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.

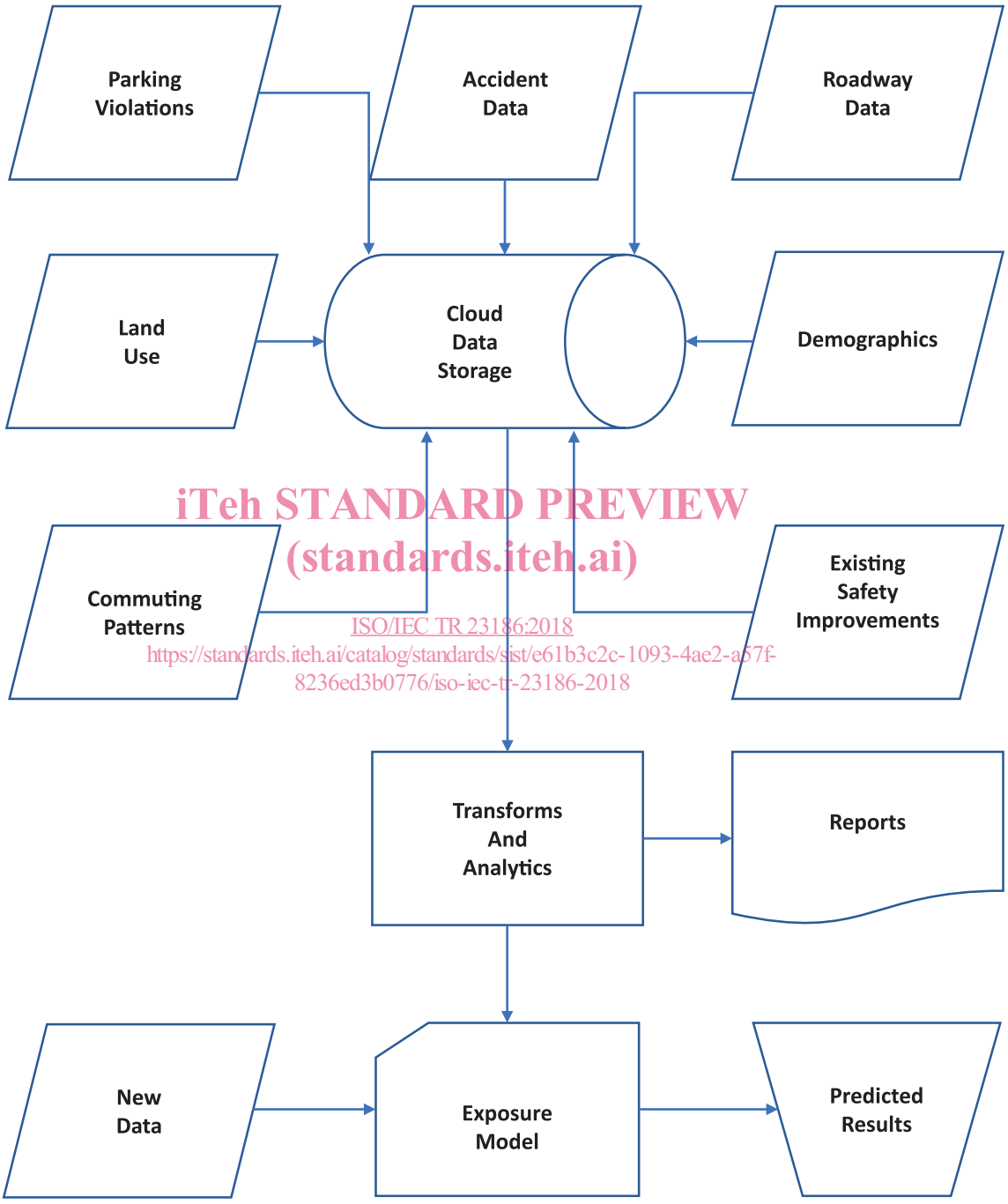

Figure 1 — Example of a system for predicting traffic accident patterns using multi-sourced data

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.