INTERNATIONAL STANDARD

ISO/IEC 19944-1

First edition
2020-10

Cloud computing and distributed platforms — Data flow, data categories and data use —

Part 1: **Fundamentals**

Informatique en nuage et plates-formes distribuées — Flux de données, catégories de données et utilisation des données —

Partie 1: Principes de base

Reference number
ISO/IEC 19944-1:2020(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

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.

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) or the IEC list of patent declarations received (see https://patents.iec.c).

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

For an explanation of 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 www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information Technology,* Subcommittee SC 38, *Cloud Computing and Distributed Platforms.*

This first edition of ISO/IEC 19944-1, along with ISO/IEC 19944-2[1)] cancels and replaces ISO/IEC 19944:2017, which has been technically revised.

The main changes compared to the previous edition are as follows:

— provides additional material which principally deals with organizational data and the need to treat some organizational data in particular ways in order to ensure confidentiality, integrity and so on,

— the new concept of data facets is introduced and data facets are used to extend the expressiveness of data use statements, including adding the concept of which individuals or organizations have control over data,

— the new data use categories are introduced, including some that address the newer uses of data associated with artificial intelligence systems.

A list of all parts in the ISO/IEC 19944 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

1) Under preparation. Stage at the time of publication: ISO/IEC PWI 19944-2:2020.

Introduction

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 are 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 may be used in several ways including, but not limited to:

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 organisations 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 organisation, 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 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 are used in an ecosystem of devices and cloud services.

This document also 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.

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 are 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.

This version of ISO/IEC 19944 contains additional material which principally deals with organizational data and the need to treat some organizational data in particular ways in order to ensure confidentiality, integrity and so on.

To assist with this, the new concept of data facets is introduced and data facets are used to extend the expressiveness of data use statements, including adding the concept of which individuals or organizations have control over data.

New data use categories are introduced, including some that address the newer uses of data associated with artificial intelligence systems.

Cloud computing and distributed platforms — Data flow, data categories and data use —

Part 1: Fundamentals

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 organisation involved in legal, policy, technical or other implications of data flows between devices and cloud services.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 17788:2014, *Information technology — Cloud computing — Overview and vocabulary*

ISO/IEC 17789:2014, *Information technology — Cloud computing — Reference architecture*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 17788 and the following apply.

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

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

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

3.1 Terms related to data categories

3.1.1
account data
class of data specific to each CSC that is required to administer the *cloud service*

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 the CSC, such as name, address, telephone, etc.

3.1.2
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

3.2 Terms related to cloud services and devices ecosystem

3.2.1
device
physical entity that communicates directly or indirectly with one or more *cloud services*

3.2.2
application marketplace
set of *cloud services* providing a digital marketplace intended to offer applications and other digital content for a particular *device platform* (3.2.4) 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* (3.2.1) can use more than one application marketplace.

3.2.3
application cloud service
cloud service that supports applications running on a given *device* (3.2.1), where the cloud service is provided by a party other than the *device platform provider* (3.2.5)

3.2.4
device platform
operating system and related feature set that provides the core capabilities for a *device* (3.2.1)

Note 1 to entry: An *application marketplace* (3.2.2) is specific to a device platform.

3.2.5
device platform provider
device platform cloud service provider
cloud service provider that provides *cloud services* necessary to support a *device platform* (3.2.4) including managing needed digital identities

Note 1 to entry: The cloud service provider that offers the *application marketplace* (3.2.2) is typically the same as the device platform provider, but it is not required to be.

3.2.6
device platform cloud service
cloud service offered by the *device platform provider* (3.2.5) to support the *device platform* (3.2.4)

Note 1 to entry: An *application marketplace* (3.2.2) can be an example of device platform cloud service.

3.3 Terms related to privacy

3.3.1
personally identifiable information
PII
personal data
any information that (a) can be used to establish a link between the information and the natural person to whom such information relates, or (b) is or can be directly or indirectly linked to a natural person

Note 1 to entry: The "natural person" in the definition is the PII principal (3.3.3). To determine whether a PII principal is identifiable, account should be taken of all the means which can reasonably be used by the privacy stakeholder holding the data, or by any other party, to establish the link between the set of PII and the natural person.

Note 2 to entry: This definition is included to define the term PII as used in this document. A public cloud PII processor (3.3.2) is typically not in a position to know explicitly whether information it processes falls into any specified category unless this is made transparent by the cloud service customer.

[SOURCE: ISO/IEC 29100:2011/Amd1:2018, 2.9]

3.3.2
PII controller
privacy stakeholder (or privacy stakeholders) that determines the purposes and means for processing *personally identifiable information* (PII) (3.3.1) other than natural persons who use data for personal purposes

Note 1 to entry: A PII controller sometimes instructs others, e.g. *PII processors* (3.3.4) to process PII on its behalf while the responsibility for the processing remains with the PII controller.

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

3.3.3
PII principal
natural person to whom the *personally identifiable information* (PII) (3.3.1) 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]

3.3.4
PII processor
privacy stakeholder that processes *personally identifiable information* (PII) (3.3.1) on behalf of and in accordance with the instructions of a *PII controller* (3.3.2)

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

3.4 Terms related to organizational data

3.4.1
individual data
class of data objects under the control, by legal or other reasons, of a natural person

Note 1 to entry: Individual data can be a mixed dataset (3.4.6).

Note 2 to entry: Customer content data is individual data when the CSC is a natural person.

3.4.2
organizational data
class of data objects under the control, by legal, contractual or other reasons, of an organization

Note 1 to entry: An organization can be a for-profit company, a non-profit organization, a public or government agency, a non-governmental organization or an international organization, and can be small, medium or large.

Note 2 to entry: Customer content organizational data when the CSC is an organization and thus not a natural person.

Note 3 to entry: Cloud service provider data (ISO/IEC 17788) is always organizational data by nature.

Note 4 to entry: Organizational data can be a *mixed dataset* (3.4.6).

3.4.3
organizational protected data
OPD
organizational data whose protection is required based on the policies established by governance of data process

Note 1 to entry: Organizations have policies that govern the data under their control. ISO/IEC 38505-1 identifies and examines higher level governance concerns regarding the use of data which is relevant from the perspective of governance of data.

Note 2 to entry: Organizational data can contain OPD and PII.

3.4.4
public domain data
class of data objects over which nobody holds or can hold copyright or other intellectual property

Note 1 to entry: Data can be in the public domain in some jurisdictions, while not in others.

Note 2 to entry: The concept of public domain and the difference between this and "publicly available" is subtle and varies between jurisdictions. Readers should make themselves aware of the specific legal situation as it may apply to them.

3.4.5
non-personal data
class of data objects that does not contain *PII* (3.3.1)

Note 1 to entry: data objects that were originally PII and were later made anonymous are non-personal data.

3.4.6
mixed dataset
set of data objects that contain both *PII* (3.3.1) and *non-personal data* (3.4.5)

3.4.7
data principal
entity to which data relates

Note 1 to entry: The term “data principal” is broader than “PII principal” (or “data subject” as used elsewhere) and is able to denote any entity such as a person, an organization, a device, or a software application.

[SOURCE: ISO/IEC 20889:2018, 3.4]

3.5 Terms related to artificial intelligence

3.5.1
artificial intelligence
<system> capability of an engineered system to acquire, process and apply knowledge and skills

Note 1 to entry: knowledge are facts, information, and skills acquired through experience or education.

[SOURCE: ISO/IEC CD 22989[2)]]

2) Under preparation. Stage at the time of publication: ISO/IEC CD 22989:2020.

3.5.2
artificial intelligence
<engineering discipline>discipline which studies the engineering of systems with the capability to acquire, process and apply knowledge and skills

Note 1 to entry: knowledge are facts, information, and skills acquired through experience or education.

[SOURCE: ISO/IEC CD 22989]

3.5.3
artificial intelligence system
AI system
system using AI

[SOURCE: ISO/IEC CD 22989]

3.5.4
machine learning
ML
process using computational techniques to enable systems to learn from data or experience

[SOURCE: ISO/IEC CD 23053[3)]]

3.5.5
machine learning model
mathematical construct that generates an inference, or prediction, based on input data

Note 1 to entry: for supervised learning, a machine learning model results from the training or a machine learning algorithm

Note 2 to entry: for example, if a univariate linear function (y = w0 + w1(x)) has been trained using linear regression, the resulting model could be y = 3 + 7(x).

[SOURCE: ISO/IEC CD 23053]

3.5.6
model training
<machine learning> task of determining optimal model parameters from a given dataset

[SOURCE: ISO/IEC CD 23053]

3.5.7
trained model
result of model training

[SOURCE: ISO/IEC CD 23053]

3.5.8
training data
samples used to fit a machine learning model

[SOURCE: ISO/IEC CD 23053]

3) Under preparation. Stage at the time of publication: ISO/IEC CD 23053:2020.

3.6 General terms

3.6.1
lifecycle
evolution of a system, product, service, project or other human-made entity, from conception through retirement

[SOURCE: ISO/IEC 29110-4-3:2018, 3.15]

3.6.2
transparency
open, comprehensive and understandable presentation of information

[SOURCE: ISO 21931-2:2019, 3.33]

4 Abbreviated terms

AI	Artificial Intelligence
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
IoT	Internet of Things
IP	Intellectual Property
IP	Internet Protocol
ML	Machine Learning
OPD	Organizational Protected Data
PII	Personally Identifiable Information
SLA	Service Level Agreement
USB	Universal Serial Bus

5 Structure of this document

5.1 Document organization

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).

5.2 Overview and reference architecture

Overview and reference architecture are covered as follows.

— 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 to include devices and the flow of data between devices and cloud services.

5.3 Data taxonomies, data categories and data use statement structure

Data taxonomies, data categories and data use statement structure (applicable to data exchanges between devices and cloud services) are covered as follows.

— 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 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 are 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 customers and cloud service providers. This interaction also makes the discussion of data classification and access and use become 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** which can include application marketplaces. 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** 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.

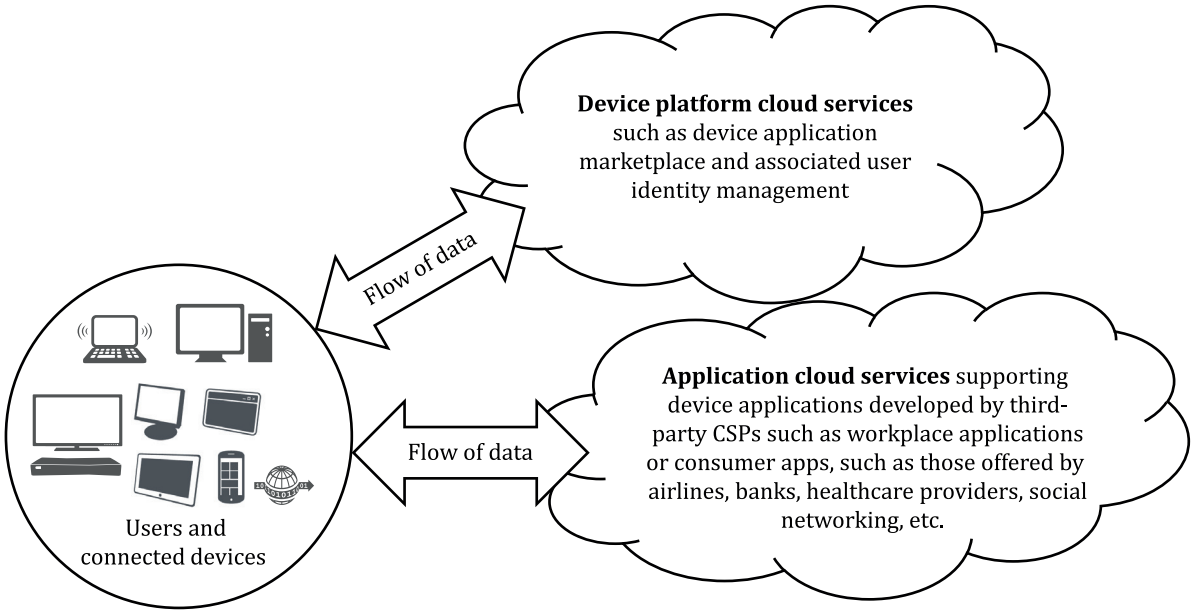


Figure 1 — Devices and cloud services ecosystem

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 are 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 are 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 in Figure 2. Examples include 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, or 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 may take the sub-role of anonymous user, where the user wants the right to use the application and cloud services in a private manner and 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.

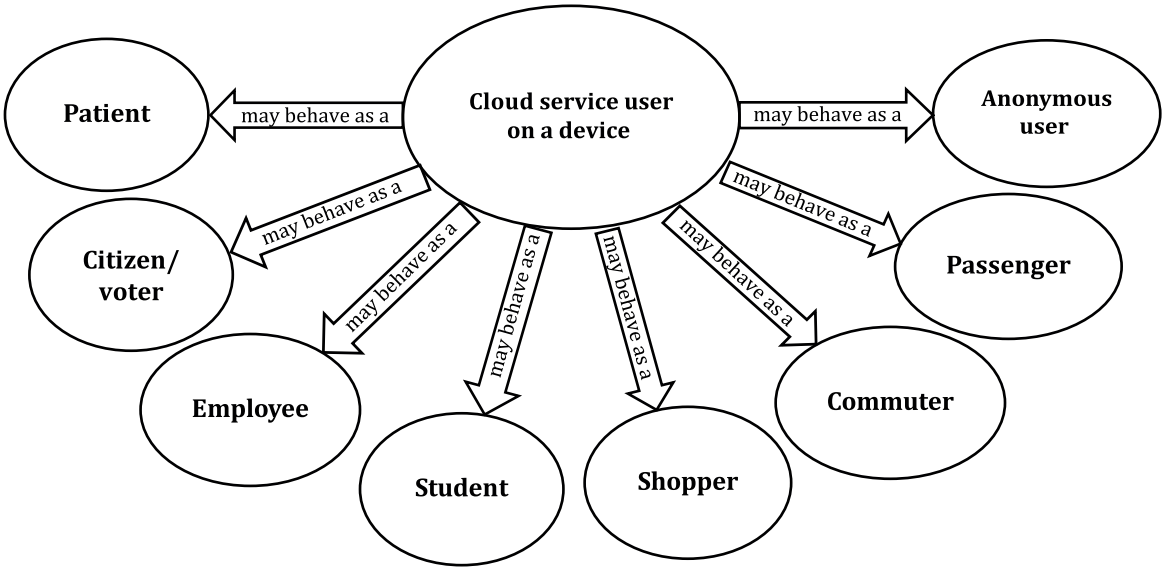


Figure 2 — Example of roles a user can assume in device use scenarios

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

— Patient: patients are subject to 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: employees should follow the organization's policies to protect the organization's confidential assets.

— Student: many students are under legal age and are therefore subject to stricter privacy and commercial advertisement laws.

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

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

— Passenger: passengers are located in public transportation hubs, like airports. Certain rules may therefore 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

"Bring your own device" (BYOD) is defined as the practice of allowing employees of an organisation 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 organisations 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. 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, as a personal device may be shared with others.

— Vulnerability of organization data which is downloaded and stored on the device, as 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 the 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.

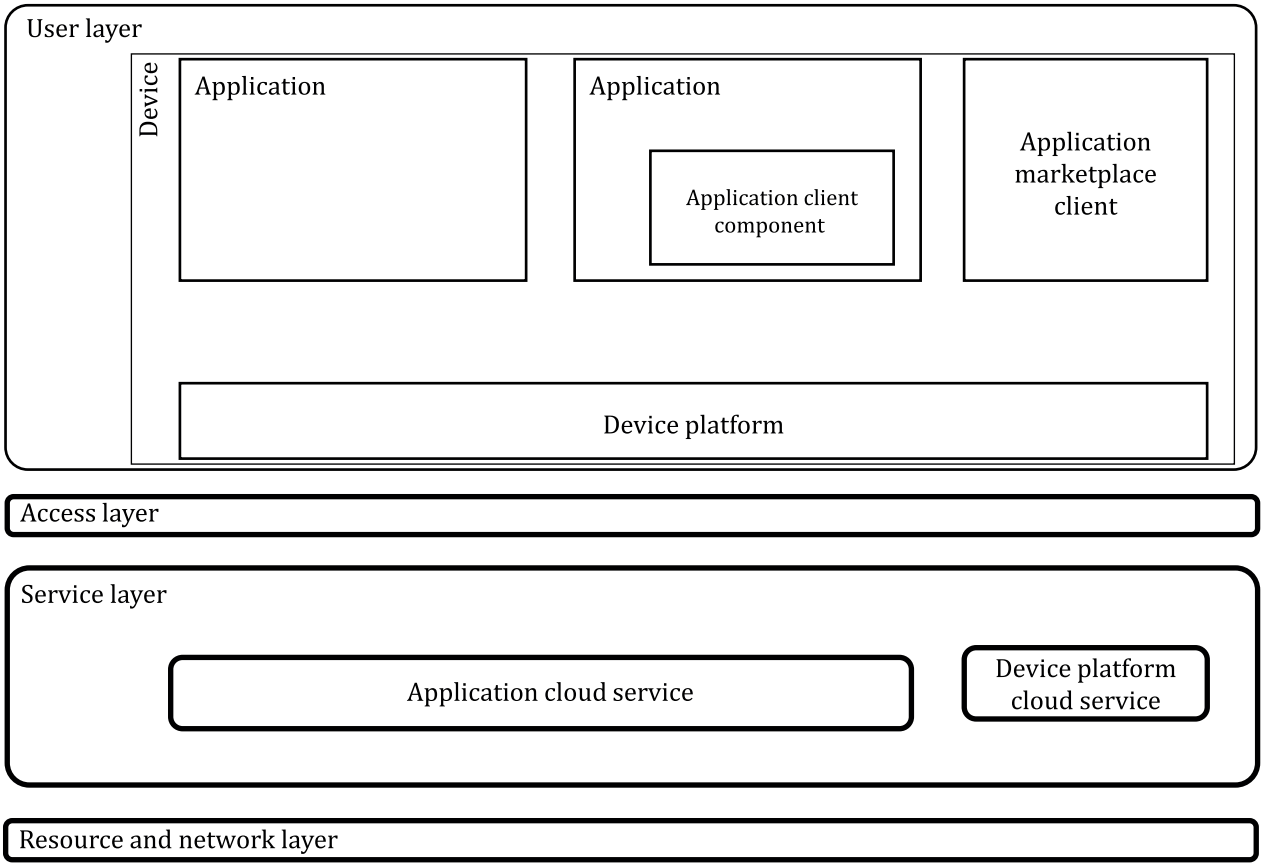


Figure 3 — Devices and cloud services functional view

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

The device platform 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 may 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 are downloaded directly without the assistance of an application marketplace. However, the application 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

The 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 the 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.

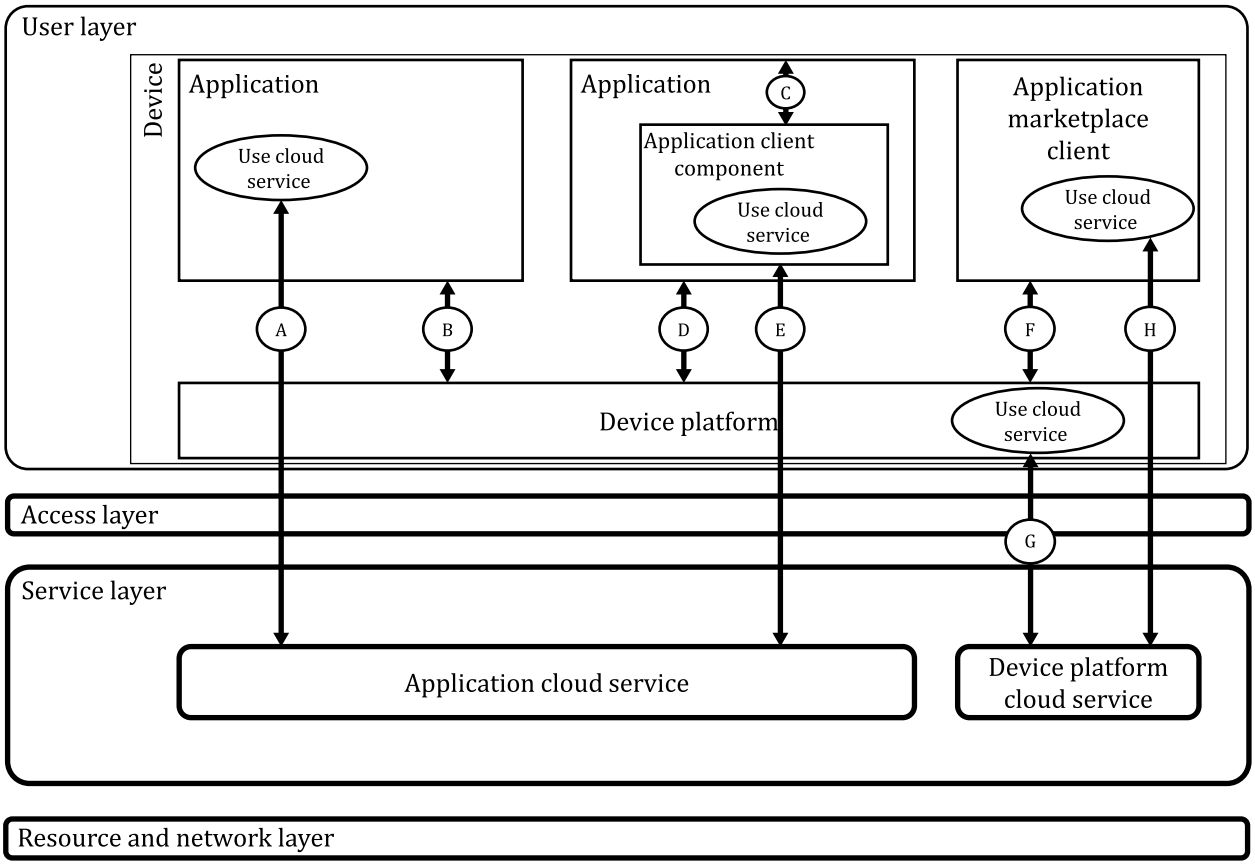


Figure 4 — Data flows between components

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 an application and the 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 an 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 an 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 an 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 are 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, it has two limits. Firstly, the data objects in the ecosystem are constantly changing as technology, devices and cloud services evolve, subjecting the list to constant revision. Secondly, 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 are actually managed by reviewing a large number of individual data objects. To facilitate transparency, cloud service providers should describe how data are 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 may 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 can be used as "data identification qualifiers" as introduced in 8.3.

As an example, a characteristic can indicate whether a particular data category contains PII, the definition of which varies between different jurisdictions and thus makes it difficult to include in a single, global, hierarchy of data categories. Additional views shall 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 that 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 are 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. The 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[20]).

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 certain 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)[19].

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)[17] 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 and credit history. This is usually regarded as sensitive personal information subject to particular regulations.

Financial details relating to an organization as organizational data include information about tax records such as invoices, accounting documents or documents supporting company registration.

8.2.2.10 Sensor measurement data

Data that has been obtained from a measurement sensor. Sensor measurement data are typically organizational data and may even exist in mixed dataset; examples are precision farming (helping to monitor and optimize the use of pesticides, nutrients and water), data about temperature or wind speed from wind turbines, data obtained from industrial robots measuring the environmental elements around them.

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

This refers to 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 the capabilities of the product or service. Telemetry data may contain information about one or more users and is a sub-type of EUII (see 9.3.2).

8.2.3.2.3 Connectivity data

This refers to 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

This refers to data provided or captured about the users' interaction with the service or products by the 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

This refers to data containing demographic information about the 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

This refers to 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

This refers to data 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 refers to data in the form of 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[18].

8.2.3.2.9 Search commands and queries

This refers to data in the form of 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 refers to data in the form of 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 refers to data in the form of 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 refers to data in the form of 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)[19].

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

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

8.2.3.2.13 End-user contact data

This refers to data in the form of 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

This refers to data in the form of 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

This refers to 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

This refers to the 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 a 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.

8.2.5.2 Account or administration contact information

This refers to the contact information for a 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

This refers to 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.

In addition to PII, which contains information related to natural persons, data exists which contains information related to the identities of non-human entities. Such non-human entities can include an agent or an IoT device, or other entity relevant to an organization. Note that in ISO/IEC 20889, human and non-human entities are collectively referred to as data principals.

Organizational data can contain information relating to the identity of non-human data principals and in some cases, this can be regarded as OPD, i.e. the organization regards the identity as confidential information which should not be divulged to unauthorized parties.

This document introduces five degrees of data de-identification in subclause 8.3 that can be achieved using techniques described in ISO/IEC 20889. These degrees of de-identification are used as attribute qualifiers for data objects described in the data use statements.

The de-identification techniques described in ISO/IEC 20889 are sufficiently generic that they can be applied in variety of scenarios involving all data principals, and this can include the identity of non-human entities such as are found in OPD.

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.

Identified data can be either PII or OPD.

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.

Pseudonymized data can be either PII or OPD.

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.

Unlinked pseudonymized data can be either PII or OPD.

8.3.5 Anonymized data

Anonymized data is data that is unlinked and for 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.

For OPD containing information related to the identity of a non-human entity, this can be made into anonymized data by unlinking and alteration of attributes in such a way that the identity of the non-human entity cannot be discovered from the data.

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.

Aggregated data can also be created from information about non-human entities such that individual-level attributes are not identifiable. Such aggregated data can be OPD.

8.4 Orthogonal facets of data

8.4.1 General

Emergence of technologies such as IoT and ML have increased the complexity of data sharing and data use. Often PII and other data are blended together in data sets generated by IoT devices or collected from CSUs during their use of cloud services. To add further precision to transparency and data use expressions supported by the data taxonomy in this document, it is beneficial to separate data under the control of individuals from that of the organizations.

Data about natural persons are often the subject of privacy considerations. Organizational data are often the subject of IP protection, value generation & confidentiality considerations.

There are, however, various orthogonal facets of data that require their own detailed consideration and description. Each facet describes a given property of data and is independent from other facets of that data. Recognizing such orthogonal facets of data and providing an extended taxonomy for their consideration and description enriches the data taxonomy presented in this document.

The subsequent clauses and Table 1 present details of facets of data including description for levels of classification, hierarchies of categorization of data, levels of de-identification of data and the stakeholders who have control over the data.

Such defined orthogonal facets of data can be described by their own respective hierarchical property sets. For example, categories of data described in this document are described in a hierarchy; similarly, there could be a classification of data defined that may include a hierarchical property set for business impact. The levels in the hierarchy and the degree of its granularity depends on each given facet.

Using various facets of data, and the associated attributes of data objects, it is possible to address two important family of data activities: 1) Data obligations and protection and 2) data utilization/value generation. The goal of a well-managed organization is to maximize data utilization while managing obligations and responsibilities relating to the data it holds. A risk-based data management framework can then be constructed to describe and report use of data in an organization.

The data use statement structure can be expanded to support use of attributes of data objects that reference elements from each of the orthogonal facets.

Table 1 presents a few commonly known orthogonal facets of data with each vertical column displaying a facet, and the horizontal rows describe the elements hierarchy for each facet.

Table 1 — Commonly known, orthogonal facets of data

Elements hierarchy for each facet (can be represented as attributes)	**Classification*** (significance of data)	**Categorization** (nature of what data describes)	**Operational control**	**Legal entity**	**Legal means**		**De-Identification**** (Degree of de-identification performed on data)	**Geo-Location/ Jurisdiction*** (where the data are)
	High business impact (HBI)	Telemetry data	Create	Individual	Intellectual property law	Trade secrets	Identified	Municipal
	Medium business impact (MBI)	Biometric data	Read	Organizational		Patents	Pseudonymized	State/Provincial
	Low business impact (LBI)	User's env sensor data	Update	Public/Other		Copyrights	Unlinked pseudonymized	National
	Or more scenario-specific granular levels	Customer contact list	Delete			Trademarks	Anonymized	Multi-jurisdictional
		Connectivity data	Copy			Designs	Aggregated	
		Political opinions	Move		Privacy law			
		Financial details			Privacy regulation			
		etc.			Contract law			
					Organizational polices sourced from organization's goals, risk appetite, ethics, local customs and culture, etc.			
					Other legal means			

* See ISO/IEC 22624.

** See ISO/IEC 20889 for detailed descriptions.

Given the orthogonality requirement, a data object can have a facet designation from each of the above facets simultaneously, as needed.

The following clauses further expand on the commonly known facets in Table 1.

8.4.2 Perspective used in the definition of data facets

When a new facet of data is defined along with the associated hierarchical elements, it is often the case that the facet, along with its elements could need to be defined from the perspective of a group of stakeholders. A common set of such stakeholders are CSPs, CSCs and CSNs. Each element in a given data facet can be defined and be applied from the point of view of a defined stakeholder, i.e. each element is defined from the perspective of a stakeholder.

8.4.3 Common orthogonal data facets

8.4.3.1 Introduction

The following are a list of commonly used facets of data. Each facet of data are described, along with associated hierarchical elements that further break down the given data facet for higher granularity and precision in data disclosures and use statements.

8.4.3.2 Data facets based on control

8.4.3.2.1 Legal control of data compared to operational control

Throughout this document, control over data is used as a central, defining concept. However, there are two types of control over data that need to have precise definitions: operational and legal.

— Operational control: provides the ability to perform operations on the data, for example Create, Read, Update, Delete. This requires having operational control over the data object on which the intended operation is being performed.

— Legal control: provides the legal or contractual basis for having the right to have operational control over data. For example, someone at the human resources department can have the legal right to access employee personnel records (i.e. exercise physical control over that data), but a security hacker can obtain operational control over the same data, by hacking into the IT systems at the company. The hacker has obtained operational control, but at no time did she have legal control.

It is important to differentiate between these two types of control, when making precise definitions. It is worth noting that having operational control does not provide for legal control, and vice versa, as exemplified by the hacking example above. Another example is when someone loses her data stored on a USB stick while traveling on a train. The person has lost physical control over her data, but she still has legal control over it (e.g. the copyright she holds over photos taken and subsequently saved on the USB stick).

The legal control is derived from laws, regulations, contracts, local customs and culture. For example, many jurisdictions grant automatic copyright protection to a person who snaps a photo.

8.4.3.2.2 Operational control data facet

The operational control data facet describes the ability to perform operations on the data, for example Create, Read, Update, Delete, Copy and Move. This means having operational control over the data object on which the intended operation is being performed.

Note that having operational control is independent of having legal control. One may have not have legal control (for example a hacker) but still obtain operational control by breaking into a system.

There could also be custom-defined operations for operational control, allowing the entity with control to perform a custom-defined operation on the data object.

8.4.3.2.3 Legal control data facet

8.4.3.2.3.1 General

The legal control data facet describes the legal, regulatory or contractual basis for having the right to have operational control (see 8.4.3.2.2) over data. For example, someone at the human resources department may have the legal right to access employee personnel records (i.e. exercise operational control over that data). Other examples of legal means of control that offer the right to have operational control are privacy laws or regulations, intellectual property laws, contract law, as well as control over the data dictated by data policies of an organization. Such data policies are typically determined based on an organization's appetite for risk, plans for value generation, local customs, ethics, and culture.

8.4.3.2.3.2 Legal entity data facet

The legal entity data facet describes the kind of entity which exercises control over data objects, including a natural person, an organization or the public at large. Examples of organization are for-profit companies, non-profit organizations or government agencies.

8.4.3.2.3.3 Legal means data facet

The legal means data facet describes the source that legitimizes legal control over data. Examples of such legal means data facets are intellectual property laws, privacy laws and regulations, contract laws and organization's data policies (see ISO/IEC 38505-1 for governance of data considerations).

8.4.3.3 Classification level data facet

Classification level describes the significance of the data, from the perspective of the organization controlling it. It is described in a given number of levels of significance *N* (*N* being scenario specific). Significance of data, and its associated degree or level, is determined by the policies and practices of an organization. As an example, one could define three levels of significance: high business impact (HBI), medium business impact (MBI), and low business impact (LBI). For more details, see ISO/IEC 22624.

The criteria for deciding which data belongs to which level of significance is determined by the organization, specifically using governance of data principles, as directed by the governing board of the organization.

8.4.3.4 Categorization data facet

The categorization data facet is about the nature of what data describes. The portions of data categorization hierarchy is described in clause 8.2. Data categories defined as sub-types of customer content data, derived data, account data and cloud service provider data are examples for this facet. For example, biometric data, political opinions and financial details.

8.4.3.5 De-identification degree data facet

This document defines five degrees of de-identification: identified data, pseudonymized data, unlinked pseudonymized data, anonymized data and aggregated data (see 8.3).

8.4.3.6 Custom data facets

Custom data facets can be defined, for example based on given application of data in a vertical sector. Custom data facets should be defined so as to be orthogonal to any other facet. An associated property hierarchy can be defined if needed.

8.4.3.7 Composite facets and associated attributes

8.4.3.7.1 General

It is possible to combine and apply two or more orthogonal facets of data to create a composite facet that embraces the original facets. This could be useful for scenarios where such short-hand, composite facets may lead to convenient descriptions. Examples are described below, where two facets (legal control and data category) are combined to build a hierarchy of data categories which are useful when customers exchange data with online service providers.

8.4.3.7.2 Customer content data

Customer content data can be thought of as having a composite facet, combining the legal entity facet, based on whether the customer is an organization or an individual, with the legal means data facet, for example contract law. A customer, be it an individual or an organization, has obtained operational control over data, based on the contract between the customer and the CSP.

8.4.3.7.3 Cloud service provider data

Cloud service provider data can be thought of as having a composite facet, combining the legal entity facet, in this case an organization offering cloud services, with the legal means data facet, for example contract law. The CSP, an organization, has operational control over data, typically because the data was always legally controlled by the CSP.

8.4.3.7.4 Derived data

Derived data can be thought of as having a composite facet, combining the legal entity facet, in this case an organization offering cloud services, with the legal means data facet, for example contract law. The CSP, an organization, has obtained operational control over data, based on the contract between the CSU and the CSP.

8.4.3.7.5 Individual, organizational, and public domain data

8.4.3.7.5.1 Introduction

The need to combine two or more orthogonal facets of data to create new composite facets is described later in this clause, where two facets, legal means data facet and legal entity data facet, are combined to build new designation for data sets useful for scenarios where individuals and organizations share data with each other and with their online service providers. Combining elements from the legal means data facet and the legal entity data facets forms composite facets that can help differentiate between what is called individual data and organizational data, as described in the following clauses.

A unified approach to data protection including both privacy and intellectual property protection requires a clear taxonomy that describes the relationships of PII, intellectual property and other properties in pools of data.

It is useful to compare the privacy of individuals with the data protection needs of organizations. While individuals (natural persons) have data and care about their privacy, organizations with data have concerns around intellectual property protection, generating value from their data, and have the need for confidentiality.

The privacy of individuals as well as IP protection and confidentiality of organizations are their respective rights. Such rights, along with other rights and obligations, are derived from applicable laws, regulations, contracts, ethics, customs and culture. Most often, such laws, regulations and contracts are local. This document however provides a core, common taxonomy that can be used internationally to help harmonize data considerations and operations to satisfy regulatory, contractual and operational needs of organizations.

8.4.3.7.5.2 Rights and obligations

Laws, regulations, contracts, ethics and local customs are sources of rights and obligations of individuals and organizations. However, to support data use statements, those rights and obligations cannot be marked into the data sets as attributes. They vary from the perspective of the CSCs, CSPs or CSNs, and also vary from region to region and contract by contract. The attributes assigned to each data object need to be independent of those regional variables.

Rights and obligations can be expressed in data use statements using the taxonomy provided in this document, as well as any extensions of it.

Note that the data use statements in this document use the attributes used in the core taxonomy, along with the data use statement structure. The rights and obligations of individuals and organizations are not captured in this document. They are described and considered outside the core taxonomy since they are derived from laws, contracts, ethics and local customs and will vary from region to region. What is harmonized however is a common taxonomy that can be used to express, with sufficient precision, the various rights and obligations worldwide, and used to negotiate contracts, educate the regulators, and use in transparency and explainability.

8.4.3.7.5.3 Building new composite facets from legal control data facet

An important concept relating to data objects concerns the entities who have control over those data objects. This concept enables the classification of data objects into:

— individual data (3.4.1)

— organizational data (3.4.2)

— public domain data (3.4.4)

Each of these classes is described in turn in what follows.

Individual data is a class of data objects under the control, by legal or other reasons, of a natural person. Note that such data may or may not contain PII.

EXAMPLE A bird hobbyist who photographs birds in the wild may have a wildlife photo collection that does not necessarily bear any of his or her PII. Similarly, someone who collects digital maps of nature trails may have compiled a large collection that he or she values, but there may not be any PII in such collection linking them to the collector.

Individual data is a composite facet that can be constructed by using the element “individual” from the legal entity data facet and combining it with a desired element from the legal means data facet, such as copyright or privacy law, depending on the scenario at hand.

Note that customer content data is individual data when the CSC is a natural person.

Organizational data is a class of data objects under the control, by legal, contractual or other reasons, of an organization. An organization can be a for-profit company, a non-profit organization, a public or government agency, a non-governmental organization or an international organization, and can be small, medium or large.

Organizational data is a composite facet that can be constructed by using the element “organization” from the legal entity data facet and combining it with a desired element from the legal means data facet such as patent, trademark, copyright or contract law, depending on the scenario at hand.

Note that customer content data is organizational data when the CSC is an organization. Cloud service provider data (ISO/IEC 17788) is always organizational data by nature.

OPD is organizational data whose protection is required based on the policies established by governance of data processes. Notice that every organization has policies that govern the data under their control.

Note that organizational data can contain both OPD and PII and both of these require protection in various ways. OPD is likely to contain material requiring confidentiality or containing intellectual property and can cover a wide range of material including financial data, trade secrets, copyright information, product designs and software. PII can concern individuals who are members of the organization (e.g. employees) as well as individuals who are not members of the organization (e.g. sales leads, customers or the population at large).

Public domain data is a class of data objects over which nobody holds or can hold copyright or other intellectual property. As a result, any natural person and any organization may make use of public domain data without restriction, license or contract.

Data becomes public domain because someone puts that data into the public domain, or because any rights over the data has expired.

8.4.3.7.5.4 Co-controlled data

Subclause 8.4.3.7.5.3 describes cases where control of data unambiguously belongs to one party. There are also cases where control of data in fact belongs to multiple parties at the same time. It is also the case that classes of data objects exist where different control regimes apply to different data elements within the data objects. Broadly speaking, these situations are termed co-control of data.

Co-control can involve multiple organizations, or it can involve multiple natural persons, or it can involve a mixture of natural persons and organizations.

One case of co-control involves an organization processing data containing PII. In many jurisdictions, the PII principal (a natural person) has some level of control over the PII in the data, in addition to the control the organization has over the data. For example, the PII principal can control what processing the organization performs on the PII. Or the PII principal can have operational control to update or delete the PII.

Another case of co-control involves data sharing between two or more organizations (see ISO/IEC 23751[4)]for more details on data sharing). In this case, the organization providing or owning the data can impose limitations on the control the receiving organization has over the data, by means of a contract in the form of a data sharing agreement.

Other examples of control over data between individuals and organizations can arise from laws and regulations such as consumer laws. Between organizations, shared data can be subject to limitations such as confidentiality obligations.

Co-control of data is becoming increasingly common. Ever more data are being created or collected and its value increasingly depends on the data being shared between individuals and organizations or between multiple organizations.

Data attributes in particular need to be associated with at least one of the parties involved in co-control, reflecting the perspective of that party with respect to that data. However, in the scenarios involving co-control of data, attribute values representing each co-controller can be applied.

8.4.3.7.5.5 Change of control over the data lifecycle

Whether data are individual or organizational, and how many organizations can claim a given data object as their organizational data, can change over the data lifecycle.

EXAMPLE When a person snaps a photo on her phone, she has created a data object that is her individual data. She may then upload that photo to a social media service, in which case depending on the service's end-user licensing agreement, the organization providing the service may obtain partial or full control of that data, making the photo organizational data. The provider is then permitted to perform data processing or analysis on that photo, for example as part of training an image recognition algorithm. The trained ML model would itself be organizational data, even though it was developed based on the user's photo, along with many others' photos.

4) Under preparation. Stage at the time of publication: ISO/IEC CD 23751:2020.

The lineage of data is important in establishing its provenance and is useful in establishing the data control regime, see Clause 11. When data are created, its provenance can readily be established by examining how it is created. If data are created by an individual, the data are individual data of if the person who created it was acting as a private individual. But if that same person was an employee of an organization and the data was created as the work done by the employee for the organization, the data are typically organizational data, under most employment contracts.

The user could share the individual data with an organization, under an implicit or explicit license, or data sharing agreement. The organization views the shared copy of data as its organizational data, again, subject to the agreement with the user who shared the data.

In the case of data created by an employee during the course of work done for the employer, the data are organizational data at the time of creation, again based on a typical employment agreement.

Data can be individual and organizations at the same time. The person who originally created the data views it as the person's individual data, with or without the PII present in it. The organization with which the person chooses to share the data under a data sharing agreement would subsequently treat its copy of the data as its organizational data, if the data sharing agreement used between the person and the organization provides the organization with a degree of control of that data, and rights to using that same data.

Two organizations can also share data with one another. For example, a movie studio could produce a movie which can then be licensed to a distributor. In this case both the movie studio and the distributor would treat the movie as their organizational data. The movie studio is the creator and has full rights and control of the movie. The distributor, being another organization, has obtained the right to distribute the movie to theatres or online streaming services, under an IP licensing agreement. The agreement gives the distributor certain rights, control and rendering the movie as organizational data of the distributor.

The movie streaming service also obtains the movie under a clearly negotiated licensing agreement. The agreement provides the streaming service certain rights and associated controls over the movie. For example, the right to stream the movie to its users for a negotiated fee, and the right to offer to download the movie for offline viewing for a limited time period. The level of control and the rights to the data are different for each organization, but the movie is the organizational data of the companies.

The provenance of the data can be traced by examining the data lifecycle from the point it was created.

8.4.3.7.5.6 Legal control of organizational data

When an organization acquires data, the copies that it holds become organizational data for which the organization assumes both control and responsibility. This acquired data may come with legal constraints, either explicit (such as a license from a copyright holder or distributor) or implicit constraints (such a regulatory obligation under privacy or data protection law). Legal constraints remain with the data throughout its lifecycle in the organization.

In the case of organizational data which contains PII, the copies held by the organization are treated as organizational data, taking into account the legal obligations applying to PII in the relevant jurisdiction. For example, the natural persons in European Union member states have certain rights and controls over their PII. The General Data Protection Regulation (GDPR)[17] gives the user the right to request update or deletion of the PII.

Both an organization and a person could have rights and control over copies of the same data, termed co-controlled data. Such data can be individual data and organizational data at the same time, although the rights and control could differ between the parties concerned.

In the case of data which contains intellectual property of another party (such as copyright), the copies held by the organization are organizational data, but are always treated according to the governing license or data sharing agreement under which the data was received.

The organization is responsible for exercising appropriate control and protection of all organizational data copies that they hold. The legal constraints applicable to any specific element of organizational data as mentioned above frequently determine the nature of this control and protection. Also, the use to which such organizational data can be put is often limited by such legal constraints.

Some data processes within the organization create new organizational data based on the original data, that is not subject to the original legal constraints, such as anonymisation or aggregation of data.

Control of data could be derived from more than one set of legal concerns, for example copyright and intellectual property laws as well as privacy laws that could govern who has control of the data, the nature of control, and restriction on their rights. It is possible to have complex scenarios involving the rights and associated controls of copies of the same data held by a person and multiple organizations, all at the same time.

There could be more than one set of laws imposing external constraints on control of data. Copyright laws and privacy laws are examples of such simultaneously applicable legal frameworks.

8.4.4 Use of data facets to describe data taxonomy

The data categorization hierarchy in this document combines two facets of data into a composite facet:

1) Cloud service provider data, cloud service customer data or cloud service derived data (based on the control facet in a provider/customer relationship).

2) Type of data based on the nature of its content (data category facet).

For point 1), there are three elements and associated attributes (e.g., CSC_DATA, CSP_DATA, and DERIVED_DATA), and for point 2) above associated attributes such as BROWSER_HISTORY and TELEMETRY can be used. Note that any categories listed in 8.2 can be used here.

NOTE When CSC is a natural person, cloud service customer data is also individual data, by definition. Similarly, when the CSC is an organization, the cloud customer data is organizational data.

Note that by definition, CSP Data is always organizational.

Derived data, by definition, is “objects under cloud service provider control” (ISO/IEC 17788:2014, 3.2.13), hence always organizational data.

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 can 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 are applied to a set of data independently or in combination with each other to achieve specific goals. 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 assumes 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 may 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 is 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 (policies may be needed on how to automatically resolve discrepancies in the event that they have different values);

— 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.

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[15].

Various technological approaches can be used for data deletion. They differ in their properties[16], 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[16].

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. It is possible to achieve re-identification 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, for example.

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 the data use categories described in this document.

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

9.3.2 Provide

9.3.2.1 General

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" can 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* also means to protect the customer content data that exists within this scope and to provide and communicate with the customer about the status and availability of the current capabilities of this scope.

9.3.2.2 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.

9.3.2.3 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 the 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/advertize/promote

9.3.6.1 General

Market/advertize/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/advertize/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/promote 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."

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 are 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 1 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.

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

Cloud service providers should use scope characteristics (see 9.4.3) for the source scope and for the receiving entity or result scope to further specify the sharing.

Cloud service providers should include a description of the network connection between scopes (see 9.4.4) following the statement structure described in data sharing (see 10.2.9).

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 processes 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 data provided by the third parties to provide the services in the scope.

9.3.8 Collect

Collect includes collecting, preparing, pre-processing and storing specified data categories from the source scope in preparation for other uses such as training machine learning algorithms.

9.3.9 Train (AI system)

Use the specified data categories from the source scope to train, retrain or test an artificial intelligence (AI) system. The AI system can use machine learning technologies, or it can use other technologies.

Cloud service providers describing the use of data to train AI systems should include a statement addressing the extent to which individuals are directly identified in the resulting AI system. The data identification qualifiers in 8.3 could be used as the basis for this description. If there is PII in the resulting AI system, then additional statements are likely required, e.g. about data retention periods.

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 shall 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.

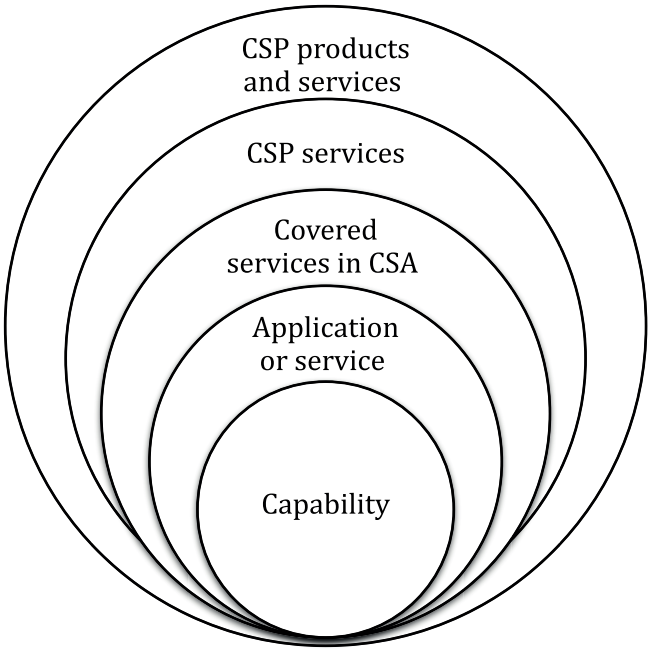


Figure 5 — Increasing levels of scope

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" may 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" may 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

This refers to 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

This refers to 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

This refers to 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" is used to denote an entity that provides data from a source scope or receives data as a result scope.

9.4.2.8 Third-party and data processors

This refers to 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.

9.4.3 Scope characteristics

9.4.3.1 General

Cloud service providers should include descriptions of additional characteristics of a scope when necessary to fully describe a use, for example, additional information about the source scope (sending) and result scope (receiving entity) are required when used in sharing statements (10.2.9).

9.4.3.2 Guaranteed capabilities of a scope

Cloud service providers should describe any guaranteed capabilities of a scope necessary to support the use statements. For example, data transfer may only be permissible if a receiving entity in a share (9.3.7) use guarantees the data will be encrypted at rest, or that the entity can respond to a request for deletion.

9.4.3.3 Controlling entity of a scope

Cloud service providers should describe the entity controlling the scope necessary to support the use statement. If no additional specification of the entity is provided, then cloud service provider is assumed to control all scopes, except for third-party product and services (9.4.2.7) and third-party and data processors (9.4.2.8). For these scopes, any third party may control the scope.

9.4.3.4 Scope location

Cloud service providers should describe where data in the scope resides. The precision on the location should be sufficient to support other descriptions such as clause 9.4.5. If no location is provided the assumption is the controlling entity of the scope is an international organization, and the data may reside in any location.

NOTE ISO/IEC 27701:2019, 7.5.2, presents an example of a need to document the location of a result scope at the precision of a specific country.

9.4.4 Network connection between scopes

9.4.4.1 General

Cloud service providers should include a description of the network connection between the source scope and the result scope when necessary to completely describe a use. For example, a statement about sharing data types (9.3.7) may include a description of the network connection between the source scope and the receiving entity (result scope).

9.4.4.2 Guaranteed capabilities of a network connection

Cloud service providers should describe any guaranteed capabilities of a network connection necessary to support the use statements. For example, a network connection may be guaranteed to

provide adequate performance, or be guaranteed to protect data crossing the network connection with encryption. The guarantee is assumed to be made by the controlling entity of the source scope unless otherwise described.

9.4.5 Control of source scope over result scope

The cloud services provider should describe any means of control it has over the use of the data types by the results scope. For example, a cloud services provider that shares data with a third-party data processor (9.4.2.8) has legal control through a contract over the processing by the third-party. In other cases, measures such as encryption may limit the physical control entities in the result scope have over the data.

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 are 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 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 CSPs, CSNs, or 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, it is possible to reduce the risk of incomplete or poorly drafted data use statements.

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 can 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 are 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.

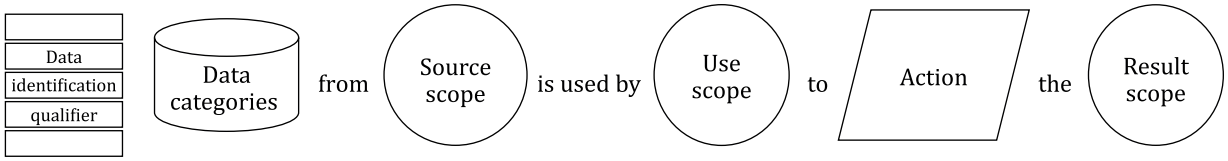


Figure 6 — Use statement structure (passive)

The following example follows the structure in Figure 6:

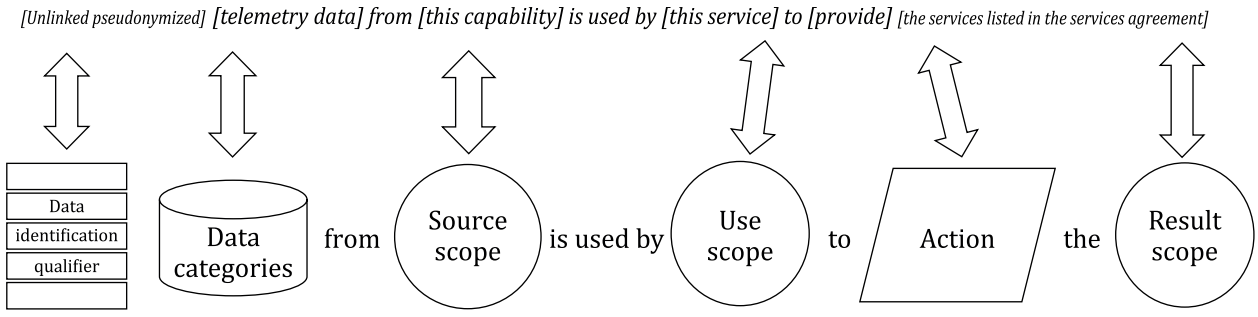


Figure 7 — Example of use statement structure (passive)

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.

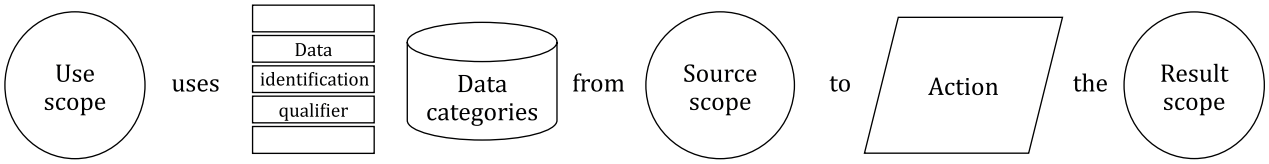


Figure 8 — Use statement structure (active)

The example given in Figure 9 follows the structure in Figure 8.

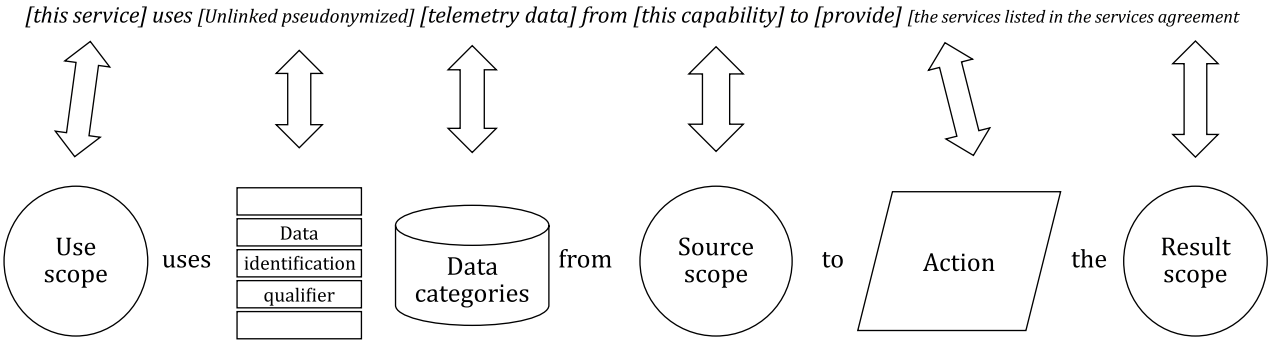


Figure 9 — Example of use statement structure (active)

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 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 are assumed to come exclusively from the use scope and the results of the data use (result scope) to apply only to that scope.

If only a source and result scope are specified, the use scope is assumed to be the same as the source scope.

NOTE It is not possible to sensibly use scopes that do not include the cloud service provider, e.g. “partner and processors” and “third party”as a single scope in communications between a cloud service provider and a cloud service customer.

Example 6 represents a data use statement with a single scope definition:

Example 6:

The cloud services covered in this agreement encrypt end user identifiable information.

Example 7 represents a data use statement with a dual scope definition:

Example 7:

The mapping service uses user location data to provide the route finder service.

10.2.3 Assumptions about when data are collected and used

Unless otherwise specified, data assumed to come from the current and any past use of the source scope and its use applies to the current and any future use of the results scope.

In some cases, the capabilities in the scope can be accessed in a single defined session of use and specifying “current use” is helpful when describing use of data that is not kept after a session is finished. When just a single scope is defined for source and results, “current” means the data type is not retained from that scope.

Example 8 represents a data use statement making use of a timing:

Example 8:

> *The recommendation service uses user location data from the current use of the mapping service to personalize the offerings proposed by the recommendation service*

It is important to have an unambiguous definition of what is meant by a single session. When using cloud services through applications, especially on mobile devices, the beginning and end of a session may be marked by specific actions in the application (e.g. "login", "logout"). In other cases, the session might be ended by a specified period of inactivity. It is assumed that a single session cannot span the "partners and processors" and "third-party" scopes and the use of "current" should be clearly defined for those scopes.

10.2.4 Defining promotion targets

Data usage that result in presenting advertising or communicating with people for commercial purposes (e.g. upgrade/upsell and promote) should specify the person or sets of people that might be contacted.

Sets of people can be defined by combining roles defined in ISO/IEC 17789 (e.g. cloud service "user" or "cloud service administrator") with an application or service scope, for example: "users of this service".

"Customer" or "you" may be used to describe the target of the promotion when multiple roles associated with a cloud service customer, e.g. customer, user, administrator, are fulfilled by a single person.

Example 9 represents a data use statement which defines a promotion target:

Example 9:

> *The restaurant search service* uses *your profiling data* to *advertise* matching establishments to you and other users of the service in your friends list.

10.2.5 Data types

The data categories defined in 8.2 are organized in a tree structure to allow a CSP to create clear use statements by defining the broadest possible category of definitions and any applicable exceptions.

Example 10 shows a set of data use statements a service might make to qualify its data use:

Example 10:

1. End user identifiable information from this service is used to provide the service.
2. Organizational identifiable information from this service is used to provide the service.
3. Anonymized telemetry information from this service is used to improve all our services.

The three statements in Example 10 are conformant to the data use statement structure in this document. However, these statements can also be stated more clearly by referring to derived data, which includes EUII and OII and stating the exception explicitly. Note that the third statement applies to all of the CSP's services.

Example 11 shows a data use statement equivalent in meaning to the three contained in Example 10, which is conformant with the data use statement structure in this document:

Example 11:

> *Derived data from the weather service is used to provide the weather service, except anonymized telemetry data from the weather service is also used to improve all of our services.*

Arranging data types in a tree structure is also intended to allow cloud service providers to make direct statements about the use of data instead of forcing them to expand definitions of data types. For

example, a CSP may treat biometric and heath data in the same manner as customer content data, with commitments to use both data types only to provide the service and by providing features that allow user control of biometric and health data. However, the CSP may not commit to the same restrictions of use nor provide the same controls for other categories of EUII. Instead of reclassifying biometric and heath data as customer content data to reflect the similar policy for both data categories, this document allows use of the biometric and health data to be specifically declared, as shown in Example 12:

Example 12:

Biometric and health data from the end user device is used only to provide this service. This service provides capabilities for cloud service users to control the biometric and health data stored in the service.

Note that unlike data categories or scope statements, use statements are distinct from each other, without a tree structure or a nested implied increase in use.

10.2.6 Data qualifiers for data types

This document does not assume any particular use of a category of data is appropriate or inappropriate. Although it is possible that in a few cases certain data categories do not make sense with specific data qualifiers, in most cases qualifiers are appropriate. Customer content data can be described with any data qualifier, for example an oil service company may upload geographic data that, while extremely confidential, does not include information linked to individuals. A cloud service provider that declares that it uses customer content data only to provide IaaS services is unlikely to know if the data provided by the customer is PII unless directly informed, but that is not assumed in this document and a specific declaration with qualifiers should be provided.

Derived data, as defined in 8.2.3.1, appears likely to be PII since many interactions with the cloud service typically require an individual account. Since direct contact information for the customer is separately considered as account data, OII may not in fact include PII. Therefore, data elements of this type may not require data identification qualifiers. On the other hand, end-user identifiable information (EUII) is by definition PII, therefore data elements of this type or its sub-types may require data qualifiers.

Account data can be assumed to contain PII information about people holding roles (e.g. cloud service administrator) as part of the CSC. Although it is natural to assume that account information is used to provide the service it can be used in other ways and in broader scopes and use should be declared using a data identification qualifier.

Cloud service provider data is defined to be unique to the cloud service and although it is exclusive of EUII, it may contain other forms of PII and use should also be declared using a data identification qualifier.

The description of data in a description of data use should include appropriate data identification qualifiers to clarify the degree that the content of the data type is linked to an individual (according to 8.3) in the specific circumstances being described.

Example 13 demonstrates the use of a data identification qualifier in a data use statement:

Example 13:

The sentiment analysis service uses unlinked pseudonymized cloud service derived data to improve the cloud service provider's products.

10.2.7 Examples of statements about data flow in the devices and cloud services ecosystem

Data taxonomy information can also be used to make specific statements about data flows in the devices and cloud services ecosystem. To make a data use statement about data flow between two elements of the ecosystem, the source scope or the result scope references a transfer.

Example 14 illustrates a statement about a location service that has data "sent from the mobile device by an application" as a source scope:

Example 14:

> *Pseudonymized precise customer location data sent from the mobile device by the mapping application is used by the platform cloud service to improve the location service.*

In Example 14, data sent from the device is de-identified on the device (note it is qualified as pseudonymized) before it is sent to the device platform cloud service. This use statement also includes the device platform cloud service as the use scope and identifies the location service as the result scope.

An alternative architectural approach that relies on an encrypted link instead of device-side de-identification can be described within a data use statement using a data processing technique as shown in Example 15:

Example 15:

> *Precise user location data sent from the mobile device by the mapping application that is encrypted while in motion using network security is used by the device platform cloud service to improve the location service.*

Data use statements can include a description of the transfer of the data ("is delivered to the mobile device" in Example 16). In Example 16, the data use statement about email data are for the device platform cloud service. The clause about delivery to the mobile device is present to clarify the data flow in the devices and cloud services ecosystem:

Example 16:

> *Email customer content data from the email service is used by the device platform cloud service to promote based on personalization third party products and services for users of the email service before the email customer content data is delivered to the mobile device.*

The data use statement in Example 16 allows the use of any customer email to define promotions for any user. If the email content is used exclusively to personalize the application on the mobile device the device becomes the result scope and declaration is as shown in Example 17:

Example 17:

> *Email customer content data from the email service is used by the device platform cloud service to personalize the mobile device email application.*

The data use statement in Example 17 does not address *how* the device platform cloud service changes the behaviour of the device, neither does it make any statements about data flow. A second statement could clarify the transfer of data as shown in Example 18:

Example 18:

> *Pseudonymized social data from the restaurant service sent from the device platform cloud service to the device using encryption of the data in motion is used by the mobile device application to personalize the mobile device application capability.*

10.2.8 Exceptional use statements

10.2.8.1 General

The data use statements described in Clause 10 are assumed to be true for the entire time of the agreement between the cloud service customer and cloud service provider and reflect an assumption that the CSP has routine access to the data categories necessary to execute the data use statements. Using a specific structure for statements to describe exceptions to stated use add clarity, conformity and transparency to the communications between CSP and CSC.

10.2.8.2 Structure

An exceptional use statement relies on the data use statements defined in this document along with roles and sub-roles described for parties defined in ISO/IEC 17789. An exceptional use statement defines who has the ability to grant permission to whom to make what use. That granted permission is the result of some action by the grantor and the permission is granted for a period of time.

Complete exceptional use statements define:

— the entity granting permission (a grantor);

— the entity making the exceptional use of the data (a grantee);

— the exceptional use [a use statement (see 10.2)];

— what can cause or is required for the grant of permission to occur (a grant trigger);

— how long the grant is in effect (a grant period).

The arrangement of these terms varies by context in where they are used. The pattern for an exceptional use statement is shown in Example 19:

Example 19:

> *The [grantor] grants permission to [grantee] to [exceptional use] by [grant trigger.] The grant is effective [grant period.]*

10.2.8.3 Grantor

The term "grantor" represents the entity, such as a person or organization that can grant permission to perform the exceptional use. If left unspecified the cloud service customer is assumed to be the grantor.

10.2.8.4 Grantee

The term "grantee" represents the entity, such as a person or organization that performs the exceptional use. If left unspecified the cloud service provider is assumed to be the grantee.

NOTE The role and sub-role definitions found in ISO/IEC 17789 may be useful to describe the grantor and grantee.

10.2.8.5 Exceptional use

An exceptional use statement provides additional information to a data use statement (see 10.2) to add transparency and precision about when data use is allowed. Example 20 shows an exceptional use statement:

Example 20:

> *The cloud service customer grants permission to the cloud service provider to provide emergency move of customer data from this service to another geographical location in case of a natural disaster. This is effective until the consequences of the natural disaster are dealt with, up to a maximum period of nine months.*

In Example 20, the cloud service customer is the grantor, the cloud service provider is the grantee and the grant trigger is "in case of a natural disaster". The grant period is specified.

Exceptional use statements can be used to describe exceptions for a narrower use than those defined in the data use categories (see 9.3). For example, a grant for exceptional use may be limited to "provide customer support" rather than the broader use "provide" (see 9.3.2). The assumption is that the more narrowly defined use is the granted use in this scenario.

Terms defined in the context of this document such as "provide" should not be expanded in scope in exceptional use statements. For example, "provide improvement" is ambiguous and will reduce the precision and transparency of the exceptional statement and the other use statements.

NOTE The exceptional data use and access triggered by authorized agencies with legal jurisdiction is outside the scope of this document.

10.2.8.6 Grant trigger

The grant trigger describes the event that causes the grant to come into effect. For example, exceptional use statements could describe the data categories required to fulfil a customer support request. For that use, the beginning of the customer support request can be the grant trigger.

The exceptional use statement describes whether the grant trigger occurs automatically as a result of some event occurring, or if the grantor needs to explicitly grant permission for the exceptional use. In Example 21 the exceptional use statement specifies that use of the location data is an automatic consequence of submitting the customer service request.

Example 21:

> *The cloud service user grants permission to the cloud service customer support and care representative to use location data from the mapping capability to provide support for a specific incident when the user submits a customer support request.*

In contrast, Example 22 specifies that access is given to a cloud service security and risk manager only when the cloud service administrator role of cloud service customer *explicitly* triggers the grant:

Example 22:

> *The cloud service administrator grants permission to the cloud service security and risk manager to access the current capability to provide emergency backup of customer content data by providing temporary access through the administrative tools interface. The grant period terminates once the backup is completed.*

Used in this way exceptional use statements can provide a formal definition of consent. Example 23 shows how this can be achieved:

Example 23:

> *The cloud service user consents to use of end user identifiable information from this service to promote third party products and services using personalization by selecting the appropriate options in the trust centre. That consent remains in force until the user revokes consent through the same capability.*

In Example 23, the cloud service user is the grantor, the cloud service provider is the grantee and "this service" is the source scope and "selecting the appropriate options in the trust centre" is the grant trigger. The grant period is defined by the last sentence.

10.2.8.7 Grant period

The grant period specifies how long the granted permission remains in effect. The start of the period can be the triggering event or a specifically defined time, including the beginning of the agreement between the CSP and CSC.

The definition of the end of the period depends on the nature of the grant trigger, but could include the conclusion of the triggering event, a fixed period of time, revocation by the grantor, or end of the agreement between CSC and CSP. Grant periods that begin with a specific activity, such as a customer support request, are assumed to end with the completion of the activity unless specified otherwise.

10.2.9 Data sharing

Transparency about scenarios such as data portability, transfer of data between jurisdictions as a part of the operation of the service or the sales of data to third parties means that additional information should be provided as part of a data use statement. The share (9.3.7) use assumes the transfer of datatypes to an entity outside the source scope, however a comprehensive description of sharing should also include a description of the network connection between the scopes and specific information about that network connection and the scopes.

Figure 10 demonstrates the addition of the network connection statement to a share data use statement structure. In addition to the network connection statement, a comprehensive data use statement for sharing should include:

a) any required guaranteed capabilities of the network connection (9.4.4.2);

b) a data use statement (Clause 10) for addressing the intended use transferred data by the result scope and a description of control of source scope over result scope (9.4.5);

c) any required guaranteed characteristics of each scope (9.4.3.2);

d) the controlling entity of each scope (9.4.5);

e) the result scope location (9.4.3.4).

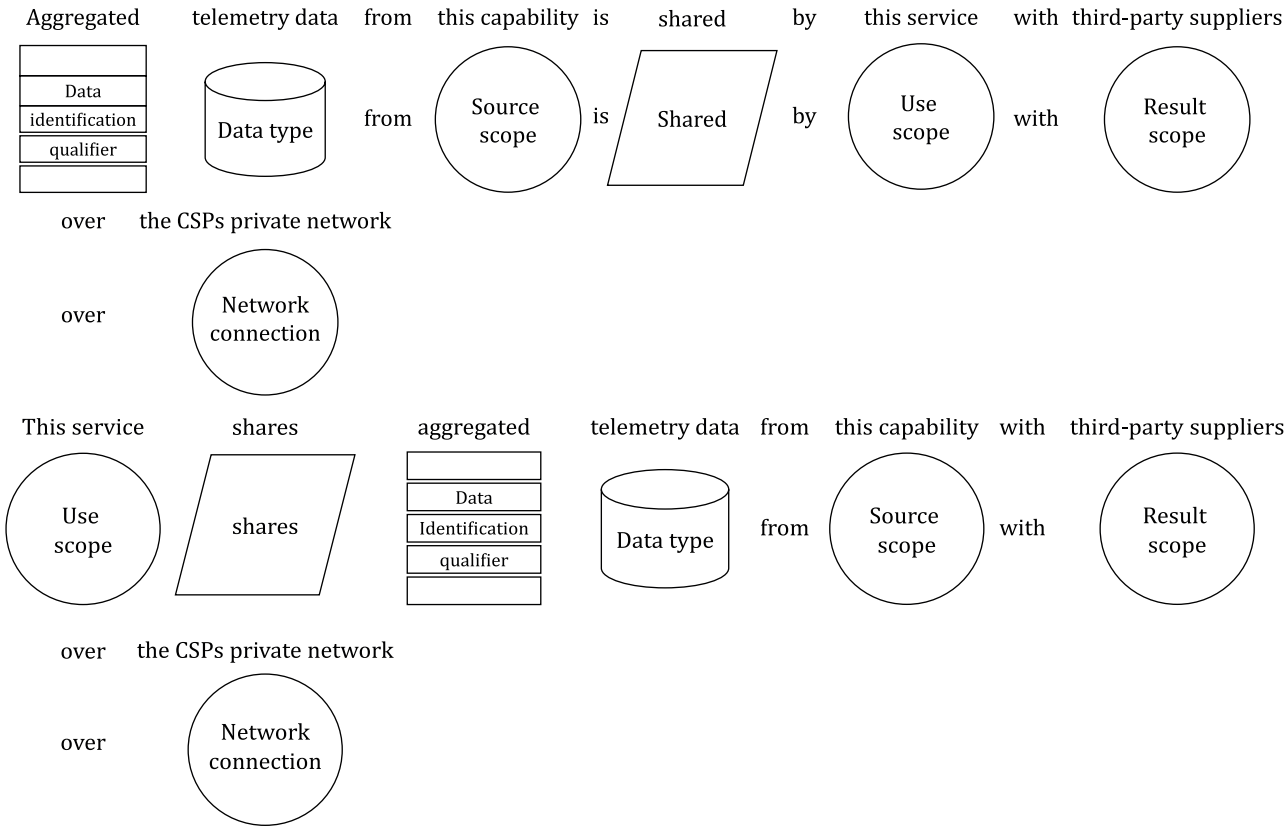


Figure 10 — Example of use statement with addition of a network connection description

Example 24:

Unlinked pseudonymized content consumption data from the application marketplace shared by the CSP with third-party application providers over the internet.

In Example 24, the content consumption data, which is processed to be unlinked pseudonymous, from an application marketplace capability provided by the CSP, is transferred to third-party scope

of application providers. The data are transferred over the internet, which implies no guaranteed capabilities for the network connection. Since no other information is provided, the statement implies that the CSP has no control over the use of the data by the third-party application developers, that their use of the data is unrestricted and that the data can be located anywhere.

Example 25:

> *Pseudonymized telemetry data from covered services in the CSA is shared by the CSP with third-party data processors over the internet using a VPN for encryption. This data are used by the third-party data processor to provide the service under contract to the CSP that ensures that the commitments made in the CSA are upheld. The third-party data processors are required to store and process the data in the same country as committed to the CSC by the CSP, and a list of their names is provided in the CSA.*

In Example 25, the pseudonymized telemetry data is shared with third party data processors (9.4.2.8) that are contractually bound to uphold the commitments in the cloud service agreement made by the cloud service provider, ensuring that the third-party data processor meets equivalent specific characteristics (9.4.3) as the CSP. The required capabilities for the network connection for this sharing is encryption and a private network. In the second sentence, an additional use statement, with a single scope of the third-party data processor, specifically states the contractual control of the CSP over the data processor for the telemetry data. The third sentence makes a commitment to the location of the third-party data processor, and references a list specifying the controlling entity in the cloud services agreement.

10.3 Use of orthogonal data facets in data use statement

10.3.1 General

This clause describes the use of elements in the data facets as attributes on data objects in the data use statement structure. The attributes, for example PII and OPD, can be used in the data use statement structure to make detailed and precise statements about how data are used. The facets provide new perspectives of data handling that enrich the statements made about data use.

10.3.2 Use of elements in the data facets as attributes

The elements in the orthogonal facets described in this document can be used as attributes when making data use statements. For example, the following set of attributes of data objects can describe a PII property:

[Individual][Privacy Law]; the [Individual] attributes are taken from the legal entity facet, noting that the data object is about an individual. The [Privacy Law] attribute signals that the individual has obtained operational control over data by the element in Legal Means of Control facet that is “privacy law”; in other words, a privacy law or regulation has enabled the individual to have operational control over the data object.

Similarly, OPD can be described as a property of data objects with the following attributes:

[Organization][IP Law] (or [Contract]). The [Organization] attributes are taken from the legal entity facet, noting that the data object is about an organization. [IP Law] or [Contract] signals that the organization has obtained operational control over data by the element in Legal Means of Control facet that is “IP law” or “contract”; in other words, an IP law or a contract has enabled the organization to gain operational control over the data object.

Any element from any facet can be used as an attribute in the data use statement structure. This allows for flexible and rich data use statements that can address many different facets of data use, depending on the application or scenario at hand.

10.3.3 Hierarchy of elements/attributes of data based on facets

This clause describes a taxonomy relationship for the elements in the legal means facet when combined with a legal entity facet. It also shows how elements from other facets can be used as well. Such taxonomy of data can be used for establishing data provenance.

There are two taxonomy relationships: 1) containment and 2) assignment of attributes.

1) contains: "A contains B" if "members of B are members of A with some additional properties not shared by other elements of A" – for ease of use we use "be" verb to say the same.

2) has attributes… (0 or more attributes): the attributes are a list of commonly needed properties and are not exhaustive.

Data contains individual data, organizational data or public domain data. Exclusiveness already follows from the definitions of individual and organizational data, depending on whose perspective is used.

Individual data can contain PII.

Organizational data can contain PII or OPD, or both.

Any data object can have 0 or more of the attributes derived from elements of various facets of data. A few examples for possible attributes are listed below:

— intellectual property law,

— a confidentiality level,

— a custom classification level,

— a custom attribute.

Intellectual property law can itself have sub-types attributes, such as trade secrets, published patents, copyrights, trademarks, designs. The list is not exhaustive.

Other facets of data could also have sub-attributes, such as levels of de-identification, or classification levels.

The following figure shows the examples of attributes derived from elements of various data facets:

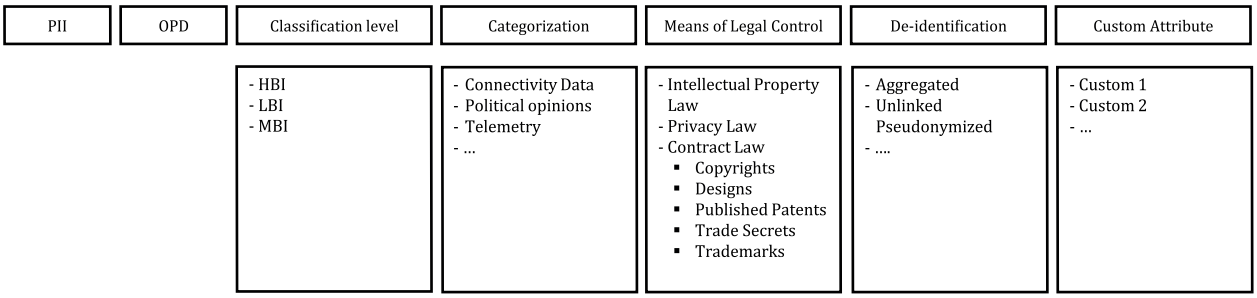


Figure 11 — Examples of attributes from data facets

Any of these attributes can be combined to provide more details about data objects in a data use statement. The only requirement is that the elements from the same data facet are not used more than once as an attribute of a given data object.

10.3.4 Use of attributes to describe PII

In this document, the attributes of PII and OPD are differentiated, and have separate attributes for OPD and PII at each de-identification level.

The following example is about an AI application developed to help lower the rate of Sudden Infant Death Syndrome (SIDS). It is believed that infants sleeping on their back have a lower rate of SIDS compared to those sleeping on their stomach. The AI application determines, by examining a live streaming video of the infant, when the infant is on her stomach, and sends an alert to the parents in the vicinity. To determine if the infant is on her back or stomach, thousands of infant photos are used to train the ML model. Such photos contain PII about the minors and are typically subject to laws and regulations.

When it comes to training an AI model for the SIDS application, to protect the minor's PII in the training data pool, de-identification needs to be performed. In the following example, the PII has been aggregated, hence the choice of "aggregated PII" qualifier attribute.

When it comes to running the SIDS application, the video streaming of the infant crib is fed to the already trained model in the application, to get a determination on whether the infant is sleeping on her back or stomach. The infant data are not stored by the service and is not used for any other purpose, and therefore does not need to be de-identified (2nd example in Figure 12).

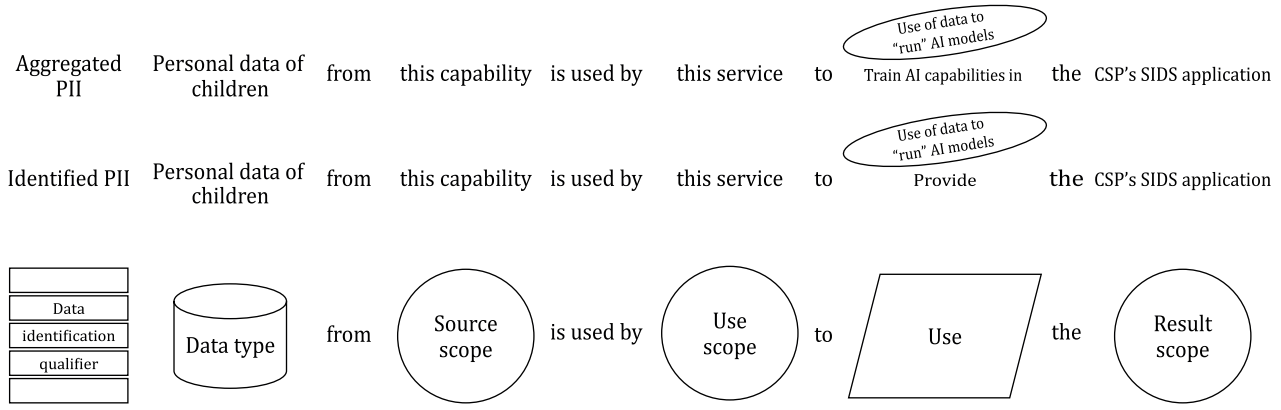


Figure 12 — Example of use statement for PII

10.3.5 Use of attributes to tag IP data

In the following example, "identified OPD" data about the innovative shape of a new generation of high-efficiency jet engine turbine blade that is collected from the aircraft engine supplier's turbine design application is used as input to the airplane manufacturer's AI-based engine predicative maintenance service. There are only two parties involves (aircraft engine supplier and airplane manufacturer), with complementary business models and mutual trust, so no data de-identification needs to take place. The data from the 3D turbine blade is used as input to a trained model for predictive maintenance service of engines, hence the use of a sub-verb of "provide" (see Figure 13).

In this example, the composite [OPD] attribute has been decomposed into its core elements: [Organization] is taken from the legal entity facet, and [Contract law] is taken from the legal means facet. [Identified] attribute is from de-identification facet.

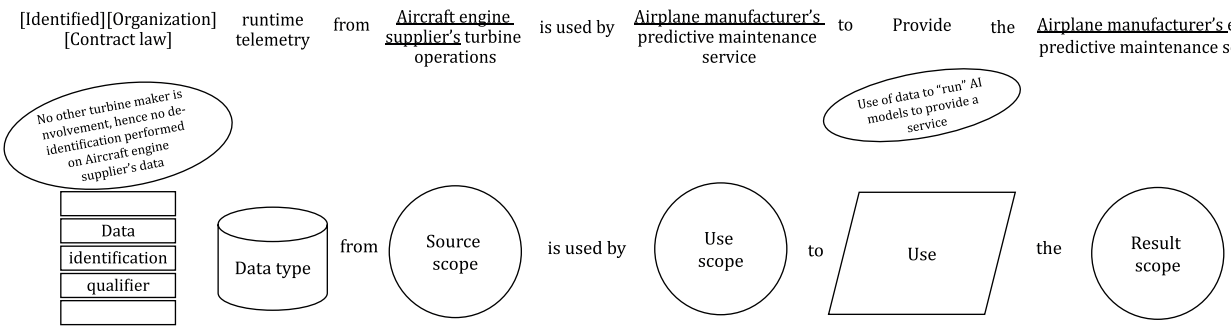


Figure 13 — Example of use statement for OPD composite attributes

10.3.6 Use of attributes to tag IP data from shared pools, while respecting partner IP

The following is an example of how shared data pools, in a multi-party partnership, can be used to train AI models, while protecting the IP of one partner from another.

In the following example, there are three parties involved, two of which are competitors (AES A and AES B) and both are aircraft engine suppliers (AES) to airplane manufacturer (AM). The aircraft engine suppliers have proprietary engine design IP, which also affects the design of the shape of the nacelle (the housing, separate from the fuselage, that holds the jet engine). Each engine suppliers needs to protect its engine design data, which includes the outer shape of the nacelle. However, the shape of the nacelle has aerodynamic ramifications for the wing design, since the shape of the nacelle hanging from the pylon under the wing directly affect the airflow around the wing and fuselage.

The shape of the nacelle for a new jet engine could provide intelligence about the design of the engine it contains, hence it is considered intellectual property in need of protection.

In order for the airplane manufacturer to use AI to design the optimum shape of its next generation wing to reduce air drag and fuel consumption, it needs to integrate the data from the shape of the nacelle of each engine, from competing suppliers. The training data for the AI algorithm for wing design will need to include the shape of the nacelle from both competitors. Therefore, de-identification in the shared pool of training data is needed in order to prevent one competitor from accessing the exact details of the design of the nacelle from the other competitor (see Figure 14).

In the following two examples, there is a four-party data sharing arrangement (see ISO/IEC 23751), two of which are competitors (engine suppliers, AES A and AES B). In this case, the engine diagnostics data from AES A or AES B engines installed on new airplane model *X* of the airplane manufacturer (AM) is used to train the ML model on top of which the AM's predictive maintenance service is built. Given that diagnostics data can have aspects to it that could reveal the proprietary engine design information that AES A and AES B want to keep from each other, before the data are added to the shared pool for ML training purposes, it is possible that AES A or AES B demand that certain engine diagnostics data be de-identified to blur out the proprietary aspects of it (see Figure 15). The de-identifications techniques used are the same as described in ISO/IEC 20889, using the de-identification qualifiers in this document.

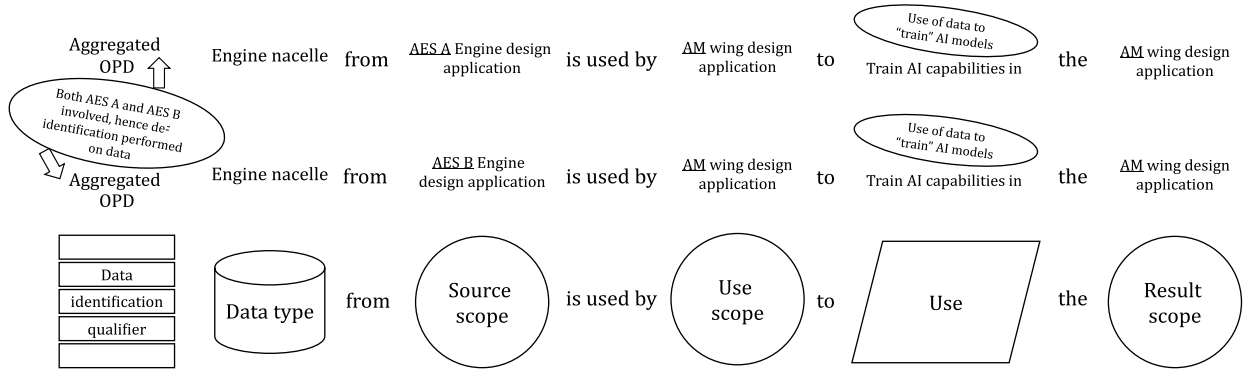


Figure 14 — Example of use statement for OPD: Training of AI model for AM wing design application. Two competitive engine nacelle designs are de-identified

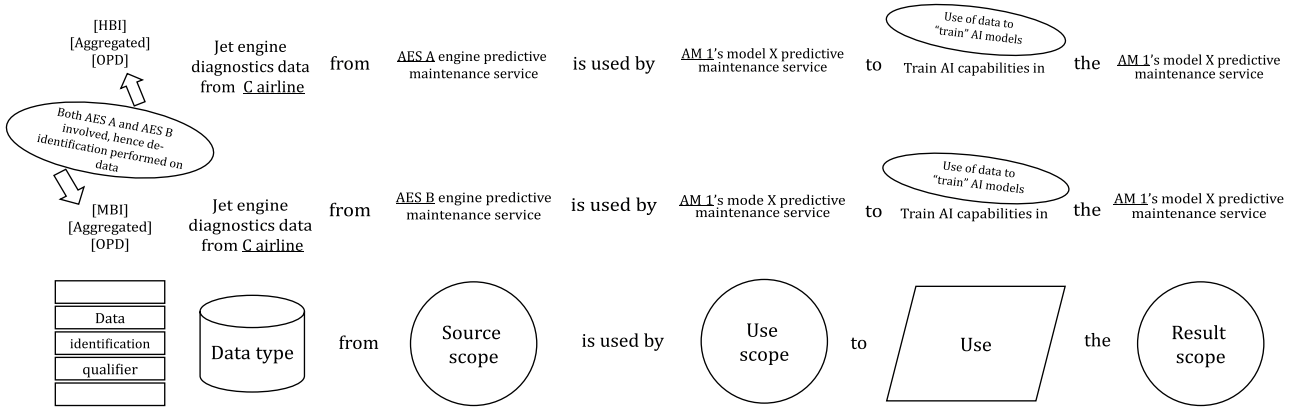


Figure 15 — Example of use statement for OPD: Training of AI model for AM 1's predictive maintenance service, using two sets of competitive diagnostics data obtained from airline C

There could also be a more complex scenario (not decoded here in terms of data use statement structure), involving a shared pool of engine diagnostics data from two engine competitors (e.g. AES A and AES B) and two airline competitors (e.g. airline C and airline D). The shared data pool could be used to train an airline predictive maintenance service by a fifth partner. Again, AES A and AES B might need to keep detailed engine diagnostics data from each other, demanding a certain degree of de-identification of certain types of engine data, as in the above examples (see Figure 16). But this time, airline C and airline D, being competitors, may not want to share detailed flight routing and dispatch data that could be decipherable from the shared pools of engine data containing flight location and time stamps. It is possible that they in turn demand a certain degree of de-identification to mask the details of the routing data that could otherwise reveal their proprietary routing strategies that improve efficiency and on-time operation.

The situation could become further complicated considering that airlines operate fleets from competing manufacturers (AM 1 and AM 2, for example), and sharing engine diagnostics data from the same or different engine on a jet built by another manufacturer could potentially reveal proprietary design information.

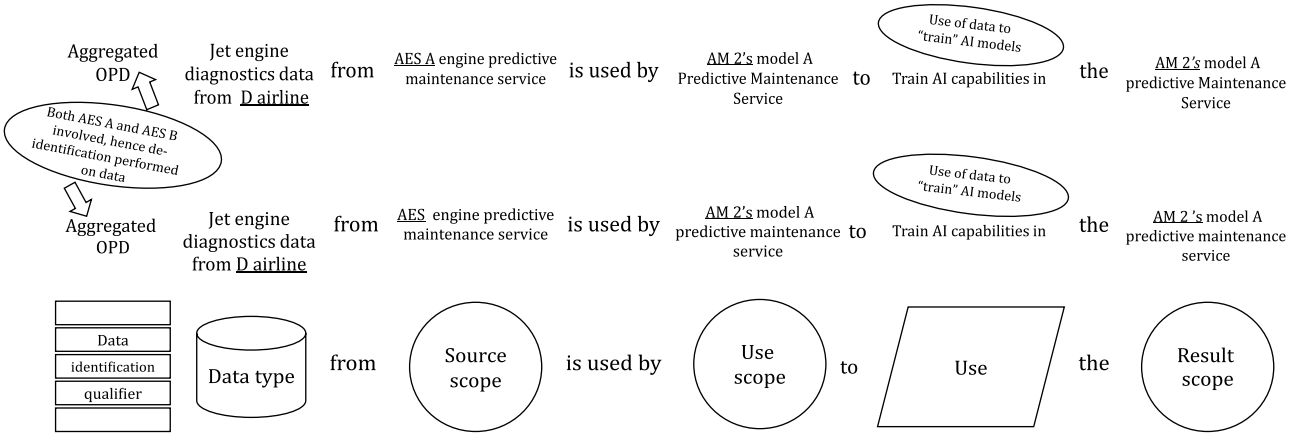


Figure 16 — Example of use statement for OPD: Training of AI model for AM 2's predictive maintenance service, using two sets of competitive diagnostics data obtained from airline D

11 Data lineage and data provenance

11.1 General

Recording data lineage and subsequently establishing data provenance help transparency and explainability of data intensive applications such as ML development.

The data taxonomy and use statements described in this document can contribute to establishing providence of data in the data driven applications where there is often the need to be accountable for use of data in privacy and IP concerns.

Data provenance provides a historical record of the data and its origins. The provenance of data which is generated by complex transformations such as workflows is of considerable value to customers, AI developers and regulators. From it, one can ascertain the quality of the data based on its ancestral data and derivations, track back sources of errors, allow automated re-enactment of derivations to update data, and provide attribution of data sources. Provenance is also essential to the business domain where it can be used to drill down to the source of data in a data warehouse, track the creation of intellectual property and provide an audit trail for regulatory purposes.

The use of data provenance is proposed in distributed systems to trace records through a dataflow, replay the dataflow on a subset of its original inputs and debug data flows. To do so, one needs to keep track of the set of inputs to each operator, which were used to derive each of its outputs.

This document offers solutions to help track lineage. Once lineage is tracked, it can be used to establish provenance. This document does not address provenance, but providing methods to track data lineage is believed to be key in establishing provenance of data.

Establishing lineage of data also helps with transparency and the explainability of the algorithms and processing that had been performed on data.

11.2 Tracing data lineage

As data objects are tagged with proper attributes from the relevant facets, lineage of data objects can be established and further maintained as they are processed through various stages of the data processing pipeline. The data object attributes are maintained and updated to reflect the effect of the processing taken place in the data lifecycle.

Use of these attributes in a data tagging system can help with establishing provenance of data, as data are processed, and new data are created from the old. Any intellectual property created as the result of processing of data can then be tagged accordingly, and the applicable contract can then be applied to determine who has the rights to what data, and the associated IP.

The hierarchical attributes as described in subclause 10.3.3 and shown in Figure 11 are based on the multi-faceted taxonomy described in Table 1. These hierarchical attributes may be used as the basis for tagging data objects as they are processed and transformed. Such attributes assigned to data objects can be used for tracking lineage and establishing provenance.

Given the various facets of data, and their associated hierarchy of properties, an attribute model can be established for tagging data objects. Each attribute is associated with a property element in a given data facet. For example [Aggregated] would be an attribute from the de-identification facet, and [OPD] would be an attribute from the legal control facet. Given the facets of data are orthogonal, only one attribute from each facet can be picked for a given data object.

Each facet of data provides an attribute group that describes that orthogonal facet, so an attribute may be picked from each facet and be applied to a given data object. Therefore a data object can have multiple attributes, one from each orthogonal facet supplying a separate set of attributes and a selection of attributes from each facet can be applied to a data object simultaneously with other selections (e.g. [aggregated][OPD] or [Unlinked pseudonymized][telemetry] are valid combinations of labels since the labels applied are taken from orthogonal facets).

When it comes to complex data analytics pipelines such as the ones used for machine learning algorithms, we need a comprehensive attribute framework. A multi-dimensional attribute framework in the data processing pipeline can be used as data are prepared for use in analytics algorithms. Such attributes can be used to help track where the data has come from, what has happened to it, what new data has been generated off of it, and which stakeholders have control over it.

Based on the above taxonomy relationship, one could envision an expanded taxonomy in a data platform that is used to attribute data objects as they are collected, processed and used. Such attribution/tagging system can be implemented using, for example, .NET attributes or equivalent Java programming language solutions in a given data platform). The advantage of having data taxonomy to include organizational data attributes is that the same, existing privacy tracking system can be used to track customer and provider's intellectual property assets.

These mechanisms are needed because all useful and interesting scenarios involving data involve shared and co-controlled data. i.e. most interesting and useful data are shared, and under control of more than one stakeholder. The above concepts, and the associated attributes are required in order to describe how data are shared, and how the control of data by multiple stakeholders is managed.

This will also help establish the provenance of data given the lineage of data can be tracked using the attributes assigned throughout the data processing platform.

PII/OPD attributes of data along with their degree of de-identification can be used in the data use statement structure.

Use of attribution of data is a way to provide evidence to help establish provenance of data, but attribution is not an end-all, and complete solution.

12 Use of taxonomy and data use statement in other computing environments

The data taxonomy, facet-based data attributes and data use statement structure described in this document are applicable to data in every distributed computing environment including cloud computing.

13 Use of data taxonomy and use statements in Artificial Intelligence scenarios

An important application of data where this document is helpful is ML applications deployed in distributed computing environments, of which cloud computing is a common example (ML applications are most likely deployed in distributed computing environments).

The data taxonomy and use statement structure in this document can assist machine learning applications with their need for transparency and explainability of data use. The data taxonomy in this document introduces verbs and attributes (see Figure 11 for examples) that can describe the data processing needed for training of ML models.

For ML scenarios, the purpose of the data used is different between "train" sub-verbs and "provide" or "improve" sub-verbs. In the latter case, the data in question is the data presented to an already trained model. The model is "run" with that data to make a prediction. This is the data that is "fed" into the model as input by the AI application that invokes the already trained model. This is very different from the training phase where large amounts of training data are used to train the model.

The taxonomy and use statement structure in this document can be used to describe data use for training ML models, as well as using them to build ML-aware applications, while helping to establish and maintain data provenance.

Data acquisition and processing pipelines in ML architectures require that the training raw data befirst collected or acquired, then pre-processed and prepared as needed based on the type of model intended for use, followed by the last stage where the prepared data are used to train the chosen model. Such stages of data processing before the data are used in training may need to be described in data use statements in order to provide transparency regarding the source of data, what data are acquired

and what has been done to it since acquisition. The data use statements would need to use the proper attributes listed in this document to establish tracing of the data and its provenance.

Once the data are ready for use in training, the training process itself may be described in one or more data use statements, with proper attributes and tags used. Such data use statements would help provide transparency and explainability of the entire AI development process.

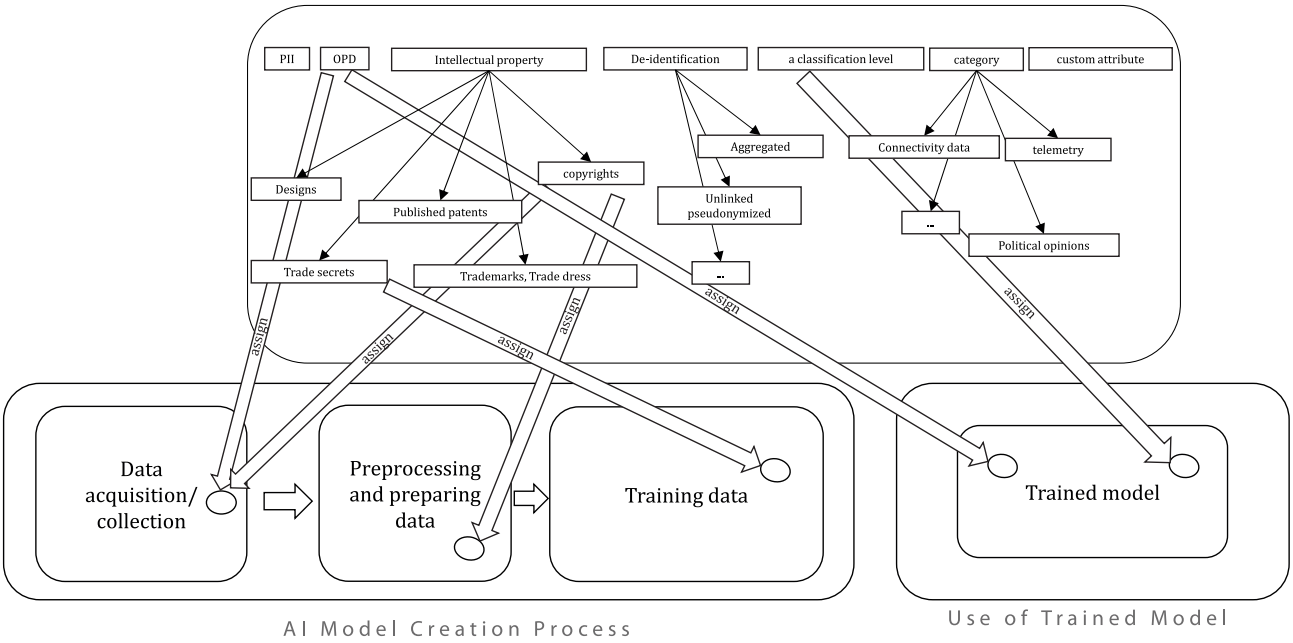
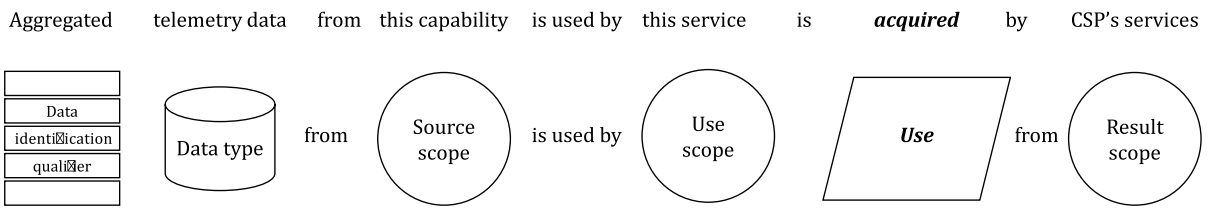


Figure 17 — Example of assignment of attributes in data processing pipeline

The attributes defined in the data model for orthogonal data facets can be used to tag various data objects as they are processed and used as training data in order to build ML models. An example of attribute assignment to the data that flows through the machine learning pipeline is shown in Figure 17. This tagging of facets of data can then be used to explain the ML development process, providing for trust and transparency.

For examples of how the data use statement structure had been expanded to cover use of data to train the ML models, see the jet aviation examples in 10.3.4 and 10.3.5.

Figure 18 illustrates a few examples of the use of the new verb “acquire”. The “acquire” verb supports statements that show where the data was sourced and what attributes were assigned to it at the time of acquisition, e.g. whether the data was PII, OPD, etc. and to what degree the data was de-identified.



The cloud services covered in this agreement ***acquire*** end user identifiable information.

The mapping service uses user location data ***acquired*** from partner social media stream.

This service shares payment instrument data with third-party partners and data processors ***acquired*** from this cloud service.

This service ***acquires*** *identified customer data from third-party partners and data processors* to provide the cloud service.

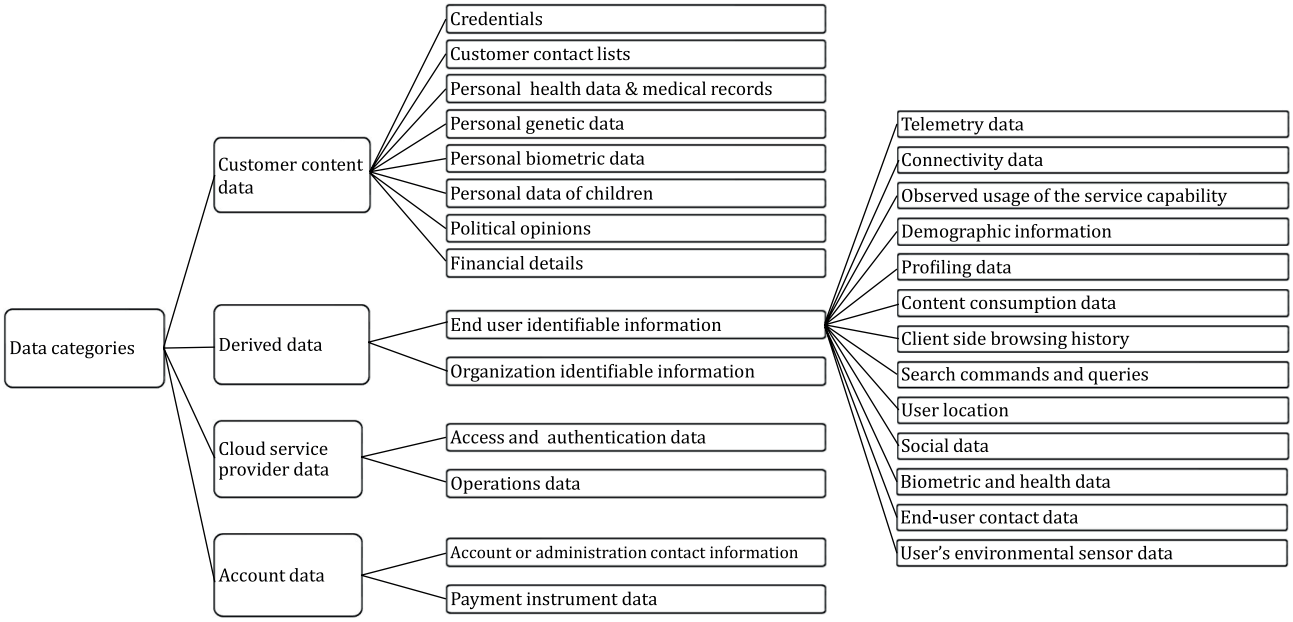
This service ***acquires*** *Anonymized PII derived data from third-party partners and data processors* to provide the cloud service.

Figure 18 — Example of use statements for data acquisition

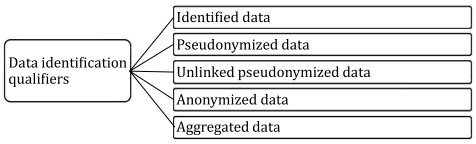
Annex A
(informative)

Diagrams of data categories and data identification qualifiers

A.1 Data categories



A.2 Data identification qualifiers



Bibliography

[1] ISO/IEC 19086-1, *Information technology — Cloud computing — Service level agreement (SLA) framework — Part 1: Overview and concepts*

[2] ISO/IEC 19944-2:—[5], *Cloud computing and distributed platforms — Cloud services and devices: data flow, data categories and data use — Part 2: Use and extension guidance*

[3] ISO/IEC 20889:2018, *Privacy enhancing data de-identification terminology and classification of techniques*

[4] ISO 21931-2:2019, *Sustainability in buildings and civil engineering works — Framework for methods of assessment of the environmental, social and economic performance of construction works as a basis for sustainability assessment — Part 2: Civil engineering works*

[5] ISO/IEC 22624:2020, *Information technology — Cloud computing — Taxonomy based data handling for cloud services*

[6] ISO/IEC 22989:—[6], *Information Technology — Artificial Intelligence — Artificial Intelligence Concepts and Terminology*

[7] ISO/IEC 23053:—[7], *Information Technology — Artificial Intelligence — Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*

[8] ISO/IEC CD 23751:—[8], *Information technology — Cloud computing and distributed platforms — Data sharing agreement (DSA) framework*

[9] ISO/IEC 27033-3:2010, *Information technology — Security techniques — Network security — Part 3: Reference networking scenarios — Threats, design techniques and control issues*

[10] ISO/IEC 27040, *Information technology — Security techniques — Storage security*

[11] ISO 27701:2019, *Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines*

[12] ISO/IEC 29100:2011, *Information technology — Security techniques — Privacy framework*

[13] ISO/IEC 29110-4-3:2018, *Systems and software engineering — Lifecycle profiles for very small entities (VSEs) — Part 4-3: Service delivery — Profile specification*

[14] ISO/IEC 38505-1, *Information technology — Governance of IT — Governance of data — Part 1: Application of ISO/IEC 38500 to the governance of data*

[15] ICO, (Information Commissioner's Office). Deleting personal data: Data Protection Act, Version 1.1, 2014. Available from: <https://ico.org.uk/for_organisations/guidance_index/~/media/documents/library/Data_Protection/Practical_application/deleting_personal_data.pdf>

[16] Reardon J., Basin D., Capkun S., SOK: Secure Data Deletion, IEEE Symposium on Security and Privacy, 2013. Available from: <http://www.ieee-security.org/TC/SP2013/papers/4977a301.pdf>

[17] General Data Protection Regulation (GDPR). 2016. Available from: <https://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf>

5) Under preparation. Stage at the time of publication: ISO/IEC PWI 19944-2:2020.

6) Under preparation. Stage at the time of publication: ISO/IEC CD 22989:2020.

7) Under preparation. Stage at the time of publication: ISO/IEC CD 23053:2020.

8) Under preparation. Stage at the time of publication: ISO/IEC CD 23751:2020.

[18] UK Investigatory Powers Act. 2016, . Available from: <http://www.legislation.gov.uk/ukpga/2016/25/contents/enacted/data.htm>

[19] Amended Act on the Protection of Personal Information in Japan. 2006. Available from: <http://law.e-gov.go.jp/htmldata/H15/H15HO057.html>

[20] The Personal Information Protection and Electronic Documents Act (PIPEDA), 2019. Available from: <https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/>

ICS 35.210
Price based on 65 pages