithms, key size, and parameters [47] for more information,
[32] For an in-depth analysis of the topic, see ENISA Privacy by Design in Big Data [9].

2.5.9 Traceability

This characteristic refers to the possibility of the PDC to trace actions performed by users or any third parties (e.g. consent granted, data modified) as well as PDC administrators (e.g. data accessed). Although it can be considered also as a security control, we particularly refer to it as a separate characteristic due to its importance in managing abuses of PDCs, as well as the subsequent legal implications. As such it is closely related to the overall accountability principle that is enshrined in the GDPR.[33]

Traceability can be primarily achieved with the use of detailed logs, as well as with more advanced measures, like "watermarks", which dynamically implement invisible marks in files, allowing to follow the transmission of documents and help spot illegal actions (unauthorized sharing of information for instance) [39].

2.5.10 Summary

The following table summarises the particular characteristics of PDCs that will be considered in the state of the art review.

	CHARACTERISTIC	DESCRIPTION
1	Data management	The ability of users to store, access, and share data within the PDC, including with regard to its level of interconnectivity and the synchronisation of data across multiple devices.
2	Privacy by design	A set of practices or measures (e.g. the use of privacy enhancing technologies) aiming to embed privacy requirements directly into the design of information systems.
3	Definition of user-centric preferences	The ability of users to define their own privacy and security preferences.
4	Privilege and access management	The ability of users to define which parties can access information stored in the PDC.
5	Privacy by default	Pre-defined behaviour and configuration of the PDC to provide the highest level of privacy and security protection.
6	Data deletion	The ability for users to freely erase their data from the PDC.
7	Data portability	The ability of users to retrieve their personal data stored on the PDC, as well as to have their data transmitted to other PDC providers.
8	Security	A set of practices or measures ensuring the confidentiality, integrity, and availability of personal data (e.g. encryption).
9	Traceability	The ability of the PDC to trace actions performed by users or any third parties.

Table 2: PDC characteristics

On top of the PDC characteristics presented below, another important element of consideration is the business model behind the PDC operation, as this may have implications on the type of service offered, as well as the control given to the users (e.g. distinction between free and subscription based models).

[33] In Article 5(2) of the GDPR, "Accountability" means an obligation to report and explain, combined with principles of transparency and traceability, with a view to identify and document the measures implemented to comply with data privacy law requirements.

2.6 The mHealth Sector: Use Case Scenarios

The mobile health sector ("mHealth") is a rapidly emerging field having the potential to play a key role in the transformation of the healthcare industry. However, due to the sensitive nature of the health data, the implementation of proper security and privacy safeguards is a very critical requirement for mHealth [40]. As shown in a survey by the EC, the uptake of mHealth is greatly dependent on security and privacy [41], with consumers aspiring to obtain safety, transparency, interoperability, and effectiveness [42]. Despite the popularity of fitness and mHealth apps and devices [43], the high sensitivity of health-related data continues to bear upon user's overall willingness to share such information, particularly when it concerns medical records [44].

The health sector stands to benefit from the introduction of user-centric, privacy-minded PDCs that help individuals control their own health information. PDCs could also enable more efficient, tailored medical care at a lower cost, allow patient empowerment, and facilitate access to medical care and information online. mHealth PDCs could also bring significant reductions in public healthcare expenditures, a large part of which is typically allocated to the gathering of data on patients.[34]

For this reason, the specific case of PDCs in the area of mHealth is considered throughout the report through the lens of the following use case scenario:

- A mHealth PDC is available for use in mobile devices.
- The PDC allows the user to centralise medical data from several sources, e.g. different sensors and wearable devices that track a variety of health indicators.[35]
- Data is sent to the PDC along one of two paths:
  - <u>Direct Path</u> (Green): the wearable sensor (e.g. heart monitor) sends information directly to the PDC (heart rate over a period of time);
  - <u>Indirect Path</u> (Blue): the wearable sensor (e.g. glucose meter) sends the information to a third party (e.g.,: the a healthcare service provider), who subsequently transfers it to the PDC;
- Once the data is stored in the PDC, the user can share it with third parties, such as his/her physician.

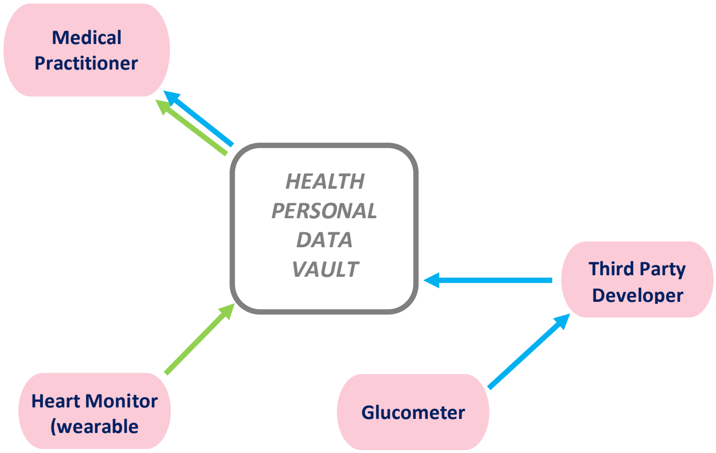

Figure 2 – Illustrative schematic of a mHealth use case scenario

[34] See [10], p. 5.
[35] Such as for example hydration levels, breathing patterns, blood glucose levels, sleep patterns, and other.

Throughout the report, we analyse the implications of each path, following the different characteristics of PDCs. It can already be noted that while the data is sent directly to the PDC in the case of the Green Path, there is an additional intermediary to be considered in the Blue Path, which raises the question of user trust in such a party. In both cases, the level of user-centricity of the PDC needs to be considered, particularly with regard to the ways in which the gathered information can be used and shared. For example, in order to allow the physician to remotely monitor the user's health, the user may wish to allow the PDC to automatically share certain types of data as they are collected.[36]

Another important dimension in this scenario is whether the personal data are stored locally on the user's device (e.g. PDC as a smartphone application) or whether the service is offered online by a PDC service provider. In the context of the report we mainly focus on the second case (PDC provider), as this is the most common business model applied today, but also due to the increased privacy and security risks that this presents. Indeed, in such a case, it is important to consider the implication of the PDC provider, as data controller, in the data processing scenario (e.g. whether the PDC provider can have access to the personal data of the users or not). Moreover, the PDC provider may have a sub processor of personal data, as is often the case in cloud-enabled applications, whose role also needs to be assessed.

Further, as mentioned above (sub-section 2.5.7), data portability often involves storing the information in an open or commonly-read format. This is especially relevant when sending data through the Indirect Path (Blue) to a third party. When adopting standard exchange formats such as Continuity of Care Record ("CCR") and Continuity of Care Document ("CCD"), a number of security considerations should be made. The security of the information exchange process depends largely on the security of the underlying transport mechanism, therefore it is crucial that appropriate measures are taken to protect this layer[37]. There are multiple approaches to this problem - Transport Layer Security (TLS) with public key infrastructure (PKI) is commonly used to ensure integrity of information, while encryption combined with some authentication scheme (e.g., API keys or passwords) provides confidentiality. However, many transport mechanisms do not explicitly address availability, although certain technologies are inherently more resilient (e.g., those allowing easy replication of resources) than others.

Lastly, it should be mentioned that both Green and Blue paths may coexist with the use of third party applications that integrate with the aforementioned wearable and process such data (as stored in the PDC) to provide user-sensitive information.

[36] This assumes that the PDC provides granularity of access to different types or levels of information stored within.

[37] Finding confirmed in ENISA's study, "Standards and tools for exchange and processing of actionable information", November 2014, p. 28, at: https://www.enisa.europa.eu/publications/standards-and-tools-for-exchange-and-processing-of-actionable-information.

3. State of the Art

3.1 Introduction

As referred to in Chapter 1, our PDC research methodology was primarily focused on the clustering of available tools around different categories with common denominators and characteristics. Such categories included e.g. cloud-based data vaults offering storage space for different types of files[38], PIMS operating as personal organisers or file management/back-up solutions[39], personal informatics (also known as quantified self) systems[40], as well as cloud solutions specialised in the management of health information[41].

Taking into account the adopted definition of PDCs, it becomes clear that none of the aforementioned categories can be used to cover the concept and functionality of PDCs. However, a PDC can fall under any of the clusters if it provides information management capabilities, offers a degree of user-centricity, and incorporates privacy-enhancing technologies.

This clustering exercise, however, permitted us to identify a list of tools and applications that approached the concept of PDCs ('identified PDCs'). This list includes both commercial and free applications, as well as research level solutions. It should be noted at the outset that although the 'ideal PDC' is not yet available in practice, different features and approaches can indeed be found in existing tools, addressing one or more of the PDC characteristics mentioned in Table 1 of Chapter 2. However, efforts are still needed both at the policy and technical levels towards the practical implementation of PDC solutions that can truly put the user in the centre of data processing and, thus, fully operate as privacy enhancing technologies.

The sample of identified PDCs in this study does not purport to be exhaustive of the current selection of tools in the market, serving only as a reference for the state of the art review of PDC characteristics. Furthermore, any reference to a specific application or tool should not be understood as an explicit or implicit exclusion of any others.

It should also be noted that the state of the art review does not delve into the analysis of the different business models that support PDC development and marketing. The distinction between free and subscription-based models is the main differentiating factor, but there are also other possibilities, such as charging access to organisations but not individual users.[42] Apart from the distinction between free and subscription-models, a PDCs may offer more advanced features in exchange for additional fees.

The next sections present in more detail the results of the review, following the specific defining characteristics of PDCs.

[38] For example, Dropbox (https://www.dropbox.com/) or OwnCloud (https://owncloud.org/).
[39] A representative example of such an application is CozyCloud (https://cozy.io/en/).
[40] For example activity trackers or health devices, such as Withings (http://www.withings.com/eu/en/), Fitbit (https://www.fitbit.com/be), and Strava (https://www.strava.com/).
[41] Examples include CareSync (http://www.caresync.com) and DiaSend (https://www.diasend.com/us/).
[42] In this model, which could be termed as a "hybrid business model", a "connection fee" may charge a set fee for organisations but not the individual users. An example of this hybrid model was found in Mydex CIC (https://mydex.org/)

3.2 State of the Art Review

3.2.1 Data Management

The capability to store and share data is a hallmark of PIMS in general, as it allows users to centralize their personal data in one location. All the identified PDCs were predicated around this feature, confirming the notion that the main functionality of PIMS stands at the core of these type of solutions.

In terms of interconnectivity, the vast majority of identified PDCs displayed the ability to interface with other applications or devices. While the majority of PDCs allowed the user to delegate the tracking of certain data to external (e.g. wearable) devices[43], a few PDCs relied mainly on the user correctly inputting and managing the corresponding information.[44]

Synchronisation of data was also widespread among the identified PDCs. While in all cases it is an essential component of the functionality of the PDC, the ability to synchronise data across multiple devices came especially into play whenever the PDC was highly dependent on data provided by wearable devices.

In terms of type of storage, cloud storage seemed to be the prevailing method.[45] However, in some cases it is possible that personal data are stored on the user's device (and not in the cloud servers of a PDC provider)[46].

In the mHealth sector, as PDCs are extremely dependent on continuous flows of accurate information from external devices, interconnectivity and synchronisation, are explicitly emphasised. User input is also a desirable feature (although only in certain cases, e.g. maintenance of a personal medical file). Another point to consider in mHealth is the sheer diversity of personal data that can be stored in a PDC. While in principle all such information will be considered health-related personal data, there is nonetheless a difference between the criticality of different types of health data, e.g. the criticality of the number of calories burned during a workout and the latest blood glucose level information. Yet no mHealth PDC permits this data differentiation as a basic data management option[47].

3.2.2 Privacy by Design

Privacy by design was contemplated across the identified PDCs, a number of which deployed different levels and types of privacy-enhancing technologies. The concrete approach to privacy by design varied considerably though, a finding which is perhaps not surprising given that privacy by design is not a set of instructions, but of principles, and therefore capable of supporting different types of implementation.

Overall, though, the state of the art review leads to the conclusion that, despite the fact that PETs are a characterizing feature of PDCs, privacy is still not yet taken into consideration from the outset of the development of these products. This finding is due to the fact that a number of key PETs where not integrated into existing PDCs, such as for example client-side encryption, built-in policy of data minimization,

[43] For example, Microsoft Health Vault (https://www.healthvault.com/be/en) offers a platform where users' data can be uploaded from wearable devices, collecting different types of health indicators, e.g. blood pressure indicators or glucose levels.
[44] For example, Mydex CIC, TeamData (https://teamdata.com/), and OpenPDS (http://openpds.media.mit.edu/).
[45] In some cases the PDC provider would allow for more storage space if the user linked the service with that of another cloud provider (e.g. TeamData).
[46] Examples of such implementations are OpenPDS, as well as CyberAll (https://www.microsoft.com/en-us/research/publication/a-cyber-all-project-a-personal-store-for-everything/), which relies on the Windows File System.
[47] However, this feature could be potentiality addressed under the level of granularity of data access, where the user can himself/herself decide on the criticality of distinct datasets and accordingly set privacy preferences.

anonymization, or pseudonymisation. Data deletion was likewise entirely dependent on user input, without the assurance of permanent erasure. More details on these topics are provided in the next paragraphs.

However, on the positive side, there were some prominent examples of privacy by design. One of them was expressed in the degree of access that the PDC provider itself had to the stored data. For instance, some solutions purposefully locked themselves out of reach of user data through the use of server-side encryption. In one notable case[48], the tool did not quantitatively minimize the data that was collected, but sought to ensure that the least possible amount of information was shared. Indeed, instead of sending raw data to third parties, this data-agnostic PDC only replies to specific questions, providing a simple answer to a query, without actually allowing access to any data stored. Privacy by design, in this case, was manifest in the way that data minimisation becomes the standard, preventing the communication of full sets of raw data.

3.2.3 Definition of User-Centric Preferences, Privilege and Access Control Management

Since PDCs are predicated on one's ability to manage personal information, it is not surprising to see that almost all of the identified solutions provide users with ways to adjust their desired level of protection. Indeed, barring a small number of very specific cases[49], all identified PDCs provided at least some degree of choice in terms of what data sets to share, and with whom. For example, as a minimum, a user may allow access to his/her data by one service provider while denying access to another.

On the other hand, the ability to modulate third party access to data by sharing only specific parts of personal data sets was found to be much less prevalent. When offered, such a degree of *granularity* was, in general, very variable: some applications offered basic controls whereas others offered multiple options for defining access control to different sets of data.[50]

In terms of implementation, the majority of identified PDCs offer a client-side oriented sharing mechanism based on consent with a simple binary system ("allow/deny"). A few tools offer a granular definition of the rights for each shared attribute. In all cases, the systems rely on software-level access control that is not enforced by client-side encryption.

However, user-centricity implies not just control over who has access to the data and to what extent, but also over the purposes for which such information can to be processed. The identified solutions tended to provide ways for users to define the purposes for which their data would be ultimately used, but this was achieved indirectly: users may authorise access by a certain third party due to their acknowledgement of the service being provided (again, within a "allow/deny" framework). In other words, the identified PDCs do not provide the means to directly control the purposes for which data is processed by a third party, acting mostly as vehicles for such data. On a related note, no application fully empowered users to modulate certain aspects of the processing of data, such as deciding on the periods of data retention after deletion of data or closing of an account.

This finding is related to the importance of *technical enforcement of user preferences*. While the former is at least partially implemented in most of the identified PDCs (with regard to the choice over which data sets can be shared), the second area is still severely lacking. In short, users do not yet have the ability to technically enforce their choices, being dependent instead on the voluntary observance – by the PDC provider or any

[48] This was the case of OpenPDS/Safe Answers.

[49] These were mainly cases where the primary scope of the PDC was to obtain informed consent from the users (rather than adopting to specific privacy settings).

[50] For example, by allowing access control policies to be set on each particular data set or subfield (e.g. OpenPDS)

third party who has access to the data – of the privacy terms selected by the user (see also section 3.2.7 on security). [51]

In the mHealth sector, the identified PDCs appeared more invested in the importance of providing granularity in user choices. Due to their nature as gatekeepers for health-related data, these PDCs often enabled users to group a set of preferences under specific profiles. A user may wish to restrict access to certain types of information while simultaneously sharing other sets of data. Depending on the personal goals of the user, there may indeed be a difference between the desired level of personal data protection of a user who requires urgent medical attention and the user that merely desires to obtain a routine medical consultation. Due to the sensitive nature of the information stored in the PDC, the mHealth sector provides the best illustration of the importance of user-centricity.

3.2.4 Privacy by Default

Most identified PDCs had certain privacy settings by default, in particular settings preventing the sharing of information without the consent of the user (with a default "do not share" status). This was particularly evident in cases where the PDC serves as an intermediary tool between the user and a third party application.

An example where this concept was taken a step further was also shown by preventing all third party access to raw data by default, and providing only answers to specific queries.[52] While this feature was at the core of the design of the PDC itself, it can be regarded as an efficient implementation of the principle of privacy by default since it liberates the user from having to tweak the PDC settings in order to attain an adequate level of protection of personal data.

On the other hand, most identified PDCs did not provide specific tools to ensure that personal data is only processed to the extent necessary and stored for no longer than what it is needed (which is another dimension to privacy by default). This responsibility seems to fall upon the users themselves, which is perhaps a consequence of the inherent user-centricity that most identified PDCs display. Nevertheless, and particularly with regard to the mHealth sector or other types of sensitive personal data, such a feature would have been desirable.

3.2.5 Data Deletion

The ability for users to freely erase their data is a core functionality of a PDC, and it is difficult to conceive any information management system that does not account for it. Amongst the identified solutions, only a very small minority did not allow for deletion of data.[53]

Given the ubiquity of this basic functionality, the analysis shifted to other aspects of data deletion, namely to how accessible it is to the end user, and which guarantees does the latter have – if any – that the data has been effectively deleted once the command to do so is given and that deleted data are in no way recoverable (secure deletion). In all cases, however, it was not possible to verify how deletion was technically carried out. Moreover, while the analysed PDCs seemed to implement secure erase procedures, these operations were conducted on the server side. Since client-side encryption is scarcely used (see Section 3.2.7 on security), users had no technical guarantees that the deletion of data was effectively performed. Therefore, the user has to

[51] However, in one notable case (Mydex CIC) it was highlighted the potential of XDI as a means to implement link contracts in a PDC ecosystem (see also [35]).

[52] This was the case with OpenPDS/Safe Answers.

[53] Even in such cases, this anomaly was explained by the specific purpose of the concerned applications, which were primarily focused on obtaining informed consent as a means to allow certain types of data to flow between selected third parties (rather than providing a true data store).

trust the service provider that his/her data have been deleted after a relevant request. In some cases, the PDC has been certified and the confidence of a well implemented secure erasing procedures is higher.

3.2.6 Data Portability

The first aspect of data portability, i.e., the storing of information in a structured, commonly used, and machine readable format was offered by the large majority of identified PDCs. This means that users are typically allowed to download a copy of their data in an open source format.

In some cases, the identified PDCs offered users additional options in terms of format. In the field of mHealth, users can export and save their health information in different ways, e.g. as a spreadsheet, or as a CCR (Continuity of Care Record), CCD (Continuity of Care Document), or an HTML file.[54] Depending on the purpose, a spreadsheet or HTML page may be convenient if the users want a copy of their information on paper to bring it to a medical appointment. CCR and CCD formats, on the other hand, may prove more useful if users wish to transfer health information between clinical systems.

The availability of the second aspect of data portability, *i.e.,* the possibility of transferring data between PDCs was considerably less prevalent in the identified solutions. Generally, the market practice relating to how this aspect of data portability is implemented is at present very limited. Moreover, implementing full portability may not be as straightforward since there is no global standard for PDC data exports. Even if standards such as UMA (for authentication) have emerged, they are not yet widely used and each PDC relies on its own standards. For the most part, data exchange seems reduced to specific PDCs and specific ecosystems.

The notable exception to this situation lies in some health-related solutions, where the CCR and CCD formats are widely employed to transfer and read data across different medical practitioners. To refer to our mHealth use case scenario (Chapter 2), PDCs are designed in such a way as to accept inputs from different applications and devices and export data for immediate use in other platforms or by other parties. In this sense, affording the user with a large degree of data portability seems to be essential for attaining the goals of an mHealth PDC.

3.2.7 Security

All identified PDC solutions offer a level of basic security measures but the degree of protection and implementation may vary significantly.

With regard to authentication, all identified solutions rely on password-based systems. Two-factor authentication is not widely deployed. The use of Access Controls Lists (ACL) is widely employed for user access control. However, it seems that the enforcement of privacy features against irregular access from the PDC provider relies mostly on legal grounds rather than technical countermeasures such as client-side encryption.

Moreover, none of the PDCs implement a Digital Rights Management (DRM) system [45][55] or, more generally, any system for the technical enforcement of user preferences as outlined in Section **Error! Reference source ot found.** above.[56] This means that once the personal information is shared with a third party, the privacy

[54] For more information on these standards, see http://healthstandards.com/blog/2010/03/10/ccd-and-ccr-the-discussion-continues/.

[55] Digital Rights Management ("DRM") is currently used to restrict usage of copyrighted works. Personal data is such platforms can be encrypted and the third party would need to request an authorization each time it wants to access the information. This would allow the user to change the access control list or render data unusable, even if such data left the PDC, or render the information unreadable if it was disclosed to an unauthorized third party.

[56] See http://www.sans.edu/research/leadership-laboratory/article/ip-digital-rights for more information.

policy cannot be enforced by technical means, relying instead on the willingness of the third party to respect the original terms and conditions. The lack of such technical measures means that no identified PDC adequately implements the possibility of establishing "sticky policies" or "link contracts" (although it should be noted that such functionalities are not unknown to the developers interviewed during the course of this study).

Another aspect of security examined in this report, *encryption*, is a common feature in the majority of identified PDCs. Vendors that advertise the existence of encryption measures highlighted that personal data is stored and encrypted in such a way as to prevent access from the PDC team itself. However preventing the access from the PDC team using encryption is difficult to assess and should be trusted only if client side encryption (with keys generated and managed by the user) is adequately implemented.

Based on the analysis of identified PDCs, it appears that encryption is mostly used at link level and server side and that client side encryption (especially end-to-end encryption) is not employed. Such an approach mitigates the risk against a network interception and some types of data leakages (e.g., stolen server disk) but does not protect against unauthorised access by the service provider.

The implementation of *server side encryption* also proved to be very variable: a number of PDCs implemented it for database or backups with a key directly controlled by the service provider while others required the user to enter an encryption key in addition to their password. However as the entire encryption process takes place on the server side and little information on the technical implementation is available, it is difficult to assess the security added by these means.

On the topic of *link level encryption*, the majority of the identified PDCs rely on the TLS protocol for user-to-PDC link encryption.[57] The configuration of TLS in identified PDCs was very heterogeneous, ranging from very well hardened implementation for some PDCs to installation with security flaws for others (e.g. certificate errors). Moreover, little information was available on the link level encryption technology in use when sharing data with third party. Depending on the third party, TLS, Internet Protocol Security or "IPSEC" (used to set up a virtual private network),[58] or application level encryption can be used. Based on the review of the identified PDCs, no general trend could be identified and further analysis has to be conducted.

Lastly, no identified PDC implemented *layered encryption*, whereby different encryption keys can be used for different categories of data.

It is also worthy of note that the identified PDCs did not deploy advanced privacy-preserving computations, particularly in the field of encrypted search (such as homomorphic encryption or ORAM). This can be ascribed, however, to the fact that some of these techniques are still under development.

3.2.8 Traceability

The level of traceability is heterogeneous across the various PDCs and no general trend could be identified in the course of this study.

While most of the identified PDCs kept logs of all performed actions (user, sharing and administrators actions) and even encrypted them, few of them implemented the tools to adequately process the information contained therein, such as logs centralisation procedures and correlation. Advanced traceability options were

[57] For a guide on Transport Layer Security protocols, see https://www.sans.org/reading-room/whitepapers/protocols/ssl-tls-beginners-guide-1029

[58] See http://www.it.cornell.edu/services/managed_servers/howto/ipsec.cfm for more information on the Internet Protocol Security (IPsec).

found in a small number of PDCs, but hinged upon specific implementation of the platform.[59] Generally, PDC developers and vendors adhered to the security specifications present in certifications such as ISO 27001.

The use of administration bastions with video session recording features (to be able to review administration actions), “4-Eye control” (whereby sensitive operations must be carried out by two administrators who monitor each other), and the level of log review was not implemented in any of the solutions.

[59] For example, OpenPDS enables the centralisation of logs and the implementation of mechanisms to check that logs remain untampered. However, the existence of such features depends on each specific implementation of the tool.

4. Security and Privacy Challenges

Having reviewed the main characteristics of PDCs and their level of implementation, this section addresses the privacy and security challenges that emerged from this exercise. In some cases, the identified challenges stem from an imperfect or inexistent implementation of certain privacy enhancing technologies; in others, they may stem from the lack of implementation of recommended practices, or even from lacklustre user adoption. Throughout the present section, mitigation strategies and techniques will be proposed to address these challenges.

This chapter concludes with an exploratory reference to the interplay between PDCs and a possible data producer's right, which is closely related to the increasing output of information from instruments and devices commonly aggregated under the Internet of Things.

4.1 Privacy Challenges

The privacy challenges of PDCs are clearly associated with their inherent role as privacy enhancing technologies. Indeed, as such, PDCs should offer the highest level of protection, providing at the same time the necessary features to enhance transparency and control for end users. To this end, the main privacy challenges identified in the course of this study are described in the next paragraphs.

4.1.1 Data quality

The proliferation of sources of personal data in PDCs may cause tensions with the principle of data accuracy. According to the GDPR, personal data must be *"accurate and, where necessary, kept up to date, and every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed"*, are erased or rectified without delay.[60]

Accuracy of personal data being a general legal obligation for controllers, the ability of a PDC to accept inputs from third-party devices should not be conceived as a mere afterthought, both in terms of how it allows a PDC to fulfil its function and the requirements that it implies. For example, the choice to accept only user inputs may have repercussions on the level of the accuracy of the data. On the other hand, automatic data-gathering functionalities should not be employed by PDCs as a way to shift the burden of maintaining the accuracy of personal data to other parties.

4.1.2 Level of granularity and implementation of consent

As shown in the state-of-the art review, most of the identified PDCs offer the possibility for users to define their privacy preferences, at least to a certain degree. However, granularity of users' control (e.g. with regard to specific types of data or service providers requesting access to data) is in general absent. This is a significant privacy challenge for the PDCs, as on the one hand it restricts users' flexibility in expressing their choices, whereas on the other it makes the expression of privacy preferences dependent only on 'traditional' binary consent mechanisms. The lack of flexibility can hinder the practical use of PDCs when different parties need to have different levels of access to the users' personal data. At the same time, reliance on 'allow/deny' consent systems only adds to the problem of "consent-fatigue", where users are so accustomed to box-ticking privacy policies that the practice finally loses its significance [46].

[60] See Article 5(1)(d) of the GDPR for the principle of accuracy of data.

This fundamental challenge is, thus, closely linked to the matter of usability, in the sense that systems depending on the user setting a large number of options or asking too many questions (to obtain consent) may not be adopted by the general public. Indeed, in the course of our study, some developers concluded that, according to their view of the market, personal data management was too onerous and time-consuming for users, who had to dedicate large amounts of time to build up a PDC before it became a viable tool. This challenge reflects a tension between granularity and usability, or in other words, it shows that usability needs to be a significant concern in the design of granular consent systems[61]. Of course, this will greatly depend on the implementation of the feature, which could go beyond current consent practices, exploring practical ways for users to express their choices and preferences. The results of the state of the art review, however, show that all the tools today tend to rely on consent and privacy agreements and, thus, more research is still needed in this field.

In the mHealth sector, granularity presents an additional challenge. Asking individuals to define which data to provide and allow access to, may not be a viable solution, in part because users may have no clear notion of what data to upload in relation to certain purposes. The case of a medical practitioner who needs access to a wide range of information about the patient/user, and not just the information that the user deems relevant, was cited as an example to illustrate this issue. At the same time, however, granularity is of upmost importance in mHealth, as a way to differentiate between different types of health data (that might have different criticality and, thus, require different levels of protection).

4.1.3 Technical enforcement of privacy preferences

Another significant privacy challenge of PDCs that arises from the state-of-the-art review is the lack of existing practical implementations for the technical enforcement of privacy preferences. This challenge on one hand poses the operation of the whole system on user's trust (e.g. trust on a particular PDC provider), whereas on the other hand it contradicts the very scope of the PDC, i.e. to ensure that users are in control of their data. Indeed, the lack of technical enforcement tools, makes it difficult to guarantee that the users' privacy choices are respected, especially when personal data are transferred to third parties.

Despite the low level of implementation, there are available solutions today (e.g. in the area of Digital Rights Management that was already mentioned) that could be further explored by PDC providers to offer technical enforcement of privacy preferences. Having said that, it is important to note that such solutions do face challenges today, as they are complex to implement and no global interoperability standard is available.

4.1.4 Privacy by Design and by Default

As noted in the State-of-the-art review, there are few examples of PDCs today implementing privacy by design and default in a thorough and consolidated way. It is, thus, of particular importance that PDCs do embed privacy enhancing technologies at different levels, not only as a way of enhancing user privacy, but also as a way of supporting PDCs role in the area of PETs.

To this end, several areas of improvement can be proposed, e.g. with regard to the use of encryption, anonymization, data minimisation, as well as secure data deletion mechanisms. Some of these areas are further explored under the Security challenges below.

4.1.5 Data Portability

[61] For example, instead of requesting granular input for every set of data, broad attitude statements could be elicited – albeit without jettisoning the possibility of granularity.

As shown in the PDCs state-of-the-art review, true data portability is not yet a reality in the current PDC market (as in all markets related to the processing of personal data). The possibility of transferring data between PDCs would benefit considerably from the adoption of generalised standards for data exports. Such standards would need to encompass not just the format of the data but also its structure. In the area of PDCs standards for consent management and data sharing are of particular importance. On the basis of such standards, migration tools could subsequently be offered by PDC providers.

Regarding data export, even if several PDCs allow an export to an open standard (XML, JSON, CSV, etc.), the underlying organization of the exported data is not standardized. The main consequence is that the user is able to export his/her data but injecting it to another PDC is a complex task if no import tool was specially built for this specific case (which is rarely the case).

It is, thus, obvious that further research in this field is needed to particularly address the export/import requirements for different sets of data. Furthermore, incentive on the implementation of such standards can be encouraged, e.g. by creating a specific label for compliant solutions or by enforcing it legally for some sensitive fields such as health-related personal data.

4.2 Security Challenges

The level of security of the studied PDCs proved to be very heterogeneous, some presenting a rather robust level and others failing to implement even basic security checks (such as validating user input in order to avoid the injection of malicious inputs). No general trend could be identified at this stage considering the diversity of the PDC solutions under review in this study. Therefore, this aspect combined with the fact that there has not been yet any uniform effort to standardise security controls and measures that can be embedded in PDCs to enhance their security should be flagged as a challenge for all PDCs in the future. This becomes particularly important when PDCs are used for the storage and management of sensitive data as in the mHealth sector. The most salient security challenges identified during this study are discussed in the next paragraphs.

4.2.1 Authentication

The support of strong authentication is not pervasive among PDCs and in some cases it relies on a third party authentication provider. Considering the sensitivity of the information which may be stored in a PDC and the risk of account hijacking linked to weak passwords or password reuse, integrating strong authentication is of high importance.

Strong authentication should be implemented taking in consideration usability in order to bolster adoption. Depending of the level of security, this can range from one-time passwords to smart card based authentication.

In the case of general public usage, an approach implementing a second factor authentication can provide an acceptable security to usability ratio. The second factor is based on an one-time-password (e.g. SMS based or application based) but the user could be given the ability to identify trusted third parties for which the second factor would not need to be re-entered every time.

4.2.2 Access Control

As shown in the state-of-the-art review, the use of Access Controls Lists (ACL) seems to be widely employed for user access control within the PDCs identified. However, relying only on ACL is not sufficient.

In particular, a key challenge for access control mechanisms is reducing the impact of misconfiguration (such as misconfigurations of Role Base Access Controls (RBAC) policy rules or security vulnerabilities (such as a SQL injection, which may allow an attacker to dump the entire database and bypass logical level ACL). Design and implementation review can mitigate the risk but as long as the ACL system relies on a single layer of defence, the data can be exposed in case of a misconfiguration or security vulnerability. Such impact can be reduced by implementing several layers of independent security mechanisms, ranging from encryption to monitoring systems. For example, encryption can be used to reinforce existing access control mechanisms with an additional layer of protection. Different keys can be used to encrypt the user data (each user having his/her own encryption key). In case of a flaw in the access control mechanism (considering that the encryption and key management part is not affected) even if data are disclosed to an unauthorised party, such information will be encrypted with a key unknown to that party and thus unusable.

Another challenge in access control are business logic flaws due to, for example, weak enforcement of business logic, such as workflows or insufficient parameters validation (e.g. role-IDs, user-IDs). In general, these flaws cannot be assessed in automatic ways. In such cases penetration testing may be performed in order to test the workflows to identify whether one can bypass the supposed access controls that would protect functions or features. Depending on the sensitivity of data stored in and processed through the PDC, one layer (e.g. ACL) or more (combined) solutions may be envisaged for implementation.

4.2.3 Encryption

The state-of-the-art review of PDCs showed that the key security challenge regarding encryption is not its use per se (since encryption of some kind is already widely used) but whether it can be implemented in such a way as to be both secure and user-friendly. The challenge is compounded by the fact that encryption mechanisms that are poorly designed or implemented provide a false sense of security to the end user. They may also induce a higher privacy risk than other security countermeasures if PDCs tend to rely mostly on encryption as the first line of defence against unauthorised disclosure of personal information.

To this end, the lack of implementation of client side encryption is an additional security challenge, as this type of encryption is the only way to provide the user with true control over his/her data, while mitigating the risk of an unauthorised or unwanted assess by third parties (such as a rogue administrator or government mass surveillance programs[62]).

Having said that, it is also important to note that client-side encryption might place a practical burden on the users of the PDCs. In the course of our study, some PDC developers argued that forcing users to install certificates or programs enabling this kind of encryption could engender a lack of trust on their part or simple inaction. This may be seen as a reasonable argument given the often misunderstood nature and limits of cryptography. In general, it was also argued that users do not readily accept the need to change their system's configuration in order to allow for this kind of encryption, reinforcing the idea that privacy by design may be the preferred path towards the eventual ubiquity of this functionality.

Moreover, it should be mentioned that, while cryptography is an efficient way to protect user's privacy, it may also raise security concerns when it serves to impede police investigations or intelligence activities.[63] Some of the questions that merit further reflection on this topic are for example whether client-side encryption will be adopted in general, or whether privacy-enhancing technologies may be weakened by government backdoors or forbidden in some countries. Another key challenge in this respect is the risk of alteration of the cryptographic software as publishing new versions remains under the control of the editor (which can

[62] See for example some information on the PRISM surveillance program, https://www.washingtonpost.com/news/wonk/wp/2013/06/12/heres-everything-we-know-about-prism-to-date/

[63] See for example the recent San Bernardino shooting investigation which brought a lot of public attention to the issue of encryption and privacy: https://www.schneier.com/blog/archives/2014/10/more_crypto_war.html

backdoor or weaken the tool). There is no silver bullet for this issue, but publishing the code and having regular reviews by independent third parties can reduce the risk.

As a general recommendation for PDCs, encryption is a strong security mechanism that should be applied at all possible levels (link encryption, server side encryption, client side encryption). If possible, layered encryption (i.e. different types of encryption depending on the level of classification of the data) should be used, providing for a practical and secure implementation of granular access control, as well as technical enforcement of privacy preferences. A theoretical example of such implementation for mHealth is provided in the next paragraph. Having said that, it is important to note that the administrative burden of such a scheme is currently increasingly high and, thus, practical implementation is quite limited in practice. In all cases, even for companies employing experts in cryptographic implementation and design, it is strongly recommended to use mainstream crypto implementation reviewed by peers[64] and not rely on in-house cryptographic systems. It is also recommended to submit the design and implementation to independent security experts and consult available documentation on the matter by independent bodies (see for example relevant ENISA's work in [47].

4.2.3.1 Cryptographically based Granular Access Control: an mHealth example

An example/scenario of a mHealth-related cryptographically granular access control system could be as follows.

If we consider a PDC that contains a medical file with information on the age, blood type and detailed medical history, the user can decide that:

- Age is not sensible data and could be openly shared with any third party;
- Blood type information can be shared only with the patient's healthcare service provider;
- The full detailed medical history can be shared only with the specific medical practitioners within the patient's healthcare service provider.

Once these preferences have been set, the user can configure consent on his/her PDC, which will:

- Configure the ACL on the server side of the PDC;
- Encrypt each data set with the keys of the authorised third parties, the encryption taking place on the client side.

Consequently, in order to access the information, two layers of controls have to be bypassed: the software ACL on the server and the encryption mechanism which is controlled client side. If the server is breached or if a malicious administrator wants to access the information, sensitive information is encrypted and thus rendered unusable without the proper key (which is only available on the client end point). Likewise, if the consent sharing mechanism erroneously shares the medical history to the healthcare service provider's secretariat, the latter will not be able to read the data as it has a different decryption key (the one that was generated so it could access a different type of data, e.g. information about the blood type).

[64] For example GnuPG, NACL, keyczar, etc (open source tools).

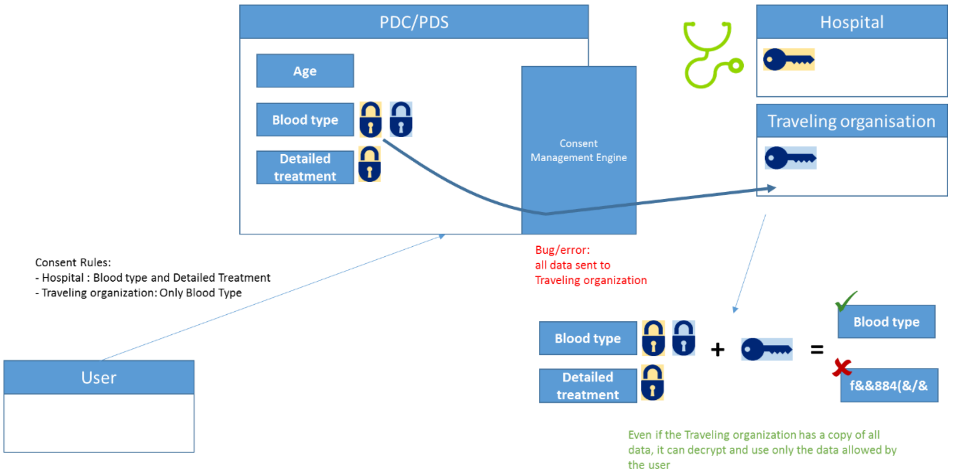
PDC/PDS
Age
Blood type
Detailed treatment
Consent Management Engine
Hospital
Traveling organisation
Bug/error:
all data sent to
Traveling organization
Consent Rules:
- Hospital : Blood type and Detailed Treatment
- Traveling organization: Only Blood Type
User
Blood type
Detailed treatment
+
=
Blood type
f&&884(&/&
Even if the Traveling organization has a copy of all data, it can decrypt and use only the data allowed by the user

Figure 3 – Cryptographically based Granular Access Control in a mHealth Ecosystem

However such an implementation currently presents several issues:

- As all data is encrypted client side, the actions of the PDC provider are limited and thus any required processing on the data cannot be performed. Homomorphic encryption, which allows a third party to perform operations on encrypted data without having to first decrypt it, may address some of these issues but, as noted earlier in the study, the technology is still being developed and is currently impractical since the required computation time for each operation increases exponentially;
- For a global application of such a solution (e.g. with different healthcare service providers), a standard protocol is required, which is not currently in place;
- This system would require a key management system with a sufficient trust level, sufficient scalability and realistic cost constraints, which is currently not available. Solutions such as a hierarchical key system (akin to the one used for website certificate) would, thus, need to be implemented or at least a peer to peer approach should be in place (for instance, the users could flash a QR code with their doctors' identifiers when they wanted to allow them access the data). However, such solutions are not always easy to set up in practice, again taking into account that global implementation among different service providers would need to be available.

Moreover, In order to allow for availability of data, a backup system for the user's encryption key as well as an escrow system in case of an emergency (for instance accessing the full health record if the patient is unconscious) has to be designed. Creating such systems in a secure way while ensuring usability is again a complex task.

Despite the aforementioned challenges, cryptographically enforced systems as the one described in our example are probably the future towards true granularity, as well as enforcement of privacy preferences in the context of PDCs (and beyond).

4.2.4 Protection against External Hacking and Data Breaches

Due to the valuable personal information stored in the PDCs, one of their biggest security challenges is the risk of compromise by an attacker. Hardening, monitoring and reacting is therefore of paramount importance, especially since an application-level flaw can allow an attacker to bypass several other security mechanisms, even strong authentication or server side encryption.

PDC developers or vendors should perform regular audits and penetration tests. These tests should be performed internally, but also by external companies specialized in information security. At the same time, it can prove difficult to identify a company with proper cybersecurity expertise which is why PDC developers should refer to list of companies accredited by EU governments or contact the corresponding competent national information security authority for advice.[65]

The practices of secure coding, secure design, and other standard measures should be implemented from the outset in a manner similar to how personal data protection should be contemplated by design.

4.2.5 Data Deletion

Since client-side encryption is scarcely used in PDCs, users have no technical guarantees that the deletion of data is effectively performed. In those cases where data is retained by the PDC for certain periods of time, it is recommended that more information be provided to the user, such as the exact retention periods, the mechanisms through which the user can be sure that the data is effectively deleted, and the purpose for which data is retained after the deletion command is given. At the very least, users should be informed when their data is effectively deleted from the PDC.

On the technical side, considering that most PDCs rely on complex cloud-based infrastructures with data replication protocols, often in conjunction with the use of flash media such as Solid State Drives,[66] ensuring a well-implemented data deletion is a complex task:

- Data replication: data is often replicated in multiple systems (such as cloud providers). This implies that a secure erase procedure must apply to all the elements in the chain. This is possible to manage in a non-complex or controlled environment but becomes increasingly difficult as third party cloud providers or services are used;
- Flash media: while a magnetic disk such as a hard-disk can be securely erased by magnetic signal re-impression (rewriting random data on the disk), flash media implements a wear-levelling mechanism which makes secure erase procedures based on rewriting hard to implement effectively[67].

Considering the aforementioned issues, as well as the complexity of implementing a secure erase mechanism in a distributed system (where data is replicated on several systems and/or where multiple services providers are used) and the difficulty to check how this is done in practice, the best approach is to rely on client side encryption to enforce secure erasing.

In the case of client side encryption, as all data is encrypted and decrypted on the client end point the server never has access to the plain text or the key. If the user wants to securely erase his/her data, he/she only has to delete the encryption key which is fully under the user's control, thus rendering the data on the server side unusable (including all possible replicas). Such an approach is efficient only if the cryptographic system is well designed and well implemented (for instance if the keys are not backed up on the server side).

[65] Some countries even provide a public list of accredited companies such as the "PASSI" certification in France (http://www.ssi.gouv.fr/entreprise/qualifications/prestataires-de-services-de-confiance-qualifies/prestataires-daudit-de-la-securite-des-systemes-dinformation-passi-qualifies/) or the "CHECK" certification in the UK (https://www.cesg.gov.uk/scheme/penetration-testing).

[66] Solid State Drives or SSD are based on flash memory, as opposed to the traditional rotating magnetic platter hard drives.

[67] For instance a 100GB flash media real size can be of 120 GB (considering 20% reserved for wear-levelling): 100GB of effective storage and 20GB of spare storage in order to "replace" defective sectors. When a sector is defective, the wear-levelling mechanism remaps a valid sector from the "reserve" and sets the invalid one as invisible to the user. In case of a secure erase procedure based on rewriting, the user will be only able to delete the 100GB visible and any information in the remaining 20GB will not be directly accessible.

If client side encryption is not applicable, server side encryption can nevertheless limit the scope to be tightly controlled to the perimeter where data is encrypted. For instance if data encryption is operated on the front end server under the control of the PDC provider and if the database is hosted on a public cloud, the risk of data leakage due to a poorly implemented erasure is limited to the perimeter where clear text information is processed (the front end). Moreover, the PDC provider does not need to trust the erasure procedures of the database provider.

4.2.6 Traceability

As presented in the state-of-the-art review, although most PDCs has some kind of logging functionality, only few implemented centralisation of logs and correlation practices. Ensuring that logs are sent to a central server and cannot be altered is of paramount importance to reduce the risk of an attacker or a rogue administrator modifying these logs.

Additional controls, such as bastion host with session recording capabilities, as well as 4-Eye control (sensitive operations must be done by two administrators who monitor each other) should also be used whenever possible. More advanced techniques, such as watermarks, can also be explored, e.g. as part of the privacy policies enforcement to different sets of data.

4.3 Data Producer's Right

We have seen above that PDCs can be highly interconnected whilst serving as information management tools. The combination of these characteristics and the exponential growth in output of data generated by connected and smart devices is likely to pose specific challenges related to the way in which ownership of data is considered.

The classic definitions of controller and processor of personal data stemming from the current regulatory framework[68] will continue to be key in determining the rights and obligations of third parties and data subjects in relation to personal data. However, due to the growth of Internet of Things, connected devices and machines, it is likely that these concepts will come under strain when attempting to respond to this new reality. Novel ways of looking at personal and non-personal data will need to be considered, but without jettisoning the fundamental rights that protect the privacy and personal data of users.

One possible way of looking at this new reality is through the introduction of a new data producer's right to data. Unlike traditional models, which seek to protect the rights of the creator and trigger the generation of more data, a data producer's right could be a form of ensuring a fair allocation of the profits generated in big data analytics [48].

While this concept is still under theoretical discussion, it is clear that PDCs, with their highly interconnected nature and their potential as privacy-enhancing tools, are poised to be the natural playground where these new forms of data ownership will be tested and analysed.

[68] See "Regulatory Framework", section 2.3 above.

5. Conclusions and Recommendations

This report seeks to examine and support the role of PDCs as privacy-enhancing technologies, putting technology at the core of the discussion and at the service of personal data protection. Based on the information collected, it would seem that existing PDC solutions still need to be more developed before they can fulfil this role, but that there are already technologies and frameworks capable of playing a role in the regaining of user trust and the drive to regain control over personal information.

All relevant stakeholders are thus encouraged to continue developing the concept of the PDC, particularly with regard to the following conclusions and recommendations, all of which are drawn having in mind the potential role of PDCs as privacy-enhancing technologies

i. PDCs and Information Management Tools

The essence of a PDC is not so much related to the technical features of an application but rather to the extent of control granted to the user while authorizing what data to share, with whom and when. A backdrop layer of privacy-enhancing technologies is thus central to the specificity of PDCs. While true PDCs are not yet in full operation, there are already tools that allow users to store, manage, and personally control their personal data. At their core, these PDCs are information management tools specifically designed to account for the particularities of personal data, and the sensitivity involved in the sharing of the same. On the positive side, an encouraging number of PDC-type solutions have begun to implement privacy-enhancing technologies, such as data minimisation and encryption mechanisms.

The research community and the developers of PDCs should continue to implement privacy-enhancing technologies in these solutions, taking into consideration that comprehensive information management tools can be combined with personal data protection mechanisms. Policy makers and regulators at national and EU levels should promote the use of PDCs as privacy enhancing technologies that can put users in control over their personal data.

ii. Degree of User Control

The study has identified a tension between granularity and usability, partly expressed in the limitations of consent as an informed basis for processing of personal data. For the most part, PDC tools still rely on the traditional and limited consent-based model, fostering binary (“allow/deny”) systems that do not easily allow to manage large quantities of data. On the other hand, full granularity of choice, for each data set, authorised party, and purpose, may engender “consent fatigue” and alienate users.

The issue may be mitigated through dynamic privacy by default policies that set a baseline approach access control, so long as the users retain the ability to subsequently set their own preferences. The role of third party applications and wearable devices also needs to be considered, since these devices and applications are responsible for gathering and updating information without any subsequent intervention or validation by the user. In some cases, where accuracy of data is of paramount importance (such as the medical information) automated collection of data might even be preferable. However, this possibility is dependent on the implementation of connectivity features in PDCs, such as the ability to synchronise data across several devices.

These emerging trends need to be kept in mind when designing systems and offering default privacy settings. Ease of use is thus a major factor in the development and adoption of PDCs, in combination with the proper embedding of privacy enhancing mechanisms.

As a positive note, none of the solutions claimed ownership or any prerogative over the data stored within. This is not to say that the level of clarity, particularly with regard to the terms of conditions of the tools, could not be improved. Engaging with the end users in a clear, transparent manner should be considered as one of the main challenges to overcome in this field, the importance of which is directly tied to the fact that individuals are required to make seemingly informed choices in order to allow the processing of their data.

The research community and the developers of PDCs must take into account the need to offer solutions that combine a robust framework for managing personal preferences and an easy to use interface or mechanism. The European Commission should promote research and development in the field of 'usable privacy', especially in the context of personal information management systems, such as PDCs.

iii. Lack of Technically Enforceable Rights

Another challenge related to control over subsequent processing of personal data lies in the complete absence of rights management mechanisms in the analysed PDCs. This represents a potential hindrance to the widespread adoption of PDCs because users have no way to ensure or enforce (in the absence of legal or contractual arrangements) that third party applications or providers will not process their personal data for other purposes. This also represents a missed opportunity as PDCs have been identified for their potential in allowing for the completion of link contracts.

PDCs should surmount this hurdle in the future by avoiding, by design, access to raw sets of data, and implementing technically enforceable privacy policies. By embedding technically enforceable instructions within the data itself, PDCs could fulfil their goal of empowering data subjects to define the conditions under which their personal information is processed.

As a key element in restoring user trust, PDC developers and the research community should place priority in implementing systems that allow users to enforce their personal choices within the PDC through the use of technical means. The European Commission and Data Protection Authorities should raise awareness to the existence and advantages of such mechanisms, as a means to facilitate the adoption technologies that are still not well understood by the general public.

iv. Lack of Standards

While PDC developers tend to rely on existing and relatively standardised protocols for identity access management and server side encryption, many other aspects of a PDC are not unified under technical standards. This leads to differing levels of technical security and limits the flow of information since it constraints the choice of users when choosing between platforms and prevents them from transferring their information to other competing solutions.

Data portability is one of the areas in which the impact of a lack of standards is most obvious. While a number of PDCs allow the user to download a copy of their data, the true challenge lies in achieving complete data portability between providers. As noted above, the GDPR will empower individuals with the right to request controllers of personal data to transfer entire sets of data to other service providers. However, currently, there are no global technical standards allowing for such a transfer. It will be the responsibility of trusted platforms to make sure that they can interoperate and form a coherent ecosystem with a positive network effect.

The European Commission, Data Protection Authorities and security-focused international bodies should promote the use of standards in the fields of encryption and data management. Standards-setting bodies may play a key role in the development of new technical specifications that will promote interoperability of PDCs with other solutions they have to communicate with, or between PDCs themselves for the implementation of *data portability. The research community and PDC developers should also strive to collectively work on the elaboration of widely-recognised standards, and to implement those, enabling users to transfer their information between different providers.*

v. Limits of Cryptography

There is little doubt that encryption is an efficient way to protect the privacy of users. However, models that depend solely on the implementation of robust cryptography as a means to protect this fundamental right of users are not sufficient. Empowerment of individuals to control their own privacy may have practical consequences on the efficiency of encryption itself. During the course of the interviews held with developers of PDCs, the justification given to account the absence of the superior protection provided by client-side encryption was not technical, but instead related to the perceived unwillingness of users to install additional applications or certificates on their platforms to ensure this type of encryption.

In any case, the limits on cryptography should not serve as a justification for developers of PDCs to abandon research on this field. On the technical side, developers need to be aware that encryption alone cannot protect data inference. Other privacy-preserving computations, such as Oblivious RAM or secure multi-party computation should also be considered as means to enhance the protection of personal data.

PDC developers should not rely only on commonly used cryptographic protocols, but actively roll out more advanced forms of encryption such as client side encryption. The research community should continue developing privacy-preserving computations and relevant key management and infrastructure processes with a view to making them functionally and commercially viable.

vi. Variable Level of Security

An attacker able to exploit a vulnerability in the PDC can successfully compromise it and access or alter private information. As usually PDCs are exposed on the internet, the probability of an attack is very high, taking also into account the value of personal information stored in the PDCs.

It has to be noted that security features, such as two factor authentication or database level encryption can be defeated if the attacker manages to compromise the infrastructure. Thus, secure coding, performing regular security audits and penetration testing as well as defining a secure architecture and implementing data breach detection systems are some example of important measures to setup.

PDC developers should combine robust code writing with standard operating procedures that ensure a high level of security. Data Protection Authorities and security-forced international bodies can provide the incentives to foster active, regular security monitoring of systems and procedures for the processing of personal data.

References

1 Meglena Kuneva, Keynote Speech, Roundtable on Online Data Collection, Targeting and Profiling, Brussels, 31 March 2009 [Online] http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm.

2 Ann Cavoukian, "Personal Data Ecosystem (PDE) – A Privacy by Design Approach to an Individual's Pursuit of Radical Control, Digital Enlightenment Yearbook 2013, IOS Press, 2013.

3 Ann Cavoukian, "Privacy by Design and the Emerging Personal Data Ecosystem", Office of the Information and Privacy Commissioner, Ontario, October 2012 [Online] https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-pde.pdf

4 European Parliament, Committee of Civil Liberties, Justice and Home Affairs, "Report on the US NSA surveillance programme, surveillance bodies in various Member States and their impact on EU citizens' fundamental rights and on transatlantic cooperation in Justice and Home Affairs", 2013 [Online] http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A7-2014-0139+0+DOC+PDF+V0//EN.

5 Neelie Kroes, European Roundtable on the Benefits of Online Advertising for Consumers, Brussels, 17 September 2010, http://europa.eu/rapid/press-release_SPEECH-10-452_en.htm.

6 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A Digital Single Market Strategy for Europe, COM(2015)192, http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015DC0192&from=EN

7 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1.

8 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 [2016] OJ L281.

9 ENISA, "Privacy by design in big data – An overview of privacy enhancing technologies in the era of big data analytics", 2015 [Online]. https://www.enisa.europa.eu/publications/big-data-protection.

10 University of Cambridge Judge Business School, Personal Data Stores, 2015, https://ec.europa.eu/digital-agenda/en/news/study-personal-data-stores-conducted-cambridge-university-judge-business-school.

11 European Data Protection Supervisor, Opinion 7/2015, 19 November 2015, Meeting the challenges of big data, 2015 [Online] https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-11-19_Big_Data_EN.pdf.

12 Sue Poremba, "Can Big Data And Mobile Make Health Care More Effective?" Forbes, 2004 [Online] http://www.forbes.com/sites/emc/2014/01/22/can-big-data-and-mobile-make-health-care-more-effective/#729edd816497

13 Pedro García López, Marc Sánchez Artigas, Cristian Cotes, Guillermo Guerrero, Adrian Moreno, Sergi Toda, "StackSync: Architecturing the Personal Cloud to Be in Sync", 2013 [Online], http://stacksync.org/wp-content/uploads/2013/11/stacksync_full_paper.pdf

14 Idilio Drago, Marco Mellia, Maurizio M. Munafò, Anna Sperotto, Ramin Sadre, and Aiko Pras, "Inside Dropbox: Understanding Personal Cloud Storage Services" [Online] http://annasperotto.org/papers/2012/imc140-drago.pdf

15 Commission Staff Working Document, Report on the Implementation of the Communication 'Unleashing the Potential of Cloud Computing in Europe', accompanying the document Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 'Towards a thriving data-driven economy', COM(2014) 442 final [Online] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014SC0214&from=EN

16 Serge Abiteboul, Benjamin André, Daniel Kaplan, “Manage your digital life in your personal info management system” [Online] http://abiteboul.com/DOCS/14.pims.pdf

17 European Union, Charter of Fundamental Rights of the European Union [2012] OJ C326/02.

18 Article 29 Working Party, "Opinion 1/2010 on the concepts of "controller" and "processor", 2010, WP 169.

19 Article 29 Working Party, "Opinion 03/2013 on purpose limitation", 2013, WP 203.

20 Article 29 Working Party, "Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU", 2014, WP 221.

21 Article 29 Working Party, "Opinion 05/2012 on Cloud Computing”, 2012, WP 196.

22 Article 29 Working Party, "Opinion 02/2013 on apps on smart devices". 2013, WP 202.

23 Jiaqiu Wang and Zhongjie Wang, “A Survey on Personal Data Cloud,” The Scientific World Journal, vol. 2014, Article ID 969150, 2014.

24 World Economic Forum, "Personal Data: The Emergence of a New Asset Class – Opportunities for the Telecommunications Industry, 2011.

25 ENISA, “Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies – Methodology, Pilot Assessment, and Continuity Plan”, 2015 [Online] https://www.enisa.europa.eu/publications/pets.

26 ENISA, “Privacy and Data Protection by Design – from Policy to Engineering”, 2014 [Online]. https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design.

27 32nd International Conference of Data Protection and Privacy Commissioners, “Resolution on Privacy by Design”, 2010 [Online] https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Cooperation/Conference_int/10-10-27_Jerusalem_Resolutionon_PrivacybyDesign_EN.pdf

28 Article 29 Working Party, "Opinion 15/2011 on the definition of consent", 2011, WP 187.

29 Bart Schermer, "Your Consent is Overrated", 2013 [Online] http://leidenlawblog.nl/articles/your-consent-is-overrated

30 Victoria Horden, "Consent - The Silver Bullet?", Privacy & Data Protection, Volume 13, Issue 3, 2013.

31 ENISA, “Recommended Cryptographic Measures – Securing Personal Data”, 2013 [Online] https://www.enisa.europa.eu/publications/recommended-cryptographic-measures-securing-personal-data/at_download/fullReport

32 ENISA, “Privacy in the Digital Age of Encryption and Anonymity Online - Speech by ENISA Executive Director, Prof. Dr. Udo Helmbrecht [Online] https://www.enisa.europa.eu/publications/ed-speeches/privacy-in-the-digital-age-of-encryption-and-anonymity-online

33 M.S. Islam, M. Kuzu, and M. Kantarcioglu, “Access Pattern Disclosure on Searchable Encryption: Ramification, Attack and Mitigation”, in Network and Distributed System Security Symposium (NDSS), 2012.

34 Drummond Reed, "XDI Contracts: An Overview", 2015 [Online] https://github.com/WebOfTrustInfo/rebooting-the-web-of-trust/blob/master/topics-and-advance-readings/xdi-link-contracts.md

35 OASIS XRI Data Interchange (XDI) TC [Online] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xdi

36 Siani Pearson, Marco Casasssa Mont, "Sticky Policies: An Approach for Managing Privacy Across Multiple Parties", 2011 [Online] https://documents.epfl.ch/users/a/ay/ayday/www/mini_project/Sticky%20Policies.pdf

37 Siani Pearson, Marco Casassa Mont, Gina Kounga, "Enhancing Accountability in the Cloud via Sticky Policies", in Secure and Trust Computing, Data Management, and Applications – Communications in Computer and Information Science, Springer, Vol 187, pp. 146-155, 2011.

38 Mydex CIC, “The Case for Personal Information Empowerment: The Rise of the Personal Data Store”, 2010 [Online] https://mydex.org/wp-content/uploads/2010/09/The-Case-for-Personal-Information-Empowerment-The-rise-of-the-personal-data-store-A-Mydex-White-paper-September-2010-Final-web.pdf

39 European Network of Excellence in Cryptology, Report on Watermarking Benchmarking and Steganalysis, 2007 [Online] http://www.ecrypt.eu.org/ecrypt1/documents/D.WVL.16.pdf

40 European Commission, Green Paper on mobile Health ('mHealth'), 10 April 2014, COM(2014) 219 final [Online] http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=5147.

41 European Commission, "mHealth in Europe: Preparing the ground – consultation results published", 2015 [Online] https://ec.europa.eu/digital-single-market/en/news/mhealth-europe-preparing-ground-consultation-results-published-today

42 European Commission, "mHealth in Europe: Next steps discussed at eHealth Week in Riga", 2015 [Online] https://ec.europa.eu/digital-agenda/en/news/mhealth-green-paper-next-steps

43 Research2Guidance, "mHealth App Developer Economics 2014 – The State of the Art of mHealth App Publishing" 2014 [Online] http://research2guidance.com/wp-content/uploads/2015/10/mHealth-App-Developer-Economics-2014-Preview.pdf

44 Boston Consulting Group, "The Value of Our Digital Identity", Liberty Global Policy Series, 2012 [Online] http://www.libertyglobal.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf

45 Reihaneh Safavi-Naini, Moti Yung (Eds.), "Digital Rights Management: Technologies, Issues, Challenges and Systems", First International Conference DRMTICS 2005, Springer Science & Business Media, 2006.

46 Article 29 Working Party, "The Future of Privacy: Joint contribution to the Consultation of the European Commission on the legal framework for the fundamental right to protection of personal data", 2009, WP 168.

47 ENISA, "Algorithms, Key Size and Parameters Report 2014" [Online] https://www.enisa.europa.eu/publications/algorithms-key-size-and-parameters-report-2014

48 Herbert Zech, "Information as Property", 2015 [Online] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2731076

ENISA
European Union Agency for Network
and Information Security
Science and Technology Park of Crete (ITE)
Vassilika Vouton, 700 13, Heraklion, Greece

Athens Office
1 Vasilissis Sofias
Marousi 151 24, Athens, Greece

Catalogue Number: TP-07-16-035-EN-N

PO Box 1309, 710 01 Heraklion, Greece
Tel: +30 28 14 40 9710
info@enisa.europa.eu
www.enisa.europa.eu

ISBN: 978-92-9204-182-3
DOI: 10.2824/24216