TECHNICAL REPORT

ISO/IEC TR 23186

First edition
2018-12

Information technology — Cloud computing — Framework of trust for processing of multi-sourced data

Reference number
ISO/IEC TR 23186:2018(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

There are many business and technical aspects relating to the processing of multi-sourced data, but trust between cloud service users, cloud service customers and the cloud service provider(s) is a significant market issue.

Cloud processing of multi-sourced data is in its early stages of development in the industry, and it is anticipated that specific customer requirements will differ and will evolve over time. Industry clouds have begun to form, and in some cases, their primary purpose is to bring multi-sourced data together from participants in specific industry or community sectors to achieve common objectives. Trust may be required in these scenarios because of regulations, agreements or policies attached to the data.

Processing of multi-sourced data will be essential to artificial intelligence applications along with machine learning on financial, transportation, energy, manufacturing, agricultural and government data. Trust in the data, in the cloud service provider(s), in the processing functions, in the outcomes and among the parties is essential to the success of these projects.

The elements of trust described in this report pertain to Personally Identifiable Information (PII), Organizational Confidential Data (OCD) or any other kind of data that can be a part of multi-sourced data.

Information technology — Cloud computing — Framework of trust for processing of multi-sourced data

1 Scope

This document describes a framework of trust for the processing of multi-sourced data that includes data use obligations and controls, data provenance, chain of custody, security and immutable proof of compliance as elements of the framework.

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*

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 http://www.iso.org/obp

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

3.1
chain of custody
demonstrable possession, movement, handling, and location of material from one point in time until another

[SOURCE: ISO/IEC 27050-1:2016, 3.1]

3.2
data
recorded information

[SOURCE: ISO 22005:2007, 3.11]

3.3
data processing
systematic performance of operations upon data

[SOURCE: ISO 2382:2015, 2121276, modified — Notes 1 to 4 to entry have been deleted and the alternate term “automatic data processing” has been deleted.]

3.4
data set
logically meaningful grouping of data

[SOURCE: ISO 8000-2:2018, 3.2.4, modified — EXAMPLES 1 and 2 have been deleted.]

3.5
multi-sourced data
data that consists of separate data sets that have been generated by multiple, diverse sources and assembled by one or more cloud services from one or more CSPs

Note 1 to entry: The data sets are then subject to combined analysis and processing with the aim of extracting insights and information not obtainable through analysis of each dataset on its own.

3.6
personally identifiable information
PII
any information that (a) can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal

[SOURCE: ISO/IEC 29100:2011, 2.9, modified — The NOTE has been deleted.]

3.7
trust
degree to which a user or other stakeholder has confidence that a product or system will behave as intended

[SOURCE: ISO/IEC 25010:2011, 4.1.3.2]

4 Symbols and abbreviated terms

PII Personally identifiable information

5 Scenarios

5.1 Using multi-sourced data to reduce traffic deaths and injuries

Worldwide, 1,25 million people die each year from traffic-related accidents and between 20 million and 50 million people suffer injuries. Data sets include accident data, roadway attributes, land use, demographics, commuting patterns, parking violations and existing safety improvements. One of the key outcomes is an "exposure model" that predicts the number of cars in a given location at a given time. Actual measurements of traffic are very expensive while predictions using machine learning are relatively inexpensive.

For example, In the US, where 34,000 people die annually in traffic-related accidents, a non-profit organization, called DataKind®[1] is using data and machine learning to develop models to predict traffic accident patterns. These patterns can then be used to determine where to focus street improvements and predict the effect on accident rates for specific improvements. Street improvements have included traffic signals and controls, bicycle lanes, road design and treatments.

DataKind® held a DATADIVE®[2] to bring data scientists together to transform the available data and develop the model.

One of the key challenges in this scenario is getting data owners to entrust their data to a group and to a third-party processor. Specific concerns include:

— How are applicable regulations, policies and other data use restrictions identified and adhered to?

— How are privacy infringements avoided?

1) DataKind is the service mark of DataKind. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO or IEC.

2) DATADIVE is the service mark of a service supplied by DataKind. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO or IEC of the service named. Equivalent services may be used if they can be shown to lead to the same results.

— What processes are employed to provide end-to-end security?

— How is data provenance maintained?

— What is the proof of compliance?

Figure 1 illustrates the system for predicting traffic accident patterns as described above.

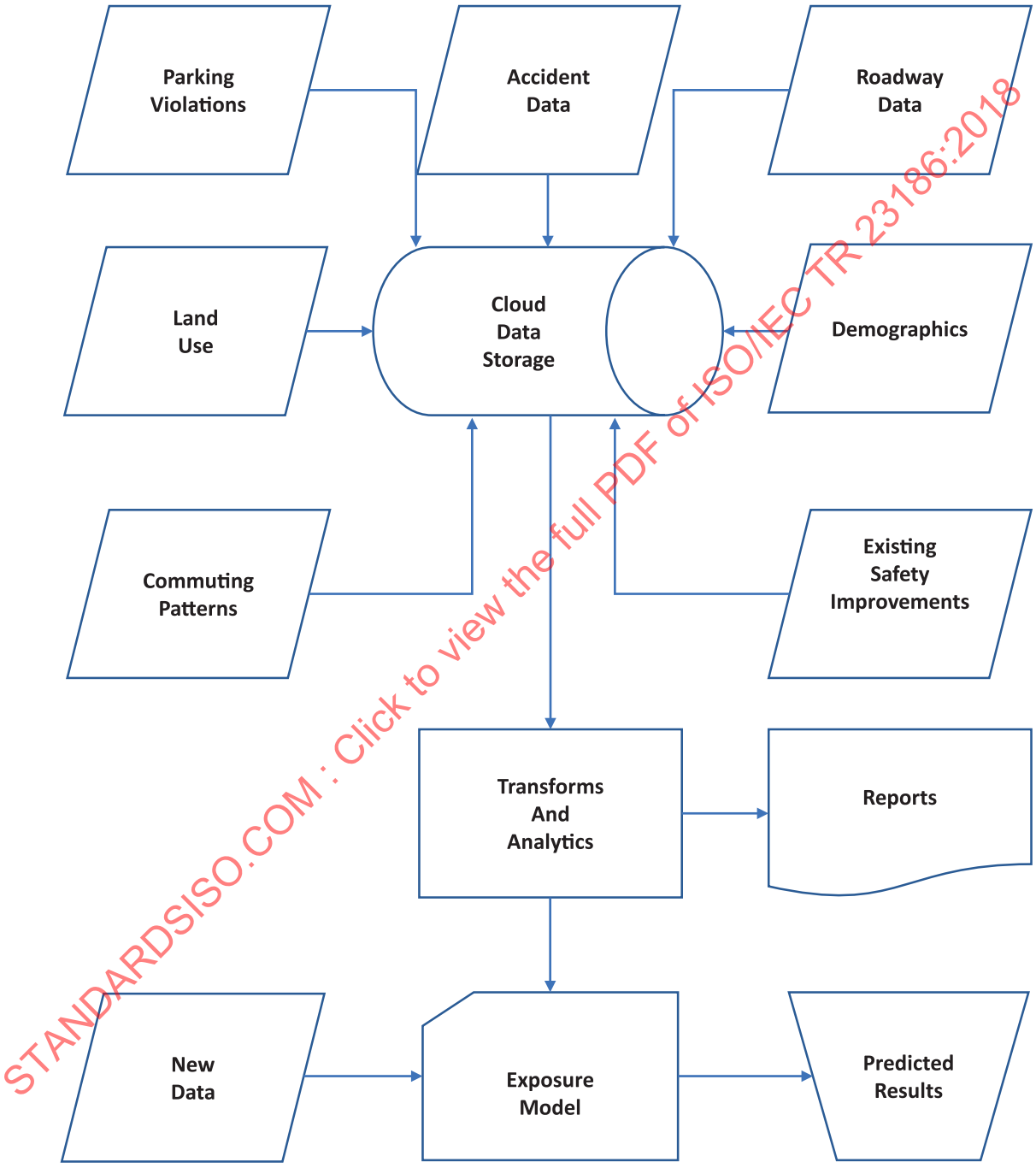

Figure 1 — Example of a system for predicting traffic accident patterns using multi-sourced data

5.2 Using multi-sourced data for home automation

A variety of emerging home automation applications could benefit from access to a larger variety of data, potentially sourced from multiple providers and processed in a coordinated and timely manner.

The general goal for any home application is to use available data to improve the efficiency and effectiveness of the home both as a part of a municipality and as a desirable place to live. To do this, home IoT/IS systems need to be informed, agile and more dynamic for its residents, both for managing the building itself and also for the quality of life inside the home.

There are many home services being developed by many service providers and manufacturers, ranging from smart entertainment systems (TV, Internet, telephone, wall art display, etc.) to emergency detection and alarms to smart electricity and water management.

Some examples of multi-sourced data for home operations processing could include:

— Time signals from public sources;

— Weather forecast information and current state for the home area;

— Neighbourhood information, e.g. alerts, fire alarms, air quality, road congestion;

— Sensor-based data from multiple multivendor systems located within the house, e.g. locks, temperature, appliance state and status, e.g. refrigerator breakdowns, occupancy status (is anyone home?), connectivity status, electricity and water status and meter readings;

— Home service maintenance, support and billing information;

— Visual and audio data sources (internal and external);

— Health emergency and intrusion alarms;

— Calendar and current location information for residents, i.e. who is expected to be home and when;

— Policies and configuration settings, e.g. water rates, electricity costs, time-of-day rules.

Information sources could be classified as:

— Public information, e.g. public databases, governments, municipalities, legal rules, supply rates;

— Home-based information, e.g. IoT sensors in and around the home. These could be distinct sources, e.g. from different vendors equipment or could be aggregated and delivered from a hub; this could include both real-time data, human user input data, and archival data;

— Related element sources, e.g. occupant vehicle data, manufacturer information, opportunity information, e.g. local events, component replacement sales;

— Historical insights and trends;

— Policies and rule settings from governments, vendors and residents.

All data would be accessed, combined, processed and managed both independently and in combination to provide an increasingly intelligent basis for home activity automation and home operations control and protection. Trust is needed to avoid accidents, spoilage, inefficiencies and false actions within and around the home.

Many point solutions using independent data sources could benefit from reliable processing in a more coordinated and orchestrated way. For example, if the home temperature control system receives input of the weather forecast, the home inside temperature and arrival time of any occupants, the energy balance could be more efficiently optimized.

5.3 Using multi-sourced data for automotive operations

The term “car” represents a wide range of vehicle types that may have very similar requirements. The manufacturers, owners, drivers and occupants of a car may be customers of the car’s cloud-based and on-board systems.

Car applications can benefit from access to data from multiple sources that is made available in a coordinated, timely and trusted manner. Two use cases are the collection of information for use:

— by the car itself (an on-board "cloudlet") or its cloud-based proxy for driving purposes; or

— by insurance companies, car manufacturers, cities, governments, and others for related and off-line services such as maintenance, usage tracking, congestion management, and many other possibilities.

General goals for any car are to optimize the user (passenger) experience, reduce transportation costs and delays, improve safety, and maximize vehicle life. To do this, car automation systems need to be informed, agile and dynamic especially if self-driving or assisted-driving systems are being used.

Car-related services range from in-car social networking to route management to emergency alarms and collision avoidance. In addition, considerable information may be collected for repair and maintenance or for defect detection. Other applications try to make the car a "home away from home" and could provide access to all the data that would be available at home.

A car could be viewed as a cluster of "IoT things," each of which may require feeds from different data sources or may interact with external systems that process data from many sources. There may be hundreds of sensors associated with a single car.

Examples of multi-sourced data for car operation could include:

— Time signals from public sources;

— Weather information and current state for areas of interest;

— Road and surroundings information, e.g. blockages, congestion, accidents, disasters, which could come from many sources including other cars;

— Sensor-based data from within the car, e.g. locks, internal/external temperature, component (e.g. engine) state and status, driver and occupant status, which can be used directly by the "car cloud" or used remotely with results fed back to the car;

— Car maintenance, support and service information (both collected and reported);

— Visual, audio and data sources (internal and external) for passenger use;

— Occupant health status, emergency and intrusion (break-in) alarms;

— Information from other ecosystems of interest, e.g. home, office.

All these data sources could be accessed, combined, processed and managed both independently and in combination to provide an increasingly intelligent basis for car operations, control and protection. From the macro perspective, information from many cars can be collected and processed to develop information for the car suppliers, city planners and regulatory agencies.

Many point solutions using independent data sources could usefully be coordinated and shared. For example, if a car knows the weather forecast, the current road conditions and the amount of fuel available, then road selection and rest/re-fuelling stops could be more effectively orchestrated and optimized.

6 Trust

Trust is a key element in the processing of multi-sourced data. Trust has a variety of meanings and forms for the various parties associated with the data and processing of the data depending on different perspectives. The parties involved include the organization(s) processing the data, the organization(s) which are the sources of the data, people whose PII is contained within any of the data, and finally people and/or organizations who use the output of the processing.

For an organization processing the data, one of the major elements of trust concerns the provenance of the data that they use: how was the data put together, how reliable is the information it contains, does the data require cleansing or filtering, how complete is the data it contains, does the data contain PII or confidential information of any kind. Other issues concern any regulations and laws that might apply to the data and any commercial terms that apply to the data that might affect the planned processing.

For an organization as the source of data, the major element of trust concerns whether the processing organization uses data as authorized. The essential questions may include:

— Does the processing organization make a clear statement about the intended uses?

— Does the processing organization sign an agreement in relation to this processing, and particularly agree to abide by any restrictions or regulations that apply to the data (both regarding commercial terms, if any, and any regulations or laws that apply)?

— Does the processing organization have appropriate certifications or equivalent proofs in relation to the processing, including appropriate security controls and PII protection?

For any individuals who have PII contained in any of the data used for processing, the major concern is that the PII is processed transparently and only for purposes that have been clearly stated to the individuals and for which consent has been obtained. A major concern relating to any PII breaches that might occur is whether all necessary measures are in place to prevent such breaches.

Finally, for the people or organizations using the output of the processing, the key element of trust concerns their ability to rely on that output, that it is correct, that it is unbiased, that it matches any claims made for the output by the processor.

7 Data access and processing rights

Establishing who can use, process and pass on data, and understanding the rights various parties have over the data is essential for a multi-sourced data ecosystem.

From the discussion of ownership of data, the example of a car ecosystem is a quintessential use case in establishing data access and the rights to view, modify, copy, or process the data considering there could be distinct sources of data from the various components, the component vendor, the car manufacturer, or the owner of the car, who can own the data after it is generated. Technically, an identity and access management (IdAM) system provides the tools to control who can access a resource and when, and thereby the IdAM system owner grants access to different parties.

It could be argued that the traditional legal notion of ownership does not fit with this new model, and either new constructs are required, or the notion of co-ownership needs to be further explored. This discussion on ownership is further complicated when discussing data aggregated from multiple sources which is then processed through AI or machine learning systems, for example.

Regardless of who owns the data, there could be legal rights for parties to access, process and collect data. For example, to ascertain the cause of a collision, police might have an automatic entitlement to all the data in a car after a collision with fatalities. Another example, related to PII, where the owner or driver of the vehicle could be entitled to copies of any data that has been collected or extracted. In Europe the General Data Protection Regulation (GDPR) and revisions to the Directive on Privacy and Electronic Communication (Directive 2002/58/EC) describe the legal basis for different types of processing of data. Another scenario is that local laws could require that the results of processed data be made publicly available, suitably anonymized, if the system has been publicly funded, e.g. real time traffic data or data relating to public transport services.

8 Framework for trusted processing of multi-sourced data

8.1 Introduction

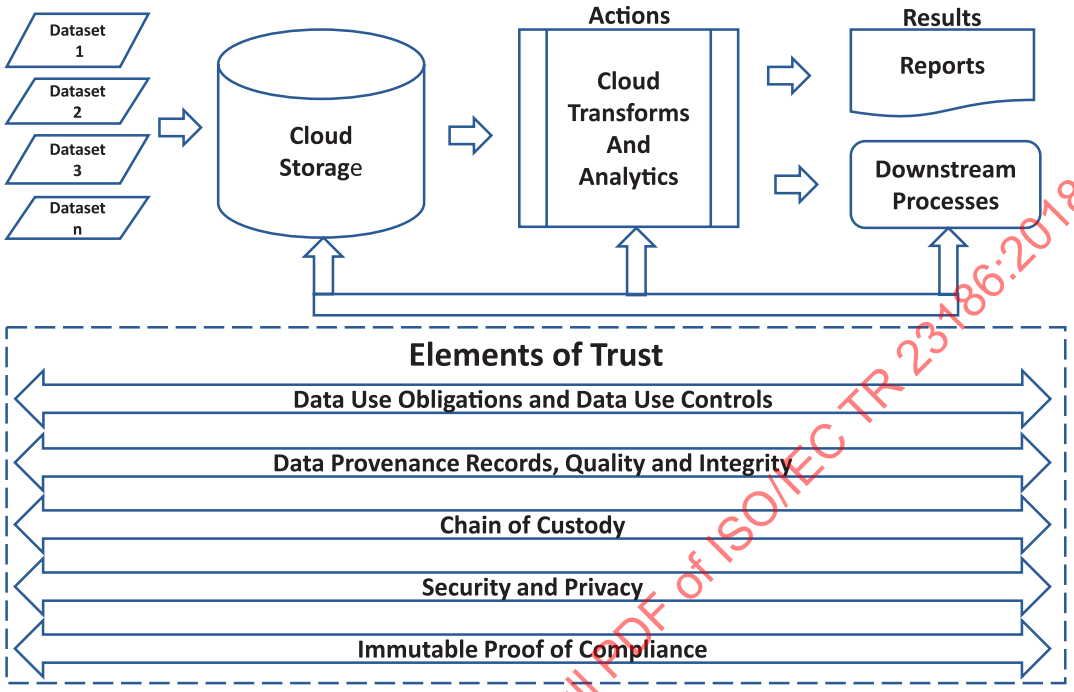

Figure 2 — Framework for trusted cloud processing of multi-sourced data

Figure 2 depicts "elements of trust" and a representative data flow for the processing of multi-sourced data. Descriptions for each part of Figure 2 are provided later in this document.

Standards to support the elements of trust will likely come from multiple JTC 1 and ISO committees such as JTC 1/SC 27, JTC 1/SC 32, JTC 1/SC 40, JTC 1/SC 41, ISO TC 307 and others. The elements of trust will likely also be informed by sector specific regulations and societal norms.

8.2 Data flow

The representative data flow shown in Figure 2 is provided only to establish a context for the elements of trust and should not be construed as the only way to architect processing of multi-sourced data. However, the various examples described in this document such as multi-sourced data for car operation have shown that it cannot be a simple direct flow where one system ends, and another begins. For a cloud computing reference architecture, see ISO/IEC 17789. For additional information on cloud computing data flows see ISO/IEC 19944. For IoT dataflows see ISO/IEC 30141.

Data can come from any number of sources, including data brokers[3)], and can include both proprietary and open data. The data can also have different formats, schema and security schemes along with different restrictions on its use. Multi-sourced data can be brought in over a network or by other means such as portable data storage appliances. The data may need to be wrangled and cleansed before it is suitable for processing.

There is industry uptake for putting computing resources at the edge of the network for applications that require real time data processing when associated with devices from cars and medical devices to consumer products and industrial machinery. Data is processed and analysed by placing servers or gateways in the proximity of the devices. This can be done to reduce latency when transporting data to a cloud service provider or data centre. The elements of trust described in this document are important to these scenarios as well.

3) Data brokers collect data from one or more sources and provide the data to one or more customers.

Several cloud data storage types are suitable for multi-sourced data. Multiple datasets can be stored in a "data lake" with each data owner retaining control of their data. "Data vaults" are another model for managing multiple datasets and have the advantage that data provenance and chain of custody are attached to each record in the vault. A store for multi-sourced data can be virtual, spanning multiple cloud storage services from multiple cloud service providers along with storage in legacy data centres.

Figure 2 shows transforms and analytics as the actions being applied to the multi-sourced data, but actions can include any other type of data processing actions such as simple queries. As the multi-sourced data is processed, the specific actions can become part of the chain of custody and provenance.

Downstream processes can include any applications that make use of actions such as recommendation systems, classification systems and predictive decision systems. Figure 2 shows a connection between data storage, actions and downstream processes as downstream processes can interact back with the actions and both will ordinarily interact with the storage service. Adding output data from the downstream process back as input data to the actions is a way to continuously train data analytics models.

8.3 Elements of trust

8.3.1 General

The elements of trust are shown end-to-end relative to the data flows in Figure 2 because an untrusted component of a system results in an untrusted system. Additionally, there can be unique trust requirements at each stage of processing. In some cases, cloud service customers demand limited distribution of reports while in other cases the reports may become part of open data. Additionally, there may be requirements to provide trusted execution environments. An element of trust may require multiple interconnected components to function end-to-end. In this representative Figure 2, the highlighted elements are detailed in subsequent clauses.

8.3.2 Data use obligations and controls

A data use control is any technical or organizational instrument whose purpose is to ensure that an obligation on data processing is met by the cloud service provider or an associated third party.

Obligations of a cloud service provider on the processing of data, e.g. company data, personally identifiable information, data with relevance to public security and safety, health data, can be derived from a number of sources:

— **General regulatory acts**, such as the General Data Protection Regulation (Europe), the Personal Information Protection Act (Japan), the Personal Information Protection and Electronic Documents Act (**PIPEDA**) in Canada, or the Cybersecurity Law (China). On the protection of personally identifiable information, the United States follows a 'sectoral' approach relying on a combination of legislation, regulation, and self-regulation rather than government regulation alone;

— **Specific legislations**, such as the German eHealth Law, the US Sarbanes-Oxley Act on the accuracy and reliability of corporate disclosures, etc.;

— **Governmental mandatory policies and standards**, such as Federal Trade Commission (US), Federal Communications Commission (US), BSI Baseline Protection (Germany);

— **Directives, guidelines and recommendations**, such as the European Directive 2003/98/EC on the re-use of public sector information. Directives, guidelines and recommendations can be issued by government entities. Guidelines and recommendations can also be developed by industries;

— **International standards**, such as ISO/IEC 27000 or ISO/IEC 19944;

— **National standards**;

— **Company and project specific policies**.

These sources define either:

— Direct obligations on the processing, transmission, or storage of data such as requirements on data security, e.g. encryption, access control;

— Indirect obligations derived from the requirement to respect rights of customers, employees, or data subjects, e.g. information rights, data avoidance, lawful requests for data deletion.

Controls associated with direct obligations usually relate to the security and privacy of data processing by a single data processor and thus are agnostic to the specific challenge of trusted processing of multi-sources data. Examples are access controls such as:

— Control of user access to corporate IT systems, networks, applications and information;

— Controls on security log entries, security alerts, locking of user;

— Rules on the length and complexity of pass phrases;

— Security of authentication information;

— Rules on privileged access rights;

— Rules on the use of removable media, e.g. USB drives, CD/DVD writers.

Examples for controls that are dedicated to increase the trust in the processing of PII are:

— Requirements for the encryption of PII — in motion, when stored and on any removable physical media;

— The deletion of PII within a specified period once the data is no longer required;

— That PII is processed only for the purposes expressly stated in the cloud service agreement;

— To cooperate in dealing with the rights of PII principals in inspecting and correcting their PII, something that is mandated by many regulations.

Controls derived from indirect obligations relate to the whole data processing chain involving multiple data controllers and processors and therefore require an end-to-end approach for definition and implementation. Examples are controls associated with data subject rights such as:

— Controls allowing data subjects to obtain information on the purpose of the processing of related data, the parties to which those data are transferred, the time period of the data processing, etc.;

— Controls allowing data subjects to validate whether stored data are correct and to request correction of incorrect data;

— Controls allowing data subjects to request the overall deletion of relevant identifiable information;

— Controls relating to the transfer of relevant data from one (cloud) service provider to another.

The so-called “Final Privacy Report” of the US Federal Trade Commission compiles a number of proposed obligations (called “principles” in the report) for data processing by companies which are not specific to PII but apply to all types of consumer data:

— Privacy by design: companies can build in consumers' privacy protections at every stage in developing their products. These include reasonable security for consumer data, limited collection and retention of such data, and reasonable procedures to promote data accuracy.

— Simplified choice for businesses and consumers: companies can give consumers the option to decide what information is shared about them, and with whom. This can include a Do-Not-Track mechanism that would provide a simple, easy way for consumers to control the tracking of their online activities.

— Greater transparency: companies can disclose details about their collection and use of consumers' information and provide consumers access to the data collected about them.

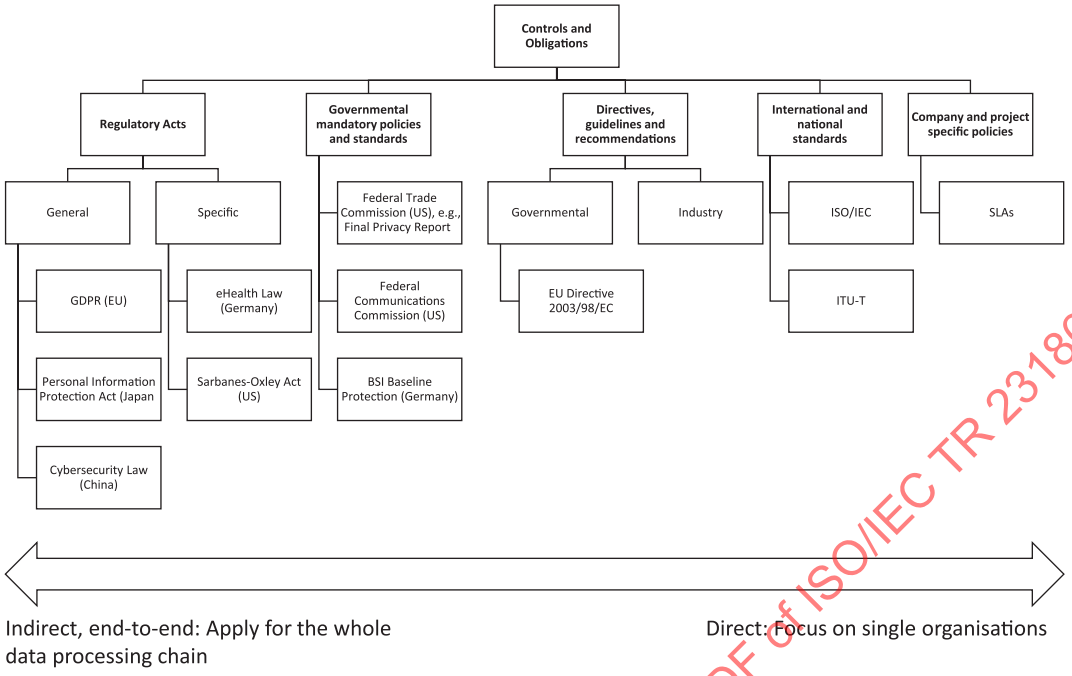

Figure 3 — Representative classification of sources of obligations and controls (dark shade) with examples (light shade)

8.3.3 Data provenance records, quality and integrity

For trusted cloud processing of multi-sourced data, the origin of each piece of data is required to establish trust in the results of data analysis and related downstream processes. Conclusions drawn from analysis, or decisions made based on analysis rely on traceability of data back to its origin.

ISO 8000-2:2018, 3.8.4 defines a "data provenance record" as a "record of the ultimate derivation and passage of a piece of data through its various owners or custodians". Furthermore ISO 8000-2:2018, 3.8.4 notes that "a data provenance record can include information about creation, update, transcription, abstraction, validation, and transferring ownership of data". The role of data custodian is particularly important in processing multi-sourced data as the custodian may manage the collection of multi-sourced data, the processing and the output. ISO 15836 and the Dublin Core Metadata Initiative (DCMI) can be useful for creating data provenance records.

Data portability can also be considered as it may be necessary to move the data to a different set of resources over the lifecycle of a project. ISO/IEC 19941 can be used to establish data portability requirements.

The quality of the output of the actions on multi-sourced data will, at best, be no better than the quality of the data itself. ISO 8000-2:2018, 3.8.1 defines "data quality" as the "degree to which a set of inherent characteristics of data fulfils requirements". This means that the participants in a multi-source data project need to determine their own characteristics and the requirements.

Data integrity is closely related to data quality and will affect the outcome of the actions on the data. ISO/IEC 13156:2011, 4.5 defines "data integrity" as an "assurance that the data has not been modified from its original form". In practical terms data integrity can mean that the meaning of the data has not been changed. For example, some transforms such as splitting columns can appear as a modification of the data that does not change the meaning.

8.3.4 Chain of custody

For data and its processing to be trusted, all the actions on the data need to be recorded (using trusted mechanisms) from its creation to its final disposition. Actions in the data chain of custody can include creation of digital records or files, copy, transfer, update, transform, analyse, report, archive and delete. For example, transforms may alter the presentation of data (e.g. extracting the Year from a Date field) or create data (e.g. interpolating values to fill in blank fields).

Actions can take place in many places including one or more cloud services, on premises data centres and devices. For trusted processing of multi-sourced data, the chain of custody does not guarantee that a given piece of data has not been altered, but it does mean that any alteration is recorded. For example, if data is anonymized before being used to train a machine learning model the data has been altered by removing certain information, but the remainder of the record is valid. However, to validate downstream results from application of the model, it is important for auditors and other persons or processes in an oversight role to understand that the original data in the training data had been anonymized.

Recording the actions related to the chain of custody can be included in the immutable proof of compliance (see 8.3.5).

For additional information on chain of custody, see ISO/IEC 27050-1.

8.3.5 Security and privacy

As described in ISO/IEC 17789, security is a cross-cutting aspect of cloud services. Cloud services used to process multi-sourced data need security capabilities including access control, confidentiality, integrity and availability. Security requirements for trusted processing of multi-sourced data include authentication, authorization, availability, confidentiality, non-repudiation, identity management, integrity, audit, security monitoring, incident response, and security policy management. This ensures security is a shared responsibility between the cloud service provider, cloud service customers, and cloud service partners. ISO/IEC 27017 and ISO/IEC 27000 can be used to establish security for multi-sourced data processing projects and ISO/IEC 19086-4[4)] can be used to include security in related cloud service agreements.

Processing of multi-sourced data requires a multi-lateral agreement between the cloud service customers and the cloud service provider(s) on the specific security technologies, techniques and standards that will be used during a given project.

Secure multi-party computation (MPC) is an area of research that's beginning to be commercialized that may be useful for processing multi-sourced data. Secure MPC can be used to derive mutual outputs from independent, encrypted data sets where the data rights holders only know what is in their own data.

8.3.6 Immutable proof of compliance

A requirement of trusted cloud processing of multi-sourced data is that proof of compliance with the other elements of trust (data use controls, data provenance, chain of custody) needs to be documented in a tamper resistant way. Immutable means that even users with the highest system privileges cannot change the document. An "append only" scheme is one way to protect records from change while not necessarily preventing the insertion of corrupt or fraudulent records. Distributed ledger technologies are examples of a technology that can be used to protect records from change and to verify the authenticity of new records.

Additionally, access to the proof document needs to be limited to authorized parties. Where required by law, or by agreement, authorized parties may include auditors or other oversight organizations.

Proof of compliance schemes for cloud services needs to scale to millions of transactions per second and can be interoperable between cloud services.

4) To be published. Current stage: 50.20.