Supporting Data Portability in the Cloud Under the GDPR

Yunfan Wang and Anuj Shah

Carnegie Mellon University

Abstract

The right to data portability under the European Union's General Data Protection Regulation (GDPR) extends beyond existing privacy frameworks and empowers individuals to transfer their personal data between data controllers. While the Working Party for Article 29 of the Data Protection Directive has issued guidance on how to respond to portability requests, the European Commission has expressed a different interpretation of this right. Data portability therefore brings new and significant challenges to data-driven enterprises, especially those with systems that are distributed across cloud infrastructure. We attempt to clarify how this right translates to the operations of cloud service providers in their roles as either data controllers or data processors. Specifically, we outline the various technical methods available for porting data in the cloud, and then consider how the recipient of data from a portability request and the cloud service level govern which compliance solution a cloud provider can put forward. The solutions we describe here are simple extensions of existing services and do not prescribe a specific legal interpretation. We encourage cloud providers to take a competitive stance on GDPR compliance by offering these solutions to their customers.

Keywords: Data portability, GDPR, privacy

1. Introduction

The General Data Protection Regulation (GDPR), jointly drafted by the Council of the European Union and the European Commission (EC), aims to strengthen data protection for all individuals within the European Union (EU) and give greater control to citizens and residents over their personal

data [1]. When drafting the GDPR, the EC pushed for the enactment of a regulation rather than a directive because regulations are binding in their entirety and directly applicable in all EU member states [2]. Therefore, any company handling personal data from people within or from the EU or processing personal data within the EU must consider how the GDPR will affect their operations once it comes into effect on May 25th, 2018.

A notable extension of user rights beyond the EU's Data Protection Directive of 1995 (the Directive) is the right to data portability under GDPR Article 20:

> "The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided." [GDPR Article 20(1)] [3]

While an individual's right to access under the Directive constrained users to receive data in a format of the data controller's choosing [4], the right to data portability ensures that individuals receive data in a format that enables transfer to another controller. Some countries are following suit: while GDPR Article 20 extends beyond the access rights provided in most other countries' and regions' privacy laws (see Table 1), Argentina and the Philippines have proposed or implemented privacy legislation that also provides their citizens and residents with the right to data portability [5, 6]. Thus, GDPR Article 20 and similar laws in other regions could serve as a crucial tool for moving toward a user-centric Internet [7].

Data portability has also generated immense concern among data-driven companies. A 2017 privacy governance report indicates that privacy professionals rate data portability as the most difficult compliance obligation under the GDPR. Firms with annual revenue of $25 billion or more report higher than average difficulty ratings for this right, potentially because they perceive themselves to be primary targets for enforcement [8]. Part of the reason for this concern could be the lack of clarity in Article 20's legal interpretation, which in turn prevents companies from knowing exactly how to operationalize compliance [9]. The same privacy governance report also shows that privacy professionals will heavily consider GDPR compliance when selecting a cloud service provider [8]. Given the pressing need for data portability solutions

Country or Region	Regulation	Rights Similar to Portability
USA	Health Insurance Portability and Accountability Act (HIPAA)	Portability of health insurance, not personal data [10]
USA	Gramm-Leach-Bliley Act (GLBA)	No rights related to access or portability [11]
Canada	Personal Information Protection and Electronic Documents Act (PIPEDA)	Access and correction [12]
Canada	Canada Health Act	Portability of health insurance, not personal data [13]
Asia-Pacific	APEC Privacy Principles	Access and correction [14]
Mexico	Federal Law on the Protection of Personal Data Held by Private Parties	Access and correction [15]
Dubai	Data Protection Law (DIFC Law No. 1)	Access and correction [16]
Japan	Personal Information Protection Act (PIPA)	Access and correction [17]
Hong Kong	Personal Data Ordinance	Access and correction [18]
Singapore	Personal Data Protection Act (PDPA)	Access and correction [19]
South Korea	Personal Information Protection Act (PIPA)	Access and correction [20]
Australia	Australia Privacy Act	Access and correction [21]
Argentina	(Draft) Personal Data Protection Act	Access, correction, and portability [5]
Philippines	Data Privacy Act	Access, correction, and portability [6]

Table 1: Rights Under Other Data Privacy Frameworks

and the rapid adoption of cloud services across industries, we explore how cloud providers and their customers can support this new right.

The remainder of the report is organized as follows: Section 2 covers what is and is not known about legal requirements under GDPR Article 20; Section 3 provides a brief overview of cloud computing service models and the types of data implicated in each; Section 4 explains the technical methods that enable data portability; Sections 5, 6, and 7 describe whether and how cloud providers can support or fulfill portability obligations for infrastructure-level, platform-level, and software-level cloud products (respectively); and Section 8 concludes the report.

2. The Legal Landscape

2.1. The Role of Data Controllers and Processors

Before discussing the legal complexities of data portability, it is useful first to cite definitions of relevant terms from the GDPR and explain the roles of controllers and processors.

> *Personal data* refers to "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or

more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." [GDPR Article 4(1)] [3]

A *controller* is the "natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data." [GDPR Article 4(7)] [3]

A *processor* is "a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." [GDPR Article 4(8)] [3]

Processing involves "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction." [GDPR Article 4(2)] [3]

The GDPR largely places the burden of compliance on data controllers. They must ensure that all processing activities align with the regulation; they are responsible for the protection of data subject rights; and they are liable for any damage resulting from processing that infringes upon the GDPR (unless a processor goes beyond the contract signed with the controller or acts as a controller) [22]. There are additional portability obligations specific to sending and receiving controllers. Sending controllers, while not responsible for ensuring the recipient's compliance with the GDPR, must check that the data extracted for transmission match the data requested by the data subject. Receiving controllers must delete data that is not relevant to their processing needs. Furthermore, they are prohibited from using data on third party data subjects for their own purposes. For example, the receiving controller cannot enrich profiles of third party data subjects simply because the individual requesting portability transferred a photo containing social media tags of those other data subjects [4].

Recital 68 of the GDPR encourages all controllers "to develop interoperable formats that enable data portability" but does not require fully compatible services [4]. If technical barriers arise when fulfilling a request for data

portability, the sending controller must explain such barriers to the data subject in an intelligible manner so they understand and can act upon their options [4]. That the GDPR offers some flexibility in the implementation of portability suggests that companies may want to create industry portability standards. A recent EC proposal for the regulation of *non*-personal data similarly advocates for a self-regulatory approach to data portability, but it also suggests that the EC may impose heavier rules if companies fail to develop their own codes of conduct within a reasonable time [23]. Companies might therefore expect the same warning for interoperability under the GDPR.

While Article 20 does not explicitly mention any responsibility for data processors, Article 28 specifies that controllers may only share personal data with processors that provide sufficient guarantees for GDPR-compliant processing [3]. Thus, entities such as cloud providers that provide data storage, availability, structuring, organization, and other functions to businesses that directly deal with personal data will similarly need to support GDPR compliance if they seek to remain competitive. Corporations who are wary of potential legal action will likely cease relationships with cloud providers who fail to do so [24]. The broader responsibilities of processors under the GDPR include processing data only as instructed by the controller; deleting and/or returning data to the controller once processing is complete; seeking permission from the controller to engage in subcontracting with other processors; and assisting the controller with the protection of data subjects' rights [25].

2.2. Difficulties in the Realization of GDPR Article 20

Guidelines from the Article 29 Data Protection Working Party (WP29) indicate that companies cannot silo data subject rights under the GDPR when formulating their compliance program [26]. Granting data portability in particular may invoke the data subject's right of access (Article 15), correction (Article 16), and erasure (the "right to be forgotten") (Article 17). Before porting personal data to another controller, a data subject would likely want to know what data the original controller has observed and collected [27]. While Article 20 does not require data controllers to check the accuracy of data before the porting process, Article 5(1) requires controllers to implement all reasonable measures to ensure that stored data is up-to-date [4]. A data subject may request a controller to erase personal data and simultaneously port the data into their own hands [28]. Therefore, companies that do not integrate data portability mechanisms with methods of access, correction, and erasure may incur unnecessary overhead.

Recent interpretations regarding how these rights handle data ownership complicate the decision of what to erase. On one hand, the WP29 guidelines on data portability clearly state that inferred data does not belong to the data subject but rather to the system that generated it [4]. In contrast, the 2014 ruling in *Google Spain vs. AEPD and Mario Costeja González* acknowledged the right of a data subject to erase an inference from Google's search algorithm, suggesting that the inference does belong to them. Ownership aside, Article 20 only grants the data subject portability rights for data they "provided", meaning that inferred data would not be included in a portability request. However, companies do not have a clear path to compliance when data subjects ask for data erasure [28].

There is additional uncertainty for the realization of Article 20 alone. Specifically, WP29 interprets the portability provision to include data both explicitly provided by the user *and* generated during a user's activity with a service. The former might cover data such as social media posts, images, and demographics, whereas the latter might cover raw data processed by a smart meter, location data, and activity logs [4]. Enforcement of this interpretation would presumably help an individual better grasp the scope of the data that the controller is observing. However, the EC recently expressed that inclusion of data generated by a user's activity in the definition of "provided" goes too far [9]. Unfortunately, companies will need to wait for legal challenges to data portability that implicate this class of personal data to know what Article 20 compliance requires.

Data Type	**Source**	**EC View**	**WP29 View**
Age	Explicitly Provided	Yes	Yes
Social Media Post	Explicitly Provided	Yes	Yes
Images	Explicitly Provided	Yes	Yes
Location	Generated During Activity	No	Yes
Browsing metrics	Generated During Activity	No	Yes

Table 2: Examples of Agreement and Conflict in Interpretation of the Right to Portability

What is certain is that data must be collected based on the consent of the data subject or in fulfillment of a contract where the data subject is a party

for those data to fall under portability obligations. To provide an example, WP29 explains that financial institutions who collect data purely to detect money laundering would not need to port those data in response to a data subject request [4]. Data must also be processed by automated means to be included in a portability request; therefore, paper records would be excluded in most cases.

2.3. Portability and Data Identifiability

In addition to the source of personal data, companies should consider its identifiability when aligning their operations to the GDPR. According to Hintze, the GDPR departs from the Directive in its establishment of a spectrum of de-identification. Specifically, it defines pseudonymization as the "processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information". According to Hintze, the GDPR recognizes that pseudonymization practically delineates between identified and identifiable data. He additionally cites GDPR Article 11, which relieves data controllers of their obligations under Articles 15-20 if "they can show that data has been de-identified and they are not in a position to identify the data subject", and Recital 26, which states that the GDPR does not regulate anonymous data [29]. Article 12(1) further specifies that controllers may not refuse a portability request unless their processing does not require identification [4].

Hintze's legal conclusion regarding this spectrum is adapted here in Table 3. He specifies that Identified and Identifiable data would be subject to an individual's portability request, while Article 11 De-Identified data and Anonynomous or Aggregate data would not. He further emphasizes that de-identifiability is relative to the entity holding the data. Pseudonymized data under the control of a data controller with the key to reverse the pseudonymization would be readily identifiable, but the recipient of the pseudonymized data would have Article 11 De-Identified data [29].

The Future of Privacy Forum presents a finer-grained spectrum, as shown in Figure 1. Three variables inform this framework: direct identifiers, indirect identifiers, and safeguards and controls. Direct identifiers can be used to identify a person without additional information or via cross-linking through other publicly available information. Indirect identifiers connect pieces of information until a specific individual can be singled out. Safeguards and controls are additional legal and technical measures that govern how data

Identifiability	Portability Required?
Identified	Yes
Identifiable	Yes
Article 11 De-Identified	No
Anonymous/Aggregate	No

Table 3: Hintze's Spectrum of De-Identification (adapted from [29])

may be obtained, used, and disseminated. The manner in which each variable is treated governs how a company might classify its stored personal data. For example, pseudonymous data is personally identifiable information (PII) that has been stripped of direct identifiers yet has indirect identifiers intact. *Protected* pseudonymous data has been further protected via additional measures that restrict access [30]. Based on Hintze's discussion, we might expect European courts to include the categories within "Degrees of Identifiability" along with "Key Coded" and "Pseudonymous" under data portability rights. A strict interpretation of the GDPR would additionally oblige companies to port "Protected Pseudonymous" data, while a more lenient ruling might allow companies to exclude this category.

For a full discussion of each of the ten categories displayed in Figure 1, please see [30].

3. Cloud Computing Service Models

Cloud computing is a widely accepted computing paradigm that delivers computing resources as utilities [31]. The most comprehensive definition of cloud computing is given by the National Institute of Standards and Technology (NIST): "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources" [32]. NIST also outlines three service models to classify different cloud services (outlined in Figure 2):

- *Software as a Service* (SaaS) allows clients to use the provider's applications running on cloud infrastructure.
- *Platform as a Service* (PaaS) allows clients to deploy their own applications on the cloud infrastructure.

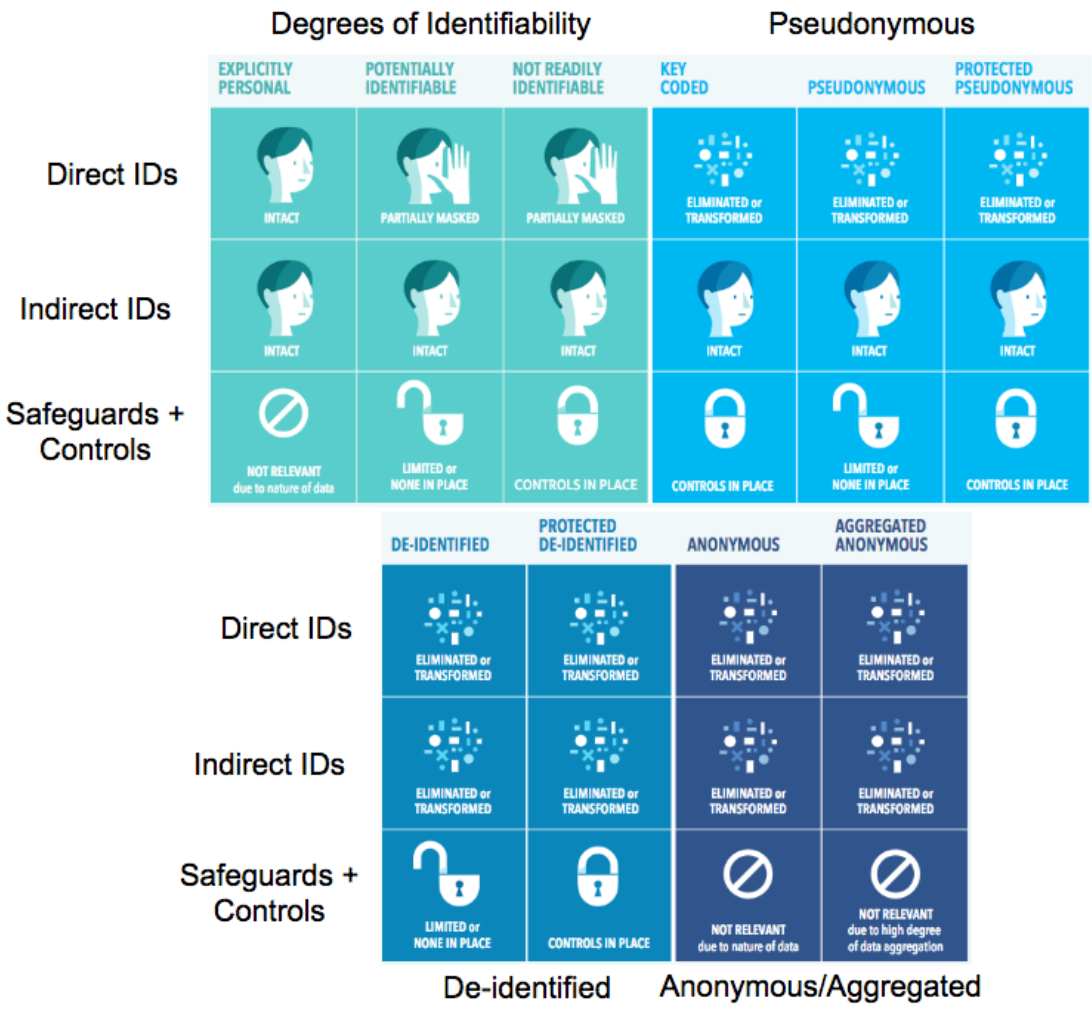

Figure 1: The Future of Privacy Forum's Identifiability Spectrum (adapted from [30])

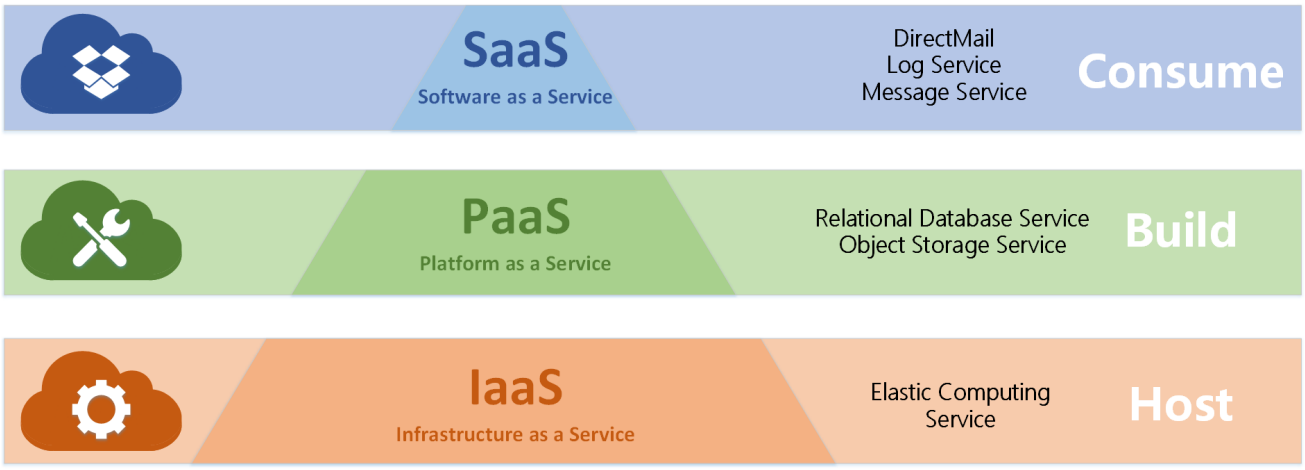

Figure 2: Cloud Computing Service Models

- *Infrastructure as a Service* (IaaS) allows clients to directly provision processing, storage, networks, and other fundamental computing resources from the cloud infrastructure.

The flexibility and portability of data generally increase from SaaS to PaaS and IaaS (see Figure 3). More specifically, IaaS consumers can modify lower level configurations and exert more control over stored data, which means they can feasibly migrate more types of data. However, IaaS products require more extensive and independent setup from consumers, while SaaS products provide more out-of-the-box functionality [33].

3.1. Data Classification

Each service model also involves distinct data types. Ranabahu and Sheth distinguish four types of data in cloud computing that may be useful in understanding where personal data potentially lies [34]:

- *Domain Data* consist of personal data, definitions for data structures, and the relationship between various structures. Files on the disk and data in the database may be included in this category.

- *Logic and Process Data* refer to the "business logic" of a program or application. For example, the program used to execute machine learning algorithms over domain data may be included in this category.

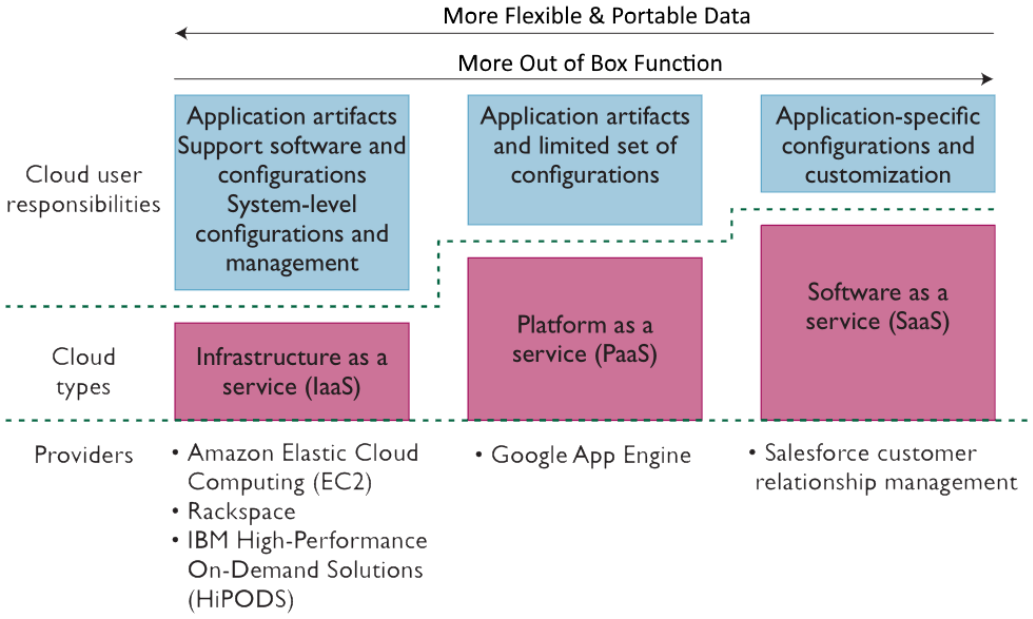

Figure 3: Portability of Data Across Cloud Service Models (from Ranabahu and Sheth, [33])

- *Configuration and Logging Data* may consist of configuration, access control and logging data generated during a user's activity with a service. The common ground of these data is their facilitation of users' interaction with a product. Thus, log information (which is used to debug or monitor running state) such as the timestamp, IP address, and location during login would fall into this category.

- *System Data* consist of the operating system and runtime environment, such as the custom system image provided by cloud customers or libraries that customers' applications rely on.

With increased abstraction of cloud computing products, customers have less visibility into those products. For example, if customers use a PaaS product like Alibaba Cloud's Relational Database Service, they will not know the operating system upon which this database is running. As another example, an email service customer does not typically know which programming language is used to build the service. If end users cannot interact with these components, they also will not provide data related to them. In short, not

Data Category	IaaS	PaaS	SaaS
Domain Data	√	√	√
Configuration and Logging Data	√	√	√
Logic and Process Data	√	√	X
System Data	√	X	X

Table 4: Data Categories in each Cloud Service Model

Data Category	ECS	RDS	Email Service
Domain Data	Files uploaded into instances	Tuples in tables	Emails
Configuration	Data center location, replication number...	cache size, encoding character set...	Forwarding policy, blocked addresses...
Logic & Process	Application running in the instances	Routines (procedures and functions)	N/A
System Data	Operating System like Linux or Windows	N/A	N/A

Table 5: Examples of Data Among Cloud Computing Products

all cloud computing products involve the above four categories of data (see Table 4.

We use Alibaba Cloud's Elastic Computing Service (ECS) as an IaaS product, Relational Database Service (RDS) as a PaaS product, and Email Service as a SaaS product to demonstrate the data involved in different cloud service models. These data are summarized in Table 5.

3.1.1. Elastic Computing Service

ECS is a typical Infrastructure as a Service (IaaS) product that can enable customers to launch new compute instances to meet real-time demand, along with a variety of basic components such as operating systems, memory,

CPU, storage, IPs, and images. It provides a lower-level infrastructure for customers and each ECS instance is highly customizable according to the user's demands. At the same time, customizability also means users need to explicitly provide many instructions and data to each ECS instance to shape them as required. These data include:

- *Domain Data:* Files uploaded into instances.
- *Configuration and Logging Data:* The configuration choices that users make during the creation process. They might contain data center location and replication number, log records generated during users' activities, and metrics like CPU usage and server liveness.
- *Logic and Process Data:* The application running in the instance, represented either in source code or as executable binary files.
- *System Data:* Operating systems like Linux and Windows. One of the advantages of ECS is that it allows users to use their own custom system image to perform their work.

3.1.2. Relational Database Service

RDS is a typical Platform as a Service (PaaS) product that directly provides users with a relational database. Users can connect to it and use it out-of-the-box without any configuration. However, they still have the capability to modify the configuration to optimize their services. Besides domain data such as tuples (the rows and columns of data), users can also insert routines into the database to perform database logic and combine multiple queries into one query. These features are provided by almost all relational databases. Customers also do not need to worry about maintenance, disaster recovery, data replication and other system level problems.

- *Domain Data:* Tuples in tables and the relationships between tables.
- *Configuration and Logging Data:* Data center location, replication number, etc.
- *Logic and Process Data:* Routines (procedures and functions), which are sets of SQL statements used to perform some task on the data in the database.

3.1.3. Email Service

Email service is a widely used Software-as-a-Service product that enables users to access their email without downloading or installing any software on their machines. Since it is not a general-purpose software, users can only use it to process emails and modify some configurations related to the email.

- *Domain Data:* Incoming and outgoing emails, drafts, contacts, etc.
- *Configuration and Logging Data:* Forwarding policies, blocked addresses, etc.

4. Technical Methods for Enabling Data Portability

There are three general methods for porting data stored in the cloud (see Table 6):

- *Application Programming Interfaces* (APIs) are usually developed by the cloud service provider, which means they are often distinct from other providers' APIs. Thus, clients who want to use APIs to extract personal data from cloud services must adapt to the platforms accordingly. APIs are advantageous because clients can extract data directly from the cloud providers, and they do not need to temporarily store data on their local disk.
- *Protocols* are usually designed by an Industry Standardization Organization, and are therefore widely accepted by many cloud platforms. This means clients do not need to adapt to each specific platform to use the protocol, which companies could argue demonstrates their efforts to support interoperability. Clients also do not need to temporarily store data.
- Some cloud platforms allow clients to download their data as a *file* in a commonly used format. It is easier to import data into new platforms when it is structured in a common format, so this method may align with requirements under GDPR Article 20. However, users must store these files temporarily on some medium like a local disk prior to import.

All three methods are widely used in the market today. In addition to enabling compliance with the GDPR, adopting industry-wide protocols

	Standardized	Direct Transport to Recipient	Example
API	X	√	RESTful API
Protocol	√	√	SMTP (Simple Mail Transfer Protocol)
File Export	√	X	Download User Data Archive

Table 6: Technical Methods for Data Portability

	Data Backup	Data Migration	Data Analysis
Receiver	Data Subject	Data Controller	Data Subject/Controller
Best Practice	File Export/CLI	Protocol	API

Table 7: Scenarios Requiring Use of Cloud Data Portability Methods

and commonly used file types can give cloud service providers a competitive edge. These standardized methods would allow for both export and import of data, in turn allowing cloud clients to offer their end users easy migration of personal data from competing services. Lowering the barrier to transfer would encourage end users to change platforms if the previous one does not satisfy their needs. Therefore, supporting these standardized methods can help cloud service providers attract more corporate clients who want to demonstrate GDPR compliance.

There is no one-size-fits-all solution for the various scenarios in which a data subject may request portability. Cloud providers and their customers should choose solutions based on the scenario's requirements. Thus, we will demonstrate the utility of each method using three scenarios: data backup, data migration, and data analysis.

4.1. Data Backup

Individuals will likely use their right to portability to backup and archive their personal data for transfer at a later time. The data receiver in this scenario is the data subject. In most cases, they do not have any technical knowledge and simply want to retrieve data from cloud storage. Therefore, the portability solution should be simple and intuitive for them to use, and the format of the ported data should also be easy to read and store.

Given these requirements, we suggest that cloud service providers or customers export data subjects' personal data as files, such as Javascript Object Notation (JSON) files or comma-separated value (CSV) files. Data subjects can interact with a Web user interface designed by the cloud provider or customer (whichever is the controller as defined by the GDPR) to download these files and store them on their personal devices.

If users want to repeat this backup process on a regular basis without submitting additional portability requests, cloud service providers can also provide a command-line interface to users and enable them to write scripts to execute this job automatically. This would unfortunately require users to have the commensurate technical skills or controllers to design this option into the user interface.

A concrete example of this scenario might be backup of resource usage records generated during the use of cloud computing services. With Alibaba Cloud, users can export these usage records from their console as a CSV file.

4.2. Data Migration

The right to data portability is partially designed to allow users to migrate their personal data from one platform to another. In this case, both the data sender and data receiver are cloud service providers, which requires our solution to be widely supported or at least supported by these two providers. Since they are different cloud service providers, we also need to consider whether their platforms or their customers' platforms are compatible. For example, users might lose some platform-specific data during the transmission process because the other platform cannot use such data. This kind of data loss might be acceptable, since end users would presumably switch services to access new features that are missing from their current service.

Based on these requirements, we suggest that cloud service providers use or develop a standardized transmission protocol to support data migration. A company that uses a standardized protocol that simplifies the transmission process lowers the barrier to transfer and could therefore attract new users.

For example, POP (Post Office Protocol) is a widely supported standardized protocol that can download email from remote servers. Through the POP protocol, users can migrate their personal email data from Gmail to Yahoo Email and receive future emails sent to their Gmail account on their Yahoo Email system.

4.3. Data Analysis

Instead of retrieving all available data, users might want to analyze this data and extract inferred information. To achieve this goal, they should be able to transfer data from the cloud service provider to their analysis program. In this case, users would want access to all stored data (even though they may not retain the full dataset), so the portability solution cannot tolerate data loss. Furthermore, given that the destination of data is an analysis program, structured and machine-readable data is necessary to save users time and energy in parsing data.

Given these requirements, we suggest that cloud service providers or their customers expose Application Programming Interfaces (APIs) to users. Users can then directly embed these APIs into their analysis programs.

We would expect users in this scenario to store their dataset in a database service like RDS. Alibaba Cloud provides both a Software Development Kit (SDK) and a HTTP API for their users to access the database. The SDK is matched to several programming languages and users can directly use them as if the functions were provided by the languages themselves. However, cloud service providers usually provide an SDK only for popular programming languages, so users cannot use the SDK to access data for other languages. In contrast, users can use the HTTP API with any language and on any platform as long as they support the HTTP protocol. This use case is very unlikely due to the technical savvy required of SDKs and HTTP APIs; however, we include it here simply to demonstrate where APIs can serve as a useful method for portability.

Cloud providers and their customers alike can use the methods described in this section when responding to portability requests for data stored in the cloud. Now that we have established how data may be ported, we next describe the unique division of responsibilities for cloud providers and customers at each service level depending on their roles under the GDPR. We also discuss whether and how cloud providers may assist with portability obligations when they are not directly responsible for responding to data subject requests.

5. IaaS-Level Portability

To demonstrate how to deal with data portability in IaaS products like ECS, we will walk through a web application hosting scenario. In this sce-

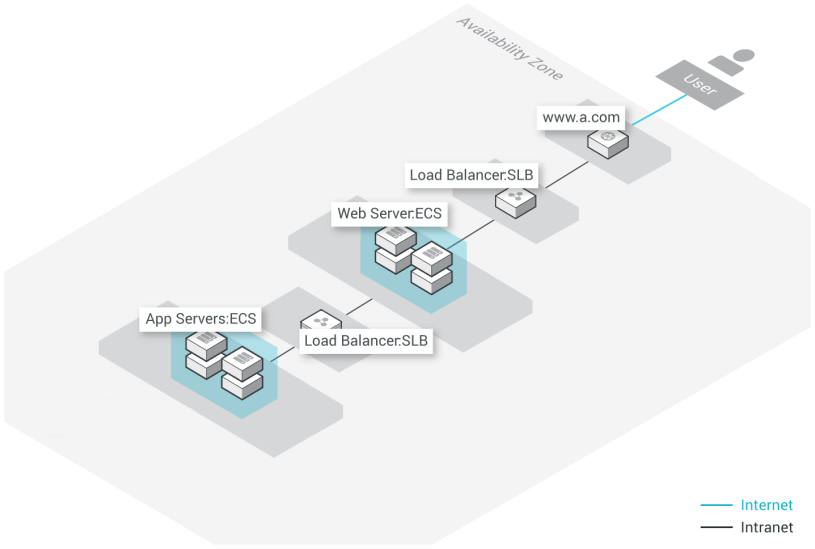

Figure 4: Web Hosting Solution Architecture (adapted from Alibaba Cloud, [35])

nario, users can utilize several popular cloud computing products like ECS and Load Balancer to host a website for themselves [35] (Figure 4).

5.1. Type of Customer Determines Cloud Providers' Responsibility

Usually, cloud service providers will serve two categories of customers: corporate customers and individuals. Corporate customers use cloud computing products to facilitate their business and provide services for their end users. Individuals, or natural persons, might use cloud computing products for personal or household purposes. These two types of customer place the cloud provider in different roles under the GDPR.

If the customer of an IaaS product is a corporate customer, the data they store in the cloud are not their personal data. The data instead describe the end users of the corporate customer. The corporate customer will make decisions on how to store and process data and the cloud service provider will only follow their instructions. This makes the corporate customer a data controller and the cloud service provider a data processor.

Since IaaS products only provide infrastructure to the customer, the customer has full control of the product and the cloud service provider only has coarse-grained visibility over the data. For example, in the ECS scenario

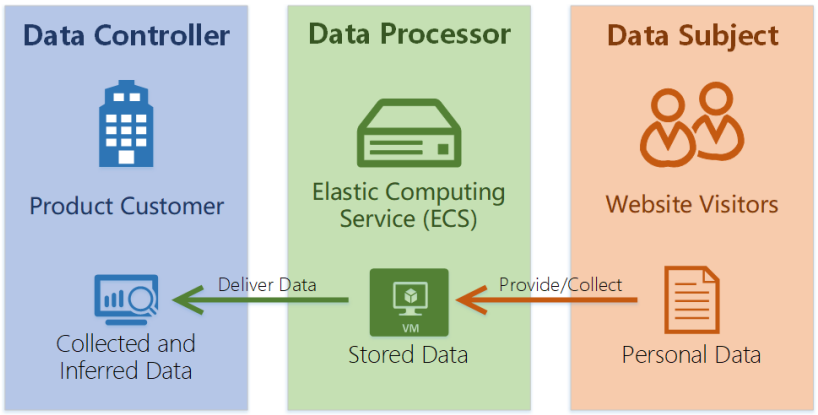

Figure 5: Corporate Customer of ECS

described above, they can view customer data as several virtual machine instances. Therefore, the cloud service provider cannot help the corporate customer implement data portability solutions. The corporate customer needs to implement the methods described in Section 4 by itself.

In the unlikely case that the customer of an IaaS product is a natural person, the roles will change. Because the data in the cloud directly belongs to the natural person, the cloud service provider becomes a data controller and therefore must directly fulfill data portability requests. In the next two sections, we introduce what kind of personal data will be stored in IaaS products and how to enable a natural person to migrate their data among different cloud service providers.

5.2. Server Provision and Related Personal Data

First, customers need to create several ECS instances to host their content. They can choose a base system provided by the service provider like Ubuntu and CentOS, or they can upload customized system images from their personal devices. These operating system images might be classified as *system level personal data*, since they are explicitly provided by customers.

After determining the base operating system, customers may decide in which data center they would like to host their web application. In most cases, they will choose a data center that is geographically close to them to reduce latency in access. They can also decide how many replications of their instances they want. Replication involves making several copies of the web

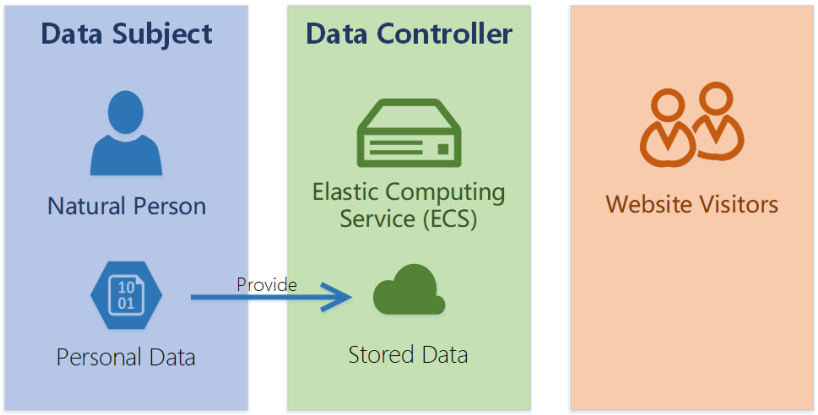

Figure 6: Corporate Customer in ECS usage

servers as backups and enables web servers to continue providing services in the event of a localized failure. Thus, it prevents single point loss. All of these *configurations* made by customers might also count as users' personal data.

After acquiring several machines on the cloud, customers set up the web servers on these machines. During this process, they can implement a self-developed program, commercial solutions, or an open source application. Some of these *programs* may belong to customers and partially comprise their personal data.

Finally, customers will put content such as text-based articles, digital images, and multimedia audio and video on this web application. This content, or *domain data*, is another part of their personal data that resides on the cloud service provider's platform.

5.3. Technical Solution to Migrate Data between Platforms

Most elastic computing service providers have multiple export options for their users. For the web application hosting scenario described above, we recommend exporting instances as a virtual machine. This exported virtual machine can contain users' base operating system, configurations, applications, and domain data like the files in the instance. In other words, users can export all four categories of data through one export method. For example, users may want to migrate their Elastic Compute Cloud (EC2) instance on Amazon Web Services' (AWS) platform to Alibaba Cloud's ECS product. We

can use the feature "VM Import/Export" (provided for EC2) to export the instance as a Virtual Hard Disk file, which is supported by both AWS's EC2 and Alibaba Cloud's ECS, and download the file to the customer's personal device. Next, they can upload this file to Alibaba Cloud's ECS to create a new ECS instance. Customers can then continue to use their instances to run their applications without losing data. The data portability solution used in this scenario belongs to the "export file" category described in Section 4.

6. PaaS-Level Data Portability

6.1. Using Metadata to Support Data Management

As described in Section 3.1.2, PaaS products such as RDS and Alibaba Cloud's Object Storage Service (OSS) support a client's ability to deploy customer-facing applications on top of the cloud infrastructure while cloud providers provision all computing resources. This division of labor at the PaaS level leaves the task of data management to cloud clients, who directly collect and store personal information from their own customers. Because clients of the cloud provider determine the processing of that data, this business relationship places cloud clients in the role of data controllers and cloud providers in the role of data processors. However, unlike the case of IaaS products, cloud providers have finer-grained visibility into customer content, and therefore have greater ability to facilitate data portability as the cloud customer responds to data subjects' requests (even though it is not legally required to do so).

Data management requires proper classification schemes and efficient inventory practices to locate and extract data when needed. However, the aforementioned uncertainty regarding personal data ownership and identifiability under the GDPR indicates that companies may struggle with GDPR-specific data classification. Furthermore, a survey of EU- and US-based companies indicates that 15-20% cannot locate all users' data in their databases [36]. This lack of sufficient data inventory would compound controllers' troubles in trying to fulfill access, correction, erasure, and portability requests in the first place.

Given these current and future challenges to data management, cloud providers can take a competitive stance by allowing controllers to classify data based on their interpretation of the GDPR. Cloud providers can enable classification for compliance by providing metadata "tags". A useful GDPR metadata scheme would need to provide tags across six categories relevant

to the dataset: 1) the specific data type, 2) the source of the data, 3) the type of identifier, 4) the identifiability of the entire dataset, 5) whether the collected data required the data subject's consent, and 6) the data subject rights that are applicable to the dataset (see Figure 7). Data type would enable companies to fulfill data subject requests pertaining to subsets of all available data. Type of identifier would inform the identifiability of the entire dataset. Source, consent, and identifiability tags would jointly determine data subject rights.

Companies would need to introduce domain-specific knowledge to expand the set of tags for data type. The earlier discussion of conflicting GDPR interpretations informs the source tags - specifically, companies must account for whether data is explicitly provided by a data subject, generated during their interaction with a service, or inferred from other data. The FPF's treatment of various identifiers indicates that controllers should consider how they handle direct, persistent, and indirect identifiers. The identifiability scales established by Hintze [29] and the FPF [30] could serve as the basis for identifiability tags. A binary "true/false" tag would effectively capture whether consent was required for a particular data item. Data subject rights under the GDPR are directly transferrable to tags for rights.

A seventh category of metadata tags would assist in classifying processes that change the identifiability of the data. The guidelines on anonymization techniques from WP29 establish that, to consider a dataset effectively anonymized, the techniques applied to it should address the risk of 1) singling out a data subject, 2) linking multiple records of a single data subject, and 3) making inferences about a data subject [37]. These three requirements would fittingly serve as tags for data processes.

6.2. How Metadata-Based Solutions Align with Processors' Obligations

While metadata-based classification would clearly assist cloud customers with responding to portability requests, they should be able to choose the extent to which cloud providers facilitate this process. We envision the following levels of metadata solutions:

- **Default**: The cloud customer uses cloud computing products and applies no GDPR-specific metadata tags.
- **Rights-Only Classification**: The cloud provider offers metadata tags that specify what data subject rights (e.g. portability, erasure, access,

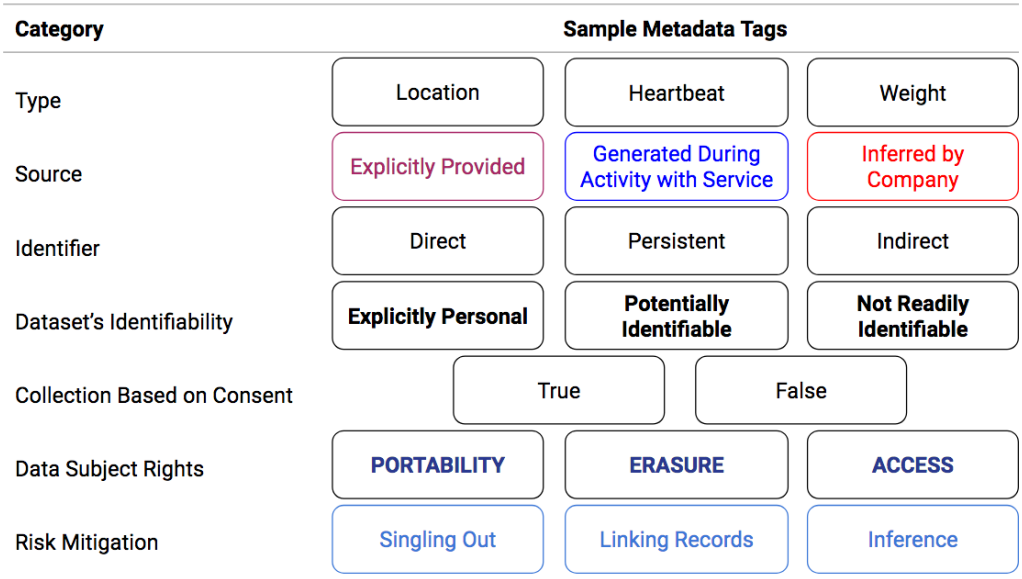

Figure 7: Proposed Metadata Categories for PaaS-Level GDPR Compliance

etc.) apply to which data. The customer uses no other part of the classification scheme outlined in Table 7. The cloud provider offers no functionality on top of these tags.

- **Portability Informed by Rights-Only Classification**: The cloud provider offers metadata tags that specify what data subject rights (e.g. portability, erasure, access, etc.) apply to which data. The customer uses no other part of the classification scheme outlined in Table 7. The cloud provider implements portability methods that read these tags to determine what to port.

- **Portability Informed by Customer-Defined Classification**: Customers design their own metadata classification scheme to support a compliance program, and they direct the cloud provider to build portability services on top of these custom tags.

- **Pre-Defined Classification**: The cloud provider gives pre-defined tags (similar to those in Table 7) to customers, who can apply the tags as they wish. The cloud provider offers no functionality on top of these tags.

- **Portability Informed by Pre-Defined Classification**: The cloud provider gives pre-defined tags (similar to those in Table 7) to customers, who can apply the tags as they wish. The cloud provider then analyzes how the combination of tags specify which data the customer believes should fall under portability obligations.

Cloud providers may be wary of pursuing these metadata solutions: the increased visibility into customer content that these strategies require would appear to bring greater legal exposure to the cloud provider. Specifically, Article 26 of the GDPR states that entities that jointly determine the purposes and means of processing personal data will be considered joint controllers. The cloud provider would then enter an agreement with its client that specifies the respective duties of each party to comply with the GDPR [3, 22]. The current approach of Amazon Web Services (AWS) best represents this reservation. In a whitepaper focused on European data protection laws, AWS states:

> "AWS has no control over what types of content the customer chooses to store in AWS and for what purposes. AWS has no insight into this content (including whether or not it includes personal data)." [38]

Thus, AWS might be averse to a metadata-based solution, which would reveal whether the customer is storing personal data in its cloud products. However, it is important to emphasize that the metadata solutions listed above require the cloud customer to decide how they want to apply metadata tags to their data. Therefore, the customer (and not the cloud provider) ultimately determines how the data will be processed, and more specifically, which data must be ported for a requesting data subject. Cloud providers would simply build portability services that respond to the way in which a customer has classified their data. Thus, cloud providers would not become data controllers under the GDPR.

In fact, Google Cloud Platform (GCP) and Microsoft Azure already offer services very similar to the classification scheme we have proposed. GCP's Data Loss Prevention (DLP) API enables automated discovery and classification of several sensitive data types, such as PII and financial data, based on identification of patterns, formats, and contextual elements. The DLP API can also use its analysis of data types to inform internal data management

and policies [39]. At the same time, GCP's Data Processing and Security Terms safely establish GCP as a processor when the GDPR applies to the cloud customer's data, and it specifies that GCP may assist with but not respond directly to portability requests [40]. Azure offers a similiar DLP tool along with an Information Protection service. Azure Information Protection operates in Microsoft's SaaS-level products and enables data classification that can be fully automated, user-driven, or based on a recommendation. Usage rights for labeled data can also be added based on administrators' policies [41].

Automated retrieval and export processes that determine what personal data must be ported based on GDPR-specific metadata is directly applicable for platform products like OSS, which allows for fine-grained, object-level metadata. In contrast, relational databases like RDS do not support distinct and fine-grained metadata entries. Rather, cloud clients must add metadata fields in each table containing personal information to describe other data fields. This means metadata would exist at the same level as the personal data it describes. Some customers may not allow a cloud provider to have such extensive visibility into their databases, precluding the use of automated processes in Portability Informed by Customer-Defined Classification and Portability Informed by Pre-Defined Classification. However, previous research suggests the creation of an entire service layer dedicated to metadata creation, placement, and editing, which may enforce enough separation between the cloud provider and the personal data [42]. In the proposed scheme, distinct metadata files stored in the "Metadata-as-a-Service" (MaaS) layer map a users query to the exact physical location where the data reside. Thus, the MaaS layer serves as a bridge between the application and PaaS storage, which operates on top of provisioned physical storage. This MaaS layer could automate retrieval of personal data that must be ported based on the tags that the controller has applied. Notably, the cited MaaS implementation also demonstrated quicker retrieval of stored files, relative to file access without metadata [42]. This would make MaaS appealing to large enterprises that may need to handle a high volume of data subject requests.

6.3. Use Case: Fitness Tracking Companies

To illustrate the utility of metadata tags for GDPR compliance, we consider a use case for companies that provide fitness tracking devices and accompanying mobile applications. In this use case, the data subject is a user

of the fitness tracking device; the data controller is the fitness tracking company; and the data processor is a cloud provider who offers capacity to the fitness tracking company for data storage and transformation.

We assume that the fitness tracking company adopts the Pre-Defined Classification approach noted above; thus, it applies pre-defined GDPR-specific metadata but implements its own portability solutions on top of those tags. We use the cloud data lifecycle [43] to organize our understanding of the collection, storage, enrichment, and transformation of a hypothetical dataset that is fully managed in the cloud (see Figure 8). Using this lifecycle also enables our discussion of when and how metadata tags can assist data controllers with GDPR compliance. We assume the same interpretation of the GDPR as declared by WP29, which encompasses more data than does the EC's understanding. To clarify what data fall under portability obligations, we additionally consider those data that a controller may need to erase upon request from a data subject. In describing the identifiability of the dataset, we utilize the categories of the FPF Identifiability Spectrum (see Figure 1). Finally, we assume that the controller employs a privacy practitioner to implement anonymization techniques in a thorough manner.

Before data collection, the controller would populate their domain-specific taxonomy of data types (Step 1) to complement the pre-defined source and identifier tags made available by the cloud provider. In this use case, the fitness tracking company creates the set outlined in Table 8 (we specify the source tags that may be associated with each data type tag to facilitate our explanation of subsequent steps in the use case). After users explicitly provide data to the device or application (Step 2) and generate data via their interaction with the company's services (Step 3), the device must encrypt and transfer these data to the controller (Step 4). Information on the fitness tracking device could be securely transferred via Bluetooth to the accompanying mobile application by employing elliptic-curve Diffie-Hellman (ECDH) public key cryptography [44]. From the mobile device, data would travel to a wireless access point through a wireless local area network (WLAN), which is typically protected by Wi-Fi Protected Access (WPA) [45]. Finally, the data would reach the controller via HTTPS, which is secured via the secure sockets layer (SSL) and transport layer security (TLS).

Upon receipt of personal data, the controller employs the taxonomy from Step 1 to apply appropriate metadata labels for data type, data source, and type of identifier (Step 5). At this stage, all direct, persistent, and indirect identifiers remain intact, and the controller can decrypt data when needed

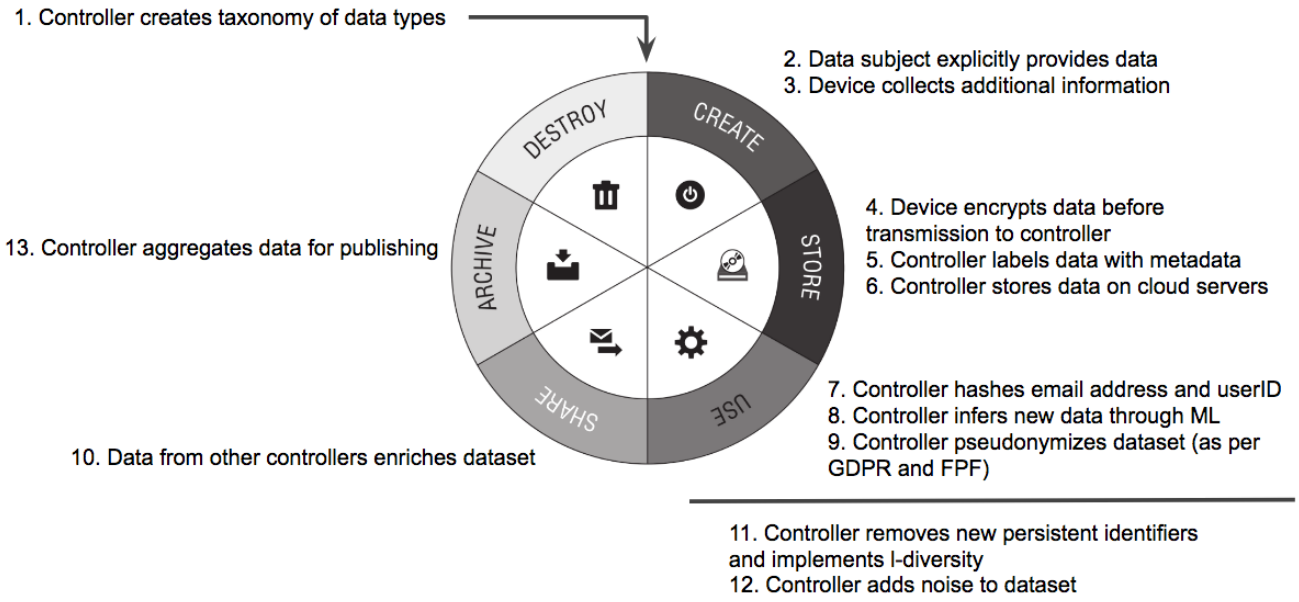

Figure 8: Transformation of Personal Information Across the Cloud Data Lifecycle (adapted from O'Hara and Masilow, [43])

Data Category	Specific Data Types		
Demographics	Age	Gender	Zip Code
Device Information	Device ID	IP Address	Internet Carrier
	Accelerometer Data	Light Absorption	
Account Information	Email	User ID	
Location Data	GPS Coordinates	Nearby Locations	Route
Health Data	Weight	Height	Steps
	Heartbeat	BPM	Calories
	Blood Pressure	Steps Per Day	Blood Glucose
	BMI	Body Composition	Gait

Table 8: Fitness Tracker Data Types

Text color in this table designates the "source" metadata tag associated with each data type. Data types in purple refer to personal data explicitly provided by the user to the company. Data types in blue refer to personal data generated during the user's interaction with the company's device and application. Data types in red refer to personal data inferred by the company based on other collected data.

for analysis. Thus, they would apply the identifiability tag of "Explicitly Personal" to this dataset. Furthermore, all data in their possession would fall under both portability and erasure obligations because the dataset consists only of those data explicitly provided by data subjects or generated during their activity with the service. The same identifiability tag and data subject obligations would hold when the controller stores data in the cloud provider's servers (Step 6).

It is important to note here that a data subject *cannot* exercise their right to portability when data has been collected without their consent [3]. For example, if the fitness tracking company collects a customer's personal data in response to an emergency involving threat of death or serious injury to any person, the company might not need to include this data when responding to that customer's portability request.

After initial storage, the controller hashes all email addresses and user IDs, both of which are direct identifiers explicitly provided by users (Step 7). While direct identifiers are now obscured, other persistent identifiers such as Device ID and IP address remain intact and thus could be used to single out an individual. The identifiability tag for the dataset would change to "Potentially Identifiable" and the full set would still be subject to portability and erasure. The controller may then infer new data through machine learning (Step 8). The identifiability of the dataset would remain the same, but the applicability of GDPR rights would diverge. While the fitness tracking company may have to erase the full record for a data subject, it would not be required to port inferred information such as the data subject's blood pressure readings or walking routes.

After inferring new data, the controller pseudonymizes the dataset according to the text of the GDPR and guidelines from the FPF (Step 9). According to the GDPR:

> *Pseudonymization* is "the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person." [GDPR Article 4(7)] [3]

The FPF further specifies that pseudonymous data has direct identifiers removed or transformed and indirect identifiers intact [30]. One way the

controller may realize pseudonymization is through placement of indirect identifiers into a bucket separate from direct and persistent identifiers. Data scientists of the fitness tracking company would have access to the former but would require approval from administrators to access and link the latter. The "Pseudonymous" tag would best describe the identifiability of the dataset after segregation of identifiers. While the GDPR highlights pseudonymization as an integral component of Privacy-by-Design [3], it does not relieve the controller of obligations to the data subject because the controller is still in a position to re-identify them. Thus, pseudonymization is not equivalent to de-identification or anonymization, and the applicable data subject rights would not change from Step 8.

The company may enrich the set of indirect identifiers using additional information from other controllers (Step 10). We consider the case in which the company can link the two datasets using age, gender, and zip code, and we assume the additional information contains Device IDs. Under these conditions, the identifiability tag would revert back to "Potentially Identifiable" given the presence of persistent identifiers. A data subject's right to erasure would apply to the full dataset, including the newly acquired information. However, their right to portability would still apply only to that which they explicitly provided or generated through activity with the fitness tracking device and application. Portability would not apply to the inferred data from Step 9 or the newly acquired data.

Once the purposes of the data have been fulfilled, a controller may wish to work toward a de-identified dataset. In this scenario, the fitness tracking company removes the Device IDs and implements l-diversity (Step 11). This anonymization technique builds upon k-anonymity, which generalizes an attribute in the dataset such that k individuals have the same value for that attribute. Specifically, l-diversity prevents deterministic inferences by creating equivalence classes in which every attribute has at least l different values. A thorough implementation of l-diversity would aim for a sufficiently high value of k and l and would account for all potential indirect identifiers [37].

According to WP29, l-diversity addresses the risk of singling out an individual and might prevent some sensitive inferences. The privacy practitioner could therefore apply the corresponding risk mitigation tags. However, l-diversity does not address the problem of linking records across datasets [37]. Thus, the controller's obligations to provide portability and erasure would not change from Step 10. Given that the dataset is not sufficiently

de-identified according to WP29 criteria, the controller might again use the "Pseudonymous" tag to describe its identifiability.

Following implementation of l-diversity, the company adds noise to the dataset (Step 12). Noise addition falls under the umbrella of randomization techniques, which alter the veracity of data to remove the strong link between the data and the individual [37]. In this use case, an appropriate level of noise would break the link between records of the same individual across multiple datasets held by the company (e.g. two distinct tables containing data collected with and without consent). The privacy practitioner could therefore apply the risk mitigation tag for "Linking Records". This in combination with l-diversity might prevent employees from re-identifying individuals if the controller cannot directly reverse the anonymization techniques and the unperturbed datasets are permanently deleted. In turn, the company may be able to apply the "De-Identified" tag to describe the dataset's identifiability. This would mean the company is free of obligations to fulfill portability and erasure requests.

Finally, the company may decide to maintain only aggregate statistics of its dataset (Step 13). If it additionally deletes the de-identified dataset, then it would apply the "Aggregate Anonymous" tag for its identifiability and it would have no obligations to provide portability and erasure.

7. SaaS-Level Portability

7.1. Cloud Providers as Controllers

In SaaS products such as email, document editing programs, and backup storage, cloud providers offer clients an out-of-the-box application to use on the cloud infrastructure. In this case they might operate as data controllers, since cloud providers own their applications and in many cases will determine the processing of information collected directly from data subjects. They would consequently have full access to both stored data and the associated metadata. While this level of visibility incurs greater responsibility under the GDPR, it also provides the flexibility to build APIs that can streamline compliance.

We offer examples of metadata-driven compliance measures through a use case very similar to that in Section 6.3, but we focus instead on personal assistants. Alibaba Tmall Genie, Microsoft Cortana, Google Home, and Amazon Echo demonstrate that major cloud providers are keen to offer

Data Category	Specific Data Types		
Demographics	Age	Gender	Zip Code
Device Information	Device ID	IP Address	Internet Carrier
Account Information	Email	User ID	Customer ID
Query-Related Data	Raw Audio Input	Raw Text Input	Device Preferences
	Ambient Sound	Derived Commands	Daily Schedule

Table 9: Personal Assistant Data Types

As in Table 8, text color designates the "source" metadata tag associated with each data type. Data types in purple refer to personal data explicitly provided by the user to the company. Data types in blue refer to personal data generated during the user's interaction with the company's device and application. Data types in red refer to personal data inferred by the company based on other collected data.

personal assistants to individuals, making this a more realistic product to examine in the SaaS context. The categories outlined in Table 7 are used again in this use case, and Table 9 provides a new set of domain-specific data types. We assume WP29's interpretation of portability rights, and the presence of a privacy practitioner that can properly implement anonymization techniques. We use the same steps through the data lifecycle as described in Figure 8, only to ground our discussion of the proposed compliance techniques.

When the cloud provider first generates its taxonomy of domain-specific data types (Step 1), it could simultaneously create metadata tag groupings. For example, the type tag for "IP Address" could be grouped with the source tag "generated during activity" and the identifier tag "persistent". Similarly, the type tag for "Derived Commands" could be grouped with the tags "inferred" and "indirect". This could provide the foundation for an API that the controller can use when labeling incoming data with relevant metadata (Step 5). Importantly, WP29's interpretation of the GDPR would classify a Customer ID that is automatically created and assigned at initial registration as something generated during a user's activity. It is not explicitly provided by the user, nor is it inferred based on other provided data. Therefore, this unique identifier might need to be ported for a data subject, even if it serves

no purpose to them.

To distinguish between data that was explicitly provided by the data subject (Step 2), generated during their activity with the personal assistant (Step 3), or inferred based on provided information (Step 8), the cloud provider can register users, devices/applications, and employees with 3 distinct account groups. Each account group would also be paired with a source tag - the user group would map to "explicitly provided", the device and application group would map to "generated during user activity", and the employee group would map to "inferred". Thus, any incoming data could be tagged appropriately based on the account group of the source. This also serves as an additional check to ensure the cloud provider grouped metadata tags properly when they created type tags (Step 1).

As a coarse-grained assessment of a dataset's identifiability, the cloud provider can design another API that checks whether the dataset contains intact direct or persistent identifiers. For example, if the Email and User ID fields are present and have not been transformed (Step 6), the API would read their identifier tags and subsequently apply the "Explicitly Personal" tag to the dataset (see Figure 9). If the Email and User ID were hashed but the Device ID was still intact (Step 7), the API would detect a data type that has been labeled as a persistent identifier and subsequently apply the "Potentially Identifiable" tag to the dataset.

A finer-grained approach would be useful once the cloud provider seeks full de-identification of the dataset (Steps 11-12). The cloud provider might design a program to determine whether anonymization techniques have been implemented correctly (Figure 10). In this case, the cloud provider would establish a new metadata category to cover anonymization techniques (e.g. "l-diversity", "k-anonymity", and "noise addition"). The privacy practitioner could then apply these process tags, which would have associated risk mitigation tags (see Table 7), after transforming the dataset. The program would read this new process tag and run an automated test to determine whether the transformation sufficiently mitigated the associated risks. For example, if the privacy practitioner failed to account for an indirect identifier or aimed for too low of a value for l when implementing l-diversity, the program would return a warning message. If the program finds no obvious shortcomings with the transformation, it would then apply the "Singling Out" and "Inference" tags to the dataset. Once all risk mitigation tags have been added, the program could then change the dataset's identifiability tag to "De-Identified". An automated program cannot replace the insight of a

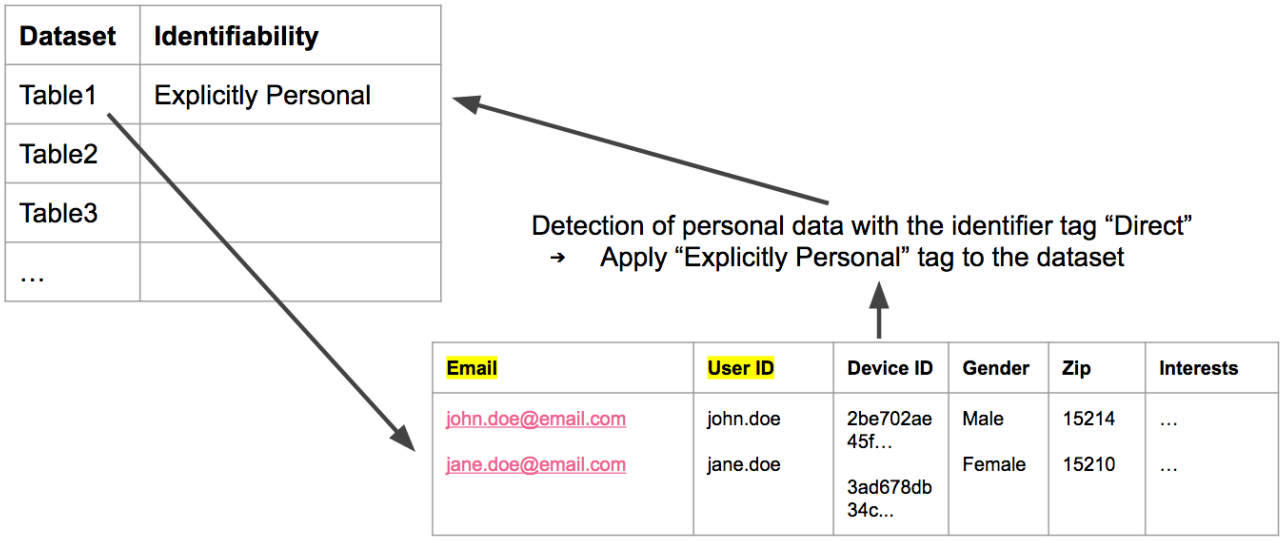

Figure 9: Automating a Coarse-Grained Assessment of Dataset Identifiability

privacy practitioner's review, but may help with detecting mistakes made clear by metadata tagging.

We take this opportunity to emphasize that our labeling of each data type in these use cases is not prescriptive. For example, we do not seek to establish that all raw audio input collected by a personal assistant will be defined in courts as data that are explicitly provided by the data subject. Such a determination is beyond the scope of this report, and we only apply this source label for demonstrative purposes. However, the fact that classification of various voice data under the GDPR remains unclear underscores the strength of a flexible compliance solution like metadata tagging. We can easily change how we label voice data to say that it is generated during a users activity with a service, and we can further revise our understanding of Article 20 to exclude this class of personal data. Changing these tags and their association with the data subject rights tags would have no effect on the technical methods for enabling portability, since those methods only need to read which data subject rights are applicable to a given dataset.

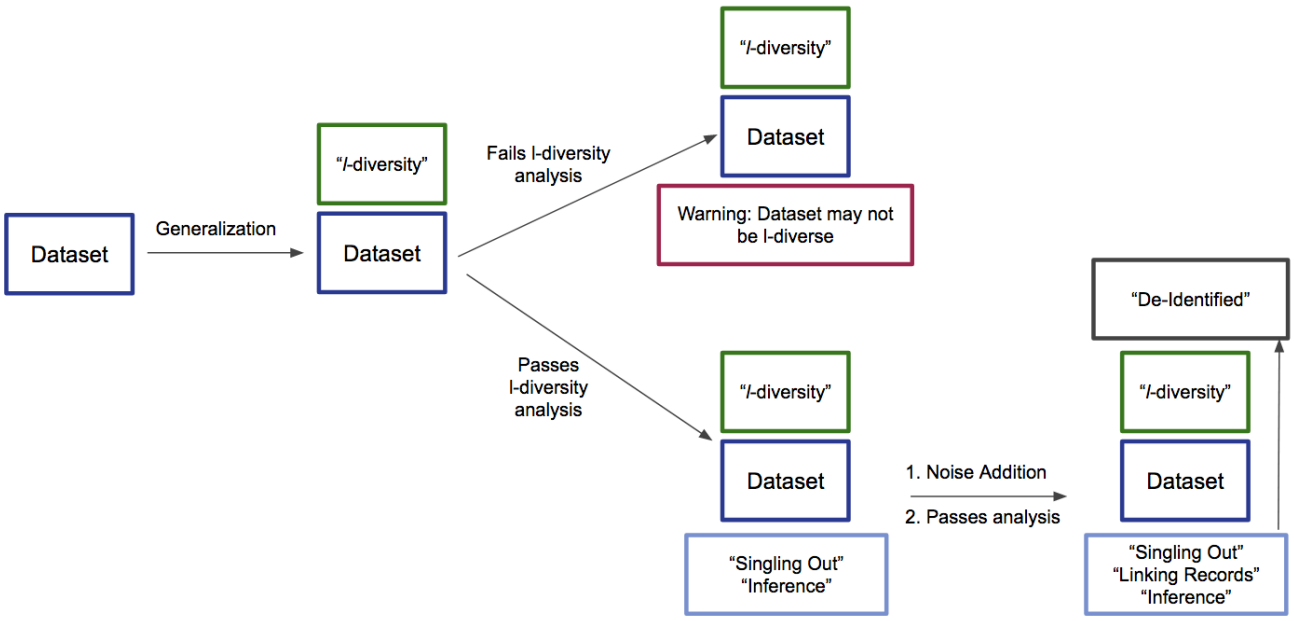

Figure 10: Automating a Fine-Grained Assessment of Anonymization Techniques

7.2. Cloud Providers as Processors

Although the users of SaaS products are typically natural persons (and thus data subjects), corporate users may also need SaaS products to support their enterprises. For example, Microsoft Excel 365 is a widely used software service that corporate customers might use for storage and processing of their customers' personal data. In this situation, the cloud provider, like Microsoft, is the data processor, the corporate user is the data controller (since it decides how to store and process the personal data), and the customers of this corporate user are the data subjects.

As stated before, data processors are not obliged to support data portability for their clients. However, with the fine-grained visibility over data stored in SaaS products, cloud providers could easily apply the same metadata-driven portability solutions described in Section 6.3. As mentioned before, GCP's DLP API and Azure's Information Protection service indicate that cloud providers already offer data classification solutions for their corporate users. Microsoft Office 365 allows corporate users to attach labels to their documents, which could be extended to classification of data based on its source, identifiability, and other relevant attributes. If GCP can discover and redact personal data for its customers, then cloud providers should be able to design an API that automatically recommend an appropriate identifiability tag for each dataset based on other applied metadata tags (as demonstrated in Section 7.1). Whether a cloud provider chooses to integrate portability-specific labeling functions into SaaS products or build a separate tool for PII detection, both methods have precedent in the market. GDPR tagging would not be a significant leap beyond currently available services.

8. Conclusion

If cloud providers want to continue providing competitive products to both their individual and corporate clients, they must consider how the right to data portability translates to their operations as controllers and processors. Their compliance solutions must also remain flexible due to the lack of certainty in GDPR enforcement. This report presents a set of recommendations for meeting this demand. The recommendations are based on careful consideration of how the recipient of data from a portability request, the cloud service level, and one's interpretation of the GDPR determine the appropriate portability solution. Most importantly, they rely on existing

technical methods and build upon precedent in the marketplace, meaning that a compliance program for GDPR Article 20 is well within reach.

[1] Proposal for a regulation of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation), Council of the European Union (2015).

[2] G. Zanfir, The right to data portability in the context of the EU data protection reform, International Data Privacy Law 2 (2012).

[3] General Data Protection Regulation (GDPR), https://gdpr-info.eu, 2016. Accessed: 2017-09-27.

[4] Guidelines on the right to data portability, Article 29 Data Protection Working Party 242 (2017).

[5] DPA of Argentina issues draft data protection bill, https://www.huntonprivacyblog.com/2017/02/09/dpa-argentina-issues-draft-data-protection-bill/, 2017. Accessed: 2017-10-01.

[6] Philippines finalizes data privacy act implementing rules, http://www.hldataprotection.com/2016/09/articles/international-eu-privacy/philippines-finalizes-data-privacy-act-implementing-rules/, 2016. Accessed: 2017-10-01.

[7] J. Mazur, M. Palinski, M. Sobolewski, GDPR: A step towards a user-centric internet?, Intereconomics 52 (2017) 207–213.

[8] IAPP-EY Annual Privacy Governance Report 2017, The International Association of Privacy Professionals (2017). Accessed: 2017-11-13.

[9] D. Meyer, European Commission, experts uneasy over WP29 data portability interpretation, https://iapp.org/news/a/european-commission-experts-uneasy-over-wp29-data-portability-interpretation/, 2017. Accessed: 2017-09-27.

[10] 29 U.S. Code 1181 - Increased portability through limitation on preexisting condition exclusions, https://www.law.cornell.edu/uscode/text/29/1181, 2015. Accessed: 2017-09-27.

[11] Gramm-Leach-Bliley Act, https://www.gpo.gov/fdsys/pkg/PLAW-106publ102/pdf/PLAW-106publ102.pdf, 1999. Accessed: 2017-09-27.

[12] Personal Information Protection and Electronic Documents Act (PIPEDA), http://laws-lois.justice.gc.ca/eng/acts/P-8.6/index.html, 2016. Accessed: 2017-09-27.

[13] Canada Health Act Annual Report 2002-2003, https://www.canada.ca/en/health-canada/services/health-care-system/reports-publications/canada-health-act-annual-reports.html, 2017. Accessed: 2017-09-27.

[14] Apec Privacy Framework, https://www.apec.org/Groups/Committee-on-Trade-and-Investment/~/media/Files/Groups/ECSG/05_ecsg_privacyframewk.ashx, 2005. Accessed: 2017-09-27.

[15] Federal Law on the Protection of Personal Data Held by Private Parties, http://media.mofo.com/files/uploads/Documents/FederalDataProtectionLaw2010.pdf, 2010. Accessed: 2017-09-27.

[16] Data Protection Law (DIFC Law No. 1), https://www.difc.ae/files/7814/5517/4119/Data_Protection_Law_DIFC_Law_No._1_of_2007.pdf, 2007. Accessed: 2017-09-27.

[17] Amended Act on the Protection of Personal Information, http://www.ppc.go.jp/files/pdf/280222_amendedlaw.pdf, 2016. Accessed: 2017-09-27.

[18] Personal Data Ordinance, https://www.pcpd.org.hk/english/data_privacy_law/6_data_protection_principles/principles.html, 2014. Accessed: 2017-09-27.

[19] Personal Data Protection Act, https://www.pdpc.gov.sg/legislation-and-guidelines/legislation, 2016. Accessed: 2017-09-27.

[20] G. Greenleaf, W.-i. Park, Korea's new act: Asia's toughest data privacy law, Privacy Laws & Business International Report (2012).

[21] Australian Privacy Principles, https://www.oaic.gov.au/privacy-law/privacy-act/australian-privacy-principles, 2014. Accessed: 2017-09-27.

[22] A. Myers, Top 10 operational impacts of the GDPR: Part 7 - Vendor Management, https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-7-vendor-management/, 2016. Accessed: 2017-10-01.

[23] S. Welbergen, Proposed EU Regulation on the free flow of non-personal data, https://www.lexology.com/library/detail.aspx?g=ccdb273e-fc02-42a2-80dc-9d5bd2733e3b, 2017. Accessed: 2017-10-29.

[24] D. S. Michael Corey, Does GDPR spell the end of the cloud as we know it today?, Big Data Quarterly 2 (2016).

[25] R. Heimes, When is a vendor a processor?, https://iapp.org/news/a/dpo-confessional-when-is-a-vendor-a-processor/, 2017. Accessed: 2017-10-01.

[26] L. Urquhart, N. Sailaja, D. McAuley, Realising the right to data portability for the domestic Internet of Things (2017).

[27] M. van Schaijck, GDPR Top Ten: #1 Data portability - legal obstacle or opportunity?, https://www2.deloitte.com/nl/nl/pages/risk/articles/gdpr-top-ten-1-data-portability.html, 2017. Accessed: 2017-09-27.

[28] L. Edwards, M. Veale, Slave to the algorithm? Why 'a right to an explanation is probably not the remedy you are looking for, Duke Law and Technology Review (2017).

[29] M. Hintze, Viewing the GDPR through a de-identification lens: A tool for compliance, clarification, and consistency, Brussels Privacy Symposium (2016).

[30] J. Polonetsky, O. Tene, K. Finch, Shades of gray: Seeing the full spectrum of practical data de-identification, Santa Clara Law Review 56 (2016) 593–629.

[31] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility, Future Gener. Comput. Syst. 25 (2009) 599–616.

[32] P. M. Mell, T. Grance, SP 800-145. The NIST Definition of Cloud Computing, Technical Report, Gaithersburg, MD, United States, 2011.

[33] A. Ranabahu, A. Sheth, Semantic modeling for cloud computing, part 1, IEEE Internet Computing 14 (2010) 81–83.

[34] A. Ranabahu, A. Sheth, Semantics centric solutions for application and data portability in cloud computing, 2010 IEEE Second International Conference on Cloud Computing Technology and Science (2010) 234–241.

[35] Web application hosting, `https://www.alibabacloud.com/solutions/hosting/Web-Application-Hosting`, 2017. Accessed: 2017-11-01.

[36] Locating customer data will be half the battle to fulfill EU GDPR's 'right to be forgotten', Database and Network Journal 47 (2017). Accessed: 2017-10-01.

[37] Opinion 05/2014 on anonymisation techniques, Article 29 Data Protection Working Party 216 (2014).

[38] Whitepaper on EU Data Protection, `https://d1.awsstatic.com/whitepapers/compliance/AWS_EU_Data_Protection_Whitepaper.pdf`, 2016. Accessed: 2017-11-12.

[39] Data Loss Prevention API Beta, `https://cloud.google.com/dlp/`, 2017. Accessed: 2017-11-12.

[40] Data Processing and Security Terms, `https://cloud.google.com/terms/data-processing-terms`, 2017. Accessed: 2017-11-12.

[41] Azure Information Protection, `https://www.microsoft.com/en-us/cloud-platform/azure-information-protection`, 2016. Accessed: 2017-11-12.

[42] R. Anitha, S. Mukherjee, MaaS: Fast retrieval of data in cloud using Metadata as a Service, Arabian Journal for Science and Engineering 40 (2015) 2323–2343. Accessed: 2017-10-01.

[43] B. O'Hara, B. Malisow, CCSP (ISC)2 Certified Cloud Security Professional: Official Study Guide, First Edition, John Wiley & Sons, Inc., 2017.

[44] J. Padgette, K. Scarfone, L. Chen, SP 800-121. Guide to Bluetooth Security: Recommendations of the National Institute of Standards and Technology, Technical Report, Gaithersburg, MD, United States, 2012.

[45] S. Frankel, B. Eydt, L. Owens, K. Scarfone, SP 800-97. Establishing Wireless Robust Security Networks: A Guide to IEEE 802.11i, Technical Report, Gaithersburg, MD, United States, 2007.