INTERNATIONAL STANDARD

ISO/IEC 19944

First edition
2017-09

Information technology — Cloud computing — Cloud services and devices: Data flow, data categories and data use

Technologies de l'information — Informatique en nuage — Services et dispositifs en nuage : Débits, catégories et utilisation des données

Reference number
ISO/IEC 19944:2017(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 38, *Cloud computing and distributed platforms*.

Introduction

Objective and target audience

This document provides a description of the ecosystem of devices and cloud services and the related flows of data between cloud services, cloud service customers, cloud service users and their devices. These are necessary to provide guidance about how data is used on the devices in the context of the cloud computing ecosystem and the associated location and identity issues that emerge from such use.

This document proposes a scheme for the structure of data use statements that can be used by cloud service providers to help cloud service customers understand and protect the privacy and confidentiality of their data and their users' data through increased transparency of policies and practices.

This document can be used in several ways including, but not limited to, the following:

a) by cloud service providers and application developers to guide them in describing what they intend to do with data in their designs, so as to simplify privacy and data use reviews and to communicate this information to non-technical departments such as internal compliance, marketing and legal teams;

b) by organizations drawing up data use statements as part of drafting cloud service agreements and application contracts, privacy statements, etc., which could apply to documents internal to an organization, in addition to public or legal documents;

c) by government regulators and agencies to advise on suitable ways of describing data flow and use;

d) by those preparing information on data flow and data use for communication to the press and the public.

This document is descriptive and not prescriptive. It cannot be used for compliance directly. Instead, it provides a set of concepts and definitions, including a data taxonomy and data use statement structure, that can be used for transparency about how data is used in an ecosystem of devices and cloud services.

Providing a clear description of data flows

This document aims to improve the understanding of the data flows that take place in an ecosystem consisting of devices accessing cloud services. It does this through an extended cloud computing reference architecture (CCRA) (based on the architecture described in ISO/IEC 17789) that describes the impact of devices on cloud service ecosystems and the impact of cloud services on devices. It also describes the data flows that take place within the extended reference architecture.

Providing transparency to all stakeholders

To maintain a relationship of trust between the stakeholders of the ecosystem of devices and cloud services and also to meet the demands of laws and regulations, it is necessary for the device platform providers and the cloud service providers to be transparent about how they make use of the various data types that flow within the ecosystem.

There is a particular need to provide simple and clear statements to end users about what is done with data that relates to them. The data may be personally identifiable information (PII) and may be sensitive, in other words, this can be a privacy issue. Cloud service customers are likely to be concerned about how their data is used, even when the customer is an organization rather than an individual. The cloud service customer may be a data controller, holding personal data about their employees or their customers; in such a role, the cloud service customer has obligations relating to the processing of that data.

To assist cloud service providers and device platform providers in being transparent about their use of data, this document defines a simple language for making statements about data use, which can be used to create clear notification to end users and other interested parties.

Information technology — Cloud computing — Cloud services and devices: Data flow, data categories and data use

1 Scope

This document

— extends the existing cloud computing vocabulary and reference architecture in ISO/IEC 17788 and ISO/IEC 17789 to describe an ecosystem involving devices using cloud services,

— describes the various types of data flowing within the devices and cloud computing ecosystem,

— describes the impact of connected devices on the data that flow within the cloud computing ecosystem,

— describes flows of data between cloud services, cloud service customers and cloud service users,

— provides foundational concepts, including a data taxonomy, and

— identifies the categories of data that flow across the cloud service customer devices and cloud services.

This document is applicable primarily to cloud service providers, cloud service customers and cloud service users, but also to any person or organization involved in legal, policy, technical or other implications of data flows between devices and cloud services.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
cloud service
one or more capabilities offered through cloud computing invoked using a defined interface

[SOURCE: ISO/IEC 17788:2014, 3.2.8]

3.2
cloud service customer
party which is in a business relationship for the purpose of using *cloud services* (3.1)

Note 1 to entry: A business relationship does not necessarily imply financial agreements.

[SOURCE: ISO/IEC 17788:2014, 3.2.11]

3.3
cloud service partner
party which is engaged in support of, or auxiliary to, activities of either the *cloud service provider* (3.4) or the *cloud service customer* (3.2), or both

[SOURCE: ISO/IEC 17788:2014, 3.2.14]

3.4
cloud service provider
party which makes *cloud services* (3.1) available

[SOURCE: ISO/IEC 17788:2014, 3.2.15]

3.5
cloud service user
natural person, or entity acting on their behalf, associated with a *cloud service customer* (3.2) that uses *cloud services* (3.1)

Note 1 to entry: Examples of such entities include devices and applications.

[SOURCE: ISO/IEC 17788:2014, 3.2.17]

3.6
device
physical entity that communicates directly or indirectly with one or more *cloud services* (3.1)

3.7
account data
class of data specific to each CSC that is required to administer the *cloud service* (3.1)

Note 1 to entry: Account data is typically generated when a cloud service is purchased and is under the control of the CSP.

Note 2 to entry: Account data consists of data elements provided by CSC, such as; name, address, telephone, etc.

3.8
cloud service customer data
class of data objects under the control of the *cloud service customer* (3.2) that were input to the *cloud service* (3.1), or resulted from exercising the capabilities of the cloud service by or on behalf of the cloud service customer through the published interface of the cloud service

Note 1 to entry: An example of legal controls is copyright.

Note 2 to entry: It may be that the cloud service contains or operates on data that is not cloud service customer data; this might be data made available by the cloud service providers, or obtained from another source, or it might be publicly available data. However, any output data produced by the actions of the cloud service customer using the capabilities of the cloud service on this data is likely to be cloud service customer data, following the general principles of copyright, unless there are specific provisions in the cloud service agreement to the contrary.

[SOURCE: ISO/IEC 17788:2014, 3.2.12]

3.9
cloud service derived data
class of data objects under *cloud service provider* (3.4) control that are derived as a result of interaction with the *cloud service* (3.1) by the *cloud service customer* (3.2)

Note 1 to entry: Cloud service derived data includes log data containing records of who used the service, at what times, which functions, types of data involved and so on. It can also include information about the numbers of authorized users and their identities. It can also include any configuration or customization data, where the cloud service has such configuration and customization capabilities.

[SOURCE: ISO/IEC 17788:2014, 3.2.13]

3.10
cloud service provider data
class of data objects, specific to the operation of the *cloud service* (3.1), under the control of the *cloud service provider* (3.4)

Note 1 to entry: Cloud service provider data includes but is not limited to resource configuration and utilization information, cloud service specific virtual machine, storage and network resource allocations, overall data centre configuration and utilization, physical and virtual resource failure rates, operational costs and so on.

[SOURCE: ISO/IEC 17788:2014, 3.2.16]

3.11
application marketplace
set of *cloud services* (3.1) providing a digital marketplace intended to offer applications and other digital content for a particular *device platform* (3.13) allowing users to browse and download applications and other content

Note 1 to entry: An application marketplace may be offered to the public, or to private groups such as a corporate environment.

Note 2 to entry: A *device* (3.6) can use more than one application marketplace.

3.12
application cloud service
cloud service (3.1) that supports applications running on a given *device* (3.6), where the cloud service is provided by a party other than the *device platform provider* (3.14)

3.13
device platform
operating system and related feature set that provide the core capabilities for a *device* (3.6)

Note 1 to entry: An *application marketplace* (3.11) is specific to a device platform.

3.14
device platform provider
device platform cloud service provider
cloud service provider (3.4) that provides *cloud services* (3.1) necessary to support a *device platform* (3.13) including managing needed digital identities

Note 1 to entry: The cloud service provider that offers the *application marketplace* (3.11) is typically the same as the device platform provider, but it is not required to be.

3.15
device platform cloud service
cloud service (3.1) offered by the *device platform provider* (3.14) to support the *device platform* (3.13)

Note 1 to entry: An *application marketplace* (3.11) can be an example of device platform cloud service.

3.16
personally identifiable information
PII
any information that a) can be used to identify the *PII principal* (3.18) to whom such information relates, or b) is or might be directly or indirectly linked to a PII principal

[SOURCE: ISO/IEC 29100:2011, 2.9]

3.17
PII controller
privacy stakeholder (or privacy stakeholders) that determines the purposes and means for processing *personally identifiable information* (PII) (3.16) other than natural persons who use data for personal purposes

Note 1 to entry: A PII controller sometimes instructs others, e.g. *PII processors* (3.19) to process PII on its behalf while the responsibility for the processing remains with the PII controller.

[SOURCE: ISO/IEC 29100:2011, 2.10]

3.18
PII principal
natural person to whom the *personally identifiable information* (PII) (3.16) relates

Note 1 to entry: Depending on the jurisdiction and the particular PII protection and privacy legislation, the synonym "data subject" can also be used instead of the term "PII principal".

[SOURCE: ISO/IEC 29100:2011, 2.11]

3.19
PII processor
privacy stakeholder that processes *personally identifiable information* (PII) (3.16) on behalf of and in accordance with the instructions of a *PII controller* (3.17)

[SOURCE: ISO/IEC 29100:2011, 2.12]

3.20
end user identifiable information
EUII
derived data associated with a user that is captured or generated from the use of the service by that user

4 Abbreviated terms

BYOD	Bring Your Own Device
CCRA	Cloud Computing Reference Architecture
CSA	Cloud Service Agreement
CSC	Cloud Service Customer
CSN	Cloud Service partner
CSP	Cloud Service Provider
CSU	Cloud Service User
EUII	End User Identifiable Information
GPS	Global Positioning System
IaaS	Infrastructure as a Service
PII	Personally Identifiable Information
SLA	Service-Level Agreement

5 Structure of this document

This document is organized to describe two topic areas.

— Overview and reference architecture (Clauses 6 and 7).

— Data taxonomies, data categories and data use statement structure (Clauses 8, 9 and 10).

Overview and reference architecture

— Clause 6 provides the foundation of the document covering the "Overview of devices and cloud services ecosystems". The clause describes the ecosystem and stakeholders where devices and cloud services operate.

— Clause 7, "Extending the cloud computing reference architecture to the devices and cloud services ecosystem" covers an extension of the architecture specified in ISO/IEC 17789[2] to include devices and the flow of data between devices and cloud services.

Data taxonomies, data categories and data use statement structure (applicable to data exchanges between devices and cloud services)

— Clause 8, "Data taxonomies" describes categories of data that can be captured, processed, used and shared. This taxonomy extends the definitions in ISO/IEC 17788[1] of cloud service customer data, cloud service derived data, cloud service provider data and account data. The taxonomy described in this clause is used in creating data use statements covered in Clause 10.

— Clause 9, "Data processing and use categories" describes the various categories of data processing and operations. "Data use categories" and related "scopes" described in this clause are required for understanding of the data use statements structure covered in Clause 10.

— Clause 10, "Data use statements" describes the syntax and statement structure for expressing how data is used by CSPs and their partners.

6 Overview of devices and cloud services ecosystems

6.1 Background and context — Impact of devices and personalized cloud services

This document builds on the foundation provided by the CCRA, ISO/IEC 17789, to accommodate data and its flow within the ecosystem of devices and cloud services.

Many kinds of devices are used as clients for accessing cloud services. These devices rely on support from cloud services which have an association between the device and the cloud service. Unique identifiers are created and maintained to enable that association. The interaction between the device and the cloud service requires an understanding of the flow of data between devices, cloud services, cloud service customer and cloud service providers. This interaction also makes the discussion of data classification, access and use more complex.

NOTE This document uses the term "device" in the context of a cloud service user as defined in ISO/IEC 17788:2014, 3.2.17, which includes natural person, or entity acting on their behalf. Examples of such entities include devices and applications. This document is written such that there is no conceptual difference between types of devices, provided the device is acting as a cloud service user using cloud services.

Cloud service providers offering device specific cloud services typically require a unique identifier and a cloud service user account in order to provide those cloud services. This identifier and user combination becomes the cloud service user's key to their own personalized cloud services which can offer an array of services, access to applications, rich advertising and retail infrastructure.

The always-on, always-with-me nature of some devices drives a new class of applications for personal use that strive to assist users with every aspect of their daily lives by making useful suggestions based on a trail of information flowing from the device and from applications running on the device.

For example, a mobile device user's interaction with the device platform cloud services may offer the device platform provider a very detailed trail of behavioural data, including user communications, contacts, calendar, whereabouts and searches and purchases.

6.2 Ecosystem of devices and cloud services

This clause describes an ecosystem of cloud-supported devices and cloud services. Figure 1 depicts a common way of how a device may operate in a cloud environment. The cloud services used by devices come in several categories. The categories of cloud services used by devices and covered in this document are as follows.

— **Device platform cloud service** (see 3.15) which can include application marketplace (see 3.11). These "core" cloud services are offered by the device platform provider and used to configure the device and register the customer (and where appropriate, the primary user of the device) with the application marketplace and associated cloud services, including online user identity management. This is depicted by the upper cloud in the diagram in Figure 1 and corresponds with the sub-role "device platform provider" defined in 7.3.1.1.2.

— **Application cloud service** (see 3.12) which supports the applications developed and supported by cloud service providers (e.g. social networking, weather, news or organization-specific applications) that are not the device platform CSP. Such applications interact with their own cloud services, distinct from the cloud services provided to support the device platform. This is depicted by the lower cloud in the diagram in Figure 1 and corresponds with the role of cloud service provider defined in ISO/IEC 17789:2014, 8.3.1.

Both categories involve interactions with the device and carry data traffic, potentially including cloud service customer data or end user identifiable information (EUII). For example, the application marketplace knows which applications have been downloaded on the device and the device platform knows how often they are invoked and how long they are used.

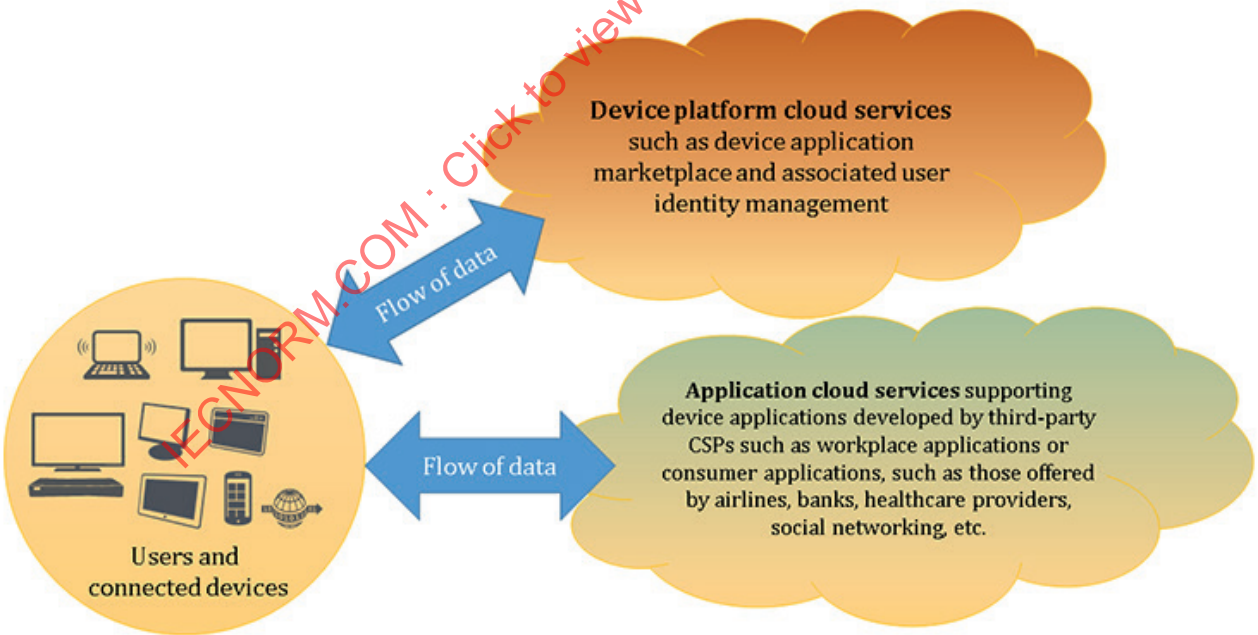

Figure 1 — Devices and cloud services ecosystem

Most tablets, smartphones and other connected devices are often connected to their device platform cloud services in order to be fully functional. This connectivity and flow of data is depicted by the arrow to the upper cloud in Figure 1, although some IoT devices may not communicate to the device platform cloud services directly. At the same time, the devices are also connected to various cloud services,

depicted in the diagram by the lower cloud, that support the applications developed and supported by cloud service providers. This connectivity and flow of data is shown by the arrow to the lower cloud in Figure 1.

6.3 Devices and multiple user sub-roles

6.3.1 General

Device users typically use the same device while assuming various roles in their daily lives, often concurrently as shown Figure 2, a citizen/voter consuming city/government services, a patient receiving medical services at a doctor's office or a hospital, a student attending school, a motorist or commuter on the road, a consumer in a mall/coffee shop/store, a passenger in the airport or train station, in addition to being an employee.

Citizens, students, patients and employees, for example, each have unique requirements and needs for data and privacy protection. Nevertheless, each user sub-role will use the same personal device including the device's local storage, which can potentially be part of the same device application marketplace(s) ecosystem and will use the same device services offered by the device's operating system.

Device provider, device services and applications, as well as cloud service providers providing the applications on the device may have visibility into the device users' actions, data and their use of applications and services. Such visibility to user data could continue as users assume multiple sub-roles throughout their use of the device and use multiple applications such as those developed for workplace use (employees), government and citizen use (voters, taxpayers, etc.), schools (students) or healthcare (patients). The user's data may be collected, stored, processed and used by the cloud service providers. In contrast, for some applications and some cloud services, the user might take the sub-role of anonymous user, where the user wants the right to use the application and cloud services in a private manner, where the user's identity and the user's personal information are deliberately not shared with the application and with the cloud service. While technologies such as application containers/sandboxes, application-specific encryption and application-specific VPNs can mitigate this, there is still a need for a data taxonomy that categorizes data in a harmonized and consistent fashion so as to enable a meaningful conversation between the cloud service customers, the cloud service providers, regulators and other stakeholders about this data.

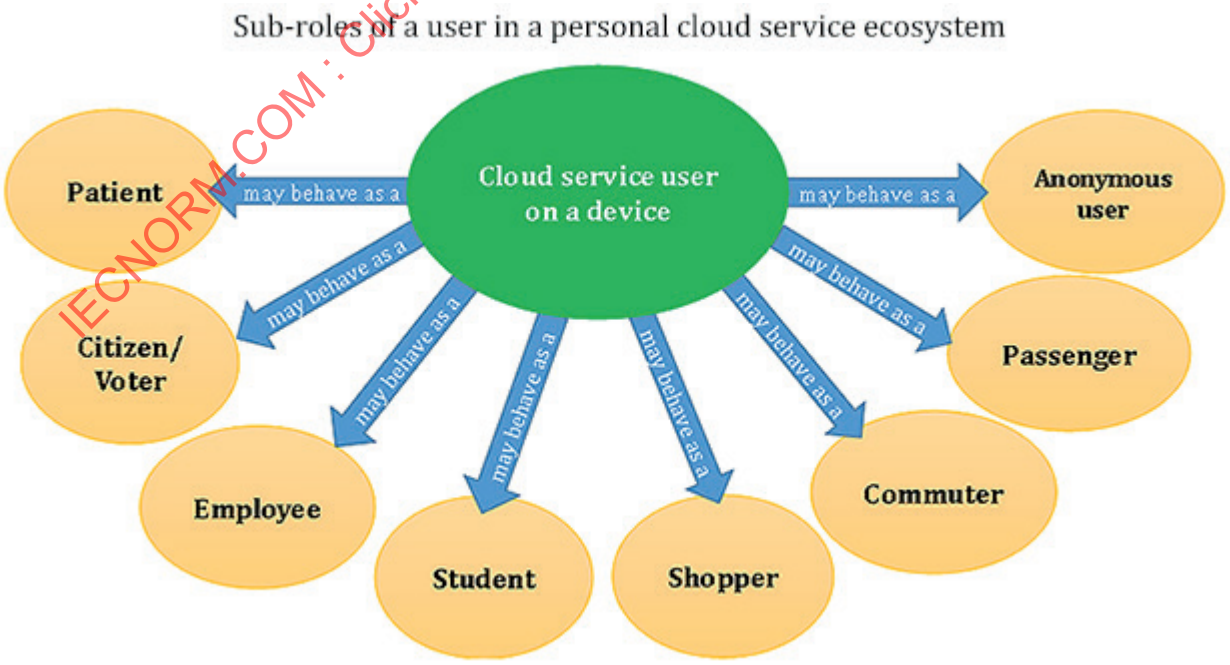

Figure 2 — Example of roles a user can assume in device use scenarios

The following is a non-exhaustive and informative list of sub-roles that help describe device scenarios and issues.

— Patient: patients are under healthcare privacy laws.

— Citizen: all aspects of an individual's relationship with government and public authorities, including voting and obligations to and benefits received from government.

— Employee: they should follow the organization's policies to protect the organization's confidential assets.

— Students: many students are under-age and therefore they are under stricter privacy and commercial advertisement laws.

— Shopper: data, such as payment instrument data, personal favourites, shopping locations, personal financial information, can be collected and processed during shopping. Such data can be relevant to privacy.

— Commuter: the flow of commuter's personal data could also be examined while the user utilizes data services offered while in transit.

— Passenger: passengers are in public transportation hubs like airports, and therefore, certain rules may apply.

— Anonymous user: where the user does not wish to share any personal data with the application and with the cloud services, including identity.

6.3.2 Bring your own device (BYOD)

"Bring your own device" (BYOD) is defined as the practice of allowing employees of an organization to use their own computers, smartphones, tablets or other devices for work purposes. BYOD is a particular case of mixing different roles when using a device where the user has the role of an employee or partner of an organization.

In the past, it was common for organizations to provide the devices that employees used mainly or even exclusively for work purposes and those devices were connected to the organization's networks and used the organization applications and systems. Organization-owned devices are typically tightly controlled in terms of the installed software, both in terms of the software that can be installed by employees and in the requirement to run a variety of management and security components including firewalls, malware checking programs, encryption of stored data and so on.

The main concern for organizations is to ensure that the organization's applications, systems and data are secure and are only used for authorized purposes, so any employee devices with access to corporate assets are controlled to ensure the integrity of organization systems.

The introduction of mobile devices such as smartphones and tablets changed the IT landscape significantly. These mobile devices are very popular and employees see them as helping them do productive work both outside the office and within the office. This leads to a demand from employees to use their personal/private mobile devices to access the organization's applications and systems. Employees do not want to have multiple different devices (one their own, another owned by the organization) since this can be burdensome and difficult to manage.

BYOD encompasses not only employees but also other users with a close relationship to the organization, such as business partners.

A mobile device user remains connected to their personalized cloud services even when they bring their own device into an organizational setting where they use organization-specific applications, systems and networks even as the device runs applications not belonging to the organization and connects with cloud services not belonging to the organization. The organization's own client applications running on the device may also use functions and rely on services from the device platform cloud services or elsewhere. That interaction is also captured and associated with the user's digital identity or the

device's identifier. Instead of a simple client-server interaction, there is the potential for intertwined flow of data between the device, the device platform cloud services, organization applications, other applications installed by the user and the organization's cloud services. The major issues are the potential for leakage of enterprise data and the potential for data of doubtful provenance to be transmitted to the organizations' cloud services and/or internal systems.

Organizational Information Technology (IT) managers need to protect intellectual property and confidential data against unauthorized disclosure or leakage and, as such, may demand tight control over a user's own device when that person is interacting with the organization as an employee or in another role. Additional information on the security threats can be found in ISO/IEC 27033-3:2010, Clause 13[5] Organizational users and their IT managers would benefit from deeper understanding of BYOD scenarios affecting security and confidentiality of organizational data when device users assume other roles when using the same device (for example, as an employee, a student, a patient, a consumer). Effectively, the need is to partition the use of the device, with organization applications and data separated by secure boundaries from other applications and data.

For organizations, BYOD brings some challenges, mostly relating to the security of organization applications and data when personal devices are used. The main risks can be summarized as follows.

— Loss of control over access to organization applications from the device, a personal device may be shared with others.

— Vulnerability of organization data which is downloaded and stored on the device, there is potential for loss, theft and unauthorized alteration of the data.

— Use of non-organization applications and cloud services on the device:

  a) to use or transmit or share or store organization data

  b) which may be used to access organization systems and applications.

— Malware on the device stealing important data including identities and credentials.

7 Extending the CCRA to the devices and cloud services ecosystem

7.1 Overview

The devices and cloud services ecosystem requires extensions to the CCRA described in ISO/IEC 17789.

Expansion of the description of the functional components in the User layer is required in order to describe a number of components which relate to mobile devices. This is particularly important to understand the data flows that take place within the ecosystem. There is an associated expansion in the cloud service customer role and its sub-roles to describe additional activities and responsibilities that exist when devices are used with cloud services. Similar extensions of the cloud service provider role and its sub-roles are also necessary.

7.2 Personal and organizational environments

The cloud services and associated applications are designed for a variety of uses. Applications and cloud services designed for the personal use of the end user form part of the "personalized cloud services" of the end user. Applications and cloud services, designed for use as part of the function of an organization to which the end user has a relationship (e.g. employee or partner), can be described as "business capabilities" or "organizational capabilities".

Personal use applications and cloud services are very likely to involve the case where the end user performs all of the roles defined for a cloud service customer, with a need for simple interfaces to allow necessary administration and management capabilities.

Organizational use applications and cloud services by contrast very likely separate out the interfaces for the different roles for a cloud service customer, since it is highly likely that the end user is not the same person as the cloud service administrator, for example.

7.3 Device impact on the CCRA: User view

7.3.1 Cloud service provider

7.3.1.1 Sub-roles

7.3.1.1.1 General

The cloud service provider role is defined in ISO/IEC 17789:2014, 8.3.1. A cloud service provider can make cloud services which are usable with any device. However, devices usually have a special relationship with one particular cloud service provider, the device platform provider; therefore, there is the need to define a new sub-role to accommodate this.

7.3.1.1.2 CSP:device platform provider

The CSP:device platform provider is a sub-role of cloud service provider that provides the set of cloud services necessary to support the device platform. The party that offers the cloud services for the application marketplace is typically the same as the party that plays the CSP:device platform provider sub-role, but this is not necessarily the case.

The CSP:device platform provider typically offers the cloud services necessary to provide identity management for the user of the device. This is usually done in conjunction with the application marketplace.

The device platform provider's cloud computing activities include:

— providing data and applications;

— sharing data with third parties;

— processing and using data;

— providing application marketplace;

— providing device platform cloud services;

— providing data related services.

7.3.1.2 Cloud computing activities

In addition to the cloud computing activities specified in ISO/IEC 17789:2014, 8.3.2, the following activities apply to the sub-roles of CSP.

— Providing data and applications: makes provider data and applications available to cloud service customers under a cloud service agreement.

— Sharing data: makes customer content data and derived data available to third party organizations under an agreement, for business purposes of the cloud service provider.

— Processing and using data: processes customer content data and derived data for certain purposes, for instance advertising, business intelligence, security and privacy, under terms stated in the cloud service agreement.

— Providing application marketplace: provide and maintain the application marketplace. This includes the applications which run on devices and the set of cloud services which support the application.

— Providing device platform cloud services: provide the set of cloud services necessary to support a device platform.

— Providing data related services: involves the providing of data related services to cloud service customers and cloud service users such as online advertisements or business intelligence.

7.3.2 Cloud service customer

7.3.2.1 Sub-roles

7.3.2.1.1 General

ISO/IEC 17789:2014, 8.2.1 and 8.2.1.1 specify the role of cloud service customer and its sub-role CSC:cloud service user. Both apply to this document.

According to ISO/IEC 17788 and ISO/IEC 17789, the cloud service customer is a party in a business relationship for the purpose of using cloud services, whereas the cloud service user, as the actual person using a particular device, is a sub-role of cloud service customer which uses the cloud services. In organization scenarios, the cloud service customer is the organization and the cloud service users are the individual employees of the organization.

There are other cases where the cloud service users may not be employed by the organization but have another type of relationship with the cloud service customer for example, the cloud service users may be customers of the cloud service customer organization.

In other cases, one person may be the customer of a cloud service but the cloud service users are a number of people who have a non-business relationship with the customer (such as a home movie streaming service).

There are consumer scenarios where the cloud service customer is the same person as the cloud service user and, in this case, devices, applications and services are all linked to the device users through their customer accounts.

The term “CSC:cloud service user” is synonymous with “device user” used elsewhere in this document

7.3.2.1.2 CSC:cloud service user

The role specified in ISO/IEC 17789:2014, 8.2.1.1 applies.

7.3.2.2 Cloud computing activities

The cloud computing activities specified in ISO/IEC 17789:2014, 8.2.2, which relate to the sub-roles of cloud service customer apply and are extended to include:

— providing customer data: makes customer data available to cloud service provider under an agreement;

— using data: uses data obtained from cloud services on their devices;

— installing applications on mobile devices: download and install applications on end user devices.

7.4 Device impact on the CCRA: Functional view

7.4.1 General

Figure 3 provides a functional view of the “devices and cloud services” ecosystem for the purposes of identifying key data flows between functional components present on the device and those of the various cloud services.

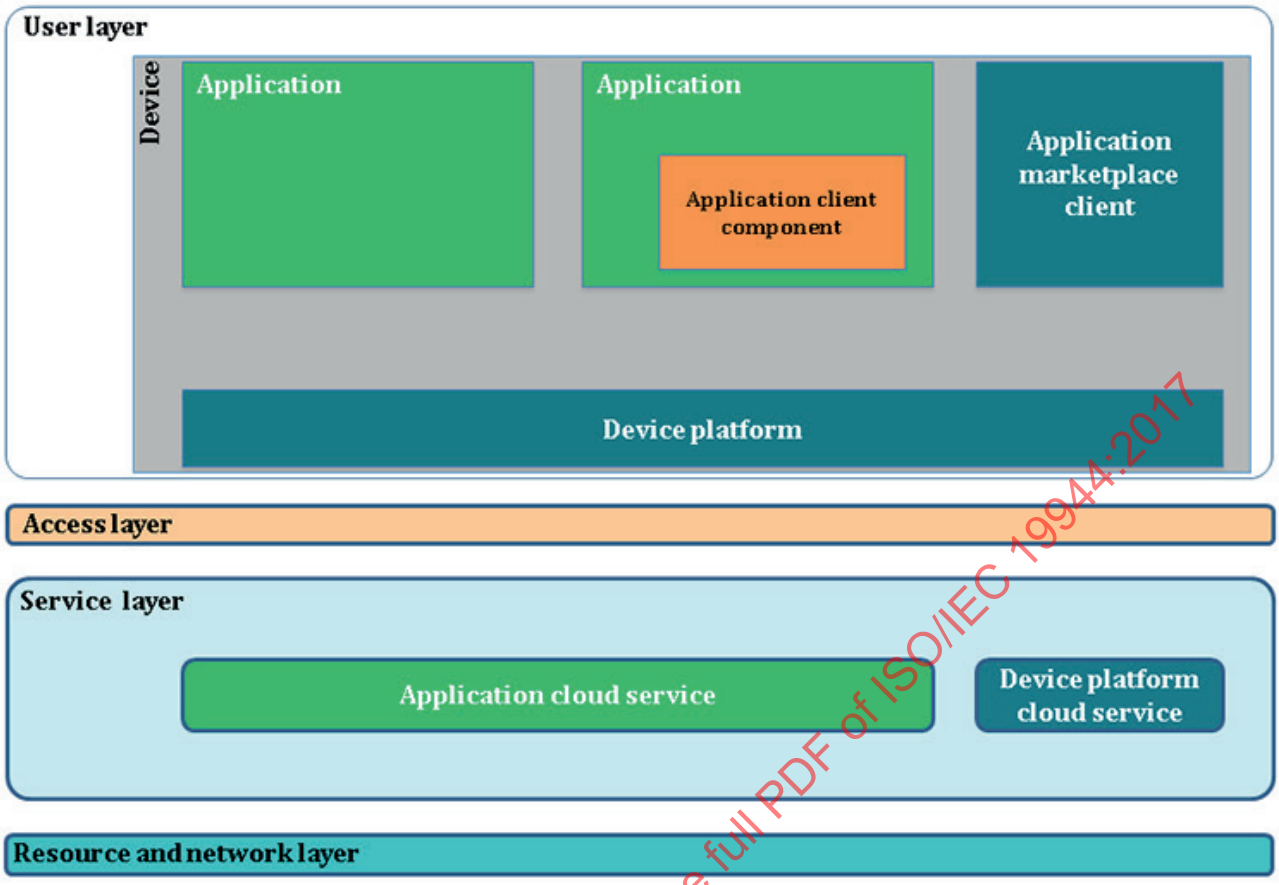

Figure 3 — Devices and cloud services functional view

The significant components of the devices and cloud services ecosystem are in the user layer and in the service layer. In the user layer, the major functional component is the device. The device embodies the user function component identified in ISO/IEC 17789 and it provides the means by which the end user interacts with the ecosystem. In the service layer are two categories of cloud services, the device platform cloud service and the application cloud services.

The device contains a number of subcomponents. There is the device platform, the application marketplace client and applications some of which may contain application client components.

It is typical for the device platform, the application marketplace and the device platform cloud service to be closely tied together, often all provided by a single cloud service provider organization. Applications running on the device typically connect with one or more application cloud services. A given application may be associated with a set of application cloud services, all owned and operated by a single organization. However, it is also common for a given application to make use of multiple application cloud services offered by multiple cloud service providers.

The application client component typically connects with a particular application cloud service, although the organization responsible for the application may be different from the organization responsible for the application client component and its application cloud service.

7.4.2 Functional components in the functional view

7.4.2.1 Device

This represents the physical device, together with any integral or attached hardware components such as memory.

7.4.2.2 Device platform

This represents the basic functionality (behaviour) of the device on which everything else depends, including the main user interface of the device. It also includes application programming interfaces (APIs) and access to hardware components, such as the screen, any buttons, network devices, GPS devices, cameras, biometric device, cryptographic functions, etc.

7.4.2.3 Application

This represents an application (app) running on the device to provide some capability to the user. It may be preinstalled on the device when delivered to the user, or installed separately. For separate installation, the application can be delivered to the device in various ways, such as being downloaded from an application marketplace, pushed to the device by an organization, or downloaded as a file.

There are also scenarios where mobile applications can be downloaded directly without the assistance of an application marketplace. The application however can always utilize cloud services or any on-board capabilities of the device platform (e.g. telemetry and environmental sensors) to collect and transmit data.

Some solutions for securing mobile operating ecosystems offer sandboxing capabilities. Such secure environments offer a parallel execution environment where the applications run in a more secure environment where data can be tightly controlled.

7.4.2.4 Application cloud service

Application cloud services are cloud services that offer capabilities to applications running on the device. The capabilities are typically offered by means of an API which the application can invoke as required.

Some application cloud services are specific to a particular application, while others can be used by many applications and are offered through public APIs.

It is typical for application cloud services to be independent of any particular device platform.

7.4.2.5 Device platform cloud service

The device platform cloud service supports capabilities that are unique to the device platform such as device customer and/or user identity, authentication, authorization, accounting, device setup and provisioning, firmware maintenance and application marketplace functions. Significant elements of the device user's data will reside here, with storage of identity and personalisation profile metadata. The device platform cloud service is accessed with an "application marketplace ID", which links all of the device user's actions on the device platform provider's services and can be used by other developers to identify the user for other applications. Those user actions can also be transferred to an advertising cloud service as input to select, price and deliver advertisements.

7.4.2.6 Application client component

This application client component is part of the application. It simplifies the creation of the application by providing the application developer with simple access to application cloud services or to device platform services.

For example, an application can call the application client component which can call on either the device platform for GPS information, or an application platform cloud service for an IP address location lookup. An application client component can also act as a common point of integration to data stored on the device which is common to multiple applications, such as a contact database, calendar, secure credentials store, or known locations.

7.4.2.7 Application marketplace

The application marketplace functionality handles installation of applications on the device. It is usually closely tied to the device platform and relies on a catalogue of applications held in the device platform cloud service. It has a privileged position in data flows in that it has access to data on exactly which applications the user has purchased, installed, used, updated and has rights to use. It usually also knows how much memory has been used. It probably also has information about location, account status and other personal information about the user and their behaviour.

7.4.3 Functional view: Data flows

This document extends the functional view expressed in ISO/IEC 17789 to include the data flows which take place between the functional components described in 7.4.2. Figure 4 shows the data flows between the functional components shown in Figure 3.

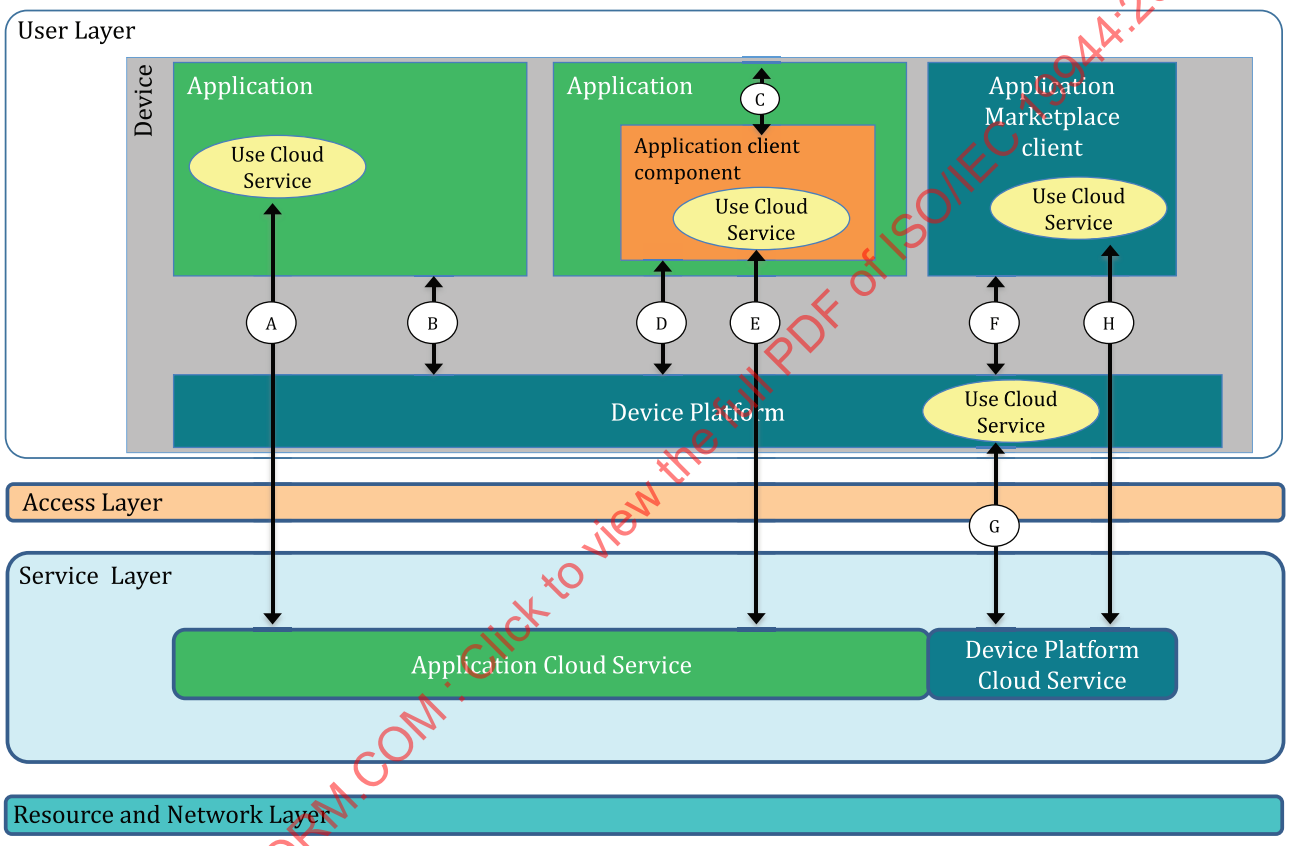

Figure 4 — Data flows between components

In Figure 4, the "Use Cloud Service" activity is shown in those places where a user layer component exchanges data with a cloud service — this indicates which functional components are interacting with one or more cloud services. The interactions between components and the associated data flows are shown with double headed arrows each labelled with a letter ("A" through "H") — the letter labels the data flow and is used in the description as follows.

Data flow A: Between an application and an application cloud service

In this flow, the communication uses no device platform-specific code in the application or the device. Use of the application cloud service is independent of the device or other functional components, except that its use may also cause data to flow between the device platform and the device platform cloud service (see Data flow G).

Data flow B: Between application and device platform

This takes place where an application requires services from the device platform or exchanges data with the device platform. An example is where camera device data flows to a camera application and then is stored as an image file on device storage. The device platform is aware of exactly which device features are being used and by which applications. In some cases, the device platform also communicates with the device platform cloud service to provide these services (see Data flow G).

Data flow C: Between an application and an application client component

The application client component may be built in to the application's own executable, be linked from an external function library, or execute in a separate operating system process on the device. Reasons for this include simplifying application development, obtaining functionality required for the application to operate, enhancing the user experience, or generating revenue. Examples of the latter include connecting the application to an advertising service, or connecting it to a payment service. Note that application use of capabilities provided by the application client component may result in data flows to the device platform (see Data flow D) and also to the device platform cloud service (see Data flow G).

Data flow D: Between application client component and the device platform

This often includes use of the user credentials or device identifier and may also include access to device sensors and functions such as biometric devices, GPS, gyroscope, microphone, speaker, light sensors, etc.

Data flow E: Between application client component and an application cloud service

An application client component may exchange data with one or more application cloud services as part of delivering functionality to the application using Data flow C. This may run securely and in isolation from other applications and the device platform, for example in order to conform with payment industry requirements.

Data flow F: Between application marketplace application and the device platform

The application marketplace application communicates with the device platform in order to obtain identity and security information about the user, to get information about the device configuration including memory and storage usage and to install and to update applications on the device.

Data flow G: Between the device platform and the device platform cloud service.

This includes associating the device identifier with a user account identifier and with an application marketplace account. It also communicates requests to support the marketplace store app, so it includes a considerable amount of data connecting the user to the applications they are installing and using. The search for, choice, purchase, download and updates of an application all result in data flows between the device and the device platform cloud service.

These flows often include EUII sub-types such as device connectivity data, user credential data and device telemetry data linked to an individual such as location (for geo-fencing of applications and content), user age information (for appropriate content control) and language choices. Processing or storage of computational or data intensive device platform capabilities, such as voice recognition or search, may be split between the device platform and the device platform cloud service. The device platform is also aware of the data flows between all applications and their respective application cloud service(s) and this may result in additional data flow to the device platform cloud service.

Data flow H: Between the application marketplace application and the device platform cloud service.

For most devices, applications are installed, uninstalled and managed on the device, separate from the underlying operating system of the device itself. This is usually done through a "marketplace application" of some kind, which provides the device-side functionality required by the application marketplace within the device platform cloud service. It is usually tightly coupled to the device platform.

Note that data flows in different directions between the components in Figure 4, depending on the particular operation taking place. For example, on a 'create' request data flows from the component making the request, whereas on a 'retrieve' request data flows back to the component making the request.

8 Data taxonomy

8.1 Overview

Transparency about the acquisition, processing and use of data by cloud services and associated applications is desired by users, regulators and cloud service customers. The data taxonomy described in this document is intended to support transparency about the types of data that are acquired by CSPs, as well as how they are used. This document provides a common data taxonomy and transparency concepts. This clause addresses the following areas:

— data categories;

— data identification qualifiers.

8.2 Data categories

8.2.1 General

This clause defines a set of data categories in the devices and cloud services ecosystem.

Any description of how data is acquired, transferred, processed and used requires, in majority of cases, clarity about the specific data categories involved. There are many different data objects in the devices and cloud services ecosystem and multiple ways to process or use those data objects. One approach to transparency is to name and define each data object and describe how it is processed and used. Although such an approach is comprehensive, this approach has two limits. First, the data objects in the ecosystem are constantly changing as technology, devices and cloud services evolve subjecting the list to constant revision. Second, the list of objects and uses would be so long, duplicative and complex that stakeholders would find it difficult to gain a useful understanding of how data is actually managed by reviewing a large number of individual data objects. To facilitate transparency, cloud service providers should describe how data is processed and used in the simplest way possible, using declarative statements that cover the largest, most abstract set of data objects.

To facilitate simple descriptions of data processing and use, a taxonomy of data categories, at the highest possible level of abstraction, is valuable. Obviously "data" is too abstract for useful description, but having to discuss data categories at the "disk access log files without customer identifiable information" level is likely to reduce actual transparency. A complete taxonomy, identifying every possible type of category together with all possible types of relations between these categories, would introduce a level of complexity beyond the requirements of this document. Instead this clause defines "data categories" in a hierarchical structure with inheritance/sub-type relationship. This hierarchy branches from the four basic data categories described in other International Standards. (i.e. ISO/IEC 17788/ISO/IEC 17789 and ISO/IEC 19086-1), cloud service customer data, cloud service derived data, cloud service provider data and account data. Each of these four categories is further divided with definitions of sub-types of related data objects, some of which are again divided into sub-types.

One use of this data taxonomy is to support broad policy statements. Although other approaches are possible to data categorization, the advantage of a hierarchy is that any statement regarding data use can apply to the broadest possible data categories, as defined in the highest appropriate branch (highest abstraction) in the taxonomy. As such, each category in the hierarchy is created to be as broad as possible, in anticipation of the requirement for granularity at various portions of the data categorization hierarchy. The data taxonomy described in this document is not intended to be exhaustive, but it is intended to be extensible. It is intended that a CSP can extend the taxonomy to define new sub-types of data to suit the needs of their cloud services. One likely data category subject to regulations, standards and contractual requirements is customer content data, particularly for application capabilities cloud services that necessarily understand the nature of the customer content data that such cloud services process.

Where a CSP does use additional sub-categories of data, it is necessary for the CSP to provide clear definitions of each new sub-type and to describe its relationship with other categories. A hierarchical

relationship is strongly recommended, based on the four topmost categories defined in this document (see Annex A for a hierarchical diagram of data categories and data identification qualifiers).

Transparency is enhanced when providers minimize the total number of statements needed to describe their overall data processing and use policy. As a result, sub-types of a data type are only defined in this taxonomy based on a perceived need to address a more specific set of data objects in descriptions of processing or use clauses of the taxonomy. For example, in 8.2.2, there are clearly data objects (e.g. an image file) which are not described by the definitions of "credentials" nor "user contact list".

This clause does not therefore propose a general purpose, comprehensive, taxonomy but instead a single view that is fit for purpose to analyse data flow and data use. A "faceted view" may be used to construct statements applying to a set of data categories sharing a single characteristic not available through a purely hierarchical view. Such characteristics could be used as "data identification qualifiers" as introduced in 8.3.

As an example, a characteristic could indicate whether a particular data category contains "personally identifiable information" (PII), the definition of which may vary between different jurisdictions and thus making it difficult to include in a single, global, hierarchy of data categories. Additional views will be developed according to specific needs of cloud service providers and customers.

Statements about data processing and use are assumed to apply to all instances of a named data type, including all sub-types. Some descriptions of processing and use may take advantage of defined sub-types to simplify statements by referring to a parent/super type but excluding one or more of its sub-types in the statement. For example, a cloud service provider may state that they encrypt "all derived data, except for telemetry", instead of naming each of the sub-types of derived data and omitting telemetry.

8.2.2 Customer content data

8.2.2.1 General

Customer content data is cloud service customer data extended to include similar data objects provided to applications executing locally on the device. Notice that the locally executing application may or may not choose to share that data with the cloud service and yet the data would still fit in this extended definition. This includes content directly created by customers and their users and all data, including all text, sound, software or image files that customers provide to the cloud service, or are provided to the cloud service on behalf of customers, through the capabilities of the service or application. This also includes data that the user intentionally creates through the use of the application or cloud service, such as documents, processed data sets, modified images, recorded sounds, etc. When customer content data local to the device is transmitted to the cloud service, it becomes cloud service customer data.

Specific types of information in customer content data may require explicit use statements by the cloud services provider to the extent the CSPs are aware of their presence. The following data categories are subsets of customer content data.

8.2.2.2 Credentials

Data provided by the customer to identify a user to the device, application or cloud service, e.g. passwords, password hints, etc., including biometric data provided for identification. The set of credentials data is a sub-type of customer content data.

8.2.2.3 Customer contact lists

Contact information for people that the cloud service customer provides, or is provided to the service on customers' behalf, through the capabilities of the service. Customer contact list data is a sub-type of customer content data.

NOTE 1 Cloud services can have a distinction between the cloud service customer and the cloud service users associated with that customer. Cloud service user contact list information provided by the cloud service customer to the cloud service provider is also customer content data.

NOTE 2 Contact information provided solely to support, to administer or to make payment for the service is account or administration contact information (see 8.2.5.2).

8.2.2.4 Personal health data and medical records

Personal health data and medical records are a form of sensitive personal data relating to an individual. This processing of this type of data is heavily regulated in many jurisdictions [e.g. Health Insurance Portability and Accountability Act (HIPAA) in the USA and Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada].

8.2.2.5 Personal genetic data

Personal genetic data is information about the genetic makeup of an individual (e.g. DNA record).

8.2.2.6 Personal biometric data

Personal biometric data is encoded data that describes characteristics of an individual (e.g. fingerprints, face geometry, iris pattern). For example, the voice prints of the human vocal cords and the posture maintained when walking (as used in Japan's Amended Act on the Protection of Personal Information)[13].

8.2.2.7 Personal data of children

Personal data relating to children is regarded as sensitive personal data and is subject to more stringent regulations and compliance rules (e.g. General Data Protection Regulation (GDPR)[11] in the European Union).

8.2.2.8 Political opinions

Political opinions of an individual are personal data that is often subject to special rules and regulations.

8.2.2.9 Financial details

Financial details relating to an individual include information about accounts, credit cards, payments, credit history. It is usually regarded as sensitive personal information subject to particular regulations.

8.2.3 Derived data

8.2.3.1 General

Derived data is cloud service derived data extended to include similar data objects derived as a user exercises the capabilities of an application executing locally on the device. When the local portion of the data is transmitted to the cloud service, it becomes cloud service derived data.

8.2.3.2 End user identifiable information (EUII)

8.2.3.2.1 General

EUII is linkable to the user but is not customer content data. EUII is a sub-type of derived data.

NOTE The term customer, user and tenant are used in the same way as cloud service customer, cloud service user and cloud service tenant in ISO/IEC 17788, with the definition of "customer" extended to include users of applications. In many services, a single individual fulfils all client-side roles, including user, customer and administrator. Customer, when used alone, is assumed to represent all three roles.

8.2.3.2.2 Telemetry data

Data collected about the capabilities of the product or service. Examples are measurement, performance and operations data. Telemetry data represents information about the capability and its use, with a focus on providing (see 9.3.2) the capabilities of the product or service. Telemetry data may contain information about one or more users and is a sub-type of EUII.

8.2.3.2.3 Connectivity data

Data that describes the connections and configuration of the devices connected to the service and the network, including device identifiers, (e.g. IP addresses) configuration, settings and performance. Connectivity data is a sub-type of EUII.

8.2.3.2.4 Observed usage of the service capability

Data provided or captured about the users' interaction with the service or products by cloud service provider. Captured data includes the records of the users' preferences and settings for capabilities, the capabilities used and commands provided to the capabilities. Usage data is a sub-type of EUII.

8.2.3.2.5 Demographic information

Data containing demographic information about end user provided or gathered though use of the capabilities of the application or cloud service. Demographic information is a sub-type of EUII.

8.2.3.2.6 Profiling data

Data provided or acquired about a users' interests and preferences relating to content, organizations or objects outside of the service, e.g. sports teams, businesses, products, etc. Profiling data is a sub-type of EUII.

8.2.3.2.7 Content consumption data

Information about media content that a customer accesses through the capabilities of the service, e.g. TV, video, music, audio or text books, applications and games. Content consumption data is a sub-type of EUII.

NOTE 1 Content consumption data is distinct from usage data collected when the user accesses customer content data.

NOTE 2 Content consumption data is distinct from client-side browsing history collected when accessing information accessed or available on the web.

8.2.3.2.8 Client-side browsing history

This data refers to records of the web browsing history when using the capabilities of the applications or cloud services stored in the service or application. Client-side browsing history data is a sub-type of EUII.

NOTE A record of the websites viewed by the user captured by a web browser is an example of a client-side browsing history. In some instances, certain legal obligations may be defined, e.g. UK Investigatory Powers Act 2016[12].

8.2.3.2.9 Search commands and queries

This data refers to records of search commands or queries provided by the user to the service or product. Search commands and queries data are a sub-type of EUII.

8.2.3.2.10 User location

This data refers to records of the location of the user within a specified degree of precision. User location data is a sub-type of EUII.

8.2.3.2.11 Social data

This data refers to records of interaction between the user, other people and organizations. This includes friends' lists and information about types of interactions (e.g. likes, dislikes, events, etc.) related to people and/or entities/ businesses which collectively encompass social graph data. Social data is a sub-type of EUII.

NOTE 1 A customer's own contact information is account or administration contact information (see 8.2.5.2).

NOTE 2 User's contact list maintained explicitly as such and entered by the cloud service user or customer using the capabilities of the service is called a "customer contact list" and is considered customer content data.

8.2.3.2.12 Biometric and health data

This data refers to metrics about the (human) user's inherent characteristics collected by the application or service's capabilities. Biometric and health data are a sub-type of EUII. For example, the voice prints of the human vocal cords and the posture maintained when walking (as used in Japan's Amended Act on the Protection of Personal Information)[13].

NOTE 1 Biometric data provided to the system or application for identification is considered credentials (see 8.2.2.2).

NOTE 2 Personal biometric data (see 8.2.2.6) entered by the user is customer content data.

8.2.3.2.13 End-user contact data

Contact information for a cloud service user. End-user contact data is a sub-type of EUII.

NOTE End-user contact data is different from customer contact lists (see 8.2.2.3) or account or administration contact information (see 8.2.5.2). This data type is captured or generated as the user interacts with the cloud service.

8.2.3.2.14 User's environmental sensor data

Data about the physical environment captured by sensors as the user exercises an application or cloud service's capabilities. User's environmental sensor data is a sub-type of EUII.

8.2.3.3 Organization identifiable information (OII)

OII is the data that can be used to identify a particular tenant (general configuration or usage data); is not linkable to a user and does not contain customer content data. This also includes data aggregated from the users of a tenant that is not linkable to the individual user. OII data is a sub-type of derived data.

8.2.4 Cloud service provider data

8.2.4.1 General

Cloud service provider data (as defined in ISO/IEC 17788) is unique to the system and under the control of the cloud service provider.

NOTE Cloud service provider data does not include customer content or derived data.

8.2.4.2 Access and authentication data

Access and authentication data are the data used within the cloud service to manage access to other categories of data or capabilities within the service. It includes passwords, security certificates and other authentication-related data. Access control data is a sub-type of cloud service provider data.

8.2.4.3 Operations data

Data which is used for supporting the operation of cloud service providers and system maintenance, such as service logs, technical information about a subscription (e.g. service topology), technical information about a tenant (e.g. customer role name), configuration settings/files.

8.2.5 Account data

8.2.5.1 General

Account data is class of data specific to each cloud service customer that is required to sign up for, purchase or administer the cloud service. This data includes information such as names, addresses, payment information, etc. Account data is generally under the control of the cloud service provider although each cloud service customer usually has the capability to input, read and edit their own account data but not the records of other cloud service customers. See ISO/IEC 19086-1[3].

8.2.5.2 Account or administration contact information

Contact information for customer of an application or cloud service and any cloud service administrators and cloud service business managers designated to administer and control the use of the service. Account or administration contact information is a sub-type of account data.

8.2.5.3 Payment instrument data

Data provided by the cloud service customer for the purpose of making payment for the services, or to pay for products or services bought through the services. Payment instrument data is a subset of account data.

8.3 Data identification qualifiers

8.3.1 General

Data in any category can provide or contribute to information that identifies or can be linked to an individual, referred to in this document as personally identifiable information (PII). The extent to which individuals are directly identified in the data and how easy it is to associate a set of characteristics in the data to an individual is important to individuals, CSCs and policy makers as they assess a use of that

data category. Therefore, the specification of data in the context of data use or data processing should include not only the type of that data, but also a description of the degree to which the data can identify an individual or associate an individual with a set of characteristics in the data.

This clause defines qualifiers that can be used with data categories to describe the degree to which an individual is directly identified by the data and how the individual is associated with characteristics (attributes) in the data.

8.3.2 Identified data

Identified data is data that can unambiguously be associated with a specific person because PII is observable in the information. Guidance on what can be considered as identifiers can be found in ISO/IEC 29100:2011, 4.4.1[4].

8.3.3 Pseudonymized data

Pseudonymized data is data for which all identifiers are substituted by aliases for which the alias assignment is such that it cannot be reversed by reasonable efforts of anyone other than the party that performed them.

This corresponds to data resulting from the process of "pseudonymization" in ISO/IEC 29100:2011, 2.24 and described as "pseudonymous data" in ISO/IEC 29100:2011, 4.4.4.

8.3.4 Unlinked pseudonymized data

Unlinked pseudonymized data is data for which all identifiers are erased or substituted by aliases for which the assignment function is erased or irreversible, such that the linkage cannot be re-established by reasonable efforts of anyone including the party that performed them.

8.3.5 Anonymized data

Anonymized data is data that is unlinked and which attributes are altered (e.g. attributes' values are randomized or generalized) in such a way that there is a reasonable level of confidence that a person cannot be identified, directly or indirectly, by the data alone or in combination with other data.

This corresponds to data defined as "anonymized data" in ISO/IEC 29100:2011, 2.3 and the process defined as "anonymization" in ISO/IEC 29100:2011, 2.2.

8.3.6 Aggregated data

Aggregated data is statistical data that does not contain individual-level entries and is combined from information about enough different persons that individual-level attributes are not identifiable.

9 Data processing and use categories

9.1 Overview

In order to understand the processing and use that is made of data which flows between devices and cloud services, it is useful to consider the various categories of data processing that can take place, the categories of data use that might occur and the scopes of the processing and use (essentially what capabilities, cloud services and parties may be involved). This clause examines each of these topics in turn.

9.2 Data processing categories

9.2.1 General

This clause describes some of the data processing techniques found in the devices and cloud services ecosystem. These data processing techniques include transformations of the data content and movement or storage without transformation of the content.

The data processing and transformation taxonomy is extensible and supports the description of the processing techniques for handling data in the devices and cloud services ecosystem and highlight areas relevant to data privacy.

NOTE Additional information about the processing techniques relevant to storage security can be found in ISO/IEC 27040.

Throughout the data lifecycle, processing techniques might be applied to a set of data independently or in combination with each other to address specific contexts. Each technique can be performed either by a single entity or by multiple stakeholders.

9.2.2 Data partitioning

9.2.2.1 General

Data partitioning refers to the approach of splitting a set of data residing in a single location or database into smaller logical units, called partitions.

Data partitioning is used within cloud services to process very large data sets by placing relevant data closer to each member of a set of distributed processors. The resultant data partitions can, for example, be stored in different datacentres running a single distributed database system, which raises issues for policy and practice that relies on a single location for a data set.

The two main approaches to data partitioning are horizontal and vertical. Hybrid partitioning refers to the method of combining horizontal and vertical partitions by applying them in any sequence to the same data set. Partitioning data vertically might be effectively used to strip sensitive information from the data before sharing the data with other parties.

9.2.2.2 Horizontal partitioning or sharding

A horizontal partition, also commonly known as sharding, is a subset of full records from the original database. The values of the attributes of each record in the partition satisfy a certain logical condition defined by the specific partitioning operation. In the relational database example, a horizontal partition would be a subset of rows from the original table satisfying a logical composition (i.e. using AND and OR logical operators) of one or more selection operators on the original table.

9.2.2.3 Vertical partitioning

A vertical partition contains all records from the original database, but with only a subset of attributes (i.e. columns) as defined by the specific partitioning operation. In the relational database example, a vertical partition would contain all rows from the original table but containing only a subset of columns.

9.2.3 Data integration

Data integration is the process of providing a unified view from multiple data sets. Information from multiple data sets can be combined in a number of ways, each of which has its own terminology. The following are a few common examples.

— Data association, where individual records from one data set are linked to data records from another.

— Data aggregation/Data consolidation, where records of the same type, but from different data sources are combined together into a single data set.

— Data accumulation, where data arising from a single source is kept over time to create a history of how the data values are changing.

These distinctions are helpful in explaining what an application or a cloud service is doing with data. For example, data linkage can create sensitive data from two seemingly innocuous data sets. Data accumulation can uncover deep trends in usage and other behaviour. Overall these processes create new insight, potentially for both the CSC and the CSP.

9.2.4 Data fusion

Data fusion is the process of combining information from multiple data sets followed by reduction or replacement, which results in a single improved data set, such as a data set with more confidence or more relevancy.

The term information fusion is synonymous with data fusion, but might imply a higher semantic level than data fusion. Other terms associated with data fusion are decision fusion and data combination.

Data fusion is used throughout the devices and services ecosystem, notably for machine learning related to users, processes and resources.

9.2.5 Data improvement

The process of improving the quality of information comes in a number of categories, including:

— data standardization: getting data into the corresponding fields in a data structure;

— data validation and correction: testing for valid values and fixing any that are not valid;

— data enrichment: filling out missing data;

— data de-duplication: matching duplicate records for the same person/thing and creating a single consolidated record from the duplicates (often they have different values so takes policies on how this is done);

— data pruning/disposal: removal of obsolete data.

9.2.6 Encryption

Encryption can be used across the devices and cloud services ecosystem to protect data. Encryption techniques that can be used include encryption of data at rest and encryption of data in motion. For more information describing these techniques, see ISO/IEC 27040[6].

9.2.7 Replication

Replication refers to the practice of creating and maintaining multiple instances of the same information typically for failure recovery. In the devices and cloud services environment, replication has also been used to speed access to information by locating instances of the same information in geographical proximity to its usage.

9.2.8 Data Deletion

9.2.8.1 General

Originally, deletion of data was designed and used mainly to allow reuse of permanent storage. Today, in the devices and cloud services ecosystem storage cost is dramatically reduced and the focus is on deletion of data as an important activity in data protection[8] and, where applicable, the data subjects "right to be forgotten"[10].

Various technological approaches can be used for data deletion. They differ in their properties[9] such as the physical granularity of data to be deleted, accessing or processing (e.g. deleting) the metadata and the latency until the complete result of the deletion operation is achieved.

An additional important aspect of data deletion includes tracking the flow of data through its lifecycle in the (distributed) system and the deletion of specific information as necessary. This can require a complex system design due to data replication, partitioning and other processes. An additional level of complexity is introduced if the deletion of information based on the identification of specified data is required.

Deletion of electronic data falls under two broad categories: “data deletion” and “secure data deletion”.

9.2.8.2 Secure data deletion

Secure data deletion refers to the process of irreversible destruction of electronic data so no party (such as the data subject, the data processor, any authorized or unauthorized third party, or any malicious actor) is capable of recovering the data from the system[9].

9.2.9 Re-identification

Re-identification is the process of linking the information from a de-identified data set to a particular data subject. Re-identification creates a new data set containing information linked to some or all of the data subject’s records in the original data set. Re-identification might be achieved by using the data integration techniques described in 9.2.3.

The resultant information about the data subjects may not be identical to or consistent with the original data due to potential distortion of data in the course of its de-identification, re-identification, or both processes.

9.3 Data use categories

9.3.1 General

Applications and services use data in complex ways to provide capabilities that appear quite simple. For example, a capability on a mobile device that provides travel directions in response to verbal commands, an everyday interaction between humans, requires a very complex interaction between the application and support services that provide speech recognition and map data. Furthermore, the data transferred and stored between the application and the services is useful in many ways beyond providing the directions, it could also be used to improve the overall performance of the speech engine, or to improve the targeting of advertising.

To increase understanding and trust, providers seek to use commonly used, non-technical, words to describe use of the data. Those common terms may not have the same meaning for the user and the provider. The following clauses define the accepted meaning of common terms in the context of the devices and cloud services ecosystem and any additional scope information needed to fully explain the use.

Using these terms in data use statements and referencing clear definitions in this document allows providers to make simple data use statements, yet provide transparency about the specifics of data use to customers, policy makers and regulators.

Unlike scope definitions, use definitions do not build on each other, e.g. use of “improve” does not imply “provide”. A more specific definition does not imply any other use, e.g. “share with third-party partners and data processors when necessary to provide the service” does not imply any other sharing of the data such as “share with partners for marketing purposes”.

Each use of data should have an explicit data use statement. A statement can include multiple uses for a specified scope and data category, e.g. “Account data is used to provide and improve the service.”

Additional “uses” and verbs can be defined to extend this document.

For definitions of data use, source scope, use scope and result scope, see 10.2.1.

9.3.2 Provide

Provide means the use of specified data categories:

— from the source scope by an applications and services scope to provide and protect the current capabilities of a results scope;

— to communicate with the customer about the status and availability of the current capabilities of the result scope;

— including providing support for the result scope and to protect at a minimum the specified data category from the source scope.

Provide can include the use of specified data categories to protect the rights and property of the cloud service provider and to prevent loss of life or serious injury to anyone. For example:

Example 1:

This cloud service uses derived data only to provide the cloud services defined in the cloud services agreement.

NOTE 1 In this example, use of derived data is restricted to provide the service contracted for in the cloud service agreement, including operational support system (OSS) and business support system (BSS) for exclusively those services. In the case of a single contracted service, "This application" or "This service" could also define the scope (see 9.4.2.3).

NOTE 2 The data use statement structure used in this example is described in Clause 10.

In the case where a single scope is involved *provide* means to protect the customer content data from this scope and to provide and communicate with the customer about the status and availability of the current capabilities of this scope.

Provide operational support for contracted service

This usage is related to the acquisition, processing and storage of data about the usage of a cloud service (derived data) contracted by a specific cloud service customer in order to operate and protect the systems and processes necessary for the provision of this cloud service. This includes:

— service usage data to be used for capacity planning;

— monitoring of user behaviour to identify potential attackers and to perform forensic analyses;

— logging data for system and network maintenance and optimization;

— correlation of service usage data and system events for fault tracking and root cause analysis.

***Improvement* of business support for contracted service**

This usage is related to acquisition, processing and storage of data on the usage of contracted services (derived data) being used for business support related to this service. This includes:

— evaluation of service usage data to determine user preference about use of the current capabilities of the services contracted for in the SLA;

— financial controlling, budgeting and resource planning.

9.3.3 Improve

Improve means to use specified data categories from the source scope to improve or increase the quality of the existing functional capabilities of the result scope.

Improve can be used with a single scope. In this case, it means that data acquired or created by applications and services in the scope is used to improve the existing functional capabilities and to add new capabilities to the scope, available to all users.

9.3.4 Personalize

Personalize means to use specified data categories from the source scope to change the presentation of the capabilities of the result scope or to change the selection and presentation of data or promotions accessed through the capabilities of the result scope to be specific to the user, based on information about the user gathered by applications and services in source scope.

The same changes may apply to multiple users, for example all users of a particular customer or all of the users sharing common characteristics may receive the same changes.

Personalize can be used with a single scope, in which case data acquired or created by applications and services in the provided scope is used to change the presentation of the capabilities of that scope or to change the selection and presentation of content by the applications and services in the scope to be specific to a user.

Example 2:

Customer content data from this service is used to personalize cloud service provider's services outside of the services listed in the cloud service agreement.

Example 2 describes personalizing of services unrelated to the contracted service based on usage of customer data regarding the contracted service to improve services that are not contracted by the customer. Since data on service usage provide information on the preferences of the cloud service user, their collection and correlation with other data sources can be used to trigger, maintain and improve a large variety of supplementary services. This includes use of other services, not explicitly contracted by user, as listed in the following examples.

— The usage of location data from mobile devices to provide location based services to the user according to his or her past behaviour.

— Add-on advertisement services based on search engine queries, combined with data on past user behaviour.

NOTE The data use statement structure used in this example is described in Clause 10.

9.3.5 Offer upgrades or upsell

Offer upgrades or upsell means to use specified data categories from the source scope to offer to the customer increased capacity or resources for the capabilities of the result scope or new capabilities currently outside of the result in exchange for compensation.

The source of new capabilities may be defined as a scope. For example: "…to upsell capabilities to customers from *any of our products and services.*"

Offer upgrades or upsell requires the definition of the person or group of people who are the target audience.

9.3.6 Market/advertise/promote

9.3.6.1 General

Market/advertise/promote means to promote specified products and services to users or customers of a results scope based on data from the source scope.

Promotion is targeted at an individual or a group of individuals. *Market/advertise/promote* requires the definition of the person or group of people who are the target audience.

9.3.6.2 Promote based on contextual information

Market/advertise based on data derived from the use of the current capability or based on the services and application scope, without the use of data derived from the user's prior use of the services.

9.3.6.3 Promote based on personalization

Use specified data categories from the source scope to change the content of a promotion to the result scope to be specific to the user. The same content may be presented to multiple users, for example, all users of a customer or all of the users sharing a profile may receive the same changes.

9.3.7 Share

9.3.7.1 General

Share means to transfer specified data categories from the source scope to an entity other than the cloud service provider of the source scope. This entity may be defined as the cloud service provider of a result scope, e.g. "… share pseudonymized operations data with cloud service providers of similar commercial cloud services."

As an example:

Example 3:

This service shares customer content data with third parties.

This example is a poor use statement in that it does not provide clarity of the purpose for which the data is being shared nor of the extent of the data being shared. CSPs are strongly encouraged to provide as much detail as possible in data use statements so that it is clear to the CSC what is being done with which data.

NOTE The data use statement structure used in this example is described in Clause 10.

Cloud service providers should specify a purpose for sharing data by including a use definition. As an example:

Example 4:

This service shares payment instrument data with third-party partners and data processors to provide the cloud service.

This example adds some clarity to how retail services provide payment instrument data (i.e. credit card information) to third parties, for example for billing purposes, for the specific purpose of providing the service.

NOTE The data use statement structure used in this example is described in Clause 10.

9.3.7.2 Share when required to provide the service

There are conditions where CSP are required to share data: by contract, applicable laws and regulations, resulting in the transfer of specified data categories to third parties to provide the service. This can include sharing data to comply with applicable law or respond to valid legal process from competent authorities, including from law enforcement or other government agencies and providing data to law enforcement to protect the service and uphold the terms governing the use of the service. This use statement only includes the use of provided data by the third parties to provide the services in the scope.

9.4 Scopes: Boundaries of collection and use of data

9.4.1 Scope concepts

The term “scope” as used here provides a way to clearly describe the boundaries of collection and use of data in the devices and cloud services ecosystem. In the example declaration given with Figure 7, the scope increases from data collected in a specific capability (for example, a single web page) to use of the collected data by any capability in the service, the results of that use may be used to provide any service agreed to in a service agreement.

The scope types in 9.4.2 are arranged to describe an increasing extent of a CSP’s products and services. Figure 5 illustrates the idea that each definition encompasses a greater extent of the services and products. CSP can simplify use statements by combining scopes with individual elements, for example by extending a scope: “…the services listed in the cloud service agreement plus our ad funded service…”, or by providing a scope with an exception “…all our services except for the following services intended for children…”.

Using a single scope type to encompass multiple scope uses can simplify use declarations, however care must be taken to ensure the statement reflects the actual use. For example, using “capability X” as the scope as a simplified scope statement, i.e. “capability X uses customer data from capability X to personalize” means the capability is restricted to use of data entered while using the capability and the personalization only applies to the capability itself. If data captured from use of the capability is used to personalize other capabilities, the correct declaration is “capability X uses customer data from capability Y to personalize the service”.

Third parties to the cloud service customer and cloud service provider relationship define a distinct scope.

9.4.2 Scope types

9.4.2.1 General

The set of scopes defined in this clause are intended to replace multiple individual descriptions of the included applications and services.

The scope definitions can be used to define the applications and services associated with data use. The definitions are listed in increasing breadth of scope and the wider scopes include the narrower scopes, except for “third-party” items which exist in an independent scope. Capabilities are parts of an application or a cloud service, which in turn may be one of the covered services listed in the service agreement.

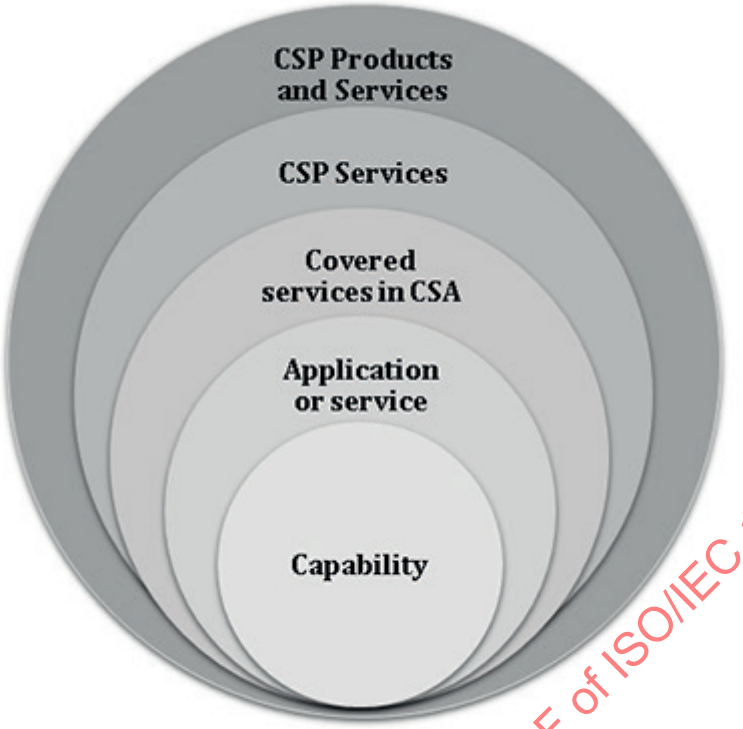

Figure 5 — Increasing levels of scope

9.4.2.2 Capability

A capability is some part of the functionality of a cloud service or associated application. Each capability shall be given a unique name and shall be clearly separated from other capabilities of the same application or cloud service, so that any data use statements which are made can clearly denote the data which is entered into the capability, acquired by the capability, processed by the capability or output by the capability, when the capability is used. The expression "this capability" can be used to specify the capability when the use of the term is unambiguous. For clarity, the name of the application or cloud service should also qualify the capability name, if there are multiple applications or cloud services.

9.4.2.3 Application or service

This scope includes the application or the cloud service that is involved in the entering or acquiring of data, the use of data or the result of use of data. Where there is more than one application or cloud service, each should be given a unique name which should be used in order to be clear about which application or cloud service is being discussed.

The expression "this service" can be used to specify the service when the use of the term is unambiguous.

9.4.2.4 Services listed in the cloud service agreement

Any of the cloud services specified in the cloud service agreement that applies to the application or service that provided the data.

9.4.2.5 Cloud service provider's cloud services

Any of the cloud services provided by the cloud service provider, including but not limited to the cloud services covered by the cloud service agreement.

9.4.2.6 Cloud service provider's products and services

This refers to any product or service from the cloud service provider.

9.4.2.7 Third-party product and services

Any product or service from entities other than the cloud service provider.

NOTE For use statements about sharing data (see 9.3.7), third party can be used to denote an entity that provides the data from a source scope or receives data as a result scope

9.4.2.8 Third-party and data processors

Third-party entities that are contractually bound to uphold the commitments in the cloud service agreement made by the cloud service provider. This includes PII processors as defined in ISO/IEC 29100.

10 Data use statements

10.1 Overview

Cloud service providers need to describe how different categories of data are used in cloud services and the associated applications. A transparent description of data use helps to resolve concerns about multi-tenancy, privacy, confidentiality, intellectual property rights and data location. There are a number of reasons why cloud service providers use data differently than is the case with on-premises IT systems. Primarily, continuous process and service improvement is an essential characteristic of mobile and cloud computing and much of that improvement is based on machine learning and automated adaptation of the services based on data as it flows through. In addition, many mobile and cloud service providers are funded though commercial use of some of the data flowing through the services.

In terms of PII processing, a CSP is a PII processor when it processes PII for and according to the instructions of a CSC. This case happens frequently in practice. However, for certain types of cloud services, a CSP could be a PII controller, in particular for cases where the CSP processes PII in order to achieve its own purposes and especially for cases where the end user is the cloud service customer for consumer-oriented cloud services.

Cloud service customers and regulators require a clear description of how the cloud service provider uses each category of data. This clause provides a structure for data use statements within the devices and cloud services ecosystem that can be used to provide consistent descriptions about the use of data. Data use statements can be extended and may use additional taxonomies of use.

The data collected from the users may be used to provide, maintain, enhance and potentially monetize the cloud services. Having a structured way to express how such data is collected, processed, stored and used will improve consistency and transparency for cloud service customers, the cloud service providers, regulators and other stakeholders. Such clarity is necessary to provide better governance of data and its usage.

NOTE ISO/IEC 38505-1[7] identifies and examines higher level governance concerns regarding the use of data which is relevant from the perspective of governance of data.

An objective of this document is to improve transparency in describing data flows and to reduce the risk of confusion. The data taxonomy, data identification qualifiers, data processing and data use categories described in Clauses 8 and 9 can be used by cloud service providers (CSPs), cloud service partners (CSNs), or cloud service customers (CSCs) to create data use statements. This document can be used to define naturally formed, complete, unambiguous and structured sentences in order to add clarity and transparency in communication between the CSP, CSN and CSC and cloud service users. There are multiple ways to achieve this. This document provides one way to define descriptions, guidance and examples for the definition of data use statements. Guided by this document the risk of incomplete or poorly drafted data use statements can be reduced.

The data flows described in 7.4.3 can provide an approach to the creation of data use statements which describe how particular categories of data are processed and used in the devices and cloud services ecosystem. Data flows can identify the source of the data and its destination or target. The functional components identified in 7.4.2 can be useful for describing the source and target. It is also important to recognize that data processing can be conducted by a particular component, but that the output from that processing could affect one or more other components.

10.2 Data use statement structure

10.2.1 Structure definition

Complete descriptions of data use should include specification for:

— data use: The data used, as a named data element that both CSP and CSC recognize, or specified as some level in the taxonomy of data categories as described in 9.3;

— source scope: The source of the data. The source may be directly specified (i.e. "video *from the camera*") or with a scope of applications and services (see 9.4.2.3);

— use scope: Applications or services that are using the data;

— result scope: The collection of elements changed, as a result of the data use.

9.4.2.3 provides definitions for collections of applications and cloud services appropriate for the specification of scope.

Figure 6 illustrates the overall structure of a data use statement. Although natural language and the context of the statement will affect word order, the basic structure is as follows.

— Data comes from a source in some part of the devices and cloud services ecosystem (a source scope).

— Data is processed or used by a part of the ecosystem (the use scope).

— In turn, that processing or use will have an effect on a part of the ecosystem (result scope.)

Since the degree to which data can be linked to a person is an important aspect of data use, the data type may be qualified using the data identification qualifier terms in 8.3.

Figure 6 — Use statement structure (passive)

The following example follows the structure in Figure 6:

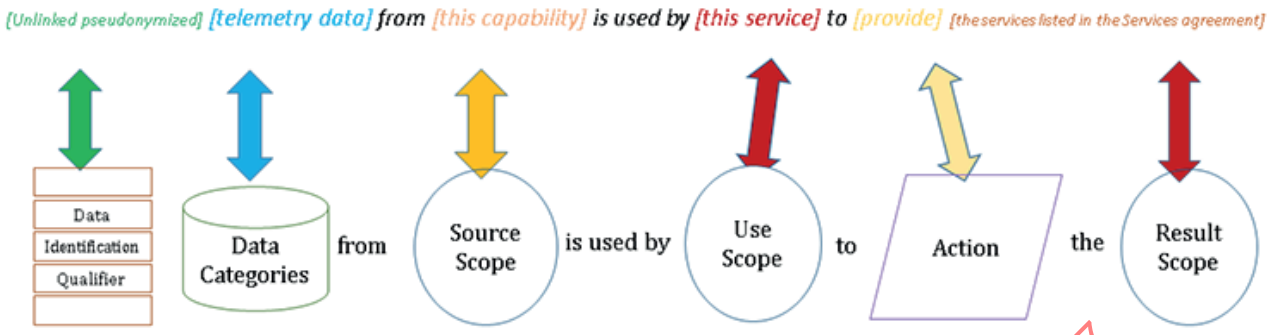

Figure 7 — Example of use statement structure (passive)

Figure 8 illustrates an alternative structure for a data use statement. It is very similar to the structure described in Figure 6 with the exception that the natural language structure used is in active form; whereas, the structure in Figure 7 uses passive form. There may be data use description scenarios and natural human languages where the use of active form is more desirable.

Figure 8 — Use statement structure (active)

The following example follows the structure in Figure 8:

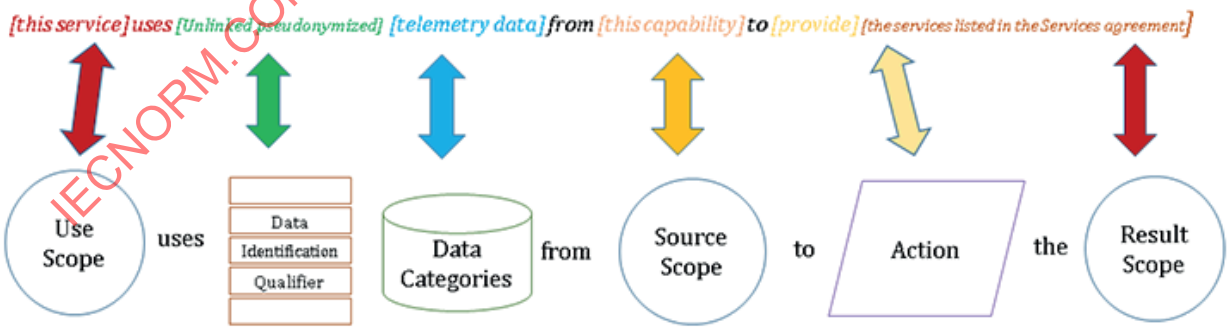

Figure 9 — example of use statement structure (active)

Some examples of data use statements follow:

Example 1:

The *services defined in the services agreement* use *account data* from the *service that provided the data* to *provide the services defined in the service agreement.*

Example 2 is very similar to Example 1 except that the source and use scopes are the same:

Example 2:

The cloud services defined in the cloud services agreement use account data from those cloud services to provide the cloud services defined in the service agreement.

Example 1 has the product and services scope, the source scope and the results scope are all the same and an actual statement is likely to be simplified to the form shown in Example 3:

Example 3:

Account Data is used to *provide* the cloud *services defined in the service agreement*.

Example 4 represents a more complex example of a statement relating to a data use:

Example 4:

The *cloud service provider services* use *unlinked* pseudonymized *customer usage data* from *the query* to *improve the cloud service provider services and products.*

Real-world use of data can be complex and the description of a data use may include multiple data categories and multiple scopes. In some cases, describing the general use of data with the widest possible scopes and then providing a list of exceptions may provide a simpler description.

10.2.2 Describing the scope of applications and cloud services that apply to use statements

Cloud service providers should describe use of data as broadly as possible to reduce the complexity of the usage statement. In addition to using the most abstract definitions of data categories, data use statements should use broad descriptions of the application and cloud services that use data or are affected by data use.

Statements that address the broadest possible set of applications and cloud services reduce the total number of necessary statements and result in statements that are more likely remain applicable as new services are offered and new data categories are added.

Example 5 shows a broad scope definition in a data use statement:

Example 5:

The cloud services covered in this agreement use user location data from these cloud services to provide the cloud services.

Generic descriptions of data use require specification of the capabilities, applications and cloud services that constitute the source of data, the capabilities, applications and services using it and where the results of the use are applied. Although the addressed capabilities, applications and services can always be listed explicitly, it is frequently clearer to describe a set of applications and services generically by defining a *scope* of capabilities, services or applications to which the statement applies.

10.2.2.1 Using single or dual scope definitions

Fully expressed data use statements have three scopes stated: the use scope, the source scope and the result scope. In some cases, where two of the scopes are the same, or where all three scopes are the same, data use statements can use a simplified format where only one or two scopes are stated, the other scopes are inferred.

If only one scope is described then it is assumed to be the same scope for the use, source and result scope. In this case, data is assumed to come exclusively from the use scope and the results of the data use (result scope) to apply only to that scope.

