INTERNATIONAL STANDARD

ISO/IEC 23751

First edition
2022-02

Information technology — Cloud computing and distributed platforms — Data sharing agreement (DSA) framework

Reference number
ISO/IEC 23751:2022(E)

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

Published in Switzerland

Contents

Page

Foreword ..... **v**
Introduction ..... **vi**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Symbols and abbreviated terms **3**
5 Overview of DSAs **3**
5.1 General 3
5.2 Data sharing scenarios 4
5.3 Role of the DSA 7
5.4 Trust as a key element in data sharing 7
5.5 Data access and processing rights 7
5.6 Data flow and DSA elements 9
5.7 Relationship between data sharing and data portability 10
5.8 Data sharing agreements (DSAs) in data lifecycles 10
5.9 Data sharing agreements (DSAs) governance 10
6 Dataset description **11**
6.1 General 11
6.2 DLOs and DQOs 11
6.2.1 Title 11
6.2.2 Domain 11
6.2.3 Data dictionary 11
6.2.4 Format 11
6.2.5 Data types 11
6.2.6 Data gathering policy 11
6.2.7 Revision history 11
6.2.8 Pre-existing transforms 11
6.2.9 Date of the dataset 11
6.2.10 Number of instances 12
6.2.11 Summary statistics 12
7 Data use obligations and controls **12**
7.1 General 12
7.2 DLOs and DQOs 13
7.2.1 Regulatory obligations and controls 13
7.2.2 Data holder obligations and controls 13
7.2.3 Allowed data uses 13
7.2.4 Disallowed data uses 14
7.2.5 Allowed uses of the data processing output 14
7.2.6 Disallowed uses of the data processing output 14
7.2.7 Data user obligations and controls 14
8 Data provenance records, quality, and integrity **14**
8.1 Data provenance records 14
8.1.1 General 14
8.1.2 DLOs and DQOs 14
8.2 Data quality 15
8.2.1 General 15
8.2.2 DLOs and DQOs 15
8.3 Integrity 16
8.3.1 General 16
8.3.2 DLOs and DQOs — Dataset integrity 16
9 Chain of custody and transfer of custody **16**

9.1 Chain of custody ... 16
9.1.1 General ... 16
9.1.2 DLOs and DQOs ... 16
9.2 Transfer of custody ... 17
9.2.1 General ... 17
9.2.2 DLOs and DQOs ... 17

10 Security and privacy ... 17
10.1 General ... 17
10.2 DLOs and DQOs ... 18
10.2.1 Data holder security requirements ... 18
10.2.2 Data user security requirements ... 18
10.2.3 Data holder privacy requirements ... 18
10.2.4 Data user privacy requirements ... 18

11 Proof of compliance ... 18
11.1 General ... 18
11.2 DLOs and DQOs — Proof of compliance mechanisms ... 19

Annex A (informative) **Governance in ecosystems ... 20**

Annex B (informative) **Examples of alternatives to bespoke data sharing agreements (DSAs) ... 21**

Annex C (informative) **ISO/IEC standards for identity, privacy, chain of custody, forensics and security ... 22**

Bibliography ... 24

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 or www.iec.ch/members_experts/refdocs).

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

This document was prepared by 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 and www.iec.ch/national-committees.

Introduction

For decades, organizations regarded data and its processing as an expense, necessary to business operations but not an opportunity. What has changed recently is the realization of the value of data and the added value that can potentially be generated by combining datasets. Artificial Intelligence (AI), Big Data, analytics, and cloud computing are making this value proposition much more obvious and the emergence of Internet of Things (IoT) is further driving the economic opportunities around data. Data is the raw material for AI, a key component of the fourth industrial revolution.

Sharing datasets to create combined datasets can have several technical, business, and regulatory challenges. One challenge is the lack of a common language to describe data sharing concepts across the entire data lifecycle and the lack of guidance for developing data sharing agreements (DSAs). This document offers standardized terminology for data sharing along with common building blocks that can be used in the development of DSAs. The aim of the project is to reduce the time and cost required to initiate data sharing projects.

Figure 1 illustrates the structure of this document, representing the Data Sharing Framework as defining both Data Qualitative Objectives (DQOs) and Data Level Objectives (DLOs) over six distinct aspects of data sharing. Each aspect is described in a separate section.

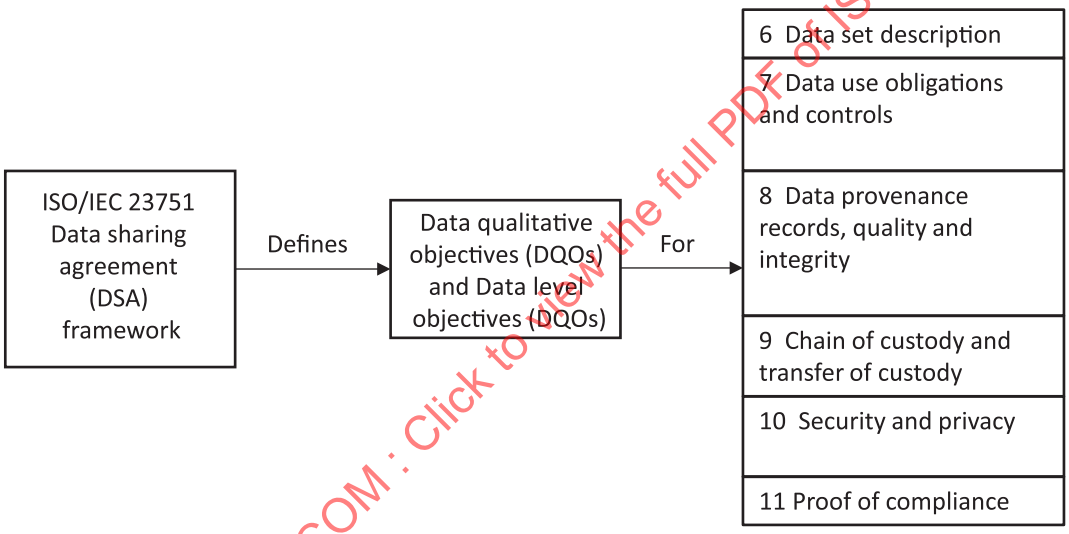


Figure 1 — Structure of this document

Information technology — Cloud computing and distributed platforms — Data sharing agreement (DSA) framework

1 Scope

This document establishes a set of building blocks, i.e. concepts, terms, and definitions, including Data Level Objectives (DLOs) and Data Qualitative Objectives (DQOs), that can be used to create Data Sharing Agreements (DSAs). This document is applicable to DSAs where the data is intended to be processed using one or more cloud services or other distributed platforms.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

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

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

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

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

3.1
party
natural person or legal person, whether or not incorporated, or a group of either

[SOURCE: ISO/IEC 27729:2012,3.1]

3.2
data originator
party (3.1) that created the data and that can have rights

Note 1 to entry: A data originator can be an individual person.

Note 2 to entry: The data originator can be distinct from the natural or legal person(s) mentioned in, described by, or implicitly or explicitly associated with the data. For example, PII can be collected by a data originator that identifies other individuals. Those data subjects (PII Principals) can also have rights, in relation to the data set.

Note 3 to entry: Rights can include the right to publicity, right to display name, right to identity, right to prohibit data use in a way that offends honourable mention.

3.3
data broker
party (3.1) that collects data from one or more sources and sells the data to one or more *data users* (3.5)

Note 1 to entry: In the context of data broker, sell means to provide data in exchange for money or other item of value.

3.4
data holder
party (3.1) that has legal control to authorize *data processing* (3.8) of the data by other parties

Note 1 to entry: A *data originator* (3.2) can be a data holder.

3.5
data user
party (3.1) that is authorized to perform processing of data under the legal control of a *data holder* (3.4)

3.6
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]

3.7
data sharing
access to or processing of the same data by more than one authorized entity

Note 1 to entry: Use of the data can be synchronous or asynchronous.

Note 2 to entry: Data can be shared, for example, (i) by allowing access to, or the execution of operations over, the original dataset, or (ii) by giving a copy of the data to the interested entity.

Note 3 to entry: The way in which data is shared fundamentally influences the available controls and the statements needed in a data sharing agreement.

3.8
data processing
systematic performance of operations upon data

[SOURCE: ISO 2382:2015, 2121276, modified — Notes 1, 2, 3 and 4 to entry were deleted.]

3.9
cloud service agreement
documented agreement between the cloud service provider and cloud service customer that governs the covered service(s)

Note 1 to entry: A cloud service agreement can consist of one or more parts recorded in one or more documents.

[SOURCE: ISO/IEC 19086-1:2016, 3.3]

3.10
data store
persistent repository for digital information

Note 1 to entry: A data store can be accessed by a single entity or shared by multiple entities via a network or other connection.

[SOURCE: ISO/IEC 20924:2018, 3.1.13]

3.11
ratio scale
continuous scale with equal sized scale values and an absolute or natural zero point

[SOURCE: ISO 3534-2:2006, 1.1.9, modified — The EXAMPLE and Note 1 to entry were deleted.]

3.12
data level objective
DLO
commitment a *data holder* (3.4) or a *data user* (3.5) makes for a specific, quantitative characteristic of a dataset, where the value follows the interval scale or *ratio scale* (3.11)

Note 1 to entry: A data level objective commitment can be expressed as a range.

3.13
data qualitative objective
DQO
commitment a *data holder* (3.4) or a *data user* (3.5) makes for a specific, qualitative characteristic of a dataset, where the value follows the nominal scale or ordinal scale

Note 1 to entry: A data qualitative objective can be expressed as an enumerated list.

Note 2 to entry: Qualitative characteristics typically require human interpretation.

Note 3 to entry: The ordinal scale allows for existence/non-existence.

3.14
public domain data
class of data objects over which nobody holds or can hold copyright or other intellectual property

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

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

[SOURCE: ISO/IEC 19944-1:2020, 3.4.4]

4 Symbols and abbreviated terms

AI	Artificial Intelligence
CSC	Cloud Service Customer
CSP	Cloud Service Provider
DLO	Data Level Objective
DSA	Data Sharing Agreement
DQO	Data Qualitative Objective

5 Overview of DSAs

5.1 General

An emerging use of cloud services and other distributed platforms is the processing of data that the CSC has acquired from a data holder. Additionally, there are cases where the CSC processes data acquired from multiple data holders (multi-sourced data) and there are cases where two or more CSCs share data among themselves including data acquired from other data holders.

Advances in cloud data storage have made it possible to create security boundaries around datasets that are then part of a larger logical dataset. Some data repositories provide customized access privileges to data users, with data provenance and chain of custody information attached to each record. These can provide an alternative approach in data sharing scenarios where the data come from multiple, independent data stores

5.2 Data sharing scenarios

A Data Sharing Agreement (DSA) can define how one or more organizations providing data to one or more third parties, several organizations pooling information and making it available to each other or to third parties. This document helps to identify and address important issues when developing DSAs between two or more entities or individuals concerning the sharing of data or information of any kind between these entities or individuals.

DSAs can be used in many different data sharing scenarios. Five representative scenarios are described below.

NOTE The arrows in the figures in this clause indicate data flow.

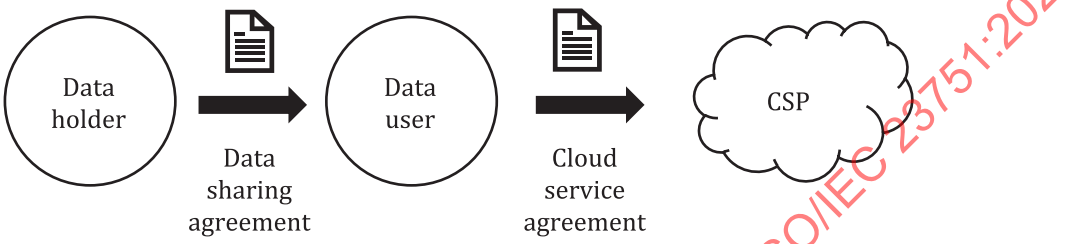


Figure 2 — Data sharing between two parties

Figure 2 shows a basic data sharing arrangement between a single data user and single data holder using a DSA. The CSP is not a party to the DSA but rather the data user is a CSC using cloud services provided by the CSP under a cloud service agreement.

EXAMPLE 1

The financial institution (Data Holder) clarifying to the financial institution bank teller (Data User) the DSA applied. The financial institution bank teller (Data User) can likewise want to understand the cloud service agreement with the CSP.

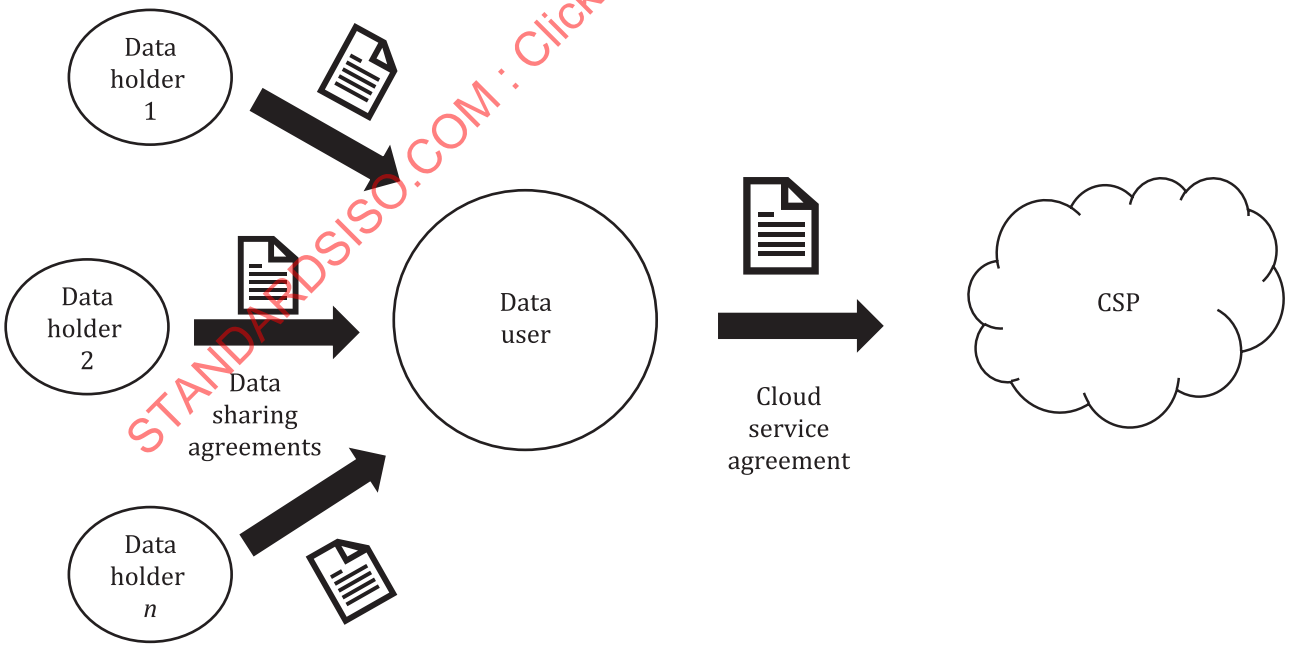


Figure 3 — Data sharing with one data user and multiple data holders

Figure 3 shows a data sharing arrangement between a single data user and multiple data holders. In this scenario the data user has a DSA with each data holder. As with the scenario in Figure 2, the CSC and CSP operate under a cloud service agreement and the CSP is not a party to the DSA.

EXAMPLE 2

An insurance broker (Data User) has a relationship with three insurance companies (Data Holders), with each having unique DSAs. The insurance broker has a single cloud service agreement with their respective CSP.

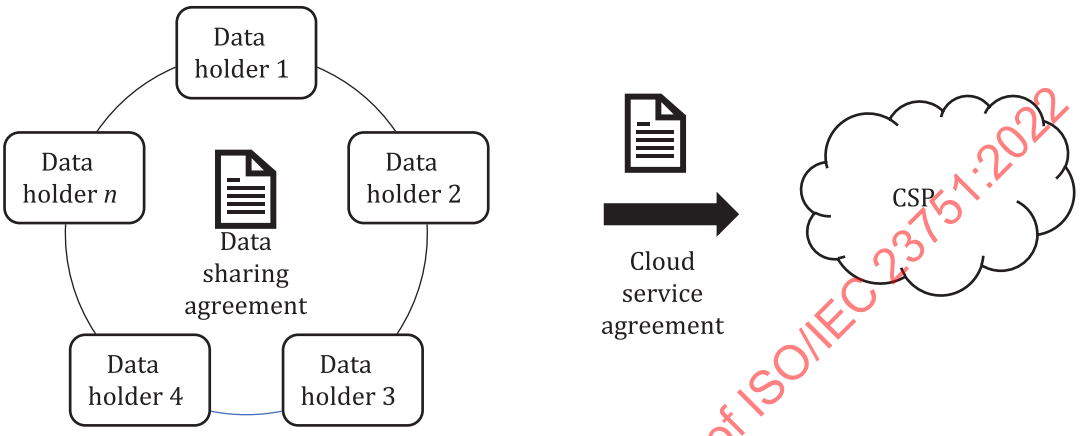


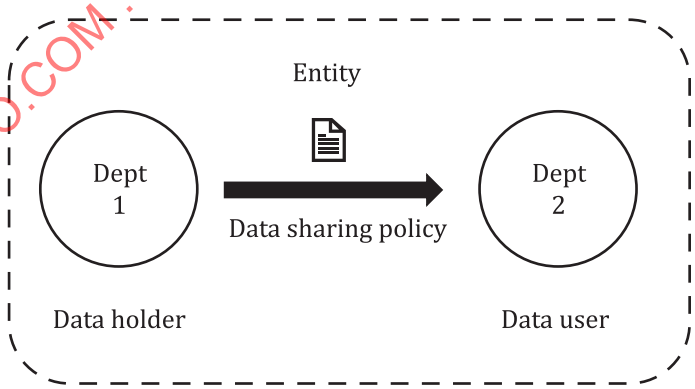
Figure 4 — Data sharing between multiple data holders

Figure 4 shows a scenario where two or more data holders share data under a common DSA and then as a group, they make use of cloud services from a CSP under a cloud service agreement.

EXAMPLE 3

A group of government agencies (Data Holders) have a mutually agreed upon DSA and have a common cloud service agreement with a CSP.

The data sharing scenarios in Figures 3 and 4 include the issues of multi-sourced data which are described in ISO/IEC TR 23186.



NOTE Entity does not include natural persons.

Figure 5 — Data sharing between departments within the same organization

Figure 5 shows data sharing between departments within the same entity where the sharing can be governed by one or more policies rather than by a contractual agreement. In some jurisdictions, it can be necessary to have a signed agreement between the data holder and the data user even if they are within the same entity. For the purposes of this document, data sharing policies can include the same elements of trust as DSAs.

EXAMPLE 4

A single financial institution offers banking and insurance from two distinct lines of business where they need clarity by means of either policies or agreements or both to govern the permitted data sharing from one line of business (Data Holder) to a Data User (such as Customer Relationship Management) in another line of business.

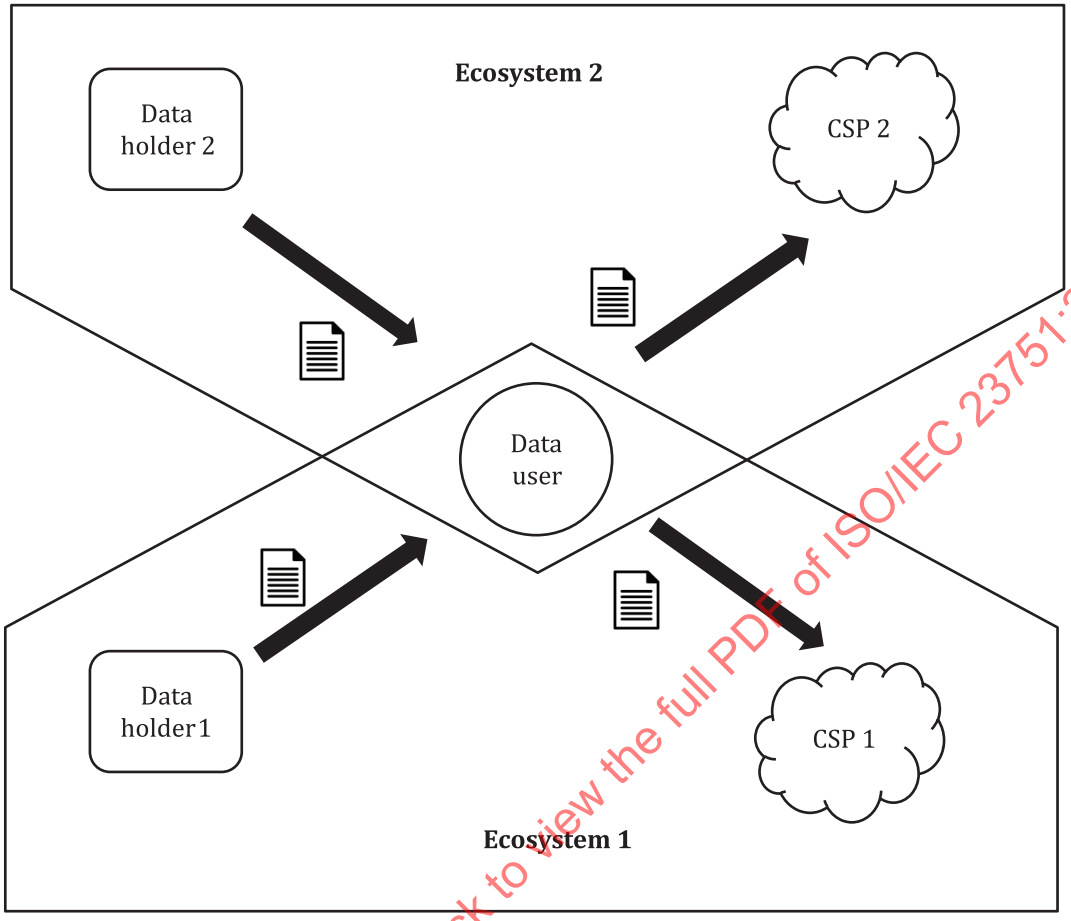


Figure 6 — Data sharing in a multiple ecosystem

As shown in Figure 6, data sharing can include more than one ecosystem. Ecosystems refer to networks of interconnected organisations which share infrastructures and services. Figure 6 displays two ecosystems. The first ecosystem includes one data holder, one data user and one cloud service provider (CSP). The second ecosystem includes the same data user working with another data holder and another CSP. The following observations can be made:

— Some business stakeholders (e.g. a data user) can have to manage DSAs from different ecosystems.

— DSAs used in a given ecosystem often include common elements, for instance.

— The introduction of policies established through a specific ecosystem governance scheme.

— The use of common cybersecurity and protection controls based on shared cybersecurity and privacy risk analysis.

EXAMPLE 5

A health application ecosystem and a financial application ecosystem have separate DSA with the same data user (a data analytics company).

A more comprehensive data sharing ecosystem can contain any combination of the scenarios described, e.g. modelling the data sharing implemented, rolled up for an overall government or corporate perspective.

5.3 Role of the DSA

A DSA helps data holders and data users document the issues and commitments for the processing of shared data and helps to establish trust between the stakeholders. A DSA can be a standalone agreement, or it can be a part of a larger agreement. This document uses the concepts of DLOs and DQOs as the components of a DSA. In addition to DSAs, there are existing agreement models that can be used for data sharing, e.g. the Creative Commons license, public domain, and open data.

NOTE Examples of alternatives to bespoke DSAs can be found in Annex B.

5.4 Trust as a key element in data sharing

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 can include data holders, data users, 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 data users and organizations 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?

— Does the data contain PII or confidential information of any kind?

Other issues concern any regulations and laws that can apply to the data and any commercial terms that apply to the data that can affect the planned use.

For an organization that is the data originator, the major element of trust concerns whether the data is used as authorized. The essential questions can include:

— Does the data user make a clear statement about the intended uses?

— Does the data user sign an agreement in relation to this processing, and 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 data user 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, a major concern can be 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. In many jurisdictions there are, however, legitimate forms of PII processing which are not based on notice and consent. Nonetheless, a major concern relates to any PII breaches that can occur and 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.

5.5 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 data sharing ecosystem.

For example, in some jurisdictions, copyright has a property aspect and a personal aspect, and the collective right to protect that personal aspect is called the moral rights of the author. The data originator holds the moral rights of the author. The moral rights include the right to publish, the right to display the name, and the right to maintain identity. In addition, since it is a right based on personality, it is a personal right that cannot be transferred or inherited. On the other hand, a data holder can have an interest in the copyright that can be transferred to another party.

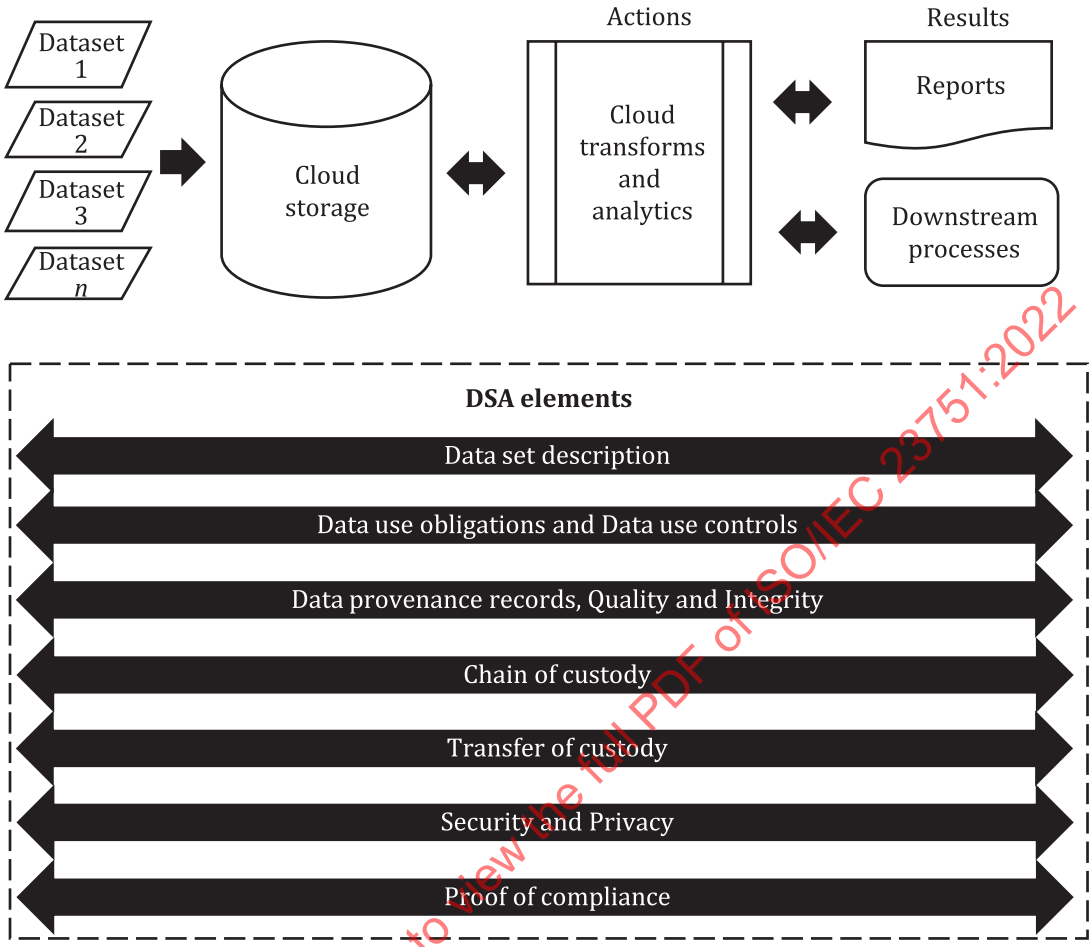
From the discussion of ownership of data, the example of a car ecosystem is a quintessential use case in establishing data access and the rights to view, modify, copy, or process the data. There can be distinct sources of data from the various components, e.g. the component vendor, the car manufacturer, or the owner of the car. Deciding who has rights to the data after it is generated can be a complex issue. An identity and access management (IdAM) system provides the tools to control when and who can access the resource, and thereby it can be used to grant access to appropriate parties.

It can be argued that the traditional legal notion of ownership does not fit with digital data, 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 can be legal rights for parties to access, process and collect data. For example, to ascertain the cause of a collision, police can have an automatic entitlement to all the data in a car after a collision with fatalities. Another example, related to PII, is where the owner or driver of the vehicle can 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 (e.g. real time traffic data or data relating to public transport services).

Another scenario is that local laws can require that the results of processed data be made publicly available, suitably anonymized, if the system has been publicly funded.

5.6 Data flow and DSA elements



NOTE The arrows in Figure 7 indicate data flow between the components.

Figure 7 — Representative data flow and DSA elements

Figure 7 depicts "DSA elements" and a representative data flow for the processing of shared data. The DSA elements can also be informed by sector specific regulations and societal norms. DSAs can also include other elements, such as legal terms and conditions, not included in this document.

NOTE 1 Expanded discussion of each element in Figure 7 is provided in Clause 6 through Clause 12.

The representative data flow shown in Figure 7 is provided only to establish a context for the DSA elements and should not be construed as the only way to architect processing of shared 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.

NOTE 2 For a cloud computing reference architecture, see ISO/IEC 17789. For additional information on cloud computing data flows see ISO/IEC 19944-1. For IoT data flows see ISO/IEC 30141.

Data can come from any number of sources, including data brokers, 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. Shared data can be brought in over a network or by other means such as portable data storage appliances. The data can need to be prepared and cleansed before it is suitable for processing.

Computing resources can be placed at the edge of the network for applications that require real-time processing, e.g. automobiles, medical devices, consumer products and industrial machinery. Data can be 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 or an on-premises data centre. The DSA elements described in this document are important to these scenarios as well.

Several cloud data storage types are suitable for shared data. A store for shared data can be virtual, spanning multiple cloud storage services from multiple CSPs along with storage in legacy data centres.

Figure 7 shows transforms and analytics as the actions being applied to the shared data, but actions can include any other type of data processing actions such as simple queries. As the shared 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 7 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.

5.7 Relationship between data sharing and data portability

ISO/IEC 22123-1 defines data portability as the ability to easily transfer data from one system to another without being required to re-enter data. In many cases there is a relationship between data sharing and data portability in that sharing data can involve transferring a copy of the data from one system to another. The ability to transfer copied data easily, rather than having to re-enter it, is highly desirable from efficiency and cost standpoints.

There are cases where data can be shared without being moved. In this case, authorized access to a data store is granted by a data holder to a data user and the data store can support embedded analytics and other processing.

NOTE For more information on data portability see ISO/IEC 19941.

5.8 Data sharing agreements (DSAs) in data lifecycles

The lifecycle of a dataset can cover the period from inception to the point where the dataset is deleted or even destroyed. Lifecycle processes for datasets can include creation, use. archiving and disposal.

A DSA covering the dataset can be developed at any time during the dataset lifecycle prior to deletion or destruction. Since a dataset can be shared as a copy, there can be multiple, asynchronous lifecycles covering different copies of the dataset.

Data holders, data users and other stakeholders in a DSA should consider whether the requirements in the DSA and the documented lifecycle processes are appropriately aligned.

NOTE See ISO/IEC/IEEE 15288 for more information on system lifecycle processes.

5.9 Data sharing agreements (DSAs) governance

DSAs are established following policies and rules that can be provided through a governance process. This process depends on the data sharing scenario:

— Data sharing with one data user and multiple data holders (Figure 3) can have a governance scheme led by the data user.

— Data sharing between multiple holders (Figure 4) can have a governance scheme agreed upon by the data holders.

— Data sharing between departments within the same organization (Figure 5) follows the governance scheme in their organization.

— Data sharing within an ecosystem (Figure 6) follows a governance scheme in use in the ecosystem.

NOTE Additional information on governance in ecosystems can be found in Annex A.

6 Dataset description

6.1 General

It can be useful for a data holder to provide information about a dataset. This information is intended to assist the data user in determining whether the dataset is appropriate for its purposes. Not all of the attributes listed below need be provided in a dataset description. Some will not be relevant; others will not be precisely known.

NOTE The DLOs and DQOs listed for data quality, described in 8.3, can also be useful in determining if the dataset is fit for the intended purpose.

6.2 DLOs and DQOs

6.2.1 Title

The title of the dataset.

6.2.2 Domain

The domain(s) the dataset applies to.

6.2.3 Data dictionary

List of the data dictionaries available for the dataset.

6.2.4 Format

List of the formats in which the dataset is available.

6.2.5 Data types

List of the data types used in the dataset.

6.2.6 Data gathering policy

List of any policies (e.g. frameworks, standards, regulations, organizational guidelines) employed in the gathering and preparation of the dataset.

6.2.7 Revision history

The revision history of the dataset.

NOTE The revision history can include the data originator name, the data holder name and a time stamp.

6.2.8 Pre-existing transforms

List of any pre-existing transforms on the dataset and any resulting limitations to use of the dataset, e.g. application of differential privacy techniques that can limit the number of future uses of the data.

6.2.9 Date of the dataset

The date the dataset was initiated and the date of the last entry.

6.2.10 Number of instances

The number of instances in the dataset.

6.2.11 Summary statistics

List of any summary statistics available for the dataset.

7 Data use obligations and controls

7.1 General

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 data holder, data user, CSP, or any associated third party. Obligations 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:

— **Jurisdictional regulatory acts**, e.g. the General Data Protection Regulation (Europe), the Basic Act on the Advancement of Public and Private Sector Data Utilization (Japan), Amended Act on the Protection of Personal Information (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 legislation**, e.g. the German eHealth Law, the US Sarbanes-Oxley Act on the accuracy and reliability of corporate disclosures.

— **Governmental mandatory policies and standards**, e.g. Federal Trade Commission (US), Federal Communications Commission (US), BSI Baseline Protection (Germany).

— **Directives, guidelines, and recommendations**, e.g. 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**, e.g. ISO/IEC 27000 or ISO/IEC 19944-1.

— **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 processing shared 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 the user account.

— 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 increasing 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 DSA.

— 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 chain of custody involving multiple data controllers and processors and therefore require an end-to-end approach for definition and implementation. Examples of controls associated with data subject rights include:

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

— 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 CSP 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.

7.2 DLOs and DQOs

7.2.1 Regulatory obligations and controls

List of any known regulatory obligations and controls that apply to the data being shared.

7.2.2 Data holder obligations and controls

List of any data holder obligations and controls that apply to the data being shared.

7.2.3 Allowed data uses

List of any uses of the dataset that are explicitly allowed.

7.2.4 Disallowed data uses

List of any uses of the dataset that are explicitly disallowed.

NOTE See ISO/IEC 19944-1 for more information on data use statements.

7.2.5 Allowed uses of the data processing output

List of any uses of the output of the processed data that are explicitly allowed, e.g. explicitly allowed uses of analytics performed on the data.

7.2.6 Disallowed uses of the data processing output

List of any uses of the output of the processed data that are explicitly disallowed, e.g. explicitly disallowed uses of analytics performed on the data

7.2.7 Data user obligations and controls

List of any data user obligations and controls that apply to the shared data.

8 Data provenance records, quality, and integrity

8.1 Data provenance records

8.1.1 General

For trusted processing of shared data, the origin of each piece of data is needed to establish trust in the results of data analysis and related downstream processes. Origin can be an intermediate or generated dataset equipped with an attestation of quality and data attributes provided by a trusted party. Conclusions drawn from analysis, or decisions made based on analysis rely on traceability of data back to its origin.

ISO 8000-2:2017 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 8000-2:2017 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 shared data as the custodian can manage the collection of shared data, the processing, and the output. The ISO 15836 series and the Dublin Core Metadata Initiative (DCMI) can be useful for creating data provenance records.

8.1.2 DLOs and DQOs

8.1.2.1 Origin

List of the organizations or persons who collected the data.

8.1.2.2 Data holders

The original data holder of the dataset.

The current data holder of the dataset.

List of any prior data holders of the dataset.

8.1.2.3 Changes

List of any updates, transcriptions, abstractions, or validations that have been applied to the shared data.

8.2 Data quality

8.2.1 General

The quality of the output of the actions on shared data will, at best, be no better than the quality of the data itself. ISO 8000-2:2017 defines data quality as the "degree to which a set of inherent characteristics of data fulfils requirements". This means that the participants in a data sharing project need to determine their own characteristics and the requirements. However, ISO/IEC 25012 provides a data quality model and characteristics that can be helpful in determining the data quality requirements to be included in a DSA.

Data integrity is one aspect of data quality and will affect the outcome of the actions on the data. ISO/IEC 19790:2012 defines data integrity as a "property that data has not been modified or deleted in an unauthorised and undetected manner". 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.

The following data quality assertions can enable the data user to determine if the dataset meets their requirements.

In addition to the data quality DLOs and DQOs listed below, refer to the ISO 8000 series of standards for more information.

8.2.2 DLOs and DQOs

8.2.2.1 Accuracy

List of standards, technologies or techniques that were used to ensure the accuracy of the data.

List of range of error or uncertainty known to exist for each numerical field in the dataset.

List of methods used to ensure that the data format specified for each field is adhered to.

8.2.2.2 Distribution

Distribution of the data.

8.2.2.3 Number of duplicate instances

The number of duplicate instances in the dataset (if known).

8.2.2.4 Number of missing entries

The number of missing entries or nulls in the dataset (if known).

8.2.2.5 Outliers

The number of samples for each numerical field in the dataset considered to be outliers.

8.2.2.6 Range

The range for each numerical field found in the dataset.

8.2.2.7 Imputed data

List of fields in the dataset that include imputed data.

List of methods used to impute data in the dataset.

8.2.2.8 Synthesized data

List of fields in the dataset that include synthesized data.

List of methods used to synthesize data.

8.3 Integrity

8.3.1 General

In the event the contents of a data set have been altered, intentionally or inadvertently, the results of further processing or analysis can be incorrect. A DSA can include a provision covering any methods used to ensure the integrity of the dataset.

8.3.2 DLOs and DQOs — Dataset integrity

List of methods used to ensure that the meaning of the data in the dataset has not been altered.

9 Chain of custody and transfer of custody

9.1 Chain of custody

9.1.1 General

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 can 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 occur in many places including one or more cloud services, on premises data centres and devices. For trusted processing of shared 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 has been anonymized in the training data.

Recording the actions related to the chain of custody can be included in the proof of compliance (see Clause 10).

For additional information on chain of custody, see the ISO/IEC 27050 series.

9.1.2 DLOs and DQOs

9.1.2.1 Data ingestion

How the data holder will transfer the shared data to the data user.

How the data user will transfer the shared data into its data storage system.

9.1.2.2 Data storage

List of the location(s) and type(s) of storage the data user will use.

9.1.2.3 Data transforms

List of data transforms that the data holder has applied to the data.

List of transforms the data user plans to apply to the data.

9.1.2.4 Combining with other data

List of any data the data user plans to combine with the shared data.

9.1.2.5 Other types of data processing

List of other types of processing the data user plans to apply to the shared data.

9.2 Transfer of custody

9.2.1 General

For the purposes of this document, transfer of custody refers to any case where the data user(s) plan to make the shared dataset available to other data users. Transfer of custody can be subject to any restrictions the data holder puts on such transfers.

9.2.2 DLOs and DQOs

9.2.2.1 Requirements or restrictions for data holder

List of data holder requirements or restrictions on the data user's ability to transfer custody of the shared data.

9.2.2.2 Transfers of custody

List of transfers of custody the data user plans to apply to the shared data.

9.2.2.3 Purpose of transfers

The purpose of any transfers of custody the data user plans to apply to the data.

9.2.2.4 Data user plans to transfer.

List of data users the data user plans to transfer the shared dataset to.

9.2.2.5 Terms and conditions for data transfer

List of the terms and conditions that will apply to any transfers of custody of the shared dataset.

10 Security and privacy

10.1 General

As described in ISO/IEC 17789, security is a cross-cutting aspect of cloud services. Cloud services used to process shared data need security capabilities including access control, confidentiality, integrity, and availability. Security aspects for trusted processing of shared data can include authentication,

access authorization, availability, confidentiality, non-repudiation, identity management, integrity, audit, security monitoring, incident response, and security policy management. Since multiple people from multiple entities can be involved in a data sharing project, it can be necessary to provide for fine-grained access authorization where access is defined per person rather than by role or some other more abstract criteria. (For more information on access authorization controls see ISO/IEC 27002) This ensures security is a shared responsibility between the CSP and CSCs. ISO/IEC 27000 or ISO/IEC 27017 can be used to establish security for shared data processing projects and ISO/IEC 19086-4 can be used to include security in related cloud service agreements.

A DSA should include the specific security technologies, techniques and standards that will be used during a given project.

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

Differential privacy is also a possible approach that can be used to protect privacy when sharing data.

The OECD Privacy Guidelines [26] provides an internationally agreed upon set of fundamental privacy principles for the free flow including data sharing of personal data across borders.

NOTE A list of standards for security and privacy can be found in Annex C.

10.2 DLOs and DQOs

10.2.1 Data holder security requirements

List of security requirements the data user shall implement to protect the dataset.

10.2.2 Data user security requirements

List of security requirements the data holder intends to implement to protect the data.

NOTE Data user security requirements can meet or exceed the data holder security requirements, if any.

10.2.3 Data holder privacy requirements

List of privacy requirements the data user shall implement to protect the privacy of data subjects covered by the dataset.

10.2.4 Data user privacy requirements

List of privacy requirements the data holder intends to implement to protect data subjects covered by the dataset.

NOTE Data user privacy requirements can meet or exceed the data holder privacy requirements, if any.

11 Proof of compliance

11.1 General

Another aspect of shared 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. 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 can include auditors or other oversight organizations.

11.2 DLOs and DQOs — Proof of compliance mechanisms

List of mechanisms the data user will employ to provide proof of compliance with the terms of the DSA.

Annex A
(informative)

Governance in ecosystems

ISO/IEC TS 27570 describes several processes, one of which is the data sharing process. This process includes two parts, the ecosystem process which establishes coordination schemes, including measures for compliance, assurance and audit of practice, and the organization process which applies the recommendations. Possible recommendations are:

— integration of common information on privacy impact assessment in DSA (e.g. following ISO/IEC 29134).

— requirements on consent information in DSA (e.g. following ISO/IEC 29184).

— requirements on privacy preference information in DSA (e.g. following ISO/IEC 27556[1)]).

1) Under preparation. Stage at the time of publication: ISO/IEC DIS 27556.

Annex B
(informative)

Examples of alternatives to bespoke data sharing agreements (DSAs)

B.1 General

While bespoke DSAs can be negotiated and written according to the specific needs of the data holder and data user, there are alternatives. Examples of alternatives to bespoke DSAs include, but are not limited to, the Creative Commons™ [2] Licenses, datasets placed in the Public Domain and datasets declared as Open Data.

B.2 Creative Commons™ license[7]

The Creative Commons™ license can be used on datasets under the following license types:

— Attribution: Allows use of the work if attribution is given to the licensor.

— ShareAlike: Allows the work to be copied, distributed, displayed, performed, and modified if any modified work is available under the same terms and conditions as the original work. Any changes to the terms and conditions requires permission from the licensor.

— NonCommercial: Allows the work to be copied, distributed, displayed, performed and modified and otherwise used for any purpose other than commercially unless the licensor provides permission to do so. The licensor can also apply the NoDerivatives restriction to the work.

— NoDerivatives: Allows only original copies of the work to be copied, distributed, displayed, and performed unless the licensor grants additional rights to make modifications.

Terms common to all the Creative Commons™ licenses include attribution, no additional restrictions (legal or technological) and an exemption for any part of the work that is in the public domain.

B.3 Public domain

Data can be available that has been placed in the public domain. Data may be placed in the public domain by statute or by a declaration by the originator. Material in the public domain can ordinarily be used without any restrictions.

B.4 Open data

Open data is a mechanism often used by governments and institutions to make data that was created with the use of public funds publicly available. Open data can ordinarily be used with few if any restrictions. An emerging view of open data is “open as possible, closed as necessary”.

2) Creative Commons ™ is an example of a suitable product available commercially. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO or IEC of this product.