EUROPEAN UNION AGENCY
FOR CYBERSECURITY

ENGINEERING PERSONAL DATA SHARING

Emerging Use Cases and Technologies

JANUARY 2023

ABOUT ENISA

The European Union Agency for Cybersecurity, ENISA, is the Union's agency dedicated to achieving a high common level of cybersecurity across Europe. Established in 2004 and strengthened by the EU Cybersecurity Act, the European Union Agency for Cybersecurity contributes to EU cyber policy, enhances the trustworthiness of ICT products, services and processes with cybersecurity certification schemes, cooperates with Member States and EU bodies, and helps Europe prepare for the cyber challenges of tomorrow. Through knowledge sharing, capacity building and awareness raising, the Agency works together with its key stakeholders to strengthen trust in the connected economy, to boost resilience of the Union's infrastructure and, ultimately, to keep Europe's society and citizens digitally secure. 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.

CONTRIBUTORS

Isabel Barbera, Claude Castelluccia, Giuseppe D'acquisto, Marta Fydrych Gasowska, Marit Hansen, Jaap-Henk Hoepman, Meiko Jensen, Konstantinos Limniotis, Maria Raphael, Marie-Charlotte Roques Bonnet, Fernando Silva, Fatbardh Veseli, Barbara Vieira, Kim Wuyts, Christian Zimmermann, Luis de Salvador Carrasco, Peter Kraus, Stephanie Mihail, Miguel Peñalba Moldes and Prokopios Drogkaris

EDITORS

Prokopios Drogkaris (ENISA), Monika Adamczyk (ENISA)

ACKNOWLEDGEMENTS

We would like to thank Kristof Verslype for his review and valuable comments.

LEGAL NOTICE

COPYRIGHT NOTICE

ISBN 978-92-9204-602-6, DOI 10.2824/36813

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Over the last twenty years, we have experienced a steady increase in the amount of data being generated and afterwards processed in some manner. Data have evolved from being a scarce resource, difficult to gather and managed in a centralised way to becoming an abundant resource created in a decentralised way easy to replicate and to communicate. There seems to be a natural trend towards 'taking the data out of devices or organisations' and sharing data among different parties to create new value for our society, or simply to reduce operational costs.

Data sharing can be considered as disclosing data to third parties outside the organisation in order to achieve a specific purpose. Such sharing can be performed either as part of a processing operation or while attempting to provide additional utility to existing data. The recent EU legislative initiatives promoting data sharing are sectoral and cross-sectoral instruments that aim to make data available by regulating the reuse of publicly and privately held data, including personal data. They also facilitate data sharing through the creation of novel intermediaries and sharing environments where the involved parties can pool data and facilities in a trusted and secure way.

This report attempts to look closer at specific use cases relating to personal data sharing, primarily in the health sector, and discusses how specific technologies and considerations of implementation can support the meeting of specific data protection. After discussing some challenges in (personal) data sharing, this report demonstrates how to engineer specific technologies and techniques in order to enable privacy preserving data sharing.

More specifically it discusses specific use cases for sharing data in the health sector, with the aim of demonstrating how data protection principles can be met through the proper use of technological solutions relying on advanced cryptographic techniques. Next it discusses data sharing that takes place as part of another process or service, where the data is processed through some secondary channel or entity before reaching its primary recipient. Lastly, it identifies challenges, considerations and possible architectural solutions on intervenability aspects (such as the right to erasure and the right to rectification when sharing data).

1. INTRODUCTION

Data are at the very heart of our daily lives and of our economies. Over the last twenty years, we have experienced a steady increase in the amount of data being generated and afterwards processed in some manner. Data have evolved from being a scarce resource, difficult to gather, managed in a centralised way and costly to store, transmit and process, to becoming an abundant resource created in a decentralised way (by individuals or by sensors) easy to replicate, and to communicate or broadcast on a global scale. This is also manifested by the fact that in the last 20 years the capacity of internet backbone optical lines has grown almost 100 times (from 10 Gbps to almost 1 Tbps), while the cost for transmitting a single Gbps has declined at the same pace, being today about 1% of the cost incurred in early 2000 [1].

There is a natural trend towards 'taking the data out of the devices or organisations' and sharing data among different parties to create new value.

Data is considered as the new currency and organisations are sharing information about their customers with their partners, analytics platforms, public administration and other ecosystem stakeholders in order to take advantage of new technologies or the additional information they can source from sharing and correlation. There seems to be a natural trend towards 'taking the data out of devices or organisations' and sharing data among different parties to create new value for our society, or simply to reduce operational costs[1]. Sharing data is already starting to become the norm and not an exception in data processing. In order to leverage the value of data, service providers need to be able to use data, including data held by others.

Attempting to provide an accurate description of the term, and building on top of the definitions provided by the Data Protection Commission (DPC) in 2019 [2] and ICO in 2020 [3], data sharing can be considered as disclosing data to third parties outside the organisation in order to achieve a specific purpose. Such sharing can be performed either as part of a processing operation or while attempting to provide additional utility to existing data. Data sharing can be performed routinely or in response to specific or emergency situations. According to Gartner [4], data sharing is a business necessity as it can empower digital transformation and innovation.

1.1 RELEVANT EU LEGISLATIVE INITIATIVES

European legislators currently have a huge interest in data sharing. One of the key pillars of the European strategy for data [5] is to make more data available and facilitate data sharing across sectors and EU countries in order to leverage the potential of data for the benefit of European citizens and businesses. Considering only the EU 27 area, the value of data to the economy predicted for 2025 will be €829 billion, up from €301 billion (2.4% of EU GDP) in 2018[2]. Enabling data access and sharing is expected to bring major and concrete benefits in various areas, such as personalised diagnosis and telemedicine, transportations, policymaking and public administration.

The European Data Governance Act [6] foresees mechanisms to increase the availability of data in the public sector and overcome technical obstacles on the reuse of data in the public interest. These mechanisms are supported by a set of concrete measures facilitating data sharing. The measures include the establishment of data intermediaries functioning as trustworthy organisers of data and technologies within the sectoral data spaces and the creation of processing environments (e.g., data rooms), supervised by the public sector. Additional legislative initiatives focus on specific sectors, such as the EU Health Data Space Proposal [7]. For the private sector, the EU Data Act Proposal [8], aims to set the rules for creating new value from the data held by consumers and businesses, clarifying who can access such data and

[1] IDC, *Data Age 2025 - The digitization of the world, from edge to core*, Nov. 2018
[2] European data strategy: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

under what conditions. Through the EU Data Act, data may be made available for sharing between enterprises, citizens and public administrations based on specific measures that aim to increase legal certainty, prevent abuse of contractual imbalances and provide access to data for public sector bodies.

1.2 THE ROLE OF DATA PROTECTION ENGINEERING

Personal data protection is an integral element of the trust individuals and organisations should have in the development of data sharing ecosystems. As highlighted by the joint opinion of the EDPB and the EDPS [9], success will also rely on *the establishment of a strong data governance and effective safeguards for the rights and interests of natural persons that are fully compliant with the GDPR*. This is where data protection engineering has a very important role to play. The legislative initiatives promoting data sharing are sectoral and cross-sectoral instruments that aim to make data available by regulating the reuse of publicly and privately held data, including personal data. They also facilitate data sharing through the creation of novel intermediaries and sharing environments where the involved parties can pool data and facilities in a trusted and secure way.

Data protection engineering, can be a key factor for building a trusted sharing environment with an adequate level of protection.

Data protection engineering, as described in [10], can be a key factor for building a trusted sharing environment, where organisations may submit data without disclosing personal data or sensitive business information or disclosing personal data with an adequate level of protection. This lies within the spirit of the concept of data protection-by-design prescribed in Art. 25 of the GDPR; safeguards must be integrated and engineered into the processing. Data protection engineering offers the possibility to cope with the increasing capabilities of transmission, storage and processing technologies without diminishing their potential for innovation and, at the same time, to mitigate emerging privacy risks for individuals and economic risks for enterprises.

As the European Data Protection Board (EDPB) has pointed out in its guidelines on Data Protection by Design and by default [11] in an increasingly digital world, adherence to data protection-by-design plays a crucial part in promoting privacy and data protection. It is crucial that data holders understand data protection principles and the rights and freedoms of data subjects, and implement appropriate measures and the necessary safeguards to reinforce these principles and to enable the exercise of these rights. Each technical and organisational measure must produce the intended results in the specific context where the processing takes place.

A special focus is therefore needed for identifying the main data protection engineering paradigms in data sharing and for understanding the types of safeguards to be implemented in all possible scenarios. The following chapters of this document will further exemplify these paradigms and the relevant technical safeguards through practical use cases. These use cases focus mainly on the healthcare sector; however, the technologies and techniques presented are also equally applicable to other application domains.

1.3 SCOPE AND OBJECTIVES

This report attempts to look closer at specific use cases relating to personal data sharing, primarily in the health sector, and discusses how specific technologies and considerations of implementation can support the meeting of specific data protection principles. After discussing some challenges in (personal) data sharing, this report demonstrates how to engineer specific technologies and techniques in order to enable data sharing that preserves privacy. This work is meant to support policy makers, regulators and data protection practitioners and is performed in the context of ENISA's tasks under the Cybersecurity Act (CSA)[3] to support Member States on specific cybersecurity aspects of Union policy and law relating to data protection and privacy.

[3] Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act) http://data.europa.eu/eli/reg/2019/881/oj

This work builds upon the Agency's activities in the area of Data Protection Engineering [10] and is produced in collaboration with the ENISA Ad Hoc Working Group on Data Protection Engineering[4].

1.4 STRUCTURE OF THE DOCUMENT

Section 2 discusses specific use cases for sharing data in the health sector, with the aim of showing how data protection principles can be met through the proper use of technological solutions relying on advanced cryptographic techniques. Section 3 discusses data sharing that takes place as part of another process or service, where the data is processed through some secondary channel or entity before reaching its primary recipient. Lastly, Section 4 discusses challenges, considerations and possible architectural solutions on intervenability aspects (such as the right to erasure and the right to rectification when sharing data). Section 5 concludes the document.

[4] https://www.enisa.europa.eu/topics/data-protection/ad-hoc-working-group-on-data-protection-engineering

2. DATA SHARING PRACTICES IN THE HEALTH SECTOR

A field for which data sharing constitutes an opportunity is, undoubtedly, the health sector. Sharing health data can strengthen coordination and collaboration between the public and private healthcare entities towards providing effective personalised health-care assistance and achieving public health goals, as well as towards conducting scientific research (including clinical trials) [12]. Data sharing in the health sector also has cross border dimensions, as identified under the Cross Border Healthcare Directive [13] and currently under the EU Health Data Spaces proposal [7]. However, several personal data protection risks occur which stem from the sensitive nature of health data under GDPR Art. 9[5] and from the fact that ensuring the fulfilment of data protection principles such as transparency and data minimisation necessitates a very thorough assessment and a cautious implementation 'by design' approach [12].

A basic approach to ensure user's transparency is to enable the user to control who will have access to her data, as well as for how long and which part of her data.

Health data include biomedical data, electronic health records (e.g., health records being stored and further processed in a hospital), and data generated by individuals themselves e.g. data from wearable devices [14]. In the context of sharing health data for various purposes, the following properties or requirements [15] need to be efficiently addressed.

- Data for the diagnosis and treatment of individual patients should be identifiable.
- The (same) data for (possibly large-scale) medical research should be appropriately pseudonymised, so as to ensure that re-identification by a researcher is not likely (unless the user provides her explicit consent for un-pseudonymised processing, which should be revokable at any time)[6], as well as the ability to remove the link [16] between two different data sets for different purposes is present.
- The ability to handle multiple sources of patient data, including wearable devices and apps should be present.

It should be highlighted that the necessity for data minimisation spans these three requirements horizontally.

Additional data protection requirements that also need to be in place are transparency to the data subject and data security.

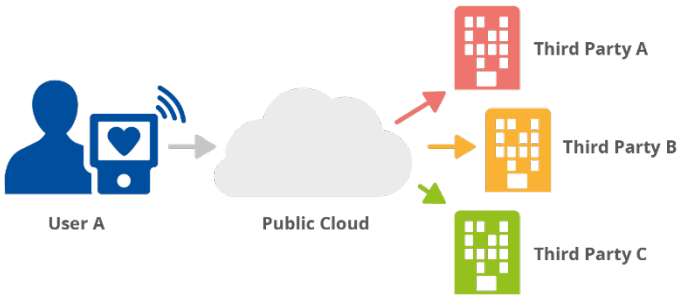
2.1 USER CONTROLLED PERSONAL DATA SHARING

A basic approach to ensure user's transparency is to enable the user to control who will have access to her data, as well as for how long and which part of her data. Hence, in such a user-controlled data sharing approach, the role of the user could be considered as a 'safeguard' towards ensuring the fulfilment of the aforementioned data protection requirements. From the point of view of a legal basis for processing, this approach is a way to **implement data sharing under the user's explicit consent in a way that cannot be overcome**; no entity would be allowed to gain access to the user's health data unless the user explicitly grants access.

[5] According to the GDPR, health data not only lie in the class of the so-called 'special categories' of data (Art. 9), but there is also a margin for the Member States to introduce, in their national legislations, further conditions including limitations (apart from those provisioned in the GDPR) with regard to the processing of these data – thus clearly illustrating the importance of their processing.
[6] From a legal point of view, the legal basis for such a processing is the user's consent.

Let us consider a scenario where User A uses a wearable device for continuous glucose monitoring (CGM), which also monitors blood pressure, caffeine levels and lactate levels7. The device uploads the data streams collected to the cloud for storage and further processing either by the user herself or by other entities, for example her family, doctors etc. as depicted in Figure 1 below. The main challenge, from a data protection point of view, is how the user is able to selectively share specific data streams generated by the device with specific parties.

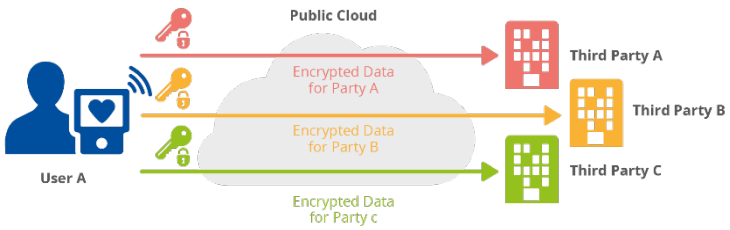
Figure 1: Generic model of user-controlled data sharing



Such an access model may not only be based on the identity of the entity requesting access but also on additional parameters such as the time period when the data was generated. For example a third party might be granted access only to data that correspond to the last three months) and/or to specific parts of the data set (e.g. blood pressure measurements).

A simple approach towards achieving the aforementioned goals is the use of asymmetric encryption. User A encrypts the data with the public key of the relevant recipient and shares the data as presented in Figure 2 below. In other words, each segment of data that is to be read by a third party is being encrypted with A's public key (similarly, if the data are to be accessed by the user herself, they are encrypted with the user's public key).

Figure 2: User-controlled data sharing through asymmetric encryption



This approach, however, has some limitations, mainly in terms of practicality and efficiency. If the same data are to be shared with multiple entities, the user needs to share the same data many times, each encrypted with the relevant entity's public key. This leads inevitably to redundancy, which becomes a predominant issue especially in cases of high volumes of data that are being constantly produced. Furthermore, the possible recipients may not necessarily be known in advance and as a consequence, for each new access that is to be granted, a new encryption would be needed.

[7] Apparently, our use case scenario could be easily adapted to describe the case where the patient uses more than one wearable – each for different purpose (e.g. CGM, holter monitoring etc.)

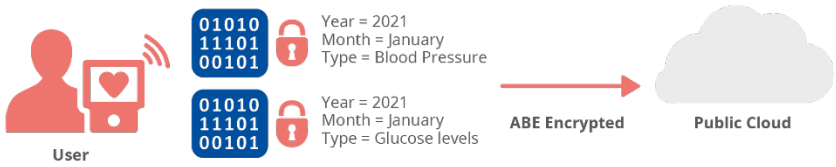
2.1.1 Attribute Based Encryption

A cryptographic technique to overcome the aforementioned limitations is Attribute Based Encryption (ABE)[8], which was first introduced in 2004 under the term Fuzzy Identity-Based encryption [17] & [18]. ABE is a special case of asymmetric encryption, in which data can be encrypted with an ABE public key but, at the same time, contrary to the 'classical' public key encryption, there may be more than one decryption key, each of them bound to small pieces of additional information related to the data, which are called attributes. The decryption keys are actually generated by a generic ABE master secret key, which should remain private.

ABE is a special type of asymmetric encryption, in which data can be encrypted with an ABE public key but there can be more than one decryption keys.

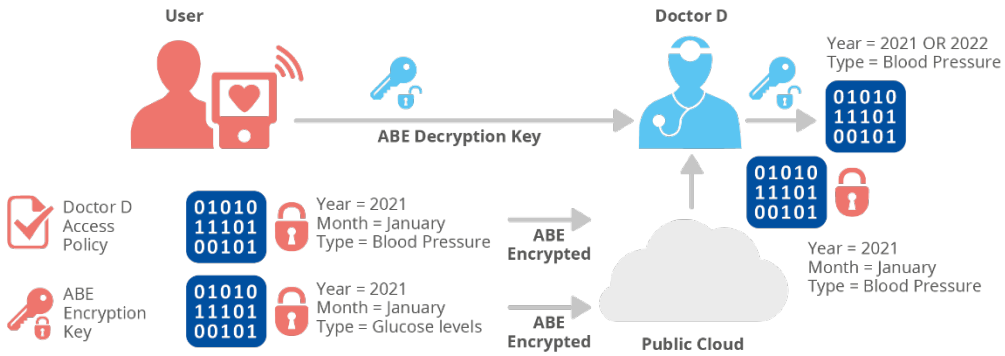
Revisiting the use case discussed earlier, we next describe how ABE can be used to implement user-controlled data sharing though a cloud service. Our scenario relies heavily on the implementation described in [19]. It should be noted however that nowadays, cloud infrastructures and services offer a wide range of possibilities and can perform parts of the processing operations, further to the transmission of the data. Such applications however are outside the scope of the use cases described in this report.

Figure 3: Storing encrypted objects to the cloud



User's data are generated by her device and are being assigned to relevant objects, describing specific attributes related to them such as the date of origination, the object type etc. These attributes will be subsequently used to define the access control mechanism to these data. Data are then encrypted with the ABE master key and are then uploaded to the cloud.

Figure 4: Sharing the ABE decryption key and encrypted data



When a third party requests access to a user's data, User A creates an access policy for that party. This policy specifies which exact properties, based on the attributes already defined, should be satisfied by the data to which she wants to grant access. Then, the user's device 'translates' the policy, for example fileType="bloodpressure" AND year>=2021 AND recipient="Doctor D", into a corresponding ABE decryption key, and sends the key to the party (the cloud provider does not have access to this decryption key). **Once doctor D receives this decryption key, she will be able to decrypt locally only the corresponding data that**

[8] ABE can be considered as a specific case of the so-called functional encryption, which is defined as a specific type of public key encryption that allows decryption keys with the property that they can only decrypt a specific function of the encrypted plaintext (regardless of the content of the plaintext)..

satisfy the access policy defined by the user, since the ABE decryption key can decrypt only a subset of the data set.

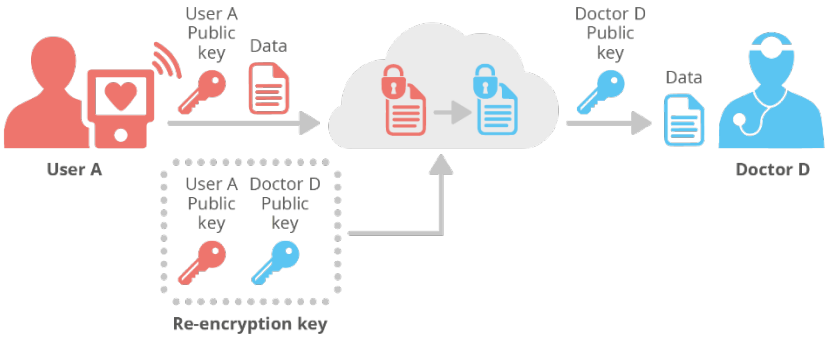
This approach however necessitates 'tagging' (i.e. labelling) of the data at an early stage, which may not be always a straightforward task, either due to the capabilities of the device or the interaction required by the user. Furthermore a suitable selection of attributes with respect to the subsequent access policies may not be always obvious. However, in a user-controlled data sharing model, it is inevitable that the user should be able to make decisions on her data. In addition, the cloud provider still collects information related to the metadata of each communication between the user and each party.

2.1.2 Proxy Re-encryption

Another advanced cryptographic technology that allows for user-controlled data sharing is Proxy Re-Encryption (PRE) [20]. This is a specific type of asymmetric encryption, which **enables the re-encryption of an already encrypted data set from one public key to another,** without the proxy having access to the unencrypted data set. It is considered a very good approach when the entity with whom data will be shared is not known at the time of the initial encryption or sharing is to be performed via untrusted infrastructures.

Elaborating further on the use case described earlier, user A encrypts her data with her own public key and uploads them to the cloud. When a third party, for example Doctor D, requests access to the data, the user generates a so-called re-encryption key, using her own private key and Doctor's D public key. The re-encryption key can be sent to the cloud provider, which is now able to transform the initial encrypted data set into a new encrypted data set corresponding to the encryption through the public key of Doctor D, as presented in Figure 5 below. By these means, **only Doctor D can now decrypt the data using her private key**.

Figure 5: Proxy re-encryption process



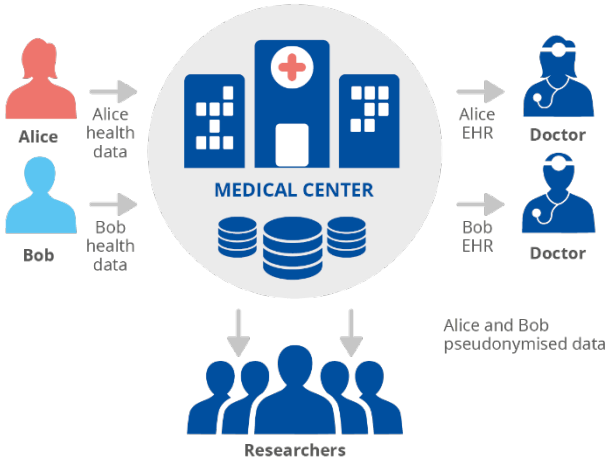
PRE allows for cryptographically-enforced access control in a user-controlled data sharing model. Furthermore, it allows the data owner to delegate access after the data is encrypted, which is important since in a typical sharing scenario it may not always be possible to identify the recipient entities beforehand.

Overall, PRE can be considered as a good solution when sharing is performed via untrusted infrastructures such as a cloud infrastructure. A number of existing applications and commercial patents are already described in [21]. However, the cloud provider still collects information related to the metadata of each communication, similarly to ABE discussed earlier. Furthermore, as the re-encryption can only be performed on the initially encrypted data set, the user must have a good understanding during the initial encryption of the data that would be shared later. It is worth highlighting though that ABE and PRE can be used in a combined way [22], thus leading to a more refined user-controlled sharing of data.

2.2 SHARING HEALTH DATA FOR MEDICAL AND RESEARCH PURPOSES BY HEALTH CARE PROVIDERS

Another typical data sharing scenario in the health sector is the management of Electronic Health Records (EHRs) by healthcare providers. Broadly, EHRs are an electronic version of a patient's medical history which contains all the key medical data relevant to that person's conditions, results of medical examinations, treatments, medications etc. EHRs are usually managed at a central repository at national level and users can authorise access to their data to treating doctors or medical institutions. During the pandemic, the need for large-scale data gathering projects became even more apparent, in an attempt to support not only the treatment of patients but also scientific research and prognosis.

Figure 6: Large scale data gathering example



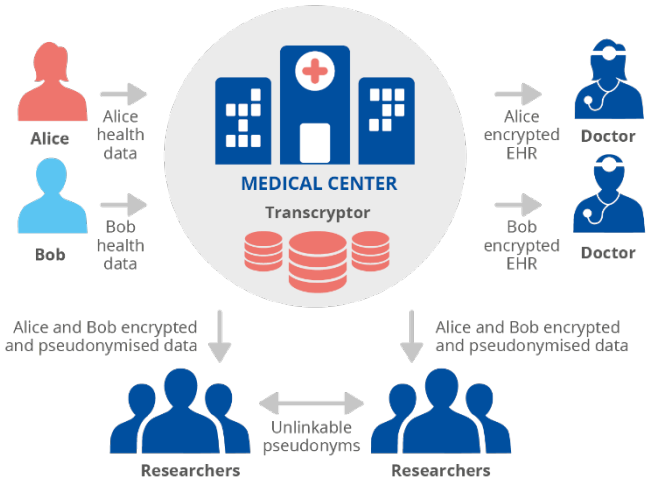
Typically, to address the applicable data protection issues, a medical centre deploys access control mechanisms, applicable for both internal and external users, to determine who will have access to the data and/or encryption of the stored data. **The aim is to ensure that only authorised health service providers will have access to personalised information** (for example doctors who need to have access to a patient's medical history). **When data is to be shared with internal or externa researchers for research purposes, appropriate safeguards should be further deployed which, in a typical scenario, include pseudonymisation in order to avoid disclosure of the identity of the patients to the recipients**, as described in [12] and presented in Figure 6 above.

2.2.1 Polymorphic encryption and pseudonymisation

Building upon the advantages of polymorphic encryption, Polymorphic Encryption and Pseudonymisation (PEP) [15] had been proposed as a means to address the challenges described earlier. The main property of polymorphic encryption is that personal data can be encrypted in such a way that there is no need to fix a priori who can decrypt the data. This can be decided later via transformations of the ciphertext which allow the decryption to be performed through different cryptographic keys. This transformation can be performed in a blind manner without the party performing this, the transcryptor, being able to access the unencrypted data set. Therefore, the encrypted data set is being 'transformed' by the transcryptor into another version so that only this recipient can decrypt [15]. The transcryptor can be either an entity within the organisation (for example the medical centre or the hospital) or outside the organisation, for example a cloud provider. With regards to the pseudonymisation, PEP utilises the transcryptor also as the pseudonymisation entity, as described in [23]. Each individual is assigned different pseudonyms for each third party that requests access to individual's data, thus preventing pseudonym linking across multiple third parties.

Going back to the use case of large-scale data-gathering projects in the health sector discussed earlier, each patient has a unique identifier. **This identifier is transformed by the transcryptor into different pseudonyms depending on the recipient and the context or purpose of sharing the data. Each pseudonym is communicated to each recipient together with the polymorphic encrypted data**. As for each recipient a new pseudonym is being generated, the pseudonyms used for the same patient cannot be linked, thus are considered as unlinkable and preserve the confidentiality of the patient's data. As depicted in Figure 7 below, when the recipient is a doctor, the transcryptor re-encrypts the health data of the relevant patient and transmits them to the related doctor.

Figure 7: Using PEP in large scale data gathering



Through polymorphic encryption data can be encrypted in such a way that there is no need to fix a priori who can decrypt the data.

It should be noted that the processing of pseudonymous data is always subject to re-identification of individuals, either through reversing pseudonyms back to the original identifiers or by re-identifying the individuals through the remaining personal information that becomes available (the so-called quasi-identifiers) as discussed in [23]. Overall, PEP constitutes a technique that is currently considered as an advanced cryptographic technique for data protection engineering and has already demonstrated its applicability in a Large-Scale Parkinson's Disease Study [24] and as a proposal for the Dutch eID scheme [25].

3. DATA SHARING USING THIRD-PARTY SERVICES

Besides the use cases of data sharing, where an entity directly shares data with another entity, which is the final recipient of the data, there are also use cases of secondary, non-direct data sharing which might also require a user to perform specific actions. In this type of data sharing, **the sharing takes place as part of another process or service, where the data is processed through some secondary channel or entity before reaching its primary recipient**. Often, this data sharing is not transparent to the users. Hence, this topic needs to be addressed by privacy engineers, architects and developers. For instance, such data sharing occurs when an application integrates a third-party service, which is also operated by an entity other than the primary recipient or sender of the data [26]. While this is an established software development practice, there are cases when this is harmful for the privacy of the users, e.g. a component or a service being used may share data with a third party – sometimes unknowingly for the system architects and developers.

Examples of secondary data sharing use-cases exist in many dimensions of software engineering. Table 1 gives an overview of a number of such use-cases structured across three main domains and a high-level description of each.

Table 1: Selected use cases of third party data sharing

Domain	Use Case	Description
Integrating third party services	Mobile Push Notifications	Using a third-party service to send push notifications to mobile phone users (apps)
	Authentication and Authorisation	Integrating a third-party authentication and/or authorisation service into an application, e.g. federated identity.
Outsourcing IT operations	Network monitoring	Whenever network monitoring is performed, especially by an outsourced company, an implicit data sharing happens.
	Data sharing between on-premises and cloud environment	When companies implement a hybrid cloud approach, data sharing between the on-premises and the cloud environment takes place.
Optimising Threat Preparedness	Sharing Threat Intelligence Information	Collaborative efforts in sharing timely and adequate information on emerging threats within a predefined community.

Within the scope of this study, we focus on the first two use cases, namely mobile push notifications and authentication, which will be discussed in more detail.

3.1 MOBILE PUSH NOTIFICATIONS

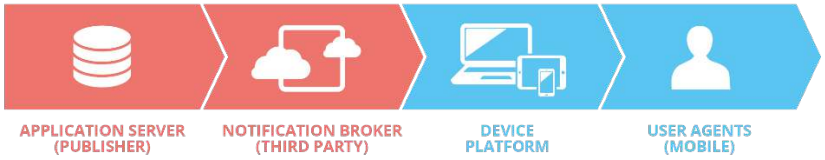
Mobile push notifications are an important communication channel between mobile applications and their users. Using push messages, a mobile application provider may instantly send a message to its users. Push notifications may provide timely information from the application provider to the user(s) and potentially prompt a reaction by the latter.

Mobile push notifications may transmit different types of content, such as text, pictures, external and in-app links etc. Further to marketing purposes, push notifications may also be used to trigger a reaction by the user(s), or to signal users to provide certain input, to proceed with a certain process or to provide a functionality. Hence, the timeliness of the push notification is often important and is usually their main advantage compared to other communication channels such as emails or text messages. Mobile push messages may be sent in bulk or individually (personalised).

In the case of personalised notifications, the information transmitted may very well include personal data or pseudonyms such as user application identifiers. Hence, additional technical and organisational measures need to be integrated to address threats to privacy as described below. Currently, one of the most widely used platforms for push notifications is Firebase Cloud Messaging (FCM) cross-platform messaging[9] which supports both major mobile OSs (iOS and Android).

From an engineering perspective, developers usually integrate code provided by third parties in their application. Although it may not be transparent to the users, the infrastructure of mobile push notifications involves at least two additional entities, which are central to their architecture in the mobile world. Hence, we have the following entities, which are depicted in Figure 8 and described below.

Figure 8: Main actors involved in mobile push notifications



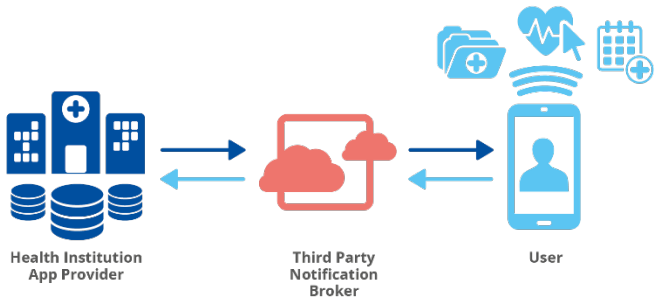
- **Application Server (Publishers):** typical application server, which sends a notification message to a mobile app user;
- **Notification Broker (Third Party):** a third party providing brokerage of push notifications to end users (user agents);
- **Device Platform / OS:** the contact from the notification broker to the application needs to run through a dedicated API provided by the Operating System of the device, such as Android, iOS, etc; the Device Platform / OS is contacted regardless of the involvement of the notification broker;
- **User Agent (Mobile):** this is usually represented by a mobile app, which is running on the corresponding device platform and represents the software interface to the mobile user.

Let us consider the following example of mobile push notifications in the health domain. A health institution offers a mobile app, which enables interaction of the user with the physicians, the institution facilities such as labs and front desk and personalised services for the user. User can receive copies of bloodwork test results, X-rays, MRIs and prescriptions in the application. She can also book appointments with physicians and receive notifications for forthcoming appointments, reminders to book an annual check-up and even daily reminders on receiving the medication she has been prescribed.

The mobile application offers push notifications for all the aforementioned services. The user is reminded of upcoming appointments, annual exams, medication received and lab results as soon as they are available and can provide replies or information back to the health institution by interacting with the push notification.

[9] Firebase Cloud Messaging (FCM): https://firebase.google.com/docs/cloud-messaging

Figure 9: Entities involved in e-Health mobile push notifications scenario



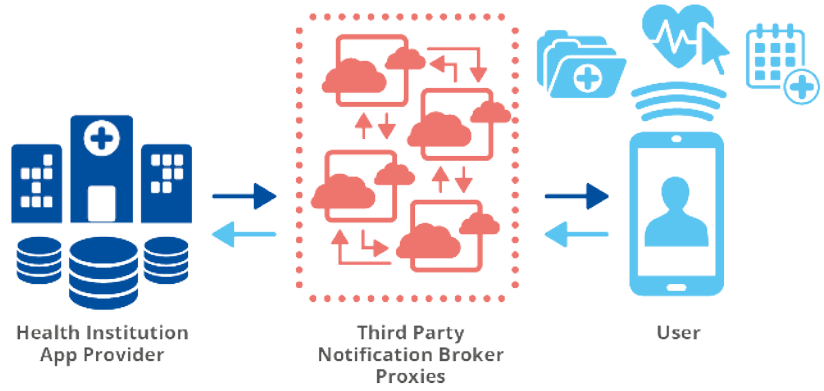
The third party operating the notification brokerage service receives notification data from the application server and delivers these to the user agent client. Hence, indirect data sharing with a third party takes place, which poses privacy threats to the user, including:

- **Linkability**: observation of the interaction between the two entities, including frequency of interaction, types of messages exchanged, etc.;
- **Identifiability**: messages that identify the user;
- **Disclosure**: the content of the messages being pushed may be disclosed, thus violating the confidentiality of the notification, since in many cases these messages are routed through the third party in clear.

3.1.1 Anonymous Notification Protocols (Using Proxies)

An approach to address the privacy threats above could be done by using private notification protocols, which enable the delivery of notification messages by using multiple anonymisation layers or proxies, as described also in [10]. This approach requires the use of a chain of proxies, through which notification messages could be mixed before reaching the end user. Such protocols enable the sending of mobile push notification services to users without the broker nodes knowing neither the original sender nor the final recipient.

Figure 10: Proxies architecture overview



Building on the example described earlier, a notification message in this approach would be sent through a chain of proxies. The application provider would prepare the message and choose a random selection of intermediate nodes (proxies), through which the message is routed. In this case, there exists not one but a mixture of notification brokerage servers. Each of them only knows the address to the next brokerage server, hiding the information about the two end entities (application server and user). Furthermore, the communication between the notification servers (proxies) itself is encrypted with their corresponding public keys, thus protecting from unintended disclosure. An example of such a protocol is AnNotify [27] which

supports the unlinking of the notification between the subscriber and publisher, the unlinking of push notifications to a subscriber and broadcast privacy, hiding whether a subscriber is subscribed to a notification or not.

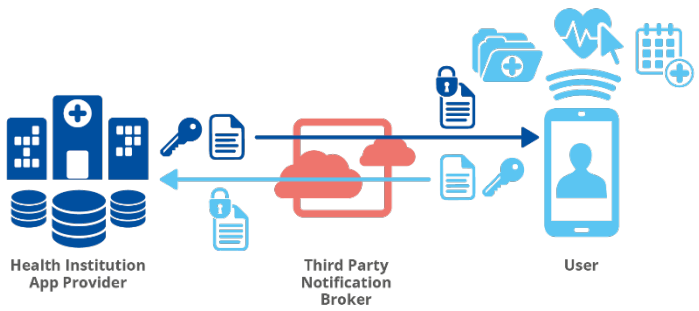
Regardless of the chosen protocol, the potential implications for both the development and the operation of the service using that protocol need to be considered, in order to provide practical viability and to be adopted by the developers. In this regard, several different criteria can be used to evaluate different options, including the ease of integration in the application, ease of maintenance, scalability, etc.

3.1.2 End-to-End Encryption

Encryption of the notification messages is the most straightforward measure to address at least some of the privacy threats mentioned above (most notably that of disclosure). Currently, the delivery of push notifications is typically not performed using end-to-end encryption. Either no encryption at all is used or only part of the communication is encrypted. In some cases, the developer(s) may decide to deliver the notification message between the application server and the notification broker using encryption at the transport layer (e.g. TLS), which provides encryption between these two entities. In addition, a notification broker may choose to encrypt the communication to the user (device). However, encryption is partial and does not address fully the threat of disclosure through each of these entities. The notification broker decrypts the message before sending it to the user. A more adequate solution for this would be the implementation of end-to-end encryption at least.

The details of this approach are depicted in Figure 11 below. The application provider uses the public key of the user and encrypts the content of the message that is being pushed. While the application server still uses the notification broker to send the message, it cannot decrypt the message. It merely delivers the notification messages without being able to read the pushed message. Next, upon receiving the notification message on her phone, the user may use her private key to decrypt it and read the pushed message.

Figure 11: End-to-end Encryption overview



Using this approach, the notification broker does not have access to the content of the notification. Nevertheless, it would still be able to observe the metadata and learn about these interactions between the two other entities. The Capillary Project[10], for instance, aims at addressing this challenge by providing end-to-end encryption of notification messages, although it is limited to Java applications on Android platforms.

3.1.3 Design Strategies

Besides the technical options described earlier, decisions on the design of the architecture that integrate privacy-by-design could be used to address some of the threats discussed earlier. One such decision is, for instance, following **a two-step strategy**. In the first step, the

[10] https://github.com/google/capillary

architecture should require that the use of using push notifications in the application should be limited. This can be achieved, for instance, by using a 'pull by default' strategy, whereby the functionality of push notifications is limited to a minimum and only used to transfer non-personal data. It suffices in such cases to notify the user that an update exists or an interaction is required. Then, in the next step, the notification message, which includes personalised content, can be directly fetched from the app through a direct communication with the application server.

Using the use-case above, the eHealth app could notify the user that a certain action is required or that there is a new message for the user, but without displaying the message itself. The user can then react proactively by clicking in the app. However, it is clear that this is complementary to the other measures presented above, since this may reduce, for example, the threat of disclosure to third parties, but not completely avoid it. Furthermore, this approach is challenging as it requires the full scope of the notification features in an application to be covered. Therefore, its use is recommended as a complementary but not as a primary measure, or where other measures are not feasible.

3.2 DATA SHARING DURING AUTHENTICATION

Authentication is a key building block for many web applications. However, in some cases it is sufficient to get evidence that a data subject meets specific criteria or has a specific property rather than revealing all the properties of her identify and then computing whether these properties are being met or not.

Privacy-enhanced ABCs allow the authentication of an entity by selectively disclosing and authenticating specific attributes, without revealing additional identity information.

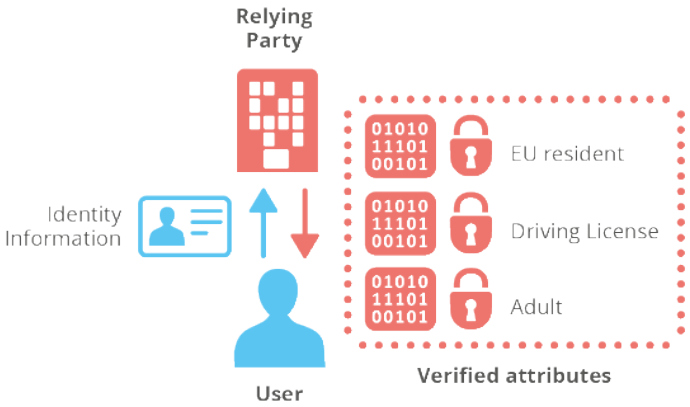
Age verification has been one of the methods used to ensure that minors are protected from possible harm in the physical world; traditional ID verification at shop checkouts is a good example. Nowadays several video-sharing platforms require age verification pursuant to the provisions of the Audio-visual Media Services Directive [28].This verification is usually performed through a self-declaration of the date of birth without any further validation. Apart from the ease with which this control may be circumvented, there is also the issue of processing more data than necessary to accomplish the specific purpose (principle of data minimisation).

Attribute Based Credentials (ABC) and more specifically Privacy-enhanced ABCs allow the authentication of an entity by selectively disclosing and authenticating specific attributes, without revealing additional identity information that is typically used and includes personal data [10]. User submits her attributes to a third party that verifies their accuracy and acts afterwards as the trusted third party. This approach has already demonstrated its potential and applicability in various scenarios under the Attribute-based Credentials for Trust (ABC4Trust) research project [29] and IRMA under the Future ID project [30].

Let us consider an online service for car rental within the EU that offers a drop off service. During the online reservation process, the user is asked to present information on whether she is over 18 years old and whether she holds a valid driver's licence. Normally, this information will be verified upon pick up of the rented vehicle and signature of the rental agreement but, with the drop off service, the user can electronically sign the lease agreement. Instead of providing evidence via mail of her age and valid driver's licence, the provider can check whether these criteria are being met by validating specific attributes of the user’s identity.

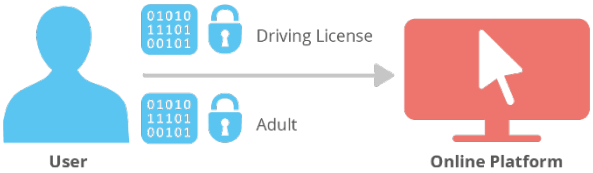
The user submits her attributes to the relying party which validates them and digitally signs them. Now the user can communicate them to third parties who can validate their authenticity and integrity via the digital signature that accompanies them as presented below in Figure 12.

Figure 12: Creation of Identity Attributes



When the user visits the website of the car rental online service, she is asked to provide the relevant attributes to prove that she is over 18 and holds a valid driver's license. The user then presents the relevant attributes to the online platform, either via her smart card or through a relevant browser extension. The platform then validates them and continues with the car rental booking process.

Figure 13: Attribute Based Authentication



In this data sharing scenario, **the user does not reveal the actual personal data but rather an attribute which is an attestation by a third party that a specific property is (or is not) met**.

3.2.1 Relevance of attribute based access to online platforms

In 2022, the French data protection authority (CNIL) highlighted in [31] that it is not possible to aim for the absolute efficiency of age control online, especially for minors. Given the increasing requirements for age verification of minors in online services, CNIL recommends the use of a trusted independent third party to promote data minimisation and unnecessary collection of personal data from service providers. CNIL provides as overview of six possible existing age verification solutions but concludes that none of them meets all the required properties in terms of reliability, data protection and security. In addition to this analysis, CNIL's Digital Innovation Laboratory (LINC) has developed a privacy-friendly age verification system[11] based on zero knowledge proof.

CNIL's recommendation for a trusted third-party and **attribute based authentication, which relates closely to the concept of zero knowledge proof, seems a rather promising concept** for the Digital Service Act Regulation [32] Art. 24 (b) (1b) provisions on online protection of minors. According to this provision, *providers of online platforms shall not present advertising on their interface based on profiling within the meaning of Article 4, point (4), of Regulation (EU) 2016/679 using personal data of the recipient of the service when they are aware with reasonable certainty that the recipient of the service is a minor.*

[11] https://linc.cnil.fr/demonstration-privacy-preserving-age-verification-process

Building on trusted third parties use case discussed earlier, the guardian or parent of the minor could support the creation of an attribute based identity for the minor using such a third-party service. The minor could then navigate into online platforms and make use of the functionalities of the attribute based identity, without having to disclose her full identity, date of birth, etc. Such an approach would satisfy the requirements of reliability, data protection and security of minor's data.

4. CONSIDERATIONS ON EXERCISING THE RIGHTS OF DATA SUBJECTS

A key element of the GDPR is the right of data subjects to be properly informed and in control of the use or other kind of processing of personal data concerning them'. These data subject rights include, among others, aspects related to the data protection principles of lawfulness, fairness and transparency (i.e.,right of information and access), unlinkability (i.e., data minimization, purpose limitation and storage limitation, integrity and confidentiality) and intervenability (i.e., right to erasure, right to rectification, right to object, as well as to lodge complaints with data protection authorities and to seek effective judicial remedies against data protection authorities; or controllers or processors).

These fundamental rights of data subjects call for a closer consideration with respect to their implementation in data sharing environments. Whereas the previous sections cover multiple examples of data minimisation techniques, removing information from data or restricting access to information to a smaller subset of stakeholders, thereby addressing the goal of protecting delinking, the other two protection goals require a different engineering approach. This chapter will address these aspects.

For the sake of an example, we chose the healthcare application domain, as discussed in Section 2. In this scenario, multiple actors participate in the collection of data, its storage, sharing and processing. Data is created in multiple locations, such as:

- at patients in hospitals or doctor's offices;
- at users of wearable devices;
- at census bureaus or government healthcare institutions;
- at healthcare insurance companies;
- at medical device manufacturers;
- at day-care institutions or retirement homes.

Each of these data sources may provide data of relevance, such as current medical conditions and medical history, statistical information on age, gender, current and recent diagnoses of diseases, ongoing and past medications, personal habits and preferences with respect to sports, nutrition, sleep, family life, etc. Though some European Member States foster a centralised approach of storing healthcare data in government-hosted institutions, there will always be relevant data sets not stored in these central repositories. Therefore, when it comes to engineering the rights of data subjects, there might be cases with a highly segregated landscape of data controllers, data processors and data storage locations.

As also discussed in Section 2, the GDPR contains specific requirements concerning personal data in the health sector and provides an extensive set of rights of data subject. Depending on the exact circumstances, this may affect personal data stored at hospitals, where purpose binding, declaration or withdrawal of consent[12], transparency requests and intervention rights have to be appropriately considered.

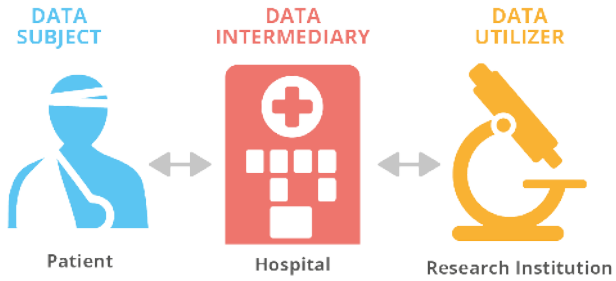
[12] If consent (Art. 6(1)(a) GDPR) is the legal basis for processing.

At the other side of the data sharing discussion, data processing organisations are interested in acquiring data for their purposes with sufficient technical support and clear guidelines on legitimate and illegitimate utilisation venues for the data. For the sake of avoiding confusion with the GDPR terms of data controller and data processor and keeping in mind that there may even be utilisation on the basis of anonymous data that is not regulated by the GDPR, we will call these entities *data utilisers* in this chapter. In GDPR terms, they can either be data controllers, data processors or data recipients. However this is yet to be clarified, depending on each scenario, as also highlighted by the EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the EU Health Data Space [9].

Third, and most relevant to consider in the foreseeable future, is the role of *data intermediaries*. These actors somehow **mediate between the suppliers of data, the data subjects, the data storage providers, and the data utilisers**. Different new and upcoming European laws define such types of entities, for instance as data intermediaries in the European Data Governance Act [6] or as stakeholders in the EU Health Data Spaces proposal [7].

Their role typically is not to use the data they share themselves or, if so, only for very limited primary purposes (such as hospitals using the personal data of patients to provide essential medical services to these patients). In the role of a data intermediary, these actors interact with data utilisers, such as pharmaceutical research institutions or statistics agencies, trying to share data with them while considering data sources and data subjects. In GDPR terms, they can be considered as data processors but their similarity to data utilisers is yet to be clarified.

Figure 14: Data sharing scenario with data intermediaries



As discussed in a past ENISA report [10], implementing the rights of data subjects can become challenging when addressing a distributed data sharing landscape with multiple actors, suppliers, sub-processors, IT service providers etc. With the new legal instruments for data spaces, this set of 'directly involved' data utilisers is extended by new actors, such as private sector platforms for sharing data, data marketplaces, government-hosted data repositories, and other types of data intermediaries as outlined in the European Data Governance Act.

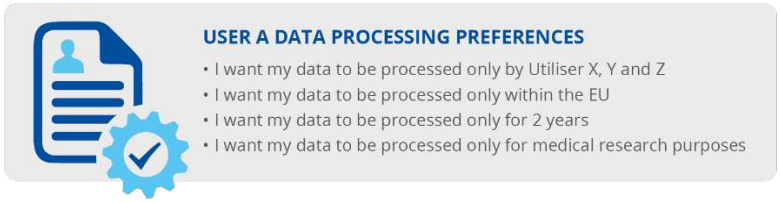
These actors do not primarily intend to process that data themselves, but merely store, host, and provide the data to data sinks on demand. According to the Data Governance Act, it is the explicit duty of these actors to cater for and enforce the rights of data subjects on the data in their domain of control. Hence, **it requires a clear definition of the permissions required to access such data, to be negotiated between data subject and data intermediary, ahead of any data sharing activities – and along with them**.

4.1 INTERACTION BETWEEN DATA SUBJECT AND DATA INTERMEDIARY

A key component of every type of processing of personal data is a valid legal basis[13]. In most cases, this tends to be the explicit consent given by the data subject. Herein, the data intermediary has to cater for managing such consent for data processing for each of the data utilisers it serves. If a data subject declares consent for processing her personal data at one organisation, for example, this does not also automatically encompass transfer to any other organisation. Data subjects may choose to restrict their consent only to certain conditions or otherwise may decide to refrain from giving consent at all. While the GDPR prescripts restrictions, e.g. stemming from the principle of purpose, Art. 5(1)(b) GDPR, the conditions under which a data subject may willingly provide personal information for some kinds of processing operations by one or more organisations may be manifold, and may even go beyond data protection criteria. Figure 15 below presents some examples that a data subject could potentially consider.

Data subjects may choose to restrict their consent only to certain conditions or otherwise may decide to refrain from giving consent at all. information.

Figure 15: Data Processing Preferences Example



It is evident that the task of declaring and enforcing such data processing policies at the data intermediary may be challenging. It requires both extensive interaction with the data subject, in order to collect and negotiate all necessary processing permissions, and interaction with the potential data sinks, in order to validate the compatibility of the demands of the data subject with the conditions of the data utiliser. Therefore an explicit policy language and data model, similar to the ones described in [33], [34] & [35] for defining such access policies would be beneficial.

What makes this challenge even more complex are the potential changes over time; a data subject may decide, at any point in time, to arbitrarily modify her data processing demands, or to revoke consent for processing or to restrict processing at certain data utilisers. Hence, a data intermediary also has to permanently keep track of ongoing processing instances of the data of each data subject, in order to respond to a change of mind from a data subject within a reasonable amount of time.

4.1.1 Purpose Limitations

A key challenge in this landscape of issues is the limitation of purpose aspect. As a data subject may arbitrarily choose to restrict processing of her own data for certain purposes, it becomes necessary to have a closer look on the domain of compatibility of purpose. How can an Information System assist in determining whether the clearance for processing personal data given for one purpose is compatible with the exact purpose of processing by the data utiliser? What can be done to provide data intermediaries with the necessary certainty and potential guarantees for a sharing of data that is data-protection compliant?

In the current setting of data production and data use, there is a plethora of different purposes, which may or may not be compatible, may be overlapping or including each other. Also, there is a large potential for conflict should the interpretation of the purposes of such permissions differ between different actors. It must be anticipated that it will take some time and effort to come up

[13] See Art. 6(1)(a)-(f) GDPR which enlists six possible legal grounds.

with a somewhat consolidated, generally accepted approach to handle such issues in consent and binding purpose in data intermediary scenarios.

4.1.2 Implementation Aspects

On the technical side of engineering these concepts, **the interaction between data subjects and data intermediaries can be realised in multiple ways**. Where cookie policy banners and similar consent management systems are common, providing and withdrawing consent is feasible, with more or less functional technical interfaces to utilise, for example in [36] and [37]. However, these systems typically are part of a website with which the data subject interacts, and cover only the services offered via this website or its service provider. Hence, this does not solve the challenges of collecting consent in different modes of interaction, e.g. in smart cars, for wearables or medical implants, or for other IoT systems that do not provide a reasonable user interface for the data subject [38] & [39].

Additionally, most of these consent management systems cover only a single, scope-specific website or service. They do not, or do not explicitly, work for cases where the initial interaction is not triggered by the data subject but merely by a data utiliser. It is obviously not possible to actively show a website-based consent banner to a user who is not currently interacting with a data utiliser's website. Hence, if the initiation of a novel data processing activity is not done by the data subject but by the data intermediary, there needs to be a different mode of interaction between data subject and data intermediary, one where the data intermediary can proactively contact the data subject for gathering a new consent for a new instance of data processing, e.g. at a new data utiliser, or for a new data processing purpose.

Withdrawing consent and communicating data processing restrictions are two other challenges in this interaction. Each of these processing operations are initially triggered by the data subject. **The common approach here is that the data is deleted by the intermediary or utiliser, along with an active notification directed to the data subject.** Hence, in terms of engineering, the withdrawal of consent, or the rights to rectification, erasure, or expression of restrictions on processing can all be handled similarly by providing a communication interface (such as a website), directed at the data subject, that provides these interactions as services.

However, the challenge for a data intermediary in such cases is that of forwarding such requests towards all concerned data utilisers. This requires some management within the data intermediary, and some interactions with the data utilisers. We start with the latter.

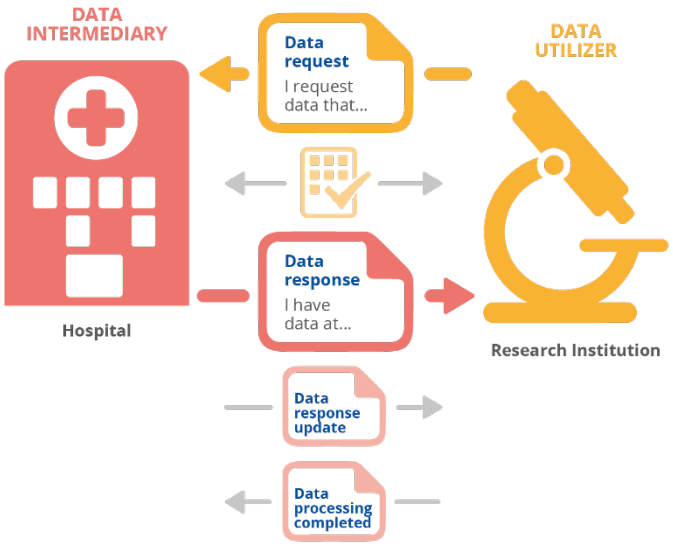
4.2 INTERACTION BETWEEN DATA INTERMEDIARY AND DATA UTILISERS

Once data is available in the repositories of the data intermediary, the details of different modes of data procurement have to be considered. Whenever a potential data utiliser shows interest in a specific subset of the data available (or the data set in total), multiple tasks have to be performed at the data intermediary's end.

4.2.1 Data Request and Data Response

Initially a data utiliser needs to be able to turn towards a data intermediary, asking for data which can be personal or non-personal in nature. In the latter case, many of the subsequent steps can be ignored. In that request for data, the data utiliser has to express the specific criteria concerning the data in which it is interested. This covers, for example, attributes of data subjects whose data is searched for aspects of data types, domains, quantities, qualities or origins, with a huge plethora of different types of filters and restrictions to consider. **The challenge here is to define the format and syntax of an interaction with a sufficient degree of expressiveness, hence allowing the data utiliser to define its demand most precisely.**

Figure 16: Interaction between Data Intermediary and Data Utiliser



Beyond expressing the search criteria for data in which she is interested, **the data utiliser also needs to precisely define the purpose and scope of the intended processing of data**. This information is needed by the data intermediary to filter out data sets where the data subject has somehow objected to such types or scopes of processing, or such types of data recipients. Hence, information on, for example, countries within which the data processing is scheduled to be performed or subsequent data utilisers on the side of the data utiliser must be disclosed with a set of attributes to the data intermediary, in order to allow for enforcing the will of the data subject by the data intermediary in a reliable way.

Based on this initial data request, the data intermediary needs to perform an internal analysis to determine the data sets available that match the requirements expressed by the data utiliser. As a result, a set of data sources is identified, which could be served to the data utiliser as a response. This data set sent in response may be quite a mixed type and quantity. Sometimes, only a subset of data from a database table might be compliant with all the restrictions that apply. Sometimes it might be a file storage location, where some files match the search criteria and policies regarding purpose. Again, the response must be able to cope with a plethora of different data types, formats, syntaxes, and access technologies.

4.3 DATA MANAGEMENT AT THE DATA INTERMEDIARY

In order to comply with the needs and rights of the data subject and data utiliser, the data intermediary needs to keep track of all data sources and data processing tasks at once. It needs to know and interact with both data subjects and data utilisers, as needed, and it needs to evaluate and update data usage policies at multiple stages throughout the data processing lifecycle.

As discussed previously, the data intermediary has to store the means of communication for each data utiliser with which it interacts, and for most data subjects as well. Each new instance of data processing must be logged, tracked, and reasonably addressed. Beyond the required types of interaction with data subjects and data utilisers as discussed above, the data intermediary has to solve some specific challenges, which will be discussed next.

4.3.1 Consent Coverage and Purpose Limitation

A key question in the selection process of a data subject's data that may or may not fulfil the criteria set up by the data utiliser is that of compliance with the purposes of the data processing. If the legal basis for processing of the data is consent, does this consent cover the scope of

processing intended by the data utiliser? Obviously, if the data utiliser was explicitly blacklisted by the data subject, the test for viability is easy; the data of that data subject cannot be part of the response to the data utiliser. However, there may be cases that the data intermediary has to determine—and to decide—whether or not to include this data in the response or not; e.g. to determine if such a transfer is compatible with the initial purpose of the data processing.

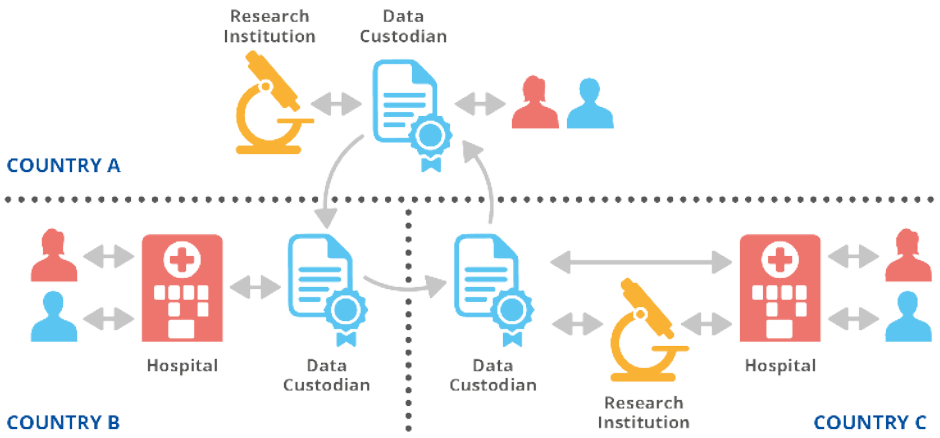
Sometimes, this may be a comparison of attributes in an explicit list of restrictions, such as the identities of data utilisers or countries of processing, but sometimes, especially for the different purposes of processing, such a decision is not that trivial. **When is the purpose defined by a specific data utiliser for a specific instance of data processing sufficiently covered by the purposes granted by a data subject in its consent?** Depending on the exact conditions, the data intermediary has to decide on one out of several different options on how to proceed:

- include the data, assuming the purposes clearly are compatible,
- exclude the data, assuming the purposes are clearly not compatible,
- contact the data subject to clarify whether the purposes are compatible or not, or to determine whether the data subject would consent to the particular processing purpose in question,
- consult relevant third parties, such as data protection authorities, for guidance on whether or not the purposes are compatible.

4.3.2 Inter-Intermediary Interaction

Sometimes there will be more than one data intermediary involved in handling a specific data request from a data utiliser. In that case, either, the data utiliser asks each data intermediary separately, resulting in scenarios as discussed previously, or there might be a need for interaction among several different data intermediaries. For instance, **the decision on whether to include a certain data set in a response may be made by one data intermediary, but the data itself may be controlled by a different data intermediary** (e.g. a data storage provider). In such cases, the data intermediaries have to interact accordingly to handle such mixed – or joint[14] – responsibility for handling the data request accordingly.

Figure 17: Cross border data exchange with data custodians



One specific scenario for such a type of interaction among data intermediaries is that of cross-border data exchange among (and beyond) member states of the European Union. In such cases, the decision on whether data may leave one country for a different country may be made, for example, by dedicated national data custodian organisations or similar data intermediary types. In such cases, the communication protocol between the data intermediaries

[14] Depending on the actual implementation, it may constitute joint responsibility according to Art. 26 GDPR.

has to cater for that specific demand and its communication aspects as well, as depicted in Figure 17 above.

4.3.3 Logging and Reporting

Whatever data request is made, whatever data is entering the control domain of a data intermediary, the data intermediary needs to keep track of it and be accountable. When a data response is sent to a data utiliser, the exact details on the decision process of what data was included must be stored indefinitely at the data intermediary, so as to be able later to inform the data subject, data utilisers, law enforcement, or other external stakeholders. This may even be a legal requirement of data intermediaries, in order to enable them to justify any data sharing activities should a lawsuit be filed.

4.3.4 Privacy-Preserving Data Selection

In order to fulfil its duties of serving data responses to the data requests put out by data utilisers, the data intermediary needs to have an in-depth knowledge of any data set it controls. This may even go into such details as disclosing individual data records to the data intermediary, e.g. to determine whether specific attributes of the data subject concerned match the filtering demands of a data utiliser's request. In that case, the problem of data privacy and confidentiality against the data intermediary becomes evident. In its response, how can the data intermediary decide whether or not to include individual data records in a data set without having access to the filtering values of attributes within the data themselves? **A possible solution could be attribute-based encryption (discussed in Section 2.1.1) or approaches of pseudonymisation, as discussed in** [16] **or even anonymisation**; these may help in hiding the identities of data subjects from both the data intermediary and the data utiliser – as long as they do not interfere with the feasibility of the processing activity itself.

4.4 DATA ALTRUISM

Data altruism, introduced in the Data Governance Act, refers to data subjects agreeing to the use of their data for purposes of general interest.

The concept of data altruism, also introduced in the Data Governance Act, refers to data subjects agreeing to the use of their data for purposes of general interest, such as scientific research or improving public services. For instance, data altruism might occur when a patient decides to allow processing of medical data collected about her not just at the hospital, but also by research institutions that develop treatments. What is interesting with respect to altruism is that typically no compensation is given to the data subject. However, this does not imply a livelong waiver of fundamental rights to privacy and data protection, both for the data subject and the organisation processing the data.

In terms of handling such data at a data intermediary, consideration could be given to flagging the data as having been released under a data altruism 'license', in order to correctly address subsequent demands from data utilisers with respect to this data. Typically, processing permissions for such data are always granted but it must be documented by the data intermediary, so as to be able to prove the rationale behind the release of the data should the data subject file a complaint at, for example, a data protection authority concerning the rights of data subjects. In that case, the data intermediary must be able to prove the origin of the data and initial permissions given by the data subject at that time.

5. CONCLUSIONS

When two or more parties decide to share their data, they become part of a larger data ecosystem where they can take advantage of the combined data set that enables the discovery, by way of computation, of new information or trends relating to individuals, groups of individuals, or to society as a whole. The easiest and most straightforward way to achieve this goal would be to exchange the raw data that each actor holds across technical interfaces putting them on a common table (i.e. a single database) but this hypothetical option is not really feasible. In reality we are pursuing trusted sharing environments that will make full use of the potential offered by a safe and secure exchange and use of personal data while respecting data protection principles.

This report attempted to look closer at specific use cases relating to personal data sharing, primarily in the health sector, and to discuss how specific technologies and considerations of implementation can support the engineering of personal data sharing in practise. The analysis ranged from user-controlled data sharing to large scale personal data gathering and data sharing using third party service.

Despite the potential of the data sharing concept and the relevant Union policy and law in the area, there are still considerations on which are the appropriate technical and organizational measures and how to engineer them into practise. The European legislative initiatives on data sharing described in Section 1.1 entail the processing of large quantities of data which will also include personal data. Therefore, in addition to the consistency of their provisions with the GDPR, it is important to remove any legal uncertainty on the roles and obligations, not only for individuals as highlighted by the EDPB and the EDPS in [9] but also for the entities involved in the data sharing. In order to leverage the potential of data sharing across the EU, practitioners could be provided with directions on which technologies and techniques can be considered, under which circumstances and which data protection principles can be met.

There are several commonly used (cryptographic) techniques (i.e. asymmetric encryption, pseudonyms, access control etc) that are already acknowledged as able to alleviate data protection risks. Some of them were discussed in Section 2, Section 3 and Section 4. In emerging concepts such as data spaces and data intermediaries, however, the risks introduced cannot always be adequately addressed only by such techniques. This is due to the fact that data subjects want to preserve confidentiality of the data they are sharing, they might not know beforehand with whom they might be sharing data with or might want to share accumulated datasets. Although there are advanced techniques that are still evolving, they should not be considered as of purely academic interest since there exist practical implementations in real use case scenarios.

Lastly, since the majority of the technologies described earlier and in previous ENISA reports [10] & [40] rely on asymmetric cryptography, the advent of quantum computing and the impact on the security of currently used asymmetric ciphers should be anticipated. Following the deployment of data sharing infrastructures and services, we cannot expect that they will cease to operate due to possible inadequacy of the asymmetric ciphers. This is where crypto agility becomes relevant as it allows for a switch between algorithms, cryptographic primitives, and other encryption mechanisms without significant changes in the overall IT system or process.

REFERENCES

1. Xenos, H.: Latest trends in optical networks- straight from NGON & DCI World. (2019)

2. Data Protection Commission (DPC): Data Sharing in the Public Sector. (2019)

3. ICO: Data sharing: a code of practice. (2020)

4. Gartner: Data Sharing Is a Business Necessity to Accelerate Digital Business. (2021)

5. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: “A European strategy for data”, COM/2020/66 final., Brussels (2022)

6. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724): Data Governance Act. (2022)

7. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space: European Health Data Space. (2022)

8. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data: Data Act. (2022)

9. EDPB-EDPS: Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space. (2022)

10. ENISA: Data Protection Engineering: From Theory to Practice. (2022)

11. European Data Protection Board (EDPB): Guidelines 4/2019 on Article 25 Data Protection by Design and by Default. (2019)

12. ENISA: Deploying Pseudonymisation Techniques: The case of the Health Sector. (2022)

13. Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients’ rights in cross-border healthcare: Cross Border Healthcare Directive. (2011)

14. Schwalbe, N., Wahl, B., Song, J., Lehtimaki, S.: Data Sharing and Global Public Health: Defining What We Mean by Data. Frontiers in Digital Health 2 (2020)

15. Hildebrandt, M., Verheul, E., Jacobs, B., Meijer, C., de Ruiter, J.: Polymorphic Encryption and Pseudonymisation for Personalised Healthcare: A Whitepaper., Cryptology ePrint Archive (2016)

16. ENISA: Pseudonymisation techniques and best practices. (2019)

17. Sahai, A., Waters, P.: Fuzzy Identity-Based Encryption. In : Advances in Cryptology – EUROCRYPT 2005, vol. 3494, pp.457-473 (2005)

18. Goyal, V., Pandey, O., Sahai, A., Waters, B.: Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data. In : 13th ACM Conference on Computer and Communications Security, pp.89-98

19. Wang, F., Mickens, J., Zeldovich, N., Vaikuntanathan, V.: Sieve: Cryptographically Enforced Access Control for User Data in Untrusted Cloud. In : 13th

20. Blaze, M., Bleumer, G., Strauss, M.: Divertible protocols and atomic proxy cryptography. In : International Conference on the Theory and Applications of Cryptographic Techniques - EUROCRYPT 1998, pp.127-144 (1998)

21. Nuñez, D., Agudo, I., Lopez, J.: Proxy Re-Encryption: Analysis of constructions and its application to secure access delegation. Journal of Network and Computer Applications 87(1), 193-209 (2017)

22. Liang, X., Cao, Z., Lin, H., Shao, J.: Attribute based proxy re-encryption with delegating capabilities. In : 4th International Symposium on Information, Computer, and Communications Security, pp.276-286 (2009)

23. ENISA: Recommendations on shaping technology according to GDPR provisions - An overview on data pseudonymisation. (2019)

24. van Gastel, B., Jacobs, B., Popma, J.: Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson's Disease Study., 19-25 (2021)

25. Verheul, E. .: The polymorphic eID scheme -combining federative authentication and privacy. (2019)

26. ENISA: Privacy and data protection in mobile applications. (2018)

27. Piotrowska, A., Hayes, J., Gelernter, N., Danezis, G.: AnNotify: A Private Notification Service., IACR eprint (2016)

28. Consolidated text: Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual: Audiovisual Media Services Directive. (2018)

29. ABC4Trust EU Project. Available at: https://abc4trust.eu/

30. Piedra, A., Hoepman, J., Vullers, P.: Towards a Full-Featured Implementation of Attribute Based Credentials on Smart Cards. In : CANS 2014: Cryptology and Network Security, pp.270-289 (2014)

31. CNIL: Online age verification: balancing privacy and the protection of minors. (2022)

32. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC: Digital Services Act. (2022)

33. McDonald, A., Reeder, R. W., Kelley, P. G., Cranor, L. F.: A Comparative Study of Online Privacy Policies and Formats. In : Privacy Enhancing Technologies Symposium, pp.37–55 (2009)

34. Drogkaris, P., Gritzalis, A., Lambrinoudakis, C.: Empowering Users to Specify and Manage Their Privacy Preferences in e-Government Environments. In : Electronic Government and the Information Systems Perspective, pp.237–245 (2014)

35. Hansen, M., Jensen, M.: A Generic Data Model for Implementing Right of Access Requests. In : APF 2022: Privacy Technologies and Policy, vol. 13279, pp.3-22 (3033)

36. Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., Holz, T.: We Value Your Privacy. Now Take Some Cookies. Informatik Spektrum, 345-346 (2019)

37. Karegar, F., Pettersson, J., Fischer-Hübner, S.: The Dilemma of User Engagement in Privacy Notices: Effects of Interaction Modes and Habituation on User Attention. ACM Transactions on Privacy and Security 23(1), 1-38 (2020)

38. Castelluccia, C., Cunche, M., Le Metayer, D.: Enhancing Transparency and Consent in the IoT. In : 2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp.116-119 (2018)

39. Murmann, P., Beckerle, M., Fischer-Hübner, S., Reinhardt, D.: Reconciling the what, when and how of privacy notifications in fitness tracking scenarios. Pervasive and Mobile Computing 77(C) (2021)

40. ENISA: Data Pseudonymisation: Advanced Techniques and Use Cases. (2021)

41. 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)

42. Alpár, G., Jacobs, B.: Credential Design in Attribute-Based Identity Management. (2013)

TP-03-22-200-EN-N

ABOUT ENISA

The European Union Agency for Cybersecurity, ENISA, is the Union's agency dedicated to achieving a high common level of cybersecurity across Europe. Established in 2004 and strengthened by the EU Cybersecurity Act, the European Union Agency for Cybersecurity contributes to EU cyber policy, enhances the trustworthiness of ICT products, services and processes with cybersecurity certification schemes, cooperates with Member States and EU bodies, and helps Europe prepare for the cyber challenges of tomorrow. Through knowledge sharing, capacity building and awareness raising, the Agency works together with its key stakeholders to strengthen trust in the connected economy, to boost resilience of the Union's infrastructure and, ultimately, to keep Europe's society and citizens digitally secure. More information about ENISA and its work can be found here: www.enisa.europa.eu.

ENISA
European Union Agency for Cybersecurity

Athens Office
Agamemnonos 14, Chalandri 15231, Attiki, Greece

Heraklion Office
95 Nikolaou Plastira
700 13 Vassilika Vouton, Heraklion, Greece

enisa.europa.eu

ISBN 978-92-9204-602-6
doi: 10.2824/36813