INTERNATIONAL STANDARD

ISO/IEC 22624

First edition
2020-02

Information technology — Cloud computing — Taxonomy based data handling for cloud services

Reference number
ISO/IEC 22624:2020(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
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 http://patents.iec.ch).

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*.

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.

Introduction

Many of the policies and practices in place for handling data in the cloud computing ecosystem need to be described based on the category of the data they address. For instance, personally identifiable information (PII) impose specific data management requirements not only in terms of security but also with regard to mechanisms that allow cloud service users to whom such data relate to exercise control on the usage and transfer of such data. Organisational data such as cloud service usage information and telemetry data from cloud services, which can be used for operational purposes such as improvement of service quality, may have to fulfil specific quality requirements to be useful for a given application.

Customer content data can be related to intellectual property rights and possibly needs appropriate protection by the cloud service provider (CSP). Certain data can be transferred from one jurisdiction to another. Depending on their data category, different instruments (multi-national laws, corporate binding rules, bilateral agreements) are applicable to enable such transfers.

When such policies and practices are to be described, it is helpful to do so in a structured and consistent way so that they can be better expressed, evaluated, analysed, and compared by the stakeholders in the cloud computing ecosystem. ISO/IEC 19944 provides a comprehensive taxonomy defining a fine-grained system of data categories that can be applied to various domains of policies for the handling of data in a cloud computing ecosystem such as cross border data transfer, data geolocation, data usage, data access and data portability, data management including data quality management and data security, or data governance, and provides guidelines on how to describe data handling policies and practices within codes of conduct (CoC).

This document describes such a structured and common approach to express any desired data handling policies and practices. It is important to emphasize that the policies and practices themselves are out of the scope of this document. This document describes a common structure and approach to express any desired data handling policies and practices. It is important to emphasize that the policies and practices are out of the scope of this document. A set of examples from data handling domains are provided in the document as guidance to understand how to use ISO/IEC 19944 regarding application of policies and analysis of policy requirements to such domains.

Information technology — Cloud computing — Taxonomy based data handling for cloud services

1 Scope

This document:

— describes a framework for the structured expression of data-related policies and practices in the cloud computing environment, based on the data taxonomy in ISO/IEC 19944;

— provides guidelines on application of the taxonomy for handling of data based on data subcategory and classification;

— covers expression of data-related policies and practices including, but not limited to data geolocation, cross border flow of data, data access and data portability, data use, data management, and data governance;

— describes how the framework can be used in codes of conduct for practices regarding data at rest and in transit, including cross border data transfer, as well as remote access to data;

— provides use cases for data handling challenges, i.e. control, access and location of data according to ISO/IEC 19944 data categories.

This document is applicable primarily to cloud service providers, cloud service customers (CSCs) and cloud service users, but also to any person or organization involved in legal, policy, technical or other implications of taxonomy-based data management in 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, *Information technology — Cloud computing — Overview and vocabulary*

ISO/IEC 19944, *Information technology — Cloud computing — Cloud services and devices: Data flow, data categories and data use*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 17788, ISO/IEC 19944 and the following apply:

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

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

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

3.1
codes of conduct
CoC
agreed set of behaviours between organisations to enhance customer and/or partner outcomes and experiences

3.2
confidentiality
property that information is not made available or disclosed to unauthorized individuals, entities, or processes

[SOURCE: ISO 24534-5:2011, 3.11]

3.3
integrity
property of being designed such that any modification of the electronically stored information, without proper authorization, is not possible

3.4
availability
property of being accessible and useable upon demand by an authorized entity

[SOURCE: ISO 22600-1:2014, 3.7]

3.5
data access
process by which a system can read published data on another system

Note 1 to entry: This data access happens over a network connection and the data typically does not persist after the connection is terminated.

3.6
data transfer
copying or moving data from one system to another

3.7
data geolocation
geographic location of a data object at rest

4 Symbols and abbreviated terms

APEC	Asia-Pacific Economic Cooperation
BCR	Binding Corporate Rules
CBPR	Cross-border Privacy Rules
CPTPP	Comprehensive and Progressive Agreement for Trans-Pacific Partnership
CoC	Codes of Conduct
CSC	Cloud Service Customer
CSN	Cloud Service partner
CSP	Cloud Service Provider
DRM	Digital Rights Management
EU	European Union
EUII	End User Identifiable Information
GDPR	General Data Protection Regulation
GPS	Global Positioning System

HBI	High Business Impact
IaaS	Infrastructure as a Service
IPR	Intellectual Property Rights
IRM	Information Rights Management
IT	Information Technology
LBI	Low Business Impact
MBI	Medium Business Impact
NASPO	National Associations of State Procurement Officials
OII	Organization Identifiable Information
PII	Personally Identifiable Information

5 Overview: The need for a structured expression of data policies and practices based on a common data taxonomy

Data policies and practices, at corporate or government level, need to be crisply expressed with the desired degree of precision and clarity. The need for varying degree of precision, along with the need to compare and analyse various policies in an efficient manner, calls for a common and structured approach to the expression of these policies and practices, an approach that is based on a common data taxonomy.

ISO/IEC 19944 provides a comprehensive set of elements which can be used to

a) assign a data category to a given data set (e.g. personally identifiable information (PII), organisational identifiable information, customer content data),

b) provide classes of actions applied to such data (e.g. use to provide a service, to optimize it, to provide marketing information),

c) include scopes explaining on what level the use of data happens (e.g. service level vs. enterprise/ organisational level vs. use by 3rd parties), and

d) define the level of de-identification (or anonymization) applied to a data set (qualifiers such as "identified", "anonymized", "aggregated").

These elements are referred to in the document as “data categories” or “data taxonomy”, “actions”, “scopes” and “qualifiers” without explicitly referencing ISO/IEC 19944. Clause 6 provides a comprehensive overview of the elements. The framework described in this document references the framework in ISO/IEC 19944.

In order to define application specific data handling policies and practices, these elements need to be applied to the application domain at hand. This includes data classifications with regards to security or risk levels that apply to data, as well as technical and organisational qualifications of data. Hence, the approach described in this document requires the considerations of data categories as described in ISO/IEC 19944 as well as orthogonal information dependent on the concrete application under consideration. Examples which are used to explain this approach therefore employ a tabular representation format emphasizing the orthogonal character of generic data categorization (rows) and application specific elements (columns). Therefore, for a person who is concerned with the development of, for example, enterprise policies for data use by a set of cloud services, all relevant cases which need to be considered are visible.

Implicitly, ISO/IEC 19944 focuses on personal data and PII, and does not explicitly cover non-personal data, or mixed sets of data that contain both PII and non-personal data. Non-personal data is defined as any data that is not personal and is not covered under PII, e.g. scientific data, sales data. Mixed data sets contain both PII and non-personal data such as human resource data that contains both organizational structures and personal employee data. It is important to recognize these different sets as different policies and regulations could apply to each. For example, the EU GDPR[9] regulates aspects of PII and the free-flow of non-personal data regulation[10] sets policies concerning the geo-location and movement of non-personal data. In line with ISO/IEC 19944, this document focuses on PII and does not delve deeper into aspects explicitly related to non-personal or mixed sets of data.

The document is structured as follows:

— Clause 6 describes the framework for the structured expression of data related policies and practices including elements of the framework building on ISO/IEC 19944. It then expands discussion on data classification (6.2.6).

— Clause 7 discusses guidance for using the framework defined in Clause 6.

— Clause 8 covers use of framework in specific areas of concern.

— Clause 9 describes the application of the framework to codes of conduct.

Examples for data handling challenges are provided throughout the document.

6 Framework for the structured expression of data related policies and practices

6.1 General

This document uses the taxonomy and data use expression structure specified in ISO/IEC 19944. Any policy or practice that conforms to this document and uses the taxonomy or data use expression shall meet the requirements of ISO/IEC 19944 as appropriate.

To handle key data management topics, Clause 6 describes a harmonized structure to express a desired policy for data management based on various data types, using data taxonomy in ISO/IEC 19944. The data management policies based on a common structure specified by this document can be expressed, compared and negotiated.

It is important to point out that this document does not define one or more data policies, rather it offers a common structure and framework for others to use in order to express their policy of choice.

Moreover, this document does not stipulate any specific format or syntax to be used to express policies and practices related to a categorization of data. Although tables are frequently employed throughout this document to illustrate the usage of the framework, the use of tabular formats is not normative or mandatory but serves for the presentation of examples only.

6.2 Framework elements

6.2.1 General

ISO/IEC 19944 defines a number of elements to express statements that describe the use of data by a CSP, namely a data categorization hierarchy, a set of qualifiers indicating the level of de-identification of data, and a hierarchy of scopes that describe at which level data are collected and processed by the CSP, a number of actions used to process data, and on which level the result of data processing is used. This clause provides an overview of the elements that are described in detail in ISO/IEC 19944.

6.2.2 Data categories

6.2.2.1 General

The data taxonomy described in ISO/IEC 19944:2017,A.1 as shown in Figure 1 below defines four main data categories, namely customer content data, derived data, CSP data, and account data

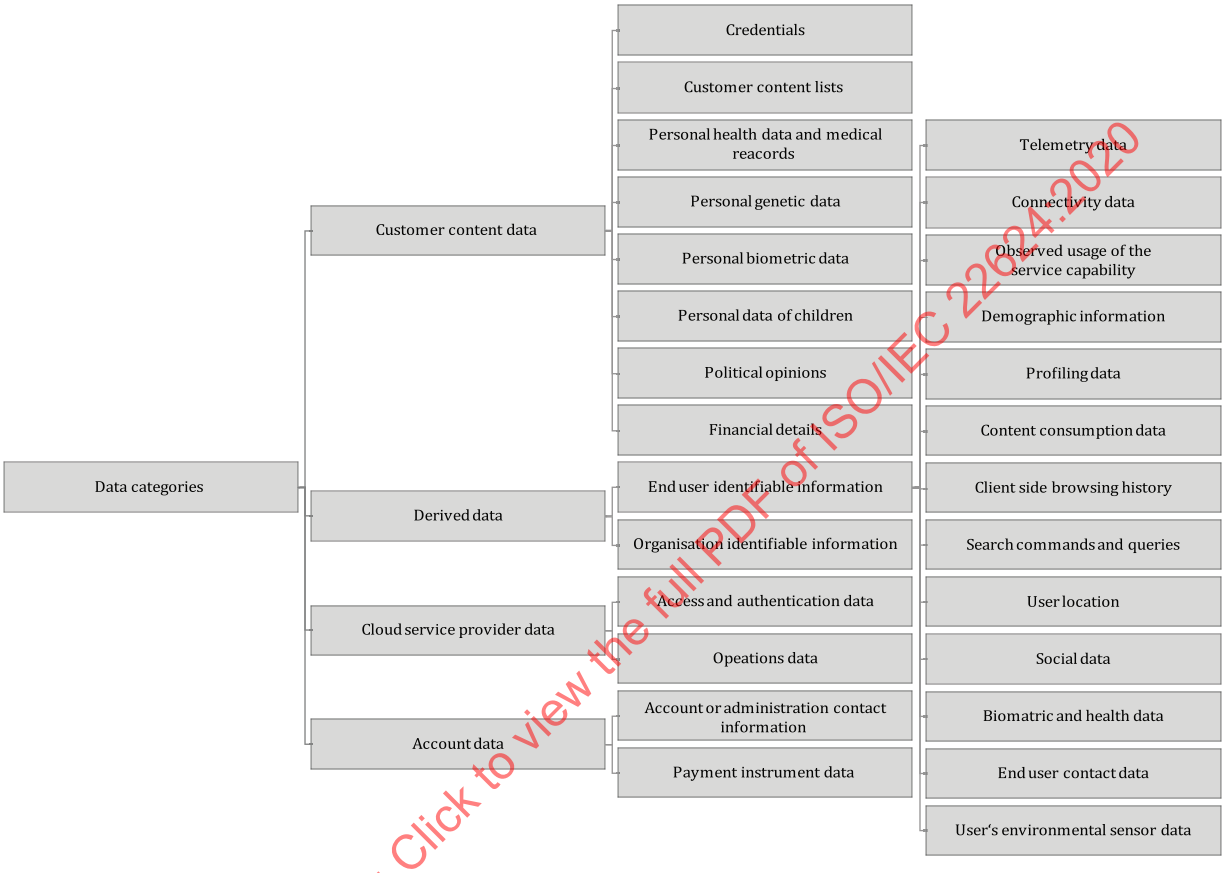

Figure 1 — Data categorization hierarchy according to ISO/IEC 19944:2017, A.1

6.2.2.2 Customer content data

Customer content data is cloud service customer (CSC) data extended to include similar data objects provided to applications executing locally on the device. This includes content directly created by customers and their users and all data 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 app or cloud service. This data category contains a large variety of sub-categories. The reader is referred to ISO/IEC 19944:2017, 8.2.2 for details.

6.2.2.3 Derived data

6.2.2.3.1 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.

6.2.2.3.2 End user identifiable information

End user identifiable information (EUII) is defined as data associated with a user that are captured or generated from the use of the service by that user; EUII is linkable to that user but is not customer content data. This data category contains a large variety of sub-categories. The reader is referred to ISO/IEC 19944:2017, 8.2.3.2 for details.

6.2.2.3.3 Organization identifiable information

Organization identifiable information (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.

6.2.2.4 CSP data

6.2.2.4.1 General

This category includes data that is exclusively under the control of the CSP. It is unique to the system and under the control of the CSP.

6.2.2.4.2 Access and authentication data

Access and authentication data is the data used within the cloud service to manage access to other categories of data or capabilities within the service.

6.2.2.4.3 Operations data

Operations data is data which is used for supporting the operation of CSPs 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.

6.2.2.5 Account data

Account data is a class of data specific to each CSC that is required to sign up for, purchase or administer the cloud service. This data includes information such as names, addresses, payment information. Account data is generally under the control of the CSP although each CSC usually has the capability to input, read and edit their own account data but not the records of other CSCs.

6.2.3 Data identification qualifiers

Data in any category can provide or contribute to information that identifies or can be linked to an individual. 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 described by the following set of qualifiers (see Figure 2):

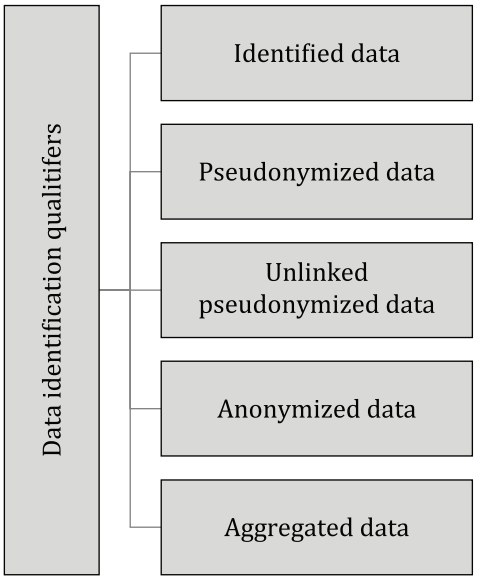

Figure 2 — Data identification qualifiers according to ISO/IEC 19944:2017, A.2

— **Identified data**. Data that can unambiguously be associated with a specific person because PII is observable in the information.

— **Pseudonymized data**. 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.

— **Unlinked pseudonymized data**. 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.

— **Anonymized data**. Data that is unlinked and which attributes are altered 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.

— **Aggregated data**. 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.

6.2.4 Data usage scopes

ISO/IEC 19944:2017, 9.4.1 defines that "scope" *provides a way to clearly describe the boundaries of collection and use of data in the devices and cloud services ecosystem*. These scopes can be used to describe the applications and services associated with data use (see Figure 3). 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 could be one of the services listed in the service agreement. These are parts of the cloud services that a CSP provides, and are a subset of the CSPs overall product and service palette.

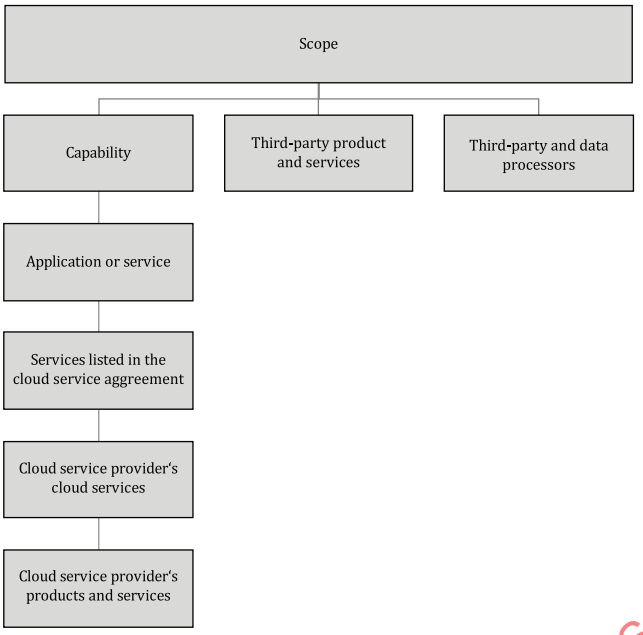

Figure 3 — Data usage scopes according to ISO/IEC 19944:2017

6.2.5 Actions

ISO/IEC 19944:2017 defines a list of actions that can be applied to data of the various categories (see Figure 4):

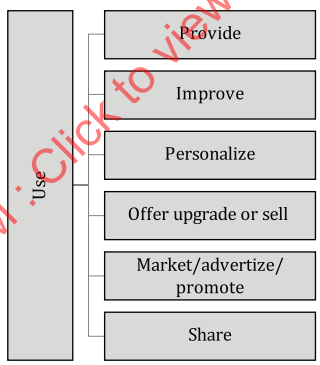

Figure 4 — Actions on data according to ISO/IEC 19944:2017

— **Provide**. The use of data to provide or protect a certain service or service capability, and to communicate with the customer about the status and availability of capabilities.

— **Improve**. To use data to improve or increase the quality of functional capabilities.

— **Personalize.** To use data to change the presentation of the capabilities or to change the selection and presentation of data or promotions accessed through these capabilities to be specific to the user, based on information about the user.

— **Offer upgrades or upsell.** To use data to offer to the customer increased capacity or resources for or new capabilities in exchange for compensation.

— **Market/advertise/promote.** To promote specified products and services to users or customers based on data.

— **Share.** To transfer data to an entity other than the CSP who originally has stored or processed those data.

6.2.6 Data classification

Data classification is the process of organizing data into specific classes to enable the secure, effective and efficient use of the data. An effective data classification process is typically an element of an enterprise's risk assessment process and risk mitigation strategies. This classification is generally based on the needs of the various stakeholders in the data being classified, which can include government regulators, external customers or suppliers, or other parties. Most organisations are involved in networks of suppliers and customers which can involve movement of data across organisational boundaries. Where this happens, measures are needed to ensure that equivalent (not inferior) protections established at source remain in force at each destination.

Data classification therefore assigns classes to data based on the level of sensitivity and the impact to the organisation should that data be used, disclosed, altered or destroyed without the proper authorization. For instance:

— The classification of data, from an information security perspective, helps to determine what the set of appropriate security controls is for safeguarding that data.

— The classification of personally identifiable data, from a data protection perspective, helps organizations to set up a privacy impact assessment to identify risks for individuals when processing their personal data.

One challenge is that data classifications can change over time (something previously considered "public" is now considered "sensitive"). Another challenge is the temptation to "over-classify the data", due to concerns related to potential derivation of PII or other sensitive data based on combining non-sensitive attributes. Evaluating the "likelihood" of disclosure and the "impact of disclosure" helps classify the data accordingly. Using appropriate encryption and tokenization techniques, further helps reduce the overall risks, however, these also will have their own considerations, such as cost and performance.

Data classification can also be used to define policies for various other data management related issues including:

— data retention periods;

— access policies;

— performance requirements regarding access and transmission speed;

— data compliance and risk management;

— data predefined storage;

— simplification of data encryption;

— data indexing;

— data protection.

Good practice in data classification is based on a risk assessment that considers critical elements for the service quality and continuity. The risk assessment is used to develop implementation policies that support those elements to ensure the secure, effective and efficient use of the data.

Each government or enterprise will have their own approach to data classification, and to the policies that apply to data that is marked in this way. Organisations wishing to establish a data classification system need to do so carefully and following best practice. For information security, standards such as BS 10010:2017[8] on information classification, marking and handling can be helpful in doing this correctly, if applied appropriately.

Exactly which data is classified at which level is a matter that should be decided by policy. Which policy applies is generally determined by a combination of "Value", "Risk", and "Stakeholders".

— Value refers to the actual value of the data to the organisation, such as a new manufacturing process, or financial information, and could also consider the potential cost of recreating the data.

— Risk refers to consequences of data loss, leakage, or corruption, such as loss of business, loss of reputation, or even criminal or civil proceedings.

— Stakeholders refers to those who care about the data remaining secure, such as other departments of the organisation, valued clients, or government regulators.

For example, an enterprise might decide that all PII should be classified as medium business impact (MBI) or higher, and "sensitive" PII (such as medical information) should always be classified as high business impact (HBI) or higher so as to avoid risk under privacy regulations. However, there will be other information classified at these levels which contains no PII at all, such as valuable financial or design information. It is normal to also have a "default" policy for handling of data that has no specific data classification assigned to it.

From a data management perspective, the type of classification is orthogonal to the other criteria discussed in this document, including the process of data categorization (see 6.2.2). Generally, when data is covered by both a policy for data management and a policy for data classification, the *most restrictive* policy will be applied. So, from a practical perspective, data management systems need to consider both the **content-based policy (as described in this document)** and the **classification-based policy** used in the organisation, and apply controls taking into account both aspects using appropriate technical mechanisms.

The selection of an appropriate scheme for data classification depends on the application context. Thus, this document imposes no restriction on the data classification scheme used in connection to the data taxonomy defined in ISO/IEC 19944.

6.2.7 Further elements specific to the application domain

The framework elements described in 6.2.2 to 6.2.4 have to be complemented by elements related to the selected application domain. For instance, policies for data security can be defined depending on overall security objectives such as confidentiality, integrity and availability of specific categories of data, taking into account application dependent data classifications as explained in 6.2.6. In some cases, further application domain specific aspects need to be taken into account. For instance, policies related to data governance can be expressed with regard to a data governance lifecycle.

This document does not impose any restriction on domain specific elements.

7 Using the framework

7.1 Modes of framework usage

The data handling framework can be used

— **analytically**, to understand which existing legal or organizational policies are to be applied when handling a specific type of data, or

— to **synthetize** data handling policies for specific data types.

Usually, both modes have to be applied: The analytical mode provides conditions, terminology and restrictions for the synthesis of data handling policies that are to be implemented by technical or organizational means.

7.2 Framework element usage

7.2.1 Data categories

If data categories (6.2.2) are used to structure the definition of policies, not the whole category tree is useful for all applications. Instead, policy analysis or definition should focus on those sub-categories which are useful for a given purpose:

— Focus on selected categories only: For instance, for the expression of policies regarding CSP data, the other three top level categories (and all sub-categories of these categories) can be ignored.

— Sub-categories can be ignored if the policies under consideration can be adequately expressed in terms of general categories.

— Adding further data categories if needed: ISO/IEC 19944 allows adding new sub-categories of the four top level categories if needed.

7.2.2 Data identification qualifiers

If required, policy expression can make use of the data identification qualifiers described in 6.2.3. For instance, if PII are considered, different data handling policies can be applied with regard to the level of de-identification of the PII.

7.2.3 Scopes and actions

ISO/IEC 19944 provides an elaborated concept of the level on which a certain action uses data.

— **Source scope**: The source of the data under consideration.

— **Use scope**: The applications or services that are using the data.

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

Policies can directly employ these different notions of scopes to express requirements and restrictions on the origin of data, the application or service that uses these data, and the result of such data processing. The list of actions explained in 6.2.5 is not exclusive but can be extended to include further actions.

7.3 Policy expressions

This document does not impose any restrictions on the way policies can be expressed. In particular, using qualifiers (7.2.2) and scopes and actions (7.2.3) are optional elements which can be useful for certain types of policies.

A very simple expression of policies can comprise a singular entry in a table (such as “locally”, “regionally” or “globally” to describe policies expressing localization requirements on data storage and processing), but can also be complex descriptions of security or privacy objectives and controls stipulated for a certain type of data.

The data use statement structure defined in ISO/IEC 19944 provides a means to express policies on data usage in a standardized way. Hence, for policies that restrict the usage of data or a certain category (taking further into account scopes or data use, data processing actions, and degrees of data de-identification as explained in 6.2 and Clause 7), this structure is available as a tool to express and to communicate data handling policies.

7.4 Example

The following table is an example for an instance of a policy for data management based on the common structure proposed by this document. Notice that the rows are data categories, and the columns are common aspects of data geolocation management practices. The intersection of each row and column

therefore describes a specific choice or setting selected to express a specific data management practice for a given type of data. The cells in this example table collectively describe an overall policy for data management on geolocation perspectives.

Table 1 — Example of expression of data geolocation policy

Data categories		Data geolocation			
		Storage requirement	Controller	Processor	Jurisdiction
Customer content data	Credentials	Locally	CSC	CSP	NA
	Customer contacts list	Locally	CSC	CSP	NA
	Personal health data	Locally	CSC	CSP	State
	Personal data of children	Locally	CSC	CSP	National
	Financial details	Locally	CSC	CSP	National
Derived data	EUII - Telemetry data	Globally	CSP	CSP	Multi-jurisdictional
	EUII – Client side browsing history	Locally	CSC	CSP	Multi-jurisdictional
	EUII – Social data	Regionally	CSP, CSN	CSP	Multi-jurisdictional
CSP data	Access and authentication data	Globally	CSP	CSP	Multi-jurisdictional
	Operations data	Globally	CSP	CSP	Multi-jurisdictional
Account Data	Customer contact info	Regionally	CSP	CSP	Multi-jurisdictional
	Payment instrument data	Globally	CSP	CSP	Multi-jurisdictional

The four sub-columns under “Data geolocation” column in Table 1 provides the set of example items that needs to be considered for the data geolocation policy. Storage requirements address where specific types of data are stored. Controller and Processor shows who is the data controller or processor for specific data types. The Jurisdiction column shows the subject of applicable regulatory requirements if it exists.

It is important to point out that this document does not define one or more data policies, rather it offers a common structure and framework for others to use in order to express their policy of choice. The result would be commonly structured, harmonious policy expressions.

8 Expression of data related policies in relation to specific areas of concern

8.1 General

The data practices and polices listed in this clause are common examples for data management practices that need to be applied based on data categories. Therefore, such considerations are examples of issues to be considered for policy definition. An actual policy for data management is defined for each data category and each data management practice.

The data management policies and practices discussed below are related to most common practices. The list is by no means a finite one, and it is possible to add additional policy aspects and practices as needed.

8.2 Data geolocation

As an example, consider the policies or practices in effect for geolocation of data controlled by globally available cloud services. Although there are many reasons CSPs could invest in multiple datacentres globally, including reducing latency and business continuity, a primary reason is to support local processing of data to comply with geolocation regulation, policy or preference, e.g. the General Data Protection Regulation (GDPR) of the European Union (EU). However, the technical and engineering constraints still apply to multiple datacentres dispersed geographically. In fact the constraints are aggravated by the distributed nature of the datacentres. The cloud services will be more cost efficient and more reliable if they can be managed centrally as a group.

Therefore, CSPs need to be transparent and pragmatic about what data is processed locally and what is allowed to cross borders to allow efficient regional and global data centre and services management. This can be helped by a common and structured method to express such polices, allowing CSPs to be precise about the geolocation of data types and use, and also defining a few broad classes of data that can used in geolocation policy and contract discussions. Therefore, data geolocation policies are a good example of the type of data polices that could benefit from this document. If such policies are expressed in a common and structured way, based on a common data taxonomy, the stakeholders

in cloud computing ecosystem will be able to understand, compare, and analyse such policies more easily and efficiently. In addition, the common and structured expression, comparison and analysis of such geolocation policies and practices would lend themselves better to the automated, and computer assisted parsing and processing, adding to the ease, precision and efficiencies of handling them.

8.3 Cross border flow of data

8.3.1 Data jurisdictions considerations

8.3.1.1 General

Cloud computing data jurisdictions is a very complex subject area that often leads to confusion. There are jurisdictional considerations pertaining to processing of data, storage of data, and transmission of data. Each has specific policies that need to be applied based on data categorization (as described in ISO/IEC 19944), desired levels of data classifications, and applicable jurisdictions.

This clause provides a model for jurisdiction relationships utilizing the dimensions "data categorization" and "data classification" as foundation (see Figure 5), where then jurisdictional policies (circles) are applied to determine what actions are permissible for the data. Data categorization systems vary and ISO/IEC 19944 provides some foundational elements for data categorizations.

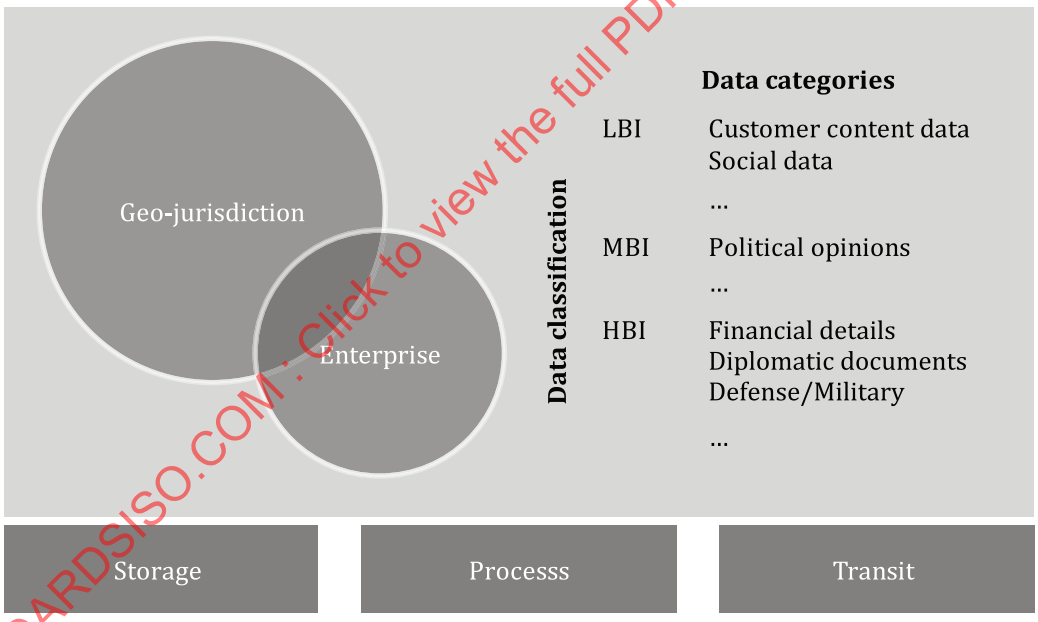

Figure 5 — Jurisdiction relationship model

The geo-jurisdictional pertains to the jurisdiction(s) that are being evaluated. The key is that geo-jurisdictional boundaries vary based on the objectives and scope.

8.3.1.2 Geo-jurisdiction

A geo-jurisdiction can be represented within four context types:

— **municipal;**

— **state/provincial;**

— **national;**

— **multi-jurisdictional.**

The latter is a more complicated type of geo-jurisdiction with several options. Examples include: EU, United Kingdom, jurisdiction defined by bi-lateral trade agreements or by multi-lateral trade agreements and others at various levels such as the National Associations of State Procurement Officials (NASPO).

A cloud service deployment can include multiple geo-jurisdictions that need to be considered.

These geo-jurisdictions can also have policies and regulations that conflict or overlap, adding to the complexity of executing within and across regions. For example, a federal country policy is that "any non-violent convictions must be retained for ten years" and at a provincial or state level the rule is these records must be forgotten after seven years. The reality is that such conflicting policies are only now coming to better light as the cloud-enabled technologies and sharing of processes and data has highlighted that such policies have been in non-compliance for many years. The cloud computing initiatives have provided the opportunity to highlight and strategically resolve these issues. In many cases, policies will require adjusting (perhaps temporary/ tactically), to be compliant.

Policies based on the geo-jurisdictions vary.

Table 2 — Examples of geo-jurisdictional policies

National	— Data policies and regulations at a Federal/National Level — National government data regulations — Defence/Military regulations — Federal/National Financial/Banking/Investment Regulations — Federal/National Telecommunication Regulations — National Privacy — National Health Care records — National PII regulations and policies — National Policing
State/Provincial	— Data policies and regulations at a State/Provincial Level — State/Provincial Government data regulations — Provincial/State Heath Care records — Provincial/State PII regulations and policies — Provincial/State Police
Municipal	— Municipal policies related to privacy of data
Multi-National	— Free trade agreements such as Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) — GDPR — North American Free Trade Agreement

8.3.1.3 Enterprise/organization

Enterprise regulations and policies can be applicable with, but not limited to:

— data policies for the enterprise (company or agency);

— employee records;

— customer records;

— contracts and other legal documents;

— financial records;

— intellectual property.

Organizations can use this model and specifically identify the pertinent policies and any specific interpretations, such that the cloud deployment can be executed in a transparent and compliant manner.

The three major foundational capabilities that need to be evaluated in conjunction with geographic jurisdictions are: storage, process and transit.

Most data jurisdiction concerns are related to privacy and protection of data sovereignty. Data includes the information values and the metadata definitions that describe the data. PII is generally the driver that impacts data jurisdiction policies and legislation. In some cases, data jurisdiction policies encourage data to be retained and processed outside of the jurisdiction where the business is conducted. For example, within jurisdictions at risk of natural disasters or under political unrest, they are better served by consuming cloud services hosted outside of their jurisdiction. The many options, objectives and constraints make this a complex subject area with no single solution.

— **Storage**. Data residency policies and regulations impact where the data may not be stored. Concerns often stem from privacy of data and access to data, because of legal disclosure orders.

— **Process**. Cloud computing processing policies and regulations impact where the data may not be processed. Concerns are often from the perspective of performance and availability constraints. Data leakage from within the process is another subject of concern.

— **Transit**. Cloud computing transit (data movement) policies and regulations impact how the data is permitted to move between network nodes of the provider and between the customer and provider. Concerns are often about potential data interception and/ or performance and availability constraints. Data leakage during transit is another subject of concern.

8.3.2 Cross border data transfer

The ability to legally transfer data from one country, region or economic or political zone to another is of critical importance in today's data driven economy. Various regulatory policies are in effect in sending or receiving data from and to CSPs and CSCs, and jurisdictions of countries they reside in influence the requirements for the movement of data. CSPs face the following challenges:

— to make sure that customer data such as Intellectual Property Rights (IPR) or company internal data are sufficiently protected when moved from one country, region or economic or political zone to another;

— to make sure that PII are sufficiently protected and the rights of data subjects remain granted in the receiving countries or zones.

The taxonomy based approach of ISO/IEC 19944 helps CSPs

— to analyse the legal or contractual requirements for cross-border data transfer,

— to express the modes of cross-border data transfer to their customers.

In principle, the data transfer from one country, region or economic or political zones to another can be based on one of the following types of policy or regulation:

— **Contracts** between the transferor and the transferee that include appropriate privacy and security controls.

— **Consent** of the data subject in case of the transfer of PII.

— **Binding Corporate Rules** (BCR) for the data transfer between affiliates (or groups of affiliates) of a corporation. For instance, the European GDPR allows companies to seek approval of such rules regarding the cross-border transfer of PII by responsible data protection authorities.

— **Cross-border rules**: Codes of conducts approved by groups of countries or zones between which data transfer occurs. An example is the Cross-border Privacy Rules (CBPR) of the Asia-Pacific Economic Cooperation (APEC).

— **Codes of conducts (CoC), certifications, seals and standards.** These mechanisms can be used to provide implementation options, technical clarifications and guidelines for sector or domain specific applications of the BCR or cross-border rules.

— **"Whitelist" arrangements based on adequacy**. Those arrangements for the liability of the transfer of PII are based on a decision of the adequacy of date protection rules of the receiving country or zone. An example is the "Safe-harbour" framework of the European Union and the "Privacy Shield" arrangement between the EU and the US.

— **Exceptions and derogations** from cross-border data transfer restrictions. For instance, the transfer of PII might be allowed if

  — the transfer is necessary for performance of the contract between CSP and data subject,

  — the transfer is necessary to protect vital interests of the data subject,

  — the transfer is necessary for a legitimate interest of the CSP and is not outweighed by fundamental rights of the data subject.

Given this collection of mechanisms to legally enable the cross-border data transfer, CSPs can use the data categories to analyse their availability and applicability as exemplified in the following Table 3 for given transferors and transferees.

EXAMPLE

Analysis matrix on the availability/applicability of regulations or policies for cross-border data transfer.

Table 3 — Hypothetical analysis of available/applicable regulations or policies for cross-border data transfer

Data categories		Regulatory types						
		Contracts	Corporate rules	Cross-border rules	Codes of conduct	Whitelis	Derogations/ exceptions	Consent
Customer content data		CSPs/CSCs have to analyse if contracts are sufficient to legitimate cross-border data transfer	CSPs/CSCs have to analyse which mechanisms are available to legitimate cross border data transfer				CSPs/CSCs have to analyse if jurisdictions in receiving counties allow for derogations or exceptions	CSPs/CSCs should seek consent of data subjects if non of the other instruments are applicable
Derived data	EUII							
	OII							
CSP data	Access and authentication data							
	Operations data							
Account Data	Customer contact info							
	Payment instrument data							

The following examples illustrate how to employ the format for data use statements defined in ISO/IEC 19944 to express cross-border data transfer policies in a structured way:

EXAMPLE

Company services outside of the United States use end-user identifiable information from this capability to improve the services listed in the service agreement.

This data use statement indicates that end-user identifiable information in general are used to improve the services of a given CSP and that the processing needed to archive this improvement is done outside of the US.

The next example indicates that certain types of data are processed differently after being transferred to a different geo-location, where different regulations may apply:

> EXAMPLE
>
> 3rd party data processors outside the European Union use customer content data except personal data of children for marketing from the CSP products and services.

This data use statement states that customer content data are used for marketing purposes by data processors outside the European Union (e.g. on the basis of an informed consent of data subjects), with the exception of PII related to children who are not in a legal position to state consent.

8.4 Data portability and data access

8.4.1 General

This clause provides a framework for expressing which data is in scope for data portability requirements, based on the data categories in the common data taxonomy, and adjusted for consideration of the linkage of the retained data to the data subject (for cases where privacy matters) and the basis for policy-driven processing of that data. The clause also outlines a way to express requirements for providing the data directly and for dealing with data portability requirements for data that contains personal data of other data subjects, such as Social Data, again for scenarios where privacy matters.

These expressions suggest an approach to providing CSCs with capabilities for accessing their content that are similar to transferring them, addressing both data access and data portability using similar structures. Note that direct transfer to "receiving" data controllers can often be supported by a "pull model" that advocates that receiving data controllers request and process these data sets at the authorized request of a data subject or CSC.

8.4.2 Data required for data portability or data access

The table in this clause outlines which data types in the data taxonomy are in scope for data portability, based on a desired data portability or data access policy or practice. As expected, most policy examples consider customer content data largely in scope for data portability requirements, with the single exclusion of credentials data.

Derived data or its sub-types are the cases where the policies can be more nuanced. For example, two of the three data categories relating to "telemetry" can often be not required for data portability or access:

— Product and Service Heath Data, and

— Device Connectivity and Configuration Data.

These two data categories do[es] not concern the data subjects since they relate to the performance of the system itself.

Such structured and harmonized expression of data portability and access could form a basis for determining when user consent is required. They can also be used in codes of conduct where an industry group comes together to develop a consensus for a set of best practices. For codes of conducts involving data portability and access, having common structures and harmonized expressions of policy facilitates the precise capture and expression of the consensus achieved with the desired level of precision.

It is worth noting that data processors could have different policies and practices from that of data controllers. For example, CSPs acting as data processors need to adopt practices that they need to

support their CSCs acting as controller for data portability or access so each policy expression needs to differentiate between the data controller perspectives and that of data processors, when applicable.

In some policies, the degree to which data is linked to the individual can also impact the requirement for data portability or access. This is when data identification qualifiers in the data taxonomy are used for policy expressions. For example, identified data and pseudonymous data can be linked to a data subject (e.g. by him or her providing the respective identifier), and therefore, can be required for data portability or access, while any data, which is anonymous or does not concern the data subject cannot be required. However, unlinked pseudonymized data cannot likely be required for data portability or access, depending on the desired policy in effect, given this data cannot be clearly linked to a data subject so in general, when policy expression involves personal data, it can be described by levels of de-identification available in the data taxonomy.

8.4.3 Formats and portability

For help on data portability requirements, see ISO/IEC 19941:2017[1] on cloud computing interoperability and portability. For example, this document describes the fundamentals of data portability and that a clear specification of the syntactic (format) and semantic (meaning) of the data is required for successful data portability between two parties.

As described in ISO/IEC 19941:2017[1], 5.2.2.6, Table 2, the example for data syntactic portability can be supported by popular formats such as XML or JSON, accessed through an exchange of files or an API. However, supporting semantic portability requires shared understanding of definition, schema and structure through additional schema documentation and developer materials. The data portability between two disparate systems will always require specific processing to map between semantic differences and preserve the most "metadata" possible. This is specifically true for the transmission of derived data, the form and content of which is specific to the service.

As a result, supporting and expressing data portability of policies for customer content data and for derived data can be different, hence the need for a taxonomy-centric expression of such policies or practices defined in this section.

Table 3 is an example of data portability policy expression based on data categories and data de-identification qualifiers.

Table 4 — Example of data portability policy expression

Data categories			Identified data	Pseudonymized data	Unlinked pseudonymized data	Anonymized data	Anonymized data
		Data portability required	Yes	Yes	No	No	No
Customer content data*	Customer content - general		Yes	Yes	Yes	Yes	Yes
	Customer contact lists		Yes	Yes	Yes	Yes	Yes
	Personal health data & medical records		Yes	Yes	Yes	Yes	Yes
	Credentials		No	No	No	No	No
Derived data	End user identifiable informaton	Telemetry data	?	?	?	No	No	No
		Connectivity data	?	?	?	No	No	No
		Observed usage of service capability	Yes	Yes	No	No	No	No
		Demographic information	Yes	Yes	No	No	No	No
		Interest and favorite data	Yes	Yes	No	No	No	No
		Content consumption data	Yes	Yes	No	No	No	No
		Client side browsing history	Yes	Yes	No	No	No	No
		Search commands and queries	Yes	Yes	No	No	No	No
		User location	Yes	Yes	No	No	No	No
		Social data	Yes	Yes	No	No	No	No
		Biometric and health data	Yes	Yes	No	No	No	No
		End user contact data	Yes	Yes	No	No	No	No
		User's environmental sensor data	Yes	Yes	No	No	No	No
Cloud service provider data	Access and authentication data		No	No	No	No	No	No
	data		No	No	No	No	No	No

*Note: Customer content data is assumed to be identifiable data

Legend/Explanation
Yes: Data portability required
No: Data portability not required
?: Data portability perhaps not reqired

8.5 Data use

Regarding the use of data, the framework described in Clause 6 defined a number of options to identify to what purpose data processing is performed. Table 5 illustrates how the framework can be used to identify policies related to the specific use of data. In addition, scopes (6.2.4) can be used to further refine the identification of relevant policies.

Table 5 — Example of expression of data use

Data categories			Action						Scope		
			Provide	Improve	Personalize	Offer upgrades/ upsell	Market/ advertise/ promote	Share	First-party	Third-party and data processors (partners)	Third-party and data processors (other)
Customer content data	Credentials		Yes	N/A	N/A	N/A	N/A	No	Yes	Yes	Yes
	Political opinions		A	No	A	A	A	A	Yes	B	No
Derived data	**End user identifiable informaton**	Telemetry data	Yes	Yes	Yes	Yes	Yes	Yes	Yes	B	B
		Demographic information	Yes	Yes	Yes	Yes	Yes	A	Yes	B	B
		User location	Yes	Yes	Yes	Yes	Yes	A	Yes	B	B
Cloud service provider data	Access and authentication data		Yes	Yes	N/A	N/A/	N/A	No	Yes	B	No
	Operations data		Yes	Yes	Yes	Yes	Yes	A	Yes	Yes	Yes
Account data	Customer contact information		Yes	No	N/A	Yes	Yes	A	Yes	Yes	C
	Payment instrument data		Yes	Yes	N/A	N/A	N/A	No	Yes	Yes	Yes

Legend/Explanation:

Y: Data use possible in principle, the usual policies on, e.g., data security apply

No: Data use not possible

A: Dep. on privacy regulation, consent

B: Consent, contracts, geoloc.

C: Contracts

N/A: Use not applicable to data category

8.6 Data management

8.6.1 Data security

8.6.1.1 General

To take into account the strategic importance of data, an effective Information Security Management system (such as described in ISO/IEC 27001:2013[2]) can be implemented. This should be extended to include third party data feeds and data management in cloud computing services (for example, ISO/IEC 27017:2015[3]). These international standards provide guidelines for information security controls, but in some cases such controls will be insufficient and the governing body will need to rely on trust and verification.

Data security is commonly defined in terms of three main security objectives, namely confidentiality, integrity, and availability.

8.6.1.2 Data security objectives

8.6.1.2.1 Confidentiality

Confidentiality refers to the property that information is not made available or disclosed to unauthorized individuals, entities, or processes (ISO/IEC 27000:2018[4]). Policies related to confidentiality will generally determine *who* (i.e. which entities, such as people or systems) have access to the data, and *what* they are permitted to do with it. This will often be enforced using Information Rights Management (IRM) or Digital Rights Management (DRM) technologies, keeping the data encrypted and only usable when all policy requirements are satisfied. When data is authorised to be moved beyond the boundaries of the organisation, it will be necessary to ensure that equivalent (or better) protection is maintained throughout its lifetime.

It is relatively rare that confidentiality related policy directly affects things like geo-restriction of the data[1]. "Who" is much more significant than "where". Security controls such as the need to use strong encryption at rest and transport, especially when crossing organisation or even international boundaries are often used to mitigate issues related to data breaches or unauthorised disclosure.

EXAMPLE

A cloud-stored document is classified and marked as "Company Secret" because it contains high-value information about a forthcoming negotiating position. Applying a geographic restriction is not practical since it will make it difficult or impossible for the negotiating team to access the information when they travel to the meeting. However, the company use IRM to ensure that this "Company Secret" information is only accessible by an authorised person on a secure device with two-factor credentials.

8.6.1.2.2 Integrity

Data integrity refers to the accuracy and completeness of data (ISO/IEC 27000:2018[4]). Policies related to data integrity will often require that the data has a higher assurance of integrity than regular data. This means that the data is maintained in the correct state, and can only be changed by deliberate and authorised means. Therefore, it is necessary to ensure that any random errors introduced due to storage medium faults or random effects are detected and eliminated. For example, multiple copies of the data can be maintained, each with checksums or other integrity checks, so that a "bad" copy can be replaced from known good data. Errors introduced by human action can be identifiable and recoverable. This will usually include change tracking, audit trails, more recently distributed ledgers, and perhaps other forensic analysis approaches.

EXAMPLE

A cloud-stored database holds student records including test results and awarded certificates. While the data is not particularly secret (beyond normal PII protection), it is very important that it is protected from unauthorised modification. As such the records are "locked" to prevent editing, and a permanent offline copy is made periodically that can be used to identify any changes that can have been introduced.

8.6.1.2.3 Availability

The property of data of being accessible and usable on demand by an authorized entity is referred to as availability (ISO/IEC 27000:2018[4]). Availability related policies will often require that high-impact data has greater availability to legitimate users than normal data. Consider that data essential to the effective operation of a business cannot be allowed to go offline when it can be needed at any moment. Therefore, additional consideration needs to be given to contingent access methods such as geographical redundancy and the provision of alternative network paths.

EXAMPLE

A cloud-based system is used to collate IoT power generation information from thousands of wind turbines and solar arrays spread across a country. This data changes all the time, and is used by energy distribution managers to adjust loading across their power networks, and to predict areas of energy shortfall in time to boost conventional power station output to compensate. If this data becomes inaccessible, there is a real risk of power brownouts or even blackouts for some areas of the country.

8.6.1.3 Framework application to data security management

The following Table 6 shows how the data categorization can be applied to derive requirements for security policies for a given organisation. It makes use of a general data classification (high/medium/low business impact) for each of the main security objectives. For instance, credentials, customer contact lists, and financial details are identified to have high business impact with regard to violation of

1) Except in matters of geographically constrained copyright licensing, where third-party content is obtained under a geographically restrictive license, such as media content or specialist publications.

confidentiality requirements. Violation of integrity requirements for financial data is of medium impact only because in many cases it is possible to reconstruct those data from other sources (and we assume that the organization performing the analysis has access to those sources).

The entries for derived data further illustrate how data identification qualifiers can be used to define security requirements: Confidentiality violations for identified or pseudonymized data has higher business impact than unlinked pseudonymized or anonymized and aggregated data.

Table 6 — Example for using the data taxonomy to derive security policies

Data categories			Policy Item								
			Confidentiality			Integrity			Availability		
			HBI	MBI	LBI	HBI	MBI	LBI	HBI	MBI	LBI
Customer content data	Credentials		X			X				X	
	Customer contact lists		X			X				X	
	Financial details		X				X				X
Derived data	End user identifiable informaton	Telemetry data	I, P	UP	A, AG		X				X
		Client side browsing history	I, P	UP	A, AG		X				X
		Social data	I, P	UP	A, AG	X					X
Cloud service provider data	Access and authentication data		X			X			X		
	Operations data			X			X		X		
Account data	Customer contact information			X			X		X		
	Payment instrument data		X			X			X		

Legend

HBI: High business impact
MBI: Medium business impact
LBI: Low business impact

I: Identified
P: Pseudonymized
UP: Unlinked pseudonymized
A: Anonymized
AG: Aggregated

X: Policies to be defined with regard to the indicated data class (HBI, MBI, or LBI)

8.6.2 Data quality

8.6.2.1 General

Data quality refers to the accuracy, validity, reliability, timeliness, relevance, and completeness of the data being used for a specific purpose. Data quality management incorporates policies, responsibilities and processes with regard to the acquisition, maintenance, disposition and distribution of data. Data quality management is related to the following issues.

8.6.2.2 Data management issues

8.6.2.2.1 Data cleansing

Data cleansing refers to the process of reviewing and correcting data to ensure data are in a standardized format (ISO 13008:2012). It involves the analysis of the quality of data in a data source, deriving measures and suggestions on how the data quality can be improved, and making changes to the data.

8.6.2.2.2 Metadata management

Metadata management can be defined as the end-to-end process and governance framework for creating, controlling, enhancing, attributing, defining and managing a metadata schema, model or other structured aggregation system, either independently or within a repository and the associated supporting processes (often to enable the management of content).

8.6.2.2.3 Data integrity

Data integrity as a security objective has already been discussed in 8.6.1.2.2, but can also be considered from the perspective of data quality management.

8.6.2.2.4 Data quality assurance

Data quality assurance refers to the confidence that one has in the managing of the data quality which can be verified through certification and auditing procedures.

8.6.2.2.5 Data provenance

Data provenance refers to the recording of the change of state of the data during its lifetime, the time of occurrence, individuals associated with the event, locations, software, reasons for why the event took place. Mechanisms providing immutable evidence of data modification, transfer, and use such as distributed ledger technologies and blockchains can be used to establish data provenance.

8.6.2.3 Framework application to data quality management

The derivation of meaningful property for data quality management depends heavily on the use case under consideration. Data quality issues do not materialize for many types of applications, but can critically impact business continuity for others. Also, responsibilities for data management can be different for CSPs and CSCs. Table 7 shows an example on how the data taxonomy of the framework can be used to understand data quality management responsibilities.

Table 7 — Example for using the data taxonomy to derive data quality policies

Data categories	Cleansing	Metadata	Integrity	Quality assurance	Provenance
Customer content data	CSC: depending on use case	CSC: depending on use case	CSC: depending on use case	CSC: depending on use case	CSC: depending on use case
Derived data	CSP: Operations and audit policies	CSP: Operations, security and audit policies	CSP: Operations, security and audit policies	CSP: Audit	CSP: Processes, technical measures (e.g., blockchain)
Cloud service provider data	CSP: Operations and security	CSP: Operations and security	CSP: Operations and security	CSP: Audit	CSP: depending on use case
Account data	CSP: Required	CSP: Required	CSP: Required	CSP: Audit	CSP: Processes, technical measures (e.g., blockchain)

8.7 Data governance

Data is a non-consumable asset with many associated attributes and aspects. These require consideration by the governing body of an organization as items that can have significant strategic impact on the organization as a whole. It is used to keep track of the business (e.g. people, accounting, inventory) and as a raw material for knowledge, innovation and insight. The accountability for data and its use rests with the governing body of the organization. Some data, such as product research or undisclosed stock market ambitions has a high business value and appropriate resources need to be applied to leverage and protect this data. The value and risk associated with managing this data is higher than other types of data and the strategies and policies should reflect this through the adoption of a data classification scheme for the data. The governing body is accountable for data and its use, including the decisions on ensuring the correct level of data stewardship.

ISO/IEC 38505-1:2017[5] applies the general governance of IT principles and model of ISO/IEC 38500:2015[6] to the governance of data. It examines areas of data use by the organization where governance policies and strategies should be in place to ensure accountabilities can be met. The examination includes understanding the potential uses of data (by the organization, its suppliers and customers as well as by its competitors). Such uses could include the buying or selling of data. For example, a manufacturer can include capabilities in their product that allow usage data to be transmitted back to the manufacturer. That data could be used to improve the product, provide technical support or preventative maintenance. It could also be used as the basis for paying for the use of the product

rather than outright purchase. The data could also potentially be sold by the manufacturer to other manufacturers as a mean for them to build products or services, or shared with suppliers to improve the overall supply chain. Additionally, the usage data could be combined with additional data such as the weather to increase the understanding of the product's use under different conditions. The data itself becomes part of a value chain.

Because data has value, it also brings risk to the organization. Additionally, there can be constraints on data use imposed by law, regulation, society norms or by the organization itself.

ISO/IEC 38505-1[5] also provides guidance for a comprehensive checklist of considerations for a governing body for developing a governance framework for data appropriate for their organization. This checklist supports the leveraging of the maximum value from data within the organization's data risk appetite while considering external and internal constraints.

In this way, the governance framework for data supports a data strategy which itself supports and influences the overall organizational strategy.

Having established the overall data strategy and supporting governance framework, ISO/IEC TR 38505-2:2018[7] describes "*what the governing body of an organization expects and requires from the data management team in order to be assured that the governing principles of IT can be implemented and are being upheld for data and its use by the organization.*" A mechanism for achieving that assurance is for the data management team to work with the governing body to derive data management policies and align with and support the data strategy as shown in Figure 6.

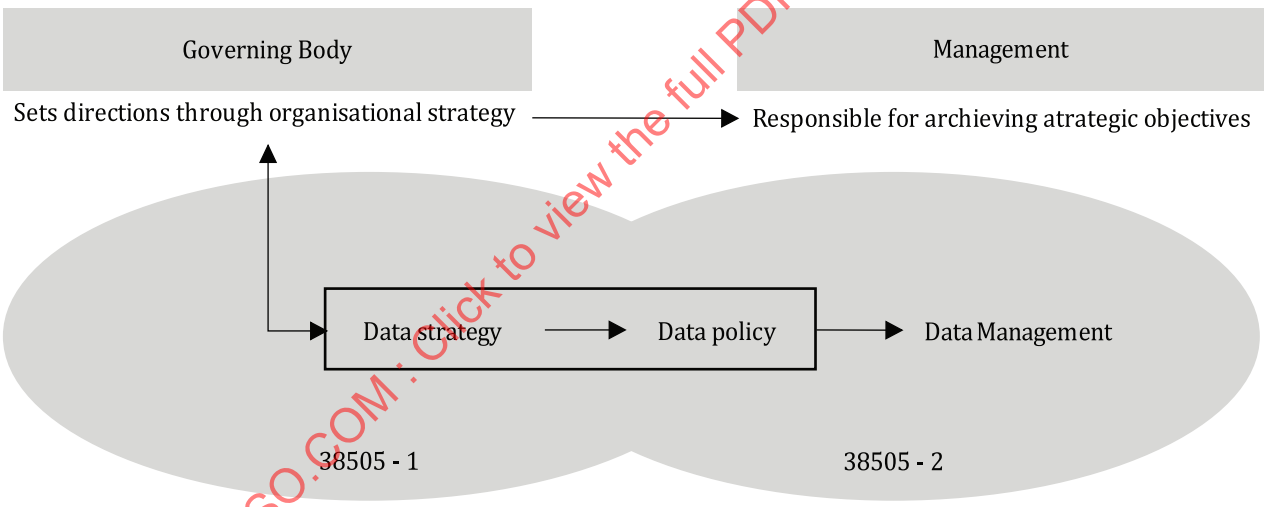

Figure 6 — Data strategy and data policies (from ISO/IEC TR 38505-2[7])

The data policies that are implemented by the data management team should apply to all data used by the organization. This will include data that is, e.g. bought, sold, shared, processed, stored. These data policies should then align with the data-related policies and practices as described in this document, i.e. we could extend part of the above diagram to encompass the data use statements as shown in Figure 7.

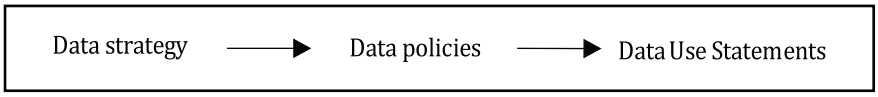

Figure 7 — Alignment of data strategy, policies and use statements

A CSP will have an organizational purpose and strategies for achieving that purpose. From there, a data strategy can be derived, along with policies that support the use of data by the organization. Data use statements can then be created by taking those policies into account. Such use statements will then align to the overall business objectives of the CSP.

Similarly, a CSC will go through this process to describe a data strategy that helps them achieve their organization's goals. At a minimum, the CSC should ensure that the Data use statements provided by the CSP align to their own Data Policies, as well as their own data use statements if they have them.

As an example, the CSP can have an organizational goal of providing a cloud service that

— is easily accessible from anywhere,

— provides virtually unlimited scale and ,enterprise-level performance, safety and reliability,

— conforms to local regulations and expectations.

These are very high-level goals and to achieve them requires a more detailed examination. Tables 8 to 10 below look at this from the CSP perspective, and the resulting policies (which are derived from a culmination of the cells in each row) are those of the CSP. The CSC should examine these policies to ensure they align with their own policies and required use of data for their organization.

These high-level governance policies also act as a guide for the other CSP data management policies explored in this document. There should be a strong alignment between the policies such that the data is managed in a way that conforms to the governance level policies and helps the organization achieve its goals.

Table 8 — Example for data governance policy worksheet (CSP perspective)

Data governance aspects	Governance of data policies			
	Value	**Risk**	**Constraints**	**Possible Policies**
Collect	CSC can collect all kinds of data from all over the world. Collect usage statistics and other telemetry. Industry data including energy use, social responsibility and certifications to ensure leading position. CSC usage required for billing. PII needed for individuals to access some services.	Valueable data could attract persistent cyber attack. Data collected could be used for illegal purposes.	Health data and PII can require special processing. CSP will process data on behalf of CSC.	CSP strategy is to provide services that allow for simple customer registration and ingestion of data. CSC agrees to CSP processing data.
Store	Improve perceived performance and save money by storing data according to frequency of access. Reduce lag by storing data close to processing. Add reliability through replication.	Natural disasters Cybersecurity risks including CSC introduced malware.	Data can be stored in-country.	Data centres to be physically located in top 10 regions based on Cloud Computing market size. CSC content Data to remain inaccessible to CSP – except in limited law-enforcement scenarios.
Report	Telemetry data needed to help improve services, including strategic expansion or alteration of services, personnel and data centres. Billing systems require CSC usage measurements. Data could be used for security incident monitoring.	Quality of metrics could be inconsistent. Measurement activities impact system performance.	Telemetry should not reveal PII.	CSP Measurement activities should remain at less than 0.05% capacity and CSC impact should be immeasurable by CSC.
Decide	Use analysis tools and AI on telemetry data to improve, expand or reduce services. Analysis for security of services – to help protect CSCs as well as the CSP operations.	Are we measuring the right things?	Not applicable	CSCs need to be protected from external attacks as well as from each other.
Distribute	Some telemetry to suppliers (capacity management, design input, reliability scores etc).	Data access management. Telemetry and billing data could be highly confidential – for CSP and CSC.	Limited CSC content data to law enforcement upon valid request and CSC wherever possible.	Telemetry, usage and billing data should be classified as confidential and high value.
Dispose	CSC to manage their data disposal.	Could ‚disposed' data be recreated?	Proper disposal of Data collected by organizations and government according to requirements from the applicable jurisdiction	Ensure that all at-rest data is encrypted so hardware can be disposed without cleansing.

Following the creation of these governance level data policies, the CSP could examine particular categories of data based on the aspects of data governance such as collect, store, report.

Again, the derived policy is a composite of the row requirements. The policies at this level could be incorporated into agreements with CSCs as well as being further detailed into data use statements.

Table 9 — Example for data category specific policies worksheet: Collect (CSP perspective)

Data categories	Governance of data policies			
	Value	Risk	Constraints	Policy
Customer content data	Not applicable	Not applicable	Not applicable	The CSC determines the data that is collected.
Derived data	Not applicable	Not applicable	Not applicable	Not applicable
Cloud service provider data	Not applicable	Not applicable	Not applicable	Not applicable
Account data	Metrics for Revenue	Not applicable	Not applicable	Account data is considered High Business Value.

Table 10 — Example for data category specific policies worksheet: Store (CSP perspective)

Data categories	Governance of data policies			
	Value	Risk	Constraints	Policy
Customer content data	CSC content data is an indirect source of CSP revenue insomuch as requires the use of CSP services.		Any CSC data used for service support investigations should be disposed of at the conclusion of the support investigation.	The CSC is responsible for the content of the data stored. The CSC indemnifies the CSP against illegal data storage (e. g. special PII, health data, illegal images). CSC data will not be used for advertising. The CSC retains all rights to the CSC data.
Derived data	Indirect revenue via service use. Behaviour to predict security signals. Telemetry to improve services.	CSC identifiable data should not 'leak' into derived data.	Not applicable	Not applicable
Cloud service provider data	Not applicable	Not applicable	Not applicable	Not applicable
Account data	Metrics for Revenue	Not applicable	Not applicable	Not applicable

9 Application of the framework to codes of conduct

Codes of conducts capture industry or regulatory consensus in the form of preferred or required policies or practices. Expression of such consensus policies or practices, when they involve data, could benefit from this document. The documented consensus captured in codes of conduct often require unambiguous and precise expressions of data handling and use to the desired level of detail and granularity. When such policies and practices involve data, they often need to be expressed in terms of the data taxonomy, using methods and techniques described in this document.

The following are a few scenarios where codes of conducts can be used:

— Codes of conduct for an industry sector in a province or country, in an economic or political zone, or for a group of nations bound together by a treaty or otherwise multi-lateral agreement, codifying the data-centric practices and policies for that industry sector (e.g. healthcare, financial, manufacturing, energy). Such codes of conducts can be required legally or created voluntarily by industry consensus.

— Codes of conduct for a regulatory or public policy practices, in a country, in an economic or political zone, or for a group of nations bound together by a multi-national agreement (e.g. codes of conduct created for adherence to one or more sections of GDPR in European Union). In such cases the codes of conduct can be created by one or more governments, or by industry stakeholders that are subject to those regulatory or public policies.

— Codes of conduct internal to an organization operating within a country, or a multi-national organization, codifying the data-centric practices and policies inside that organization (e.g. codes of conduct within the United Nations, or a multi-national corporation).

As an example, consider the flow of data across geopolitical boundaries. The following table is an example for how the structures defined elsewhere in this document that address data geolocation, cross border flow of data as well as data portability can be used in codes of conduct that addresses the flow of data and data portability in an economic zone or common market:

Table 11 — Example for data taxonomy-based codes of conduct — Free flow of data within an economic zone or common market

Data categories			Identified data	Pseudonymized data	Unlinked pseudonymized data	Anonymized data	Aggregated data
Customer content data	**Customer content - general**		1	1	3	3	4
	Customer contact lists		1	1	3	3	4
	Personal health data & medical records		1	1	2	3	4
	Credentials		1	N/A	N/A	N/A	N/A
Derived data	**End user identifiable informaton**	**Derived data - general**	1	1	3	4	4
		Telemetry data	1	1	3	4	4
		Connectivity data	1	1	3	4	4
		Observed usage of service capability	2	2	3	4	4
		Demographic information	2	2	3	4	4
		Interest and favorite data	1	1	3	4	4
		Content consumption data	2	2	3	4	4
		Client side browsing history	2	2	3	4	4
		Search commands and queries	2	2	3	4	4
		User location	1	1	3	4	4
		Social data	1	1	3	4	4
		Biometric and health data	1	1	3	4	4
		End user contact data	1	1	2	3	3
		User's environmental sensor data	2	2	3	4	4
Cloud service provider data	**Cloud service provider data - general**		4	4	4	4	4
	Access and authentication data		3	3	4	4	4
	Operations data		4	4	4	4	4

Legend/Explanation
Date flow needed between CSP, CSC and CSN that are located
1 - in the country only
2 - within the economic or political zone only
3 - select non-member countries included in (2)
4 - global flow

The following is an example for a captured agreement as part of codes of conduct for data portability requirements, based on the data taxonomy, within an economic zone or common market:

Table 12 — Example for data taxonomy-based codes of conduct — Data portability of data within an economic zone or common market

Data categories			Identified data	Pseudonymized data	Unlinked pseudonymized data	Anonymized data	Aggregated data
Customer content data	**Customer content - general**		2	2	3	4	4
	Customer contact lists		2	2	3	4	4
	Personal health data & medical records		2	2	3	4	4
	Credentials		No	No	No	No	No
Derived data	**End user identifiable informaton**	**Derived data - general**	No	No	No	No	No
		Telemetry data	No	No	No	No	No
		Connectivity data	No	No	No	No	No
		Observed usage of service capability	1	1	No	No	No
		Demographic information	1	1	No	No	No
		Interest and favorite data	1	1	No	No	No
		Content consumption data	1	1	No	No	No
		Client side browsing history	1	1	No	No	No
		Search commands and queries	1	1	No	No	No
		User location	1	1	No	No	No
		Social data	1	1	No	No	No
		Biometric and health data	1	1	No	No	No
		End user contact data	1	1	No	No	No
		User's environmental sensor data	1	1	No	No	No
Cloud service provider data	Cloud service provider data - general		No	No	No	No	No
	Access and authentication data		No	No	No	No	No
	Operations data		No	No	No	No	No

Legend/Explanation

Date flow needed between CSP, CSC and CSN that are located

1 - in the country only

2 - within the economic or political zone only

3 - select non-member countries included in (2)

4 - global flow

No - Data portability not required

The above examples demonstrate how data-centric codes of conduct could be expressed with the desired level of precision and granularity. Furthermore, the expression of the policies and practices in given codes of conduct is more effective when they are expressed based on the categories and levels of de-identification of the data involved.

Table 13 is an example for captured agreements for a data taxonomy-based codes of conduct that address data portability for application capabilities type cloud services (for example a SaaS service). This example is based on ISO/IEC 19941:2017[1], 8, cloud data portability and the three facets: syntactic, semantic and policy. Furthermore, this document describes these facets based on one of the three cloud capacities types: Application, Platform and Infrastructure.

Table 13 — Example for data taxonomy-based codes of conduct — Data portability for application capabilities type cloud services

Data categories			Data syntactic	Data semantic	Data policy
Customer content data	**Customer content - general**		JSON	Comm	Data locality, security, privacy
	Customer contact lists		OOXML	Comm	Data locality, security, privacy
	Personal health data & medical records		Custom1	H	Data locality, security, privacy, health
	Credentials		ASCII	N/A	Data locality, security
Derived data	**End user identifiable informaton**	**Derived data - general**	N/A	N/A	N/A
		Telemetry data	JSON	Comm	Data locality, security, privacy, right to access
		Connectivity data	JSON	Comm	Security
		Observed usage of service capability	CSV	Photo	Security, privacy, data sharing
		Demographic information	ASCII	SM	Security, privacy, data sharing
		Interest and favorite data	JSON	SM	Security, privacy, data sharing
		Content consumption data	JSON	Photo	Security, privacy, data sharing
		Client side browsing history	CSV	SM	Security, privacy, data sharing
		Search commands and queries	CSV	Hu	Security, privacy, data sharing
		User location	Custom2	Comm	Security, privacy, data sharing, data locality
		Social data	JSON	SM	Security, privacy, data sharing
		Biometric and health data	Custom3	H	Security, privacy, data sharing, data locality, health
		End user contact data	ASCII	General	Security, privacy, data sharing
		User's environmental sensor data	JSON	Comm	Security, privacy, data sharing
Cloud service provider data	**Cloud service provider data - general**		Portability not required	Portability not required	Portability not required
	Access and authentication data		Portability not required	Portability not required	Portability not required
	Operations data		Portability not required	Portability not required	Portability not required

Legend/Explanation: Key codes describing portability of data

Syntactic	**Semantic**		**Policy** (regulatory or enterprise)
ASCII, JSON, CSV, XML, OOXML, Custom1, Custom 2	**Domain models:** Communication (Comm), Photography (Photo), Human Resources (Hu), Social media (SM), Healthcare (H), Music (M), etc.	**Semantic languages used:** OWL, Resource Description Framework (RDF), modelling notations such as UML and Business Process Management Language (BPML), are tools that can be used to document and share understanding	data locality, right to access, data use, data sharing, security, privacy, health regulations (health)

In the above table, for each facet of data portability, several options and possibilities are selected from the available options from the key codes. For example, syntactic portability is done via JSON syntax for the case of customer content data, while credentials are delivered in plain ASCII format. For data semantic portability, the example points to the customer content data assumes semantic definitions for communication line of applications, not from photography or human resources space. Equivalently, customer content data is subject to data policies in the areas of data locality, security and privacy. But when it comes to portability of personal health data, applicable healthcare regulations in that jurisdiction should also have to be taken into account.