International Standard

ISO/IEC 5212

First edition
2024-04

Information technology — Data usage — Guidance for data usage

Technologies de l'information — Utilisation des données — Recommandations pour l'utilisation des données

Reference number
ISO/IEC 5212:2024(en)

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

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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 Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 32, *Data management and interchange*.

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

This document is a high level, principles-based advisory International Standard. It sets out a framework of two elements with the relevant concepts, that can be referenced by organizations, persons and systems that use data. The framework and concepts outlined in this standard should be read in conjunction with the terms and definitions contained in ISO/IEC 5207.

Organizations of all types (including commercial enterprises, government agencies, not-for-profit organizations), sizes and purposes depend on the use of data for day-to-day business operations and are increasingly reliant on data dependent systems such as information technology management, cloud computing, big data, Internet of Things, and artificial intelligence.

There are numerous approaches to data usage, from the most complex which includes highly sensitive, personal or confidential information to the least sophisticated data capture systems. Within each data usage scenario, there are different approaches to system integrity, data quality, data user capabilities, and organizational governance.

This document has been prepared using a principles-based approach to encourage organizations to implement data governance to manage risks at each stage of data use, exchange or sharing. This approach supports organizations seeking greater value, knowledge and insights from data while providing a framework for data users. As data are essential to a broad range of roles within an organization, it is imperative that all users have a fundamental understanding of data use to ensure appropriate data management. There is a risk that as the use of data is ubiquitous within organizations, users without appropriate knowledge, context and expertise can inadvertently lead to incorrect data usage.

The sharing or exchange of data can involve multiple individuals, systems or organizations with different processes and procedures. Furthermore, each entity involved in the sharing or exchange of data can have different approaches to security, privacy, data sensitivity or legal considerations. While data usage activities can be managed under different governance arrangements such as a formal contract or data sharing agreement, there are many steps involved in data usage that may not be formalized. This document uses a risk identification and management methodology which can be considered by any data user, be they an individual or organization. There can be an advantage for organizations operating with existing data or technology governance processes, such as those outlined in International Standards related to the governance of information technology, such as ISO/IEC 38500 or ISO/IEC 38505-1.

In addition, organizations can consider the suitability of IT systems, security and storage requirements to support governance capabilities which are addressed within ISO/IEC 27001, ISO/IEC 27701 and ISO/IEC 27040.

Information technology — Data usage — Guidance for data usage

1 Scope

This document provides high-level guidance to data users, whether organizations or individuals, to assist in realizing the benefits from data usage while managing risks.

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 5207, *Information technology — Data usage — Terminology and use cases*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 5207 apply.

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

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

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

4 Abbreviated terms

DLO	data level objective
DQO	data quality objective
DSA	data sharing agreement
IoT	internet of things
PII	personally identifiable information
SLA	service level agreement
SLO	service level objective

5 Introduction to data usage

5.1 Overview

5.1.1 The context of data usage

Data usage is any activity involving data. This includes the use, sharing and exchange of data that can occur across entities of all types, sizes, and purposes. The decision-making process around data usage requires the identification of:

— the decision to use, exchange or share data;

— the purpose for data use, exchange or sharing;

— details about the data itself including its characteristics, quality, security, and privacy;

— pathways to use, share and exchange data and the alternatives;

— acceptable risks in using, sharing, or exchanging data;

— mitigation measures for risks;

— authorization steps required;

— the policies, processes, procedures, or instruments required to provide predictability and reliability around data usage activities.

Identifying these characteristics can be complex particularly when data use is ubiquitous in an organization, or when there are multiple parties, or informal data sharing arrangements. This document proposes two perspectives to assist organizations to identify and mitigate data project risks and opportunities being:

a) the **data process** within the organization or between or among organizations or entities when sharing or exchanging data, using the data lifecycle as a framework;

b) the **data environment** to assess the surroundings or conditions which can be consequential.

These two elements are the core components in developing a *data usage framework* which provides the structure for organizations to understand the characteristics of any entity in possession of the data. Each component within the data usage framework should be captured within the *metadata* description. Understanding the data process, the data environment and developing a data usage framework can assist organizations with benefits including:

— identifying risks to the data management process and opportunities for corrective actions;

— identifying opportunities to standardize identical *data processes*;

— increasing value capture from data through improved usage practices;

— ensuring that all entities have a common understanding of the data through well documented metadata.

5.1.2 Data process

The *data process* can be assessed using the *data lifecycle* as a framework to identify the steps involved in *data usage*. This can assist organizations in better understanding their *data processes* and identify areas of greater risk or opportunity. There are many iterations of the *data lifecycle* and organizations can find it useful to develop a data lifecycle map for each data project to identify data usage activities and related changes to the data characteristics. See Figure 1 as an example of a data lifecycle and the maturity of data at each stage.

Figure 1 — Data lifecycle

For more information on data lifecycle frameworks for projects related to the development and use of AI systems, see ISO/IEC 8183.

Organizations can apply a multi-layered *data lifecycle* where steps within the process occur within other *entities* to recognize data movement beyond organizational boundaries. Data management policies and procedures should consider how these apply to other entities to maintain a consistent and shared

understanding of data management processes and where these occur, as shown in the example in Figure 2. For example, organizations that operate within a specified jurisdiction with data required to be retained, stored and used within the same location. The addition of *cloud computing* in a different jurisdiction may require specification within *cloud service* (or storage) *agreements* to address jurisdiction-based operational expectations.

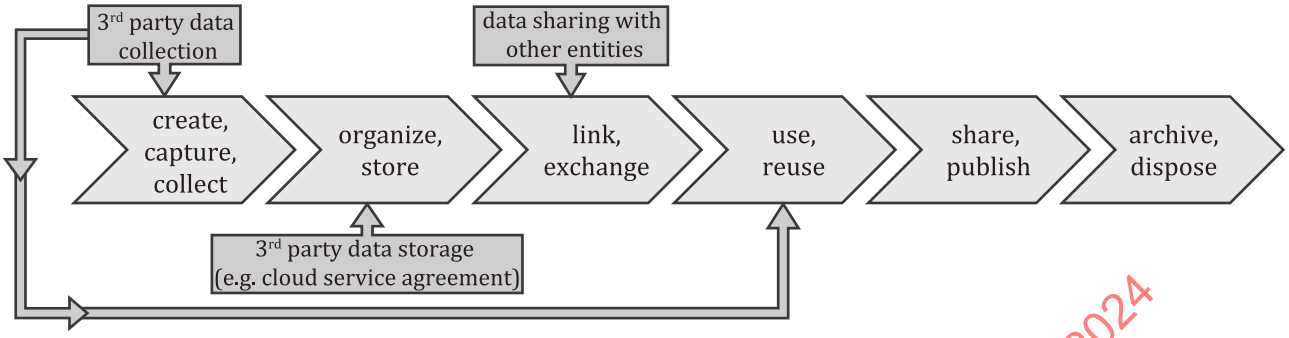

Figure 2 — Third parties and the data lifecycle

Each of the steps in the data life cycle can be considered as distinct elements however, this does not recognize the interconnected nature of data, or the ability of data to persist in perpetuity. Therefore, decisions at each stage of the data life cycle should consider the preceding and following stages, while risk mitigation measures need to extend beyond the immediate stage to explore potential risks carried forward. This is important for any action that changes the data including:

— expanding or refining the data set through the ongoing acquisition of data;

— data preparation including cleaning, decryption, transformation, data translation, representation, etc.;

— data analysis;

— exchange of data via other entities;

— sharing data;

— combining data.

The ubiquitous nature of data collection has resulted in large data sets and big data in parallel with increasing levels of data complexity. Data subjects can be far removed from the data users, who themselves can operate in disassociated and distinct parts of the data life cycle. Risk mitigation measures may require ensuring that each data user is aware of the risks preceding and following their immediate area of data use, in addition to understanding the overall data project objective including related privacy, security, and sensitivity and other fitness-for-purpose considerations. These risks need to be managed effectively to avoid high compliance costs, data breaches, non-compliance, or reputational damage. Without a framework to consider these issues, organizations can avoid sharing or exchanging data, failing to realize the value derived from data driven insights and knowledge. Preparing data users and systems appropriately is part of managing the data environment.

5.1.3 Data environment

The data environment relates to the surroundings or conditions that can influence the data usage outcomes. The *data environment* is important in considering the decision-making process related to *data usage*. The data environment can include but is not limited to the components outlined in Figure 3.

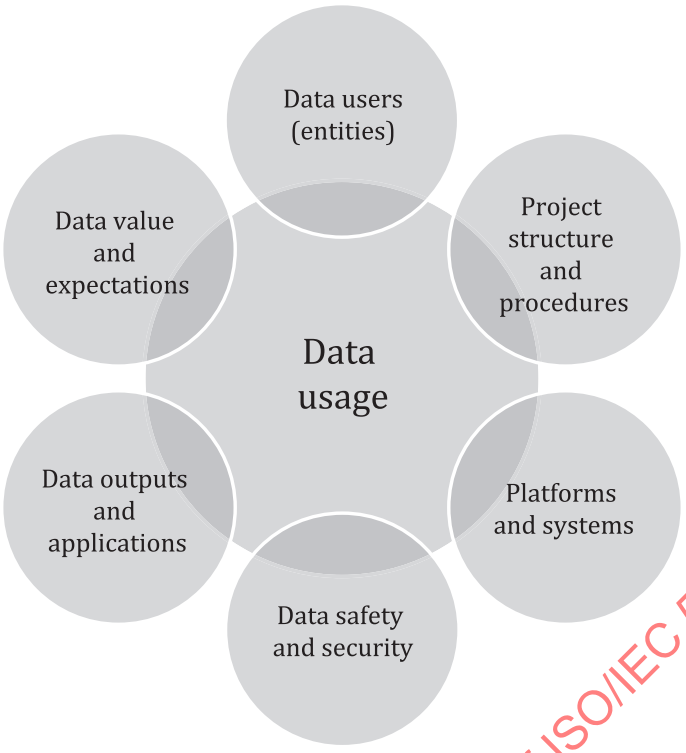

Figure 3 — The data environment

Organizations should consider each of these in the context of whether each component is:

— competent;

— structured and resourced;

— defined, understood, and agreed among all parties;

— enabled with appropriate issues awareness and access to remediation avenues;

— appropriately considered with regards to responsible use, misuse, and corruption;

— documented under an appropriate governance framework.

For example, the people or data users operating in the data environment can affect the data quality, safety, security or value of the data through their actions. Therefore, organizations should ensure that data users are appropriately trained and have a common understanding of the data usage task.

Organizations who are sharing or exchanging data should consider the data environment related to each entity and whether there are material differences that can affect the data project outcomes, risks or opportunities. This can include documenting:

— the entities or organizations holding or using the data;

— the people, users, or systems within each entity;

— the data project including objective, outcomes and expectations;

— the data management and security processes;

— the data outputs.

Each of these components presents a decision-making situation for organizations which can warrant an assessment of the risks and opportunities within each one. This can be important when entities, which are fundamentally different, share data. Each entity should consider the comparability in each component

particularly with regards to risk appetite, governance processes, internal and external accountability and data project expectations.

By considering both the *data process* and the *data environment*, organizations can identify where there are potential vulnerabilities across all data usage activities.

To support responsible data usage, entities should consider the existing data environment including legacy data, governance, competency and processes and undertake any necessary steps in preparing for data usage.

5.2 Preparing for data usage

Good data governance is important in data sharing, exchange and use. Organizations with good data governance processes can have an advantage in data project management, enabling greater value capture and effectively mitigating risks. Table 1 provides an overview of the data usage environment and the aspects that organizations should consider prior to undertaking data use, sharing and exchange.

Table 1 — Preparing for data usage

Pre data use, sharing and exchange preparation	Data use, sharing and exchange	Post data sharing, exchange, and use considerations
Organizational capability — policies — procedures	**Recorded information about the data** — data catalogue and metadata	Archived data — security
Defined data context for users	**Data transmission or receiving**	Deleted data
Data management system	**Data systems operation**	
Data user capability	**Data accessibility, security and privacy**	
Data acquisition — data capture — primary data — secondary data — externally acquired data	**Purpose of data use, sharing or exchange** — outcomes — risks — sensitivities — privacy concerns — compliance considerations — data subjects and stakeholders — data user processes	
Data preparation including: — representation — decryption — de-identification — cleaning — normalizing — transforming — labelling — enrichment — integrity verification — updating metadata	**Data use including:** — access — observation — interpretation — analysis — event detection — event detection response — consolidation, collation — recording — creating data products	

Table 1 *(continued)*

Pre data use, sharing and exchange preparation	Data use, sharing and exchange	Post data sharing, exchange, and use considerations
Data storage — access platforms — safety and security provisions — business continuity plans	**Data exchange** — storage — access — transferring — archiving	
Other data entities and parties	**Data sharing** — specific access — group access — public access — open access, publication	
	Data use conclusion — selective archiving, data removal — deleting — destroying	

5.3 Applications for this document

This document provides a framework for organizations to support the development of data use systems, tailored to their specific circumstances. It provides guidance for any data user undertaking their own risk assessment using the data life cycle as a framework. This document involves both data management and data governance for all data users. This can include data sharing or exchange between multiple platforms or entities, or the sharing of data across public domains.

This document also provides guidance to those advising, informing or assisting governing bodies which can include:

— public officers;

— board members and senior executives;

— members of data management teams within an organization;

— advisory boards;

— external business or technical specialists, retail or industrial associations or professional bodies;

— internal and external service providers (including consultants);

— auditors.

The guidance set out in this document is generic and intended to be applicable to all entities regardless of type, size or nature. This document provides guidance to organizations and individuals who use, analyse, or share data through a framework for data usage.

Organizations seeking additional guidance on data and information management systems can reference ISO/IEC 38500 and ISO/IEC 38505-1.

6 Preparing the data environment for use, sharing and exchange of data

6.1 Overview

In preparing to use, share or exchange data, organizations should consider the *data environment* within the organization. Organizations should prepare for *data usage* by developing the capability of people within the organization as well as establishing documented and reliable data procedures which operate under a robust data governance framework. This clause provides guidance for organizations to support a strong foundation for proficient data use, sharing and exchange. The basis for both the data environment, procedures and governance is to examine and prepare the organization's competence in terms of people, systems and data preparation.

6.2 Overview of organizational readiness and capability

6.2.1 General

Organizations increasingly require access to vast quantities of high-quality data to support digital transformation activities, increase operational efficiency and develop enhanced knowledge and insights. To support data related activities all entities should consider their organizational readiness, operational and system capabilities, governance, and general data competency. All data users within an organization should have a fundamental level of understanding of the data types, quality, use, and processes. Key to this is operating under a sound governance framework that supports the appropriate competence and measures regarding security, privacy and data sensitivity.

6.2.2 Understanding the organization and the data context

The organization should determine the intended use of data including:

a) the purposes for which data are used;

b) the parties involved in the *data process* including *data subjects* (stakeholders and beneficiaries);

c) data attributes such as data volume, data format, data categories, etc;

d) the alignment of interests and benefits arising from data usage

e) the sensitivity of data sets;

f) data privacy or security requirements that can be applicable to other entities, such as a different department within the organization or entity, functional units (for example from accounting to legal), or to external entities;

g) the property right of data such as data ownership;

h) the data users;

i) data systems including operating platforms, electronic data interchange (EDI), machine learning or artificial intelligence systems;

j) the data stakeholders including, but not limited to data subjects;

k) policies and procedures relevant to data usage by the entity;

l) legal, regulatory, and contractual obligations that relate to the entity's data usage.

6.2.3 Identifying uses of data

Data can be used in many ways by an entity. This document provides guidance to assist organizations in understanding the full breadth of data usage, beyond what has been previously considered within specific data processes. Data usage can include several data related activities that have not been subject to a risk-

based analysis to quantify the potential risk associated with each action. For example, the following are examples of data usage:

— for analysis: tallying, visualising, describing, diagnosing, showing relationships, predicting or modelling;

— for event detection: monitoring, alerting;

— to trigger actions: based on thresholds or as a consequence of event detection;

— for historical record: observing, recording and archiving;

— storage: files for photos, videos, programs, other recordings, presentations or spreadsheets;

— to create data products: copies, aggregates, subsets, perturbed versions, insights or outputs in other formats (e.g. printing of digital documents;)

— in transmission: broadcasting, transferring or connecting;

— deletion: destruction or rendering data inaccessible to the current entity.

Organizations should consider whether existing data practices and data sets are consistent with current governance expectations. For example, data collected prior to relevant personal privacy protections cannot be considered complaint, particularly if mass collection methods such as scraping have been used. These legacy processes and systems can be problematic when data are used, shared and exchanged as, combined with other data, data subjects can be identified. Organizations should consider whether reusing data warrants additional risk management measures particularly if the insights from the data project are to be published.

6.2.4 Data strategy within the organization

To effectively use, exchange and share data, organizations should develop a coherent and company wide data usage strategy. Central to this is the development of data capability across the organization, beyond a delegated role or area of responsibility such as a responsible data officer. The data strategy should encompass protective capabilities to understand sensitivities and opportunistic capabilities to capture value.

6.2.5 Data policy

Entities that use data should establish a data usage policy that:

— is appropriate for the entity;

— references data usage processes and procedures;

— recognizes the relevant jurisdictional and other legal obligations applicable to the entity;

— encourages the effective, safe, and responsible use of data;

— considers the ethical aspects of data usage for stakeholders and particularly data subjects;

— recognizes the value of data;

— is accessible to all actual or potential data users.

6.2.6 Leadership and commitment

Data processes, procedures and opportunities should be documented and understood by all data users in the organization. The effective, safe, and responsible use of data is a principle that should be upheld at all levels of the organization, beginning with the board or group with external accountability for strategic oversight. Data issues should be managed by leadership with relevant issues reported at board level or equivalent.

6.2.7 Organizational roles and responsibilities

The organization should ensure that all data user roles within the entity follow:

— the appropriate qualifications and experience level;

— the organization's data usage policy;

— this document.

6.2.8 Competence

The organization should:

— determine the necessary competence of individuals acting as data users;

— ensure that the required competence is specified within terms of engagement such as contracts (including but not limited to employment contracts);

— ensure competence requirements are included as appropriate in induction;

— implement a framework that identifies data sensitivity and security levels in association with the relevant processes and users;

— ensure that data users who have access to all levels of data (for example IT personnel or data scientists) are appropriately qualified, inducted and appraised of any relevant data sensitivity or security considerations;

— undertake periodic assessments as required to ensure the appropriate level of competence is maintained;

— operate a system to maintain competency in response to changes in the data usage environment.

6.2.9 Awareness

The organization should ensure that individuals, either employees or third parties responsible for data, are aware of and have the necessary skills and knowledge to action:

— the organization's data policies and processes;

— the requirements, including competence, for data users;

— governance requirements related to data exchange and sharing;

— the implications of non-compliance with the organizations data policy.

6.2.10 Tools and resources

The organization should ensure the appropriate resources are made available for data usage including:

— platforms that enable data sharing and data exchange;

— data system security provisions as required by the data policy and jurisdictional or legislative requirements.

The tools and resources used should be reliable with the appropriate means of continuity of service, redundancy and back-up to ensure that they do not compromise the level of data integrity, completeness and reliability.

6.2.11 Data sharing and exchange protocols

The organization should develop explicit protocols for data sharing and exchange with all sharing, and exchange activities operating under a framework that provides clarity around:

— Operational authority for data users;

— Approved platforms and systems for data exchange;

— Data representation and quality expectations for each data sharing or exchange project;

— Approval frameworks for external data exchange or sharing entities.

6.2.12 Data documentation

The organization should develop a data documentation system to ensure the way data are catalogued is consistent across the organization. A shared understanding of all data is vital to ensure data users operate in a consistent and predicable manner. In most organizations this involves the development of a data catalogue at the simplest level or a metadata registry to appropriately structure data about all data. Metadata registries are vital to tracking all changes related to data and can be useful in the event of a data breach or incident.

6.3 Data systems

The data systems applicable to a data project can be continuous or connected directly or indirectly. Organizations should consider all systems involved in data exchange (access, transmission, storage and archive) to ensure data integrity and provenance are not compromised, including where the data exchange takes place beyond the organization. This can include systems that operate actively undertaking analysis or exchanging data or passively by providing data storage and can include:

— cloud computing;

— electronic data interchange (EDI);

— third party processing.

Data systems should enable all data exchange to be recorded in a transaction log, data catalogue or metadata registry.

6.4 Data modelling and data design

An important step in understanding data is to model the data and the relationships among individual data elements. Such modelling also aids in the design of databases and other data stores used to store the data. The corresponding metadata should be captured during this process. In many data processing systems, the data modelling and data design processes precede the creation or acquisition of the corresponding data. However, in more exploratory data projects it is possible for data to be acquired first, then analysed and modelled.

6.5 Data acquisition

Data acquisition is the process of collecting data. This can include the digitisation of data provided via manually completed forms, through to technically sophisticated technologies such as simulations, AI system outputs, sensor networks etc. Data acquisition falls into four main categories:

— existing data within the entity;

— data captured via data collection mechanisms by the entity;

— data acquired from another entity;

— data used for a different purpose than originally specified.

In all situations, data acquisition should enable:

— mechanisms to organize and structure data or datasets;

— recording associated metadata;

— the ability to validate aspects of the data including security, quality, and privacy;

— if applicable conversion or translation tools involved;

— data provenance.

For data acquisition, the entity should consider:

— what data should be acquired;

— the reliability of data received;

— the means of acquisition;

— the metadata of the data;

— the volume of data received;

— the data representation of the data received;

— the data representation to be used for storage;

— any data transformation or data translation required.

Entities should also consider older or legacy data in the organization and whether this is suitable for current applications or compliant with contemporary privacy, security and sensitivity expectations. Additionally, the origins of data sets acquired externally (from other entities within the organization, or from external organizations) should be appropriately assessed and documented to understand:

— the data characteristics and metadata;

— conformance with the entities own data systems and standards;

— the privacy, security, authority, sensitivity and ethical implications of the data set.

6.6 Data storage

Computer data storage or information storage refers to storage devices and storage media that retain electronically stored data or information.

Data should be stored appropriately to maintain a consistent and dependable level of security. The data storage procedures should include access, security and privacy considerations of both data users and the data itself. Metadata should be stored securely and be accessible independent of its related data sets.

Data storage systems should comply with privacy and security requirements. Organizations should consider and document security and redundancy needs including the need for non-volatile storage for both internal requirements and, where required, any external parties involved in data sharing or exchange. Organizations should ensure their data storage systems consider data transfer where data users can take data offsite through both authorized (documented) and unauthorized (explicitly prohibited) activities.

6.7 Data preparation

Prior to the sharing, exchange or use of data, data can require preparation including but not limited to:

— decryption;

— cleaning: including steps such as validating relevance, deduplicating, outlier removal and bias mitigation;

— normalizing data values: modifying data so that, for instance, it fits within a predefined range; the organization of the data remains unchanged;

— normalizing data structures;

— data translation: converting data into standardized code sets;

— transforming: reorganizing the data without changing its values;

— adding labels that describe the data so that the data can be used as training data for an AI system. For example, important components of an initial data set of video clips can be labelled in order to create training data for building a visual recognition AI system;

— enrichment: running tools to link diverse data sources and to add additional context to the data. For example, unstructured data can be processed by natural language processing (NLP) tools to extract named entities. Placenames and addresses can be identified and geocoded using a gazetteer to enable location-based analysis later;

— integrity verification: applying a process specific to the kind of data to check the overall integrity of a data set. This is more likely to be applicable to structured and semi-structured data, for which there can already be a structural model (e.g. database schema, formal ontology);

— updating the provenance record of each data set to record changes and operations undertaken.

Organizations should have a process for managing de-identified data to consider:

— de-identification measures including analytical, mathematical or process means;

— the sensitivity of de-identified data;

— the risk of re-identification, particularly in the event of combining, sharing, or expanding data sets;

— measures to manage personal information and personally identifiable information (PII).

These measures should be considered independently as well as how they can be impacted during data preparation processes.

This topic is covered in detail in ISO/IEC 27559.

7 Guidance for the use, sharing and exchange of data

7.1 Overview

Data use, sharing, and exchange present both opportunities and risks to organizations. Increasingly, decisions in one organization or sector are predicated on data generated in another organization or sector. Data usage involves numerous regulatory and governance issues, which should be documented and understood by all actual or potential data users. Documentation should encompass processes within an organization as well as expectations applicable to other entities where data are shared or exchanged. It is advisable that organizations follow the guidance outlined in Clause 6 to address the preconditions for data use, sharing and exchange.

As outlined in Clause 5, the management of data use, sharing and exchange can be supported by understanding the *data process* and the *data environment*. By quantifying the risks and opportunities in the data process and the data environment, organizations can maintain a robust system of data governance that allows for additional control measures to be included in proportion to increasing complexity or sensitivity.

7.2 Data risk management

7.2.1 General

Risk management is a key concept in avoiding unexpected consequences in data projects. A risk matrix is a useful tool to apply to the *data process* and the *data environment*. to support the decision-making process for the use, sharing, or exchange of data. This guidance has been structured as follows:

a) applying a risk assessment framework to determine the likelihood, consequences and mitigating measures to ensure the appropriate use of data;

b) using a risk management approach to assess the data environment to support the use of consistent and documented processes so all parties and entities share a common understanding of the data and its characteristics;

c) understanding the risks within the data process using the data life cycle to examine risks at each stage of data use within the organization.

Entities should consider the guidance in Table 2 that provides a risk matrix that considers the *likelihood* of an issue arising and the *consequences* that can stem from this issue.

Table 2 — Risk assessment matrix

Likelihood	Risk rating (consequence)			
	Risk to parties and data subjects	**Risk to organization**	**Financial risk or loss**	**Data integrity risk**
Low	No impact	No impact	No loss	No impact
Minor	Moderate	Moderate	Lost time minimal expense	Minor lost or compromised data
Moderate	Security and identification risk	Business interruption or reputational risk	Significant financial cost or loss to the business	Moderate lost or compromised data and systems rectification required
High	Catastrophic loss of personal privacy, security or identity	Major loss of service, brand and reputation damage	Major financial impact in business interruption or penalties	Major loss of data, data integrity or data related systems

7.2.2 Risk management in the data environment

7.2.2.1 General

The risk matrix should be applied against the data environment to understand the organization's vulnerabilities beyond the establishment of a data governance process as outlined in Clause 6. The risk assessment of the data environment should consider, but not be limited to:

— the data users;

— the data processes;

— the data security and safety;

— the data setting (including entities and platforms involved with the data);

— the parties with access to the data.

7.2.2.2 Data users and custodians

Organizations should develop a documented system for all data users to ensure the appropriate levels of awareness, competency and authority are developed and applied. This system should consider:

— data access authority levels within existing employment structures;

— induction and on-boarding (including unpaid, volunteer, collaborative research partners, or seconded data users);

— training and development requirements including ongoing professional development.

The first step in managing the risks associated with data projects is to ensure that there is a shared understanding of the data project objectives commencing with a common understanding of the data characteristics.

7.2.3 Classifying data and data set information

To ensure a shared understanding amongst data users about data and to log changes to data that can occur through the use, sharing or exchange of data it is important that the data are understood and appropriately classified. Organizations should consider the use of a data catalogue or metadata registry which should be updated consistent with all relevant changes to the data.

In addition, there are several other considerations in classifying the data and data set information including developing and recording the following:

— data sources;

— data quality (see 7.3.2.1);

— data representation;

— data transmission means for sharing and exchange to ensure data quality;

— data sensitivity (see 7.3.2.2);

— the identification of personal information (see 7.3.4) and PII (see 7.3.5);

— data accessibility and security including privacy protection requirements.

7.2.4 Data attributes

Organizations should document the data attributes to ensure a consistent and clear understanding of the potential applications for the data. The data attributes can be important in identifying reasons for particular data actions. For example, gaps in the data can precipitate the acquisition of additional data or the combination of datasets.

7.2.5 Data process and the data environment

Whilst the data lifecycle is broadly considered as the stages involved from the creation of data through to its destruction with all other activities considered in between, there is no universally agreed version. Each organization should consider the data life cycle in the context of the organization as a whole, or as it relates to each specific data project. Each step in the data process can be influenced by the data environment as illustrated in the example provided in Figure 4. Organizations should consider the physical, operational, behavioural and organizational factors which are applicable to each step in the data lifecycle. The data environment should be considered for all entities involved in each stage of the data process, avoiding any assumptions that different entities will operate under the same data environmental standards. Each data project should develop a documented understanding of the data process steps and the risk mitigation measures required to maintain the desired outputs and project integrity.

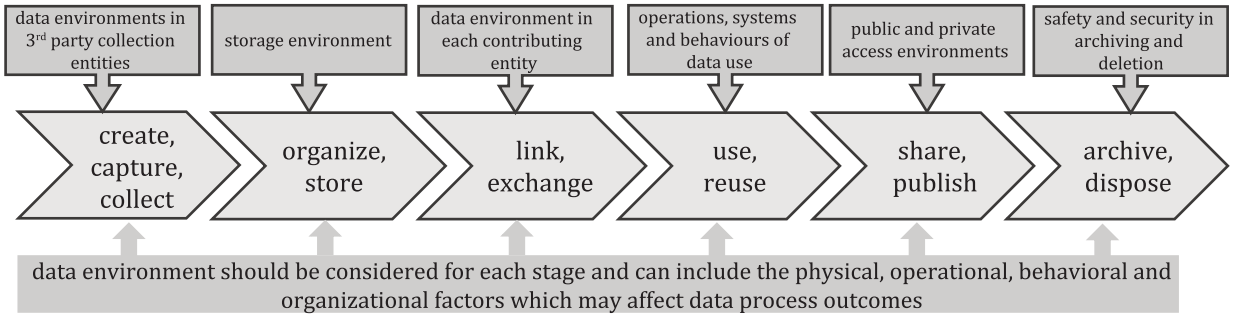

Figure 4 — Relationship between the data environment and data process

7.3 Managing data usage

7.3.1 General

Data usage in many instances can involve multiple parties only some of which own or provide data or have direct control of the data outputs. Parties can have differing standards of governance and are not necessarily the ultimate user of the information or insights developed at the end of a data project. It is vital therefore that all parties operate with a consistent and auditable standard of data management and documentation. This subclause sets out a framework for data usage.

7.3.2 Establishing a data catalogue or metadata registry

7.3.2.1 General

Data usage requires users in any stage of the data lifecycle to understand the structure, context and categorization of data through a process that is consistent and auditable. A shared understanding of the data attributes is vital to enable sharing and to maintain standards that relate to data structure, format, language, and quality. To do this, the entity should develop an organized inventory of the data assets that provides a consistent framework for every data user to understand the structure, meaning, categorization, sensitivities, and value of the data sets. This data, that defines or describes other data is known as metadata. The simplest framework for metadata is a data catalogue.

The metadata of data should be managed as a record in that it should be protected from loss or unauthorized deletion and retained or destroyed in accordance with requirements identified in appraisal. Access to metadata should be controlled using authorized access and permissions rules.

All additions, deletions and changes including data preparation should be recorded in the metadata.

7.3.2.2 Metadata registries

Information about the data, known as metadata, should be recorded in a metadata registry. This document provides an overview of the importance of metadata and what should be included in the metadata registry including:

a) one or more unique identifiers for the data set;

b) the designation or title of the data set;

c) a definition or description of the data set that provides sufficient detail to enable a user to quickly understand whether this data set is of interest;

d) the date the data set was issued and, if appropriate, the date that subsequent versions of the data set were, or will be, issued;

e) the access level, organizational boundaries and rights associated with the data set;

f) sensitivity and privacy factors associated with the data set;

g) the provenance of the data set, i.e. information about the place and time of the origin of the data set, its ownership, and the method of the generation of the set;

h) a set of keywords or tags that help to explain the data set;

i) the language or languages used to describe the data set;

j) the temporal and spatial coverages of the data set;

k) jurisdictional considerations of the data set including those related to:

   1) the base of operations for the data entity;

   2) relevant jurisdiction(s) for the data subjects;

   3) relevant jurisdictions in which the data will be used;

l) the accrual periodicity of the data set, i.e. the frequency at which new, revised or updated versions of the data set are made available;

m) the details of the distributions of the data set, including the identifier, the title, a description, the media type or file format, the size, the issue date, languages, access level and rights and access and download URLs;

n) annotations drawn from a concept system, such as an ontology, to describe the theme or category of the data set or the collection of data sets;

o) the details of any contexts, such as a programme, project or business area that use the data set;

p) the details of any quality, fitness for role or risk assessments made in respect of the data set;

q) any additional descriptions of the data set, including:

   1) any data elements that are already registered that are included in the data set;

   2) any information models that describe the structure of the information in the data set;

   3) any documents which describe aspects of the data set, such as technical information about the data set or developer documentation such as a graphical representation of the data model of the data set;

   4) the details of any superset or subset hierarchies containing the data set;

   5) the details of any replacement data set if this data set is superseded;

   6) the details of any collection of data sets of which this data set is a part, including the identifiers, the designation or title, a definition or description, issue dates, languages, access level, rights, the spatial coverage, the provenance and any quality assessments of the collection.

NOTE Further guidance regarding metadata registries can be found in ISO/IEC 11179 series of standards, especially:

— registry common facilities in ISO/IEC 11179-3:2023;

— data set registration in ISO/IEC 11179-33.

7.3.3 Data quality, sensitivity and security

7.3.3.1 General

A shared understanding of data quality, sensitivity and security is important to ensure that parties do not inadvertently add risk or compromise the value by changing these parameters over the life of a data project.

7.3.3.2 Data quality

Maintaining an understanding of the data quality is fundamental to the ability to create, collect, store, maintain, transfer and present information and data to support industrial and societal processes.

The following fundamental principles apply to managing the quality of data.

a) process approach

the processes that use, create and update data are defined and operated. These processes become repeatable and reliable by also defining and operating processes for managing data quality.

b) continuous improvement

data are improved through effective measurement and correction of data nonconformities that arise from data processing. Such improvements, however, do not prevent the same nonconformities occurring repeatedly. Improvements can be derived from analysing, tracing and removing the root cause of poor data quality, which can lead to improvements in data management processes.

c) involvement of people

specific responsibilities for data quality management exist at different levels of the entity. End users have the greatest direct effect on data quality through data processing activities. In addition, data quality specialists perform the necessary intervention and control, implement and embed processes for improvement of data quality across the organization. Finally, oversight by top management ensures the necessary resources are made available and directs the entity towards achieving the vision, goals and objectives for data quality.

NOTE 1 For more information about the importance of data quality in the development of information technology systems, see ISO/IEC 25012.

NOTE 2 For more information on the importance of data quality in the field of data quality for analytics and machine learning (ML), see ISO/IEC 5259.

7.3.3.3 Data sensitivity

A key aspect of data usage is the use of a documented process to understand data sensitivity. This is particularly important when considering data that relates to individuals (personal information), data that could be used to identify individuals (PII), and sensitive data subjects (e.g. children).

Regardless of whether the data usage process is undertaken by an individual or entity, the data sensitivity should be documented and recorded within the metadata. As some data processes can take place at a very granular level, data users should understand the overarching context of data in the organization. For example, a specific data process could obfuscate the fact that the data are part of a larger data set which includes personal information.

In addition to personal information, organizations will typically classify data (see 7.2.2) into one of several classifications, such as:

— confidential – access restricted to senior management;

— restricted – access limited to a controlled range of employees;

— internal – all employees have access but not public access;

— public information – unrestricted access.

Classification terms relating to data confidentiality may vary based on the organization, jurisdiction or other facet and data users should familiarise themselves with the terms applicable to their data environment to ensure the appropriate classification is applied.

The subject matter of the data also has a bearing on who can access it. For example, medical personnel can access highly personalized medical information about a patient, but not their financial information. This

can be important when considering appropriate controls to manage the classification and sensitivity of data shared between different organizations or areas within an organizations, with access to specific data subject matter.

7.3.3.4 Data security and operability

Data security and operability considers the security, sensitivity, and operational requirements in data usage. The first principle to consider in responsible data usage is the principle of do no harm (this might be with reference to privacy, security or access for bad actors etc.). The second principle is to maintain consistent standards within the data.

Data security and operability processes should include legal requirements and legal instruments such as contracts.

Data security and operability requirements should also be considered with the organization's business continuity plans as well as any third-party business processes.

7.3.4 Privacy protection requirements

The decision-making process for sharing, exchange, or use of data should consider the privacy protection requirements (PPR). These can include several overlapping obligations including, but not limited to legal requirements, contractual obligations, operational policies, public policy requirements and external treaties.

Fundamental to privacy protection requirements is recognizing the degree of personal information or PII within the data.

NOTE Further information regarding privacy protection requirements can be found in ISO/IEC 15944-12.

7.3.5 Personal information

A sufficient amount of personal information can be used to assume an individual's identity. In order to protect individuals from the risk of fraud and other issues arising from the misuse of personal information, organizations should ensure they have appropriate measures in place to protect data which includes personal information. These measures should ensure compliance with relevant legal and jurisdictional requirements and assess and manage vulnerabilities in both the *data environment* and the *data process*.

Data users within an organization should understand the type of personal information present in the data including within de-identified data sets or when involved in a granular level of data use such as developing algorithms, platforms or programs. The organization should consider the sensitivities of personal information such as:

— name;

— date of birth;

— national identification number (e.g. social security, national identity);

— photo identification records (e.g. Passport, Drivers Licence);

— individual image (e.g. from surveillance data);

— other government identification record (e.g. Healthcare number);

— birth certificate information;

— tax related identifiers (e.g. tax file number);

— bank records (e.g. account number and bank);

— address;

— government agency identification (e.g. employment record);

— health record;

— genetic and biometric data;

— criminal records;

— penalty records;

— vehicle registration record;

— student identification record;

— school enrolment record;

— employment record;

— electoral record;

— IP Address;

— telephone number;

— asset records (e.g. land title);

— mortgage;

— utility bill;

— union or trade association membership record;

— place of employment;

— racial and ethnic origins;

— religious data.

7.3.6 PII

PII is any data or information that when used in combination could be used to identify an individual. The data itself are often not considered personal information but can be able to link to personal information. For example, transit data associated with a registered contactless smart card can enable the identification of an address (journey origin) and a school (journey destination). Examples include:

— location (geospatial) information;

  — public transport trip data;

  — vehicle toll records;

  — security camera images (non-identifiable);

— time based information;

  — work shift records;

  — movement data;

— activity based information;

  — browser history;

— credit history report;individual recorded information;

  — examination paper.

8 Data use, exchange and sharing

8.1 Overview

This clause has been structured with data use as being all data activities, exchange being the movement of data from entity to entity and data sharing being the provision of data to other entities including publishing data. In all situations, all actions, analysis and changes to the data should be recorded in the *metadata*.

8.2 Data use

Data use involves the handling or dealing with information for a specific purpose. This can include highly repeatable business operations such as accounting or specific analysis to develop knowledge and insights. Data use can also involve a series of sequential actions each of which increases the amount of knowledge developed from the data or enables processing via AI systems. Each action can result in changes to the data and should be recorded in the metadata to ensure repeatability and analysis particularly in relation to the knowledge gained through the process. The process of recording data changes in the metadata is important for the purposes of audit.

Increasingly organizations are using data to improve insights and knowledge and to develop AI systems.

Data use can include:

— access;

— observation;

— interpretation;

— analysis;

— event detection;

— event detection response;

— consolidation or collation;

— recording;

— creating data products.

8.3 Data exchange

Data exchange involves the retrieval, data transformation, optional data translation, transference and receipt of data. Within an organization data can be exchanged many times from steps within the data lifecycle (e.g. data collection to data storage) and across different platforms or systems.

Organizations should consider necessary steps in the exchange of data to ensure data integrity, security, and privacy. These steps may include:

— de-identification of personal data;

— removal of sensitive or irrelevant data;

— development of data products to constrain data analysis;

— deletion or archiving of data;

— activities of data users that can take place outside the controlled environment of the organization such as:

  — remote working;

  — offsite access to higher order processing platforms (e.g. super computer).

8.4 Data sharing

8.4.1 General

Data sharing involves access to or processing of the same data by more than one authorized entity, making data available to, and findable by, other authorized entities. This can involve:

— data shared from one data analyst within the same department of an organization;

— data shared from one department to another department in an organization;

— data shared from an organization to another organization under a two-party contracted arrangement (controlled environment);

— data shared to numerous organizations but in a controlled way (multiparty contract).

Figure 5 illustrates the increasing complexity in data management and the associated decreasing level of direct control which can occur when more data sharing parties or data domains are applied.

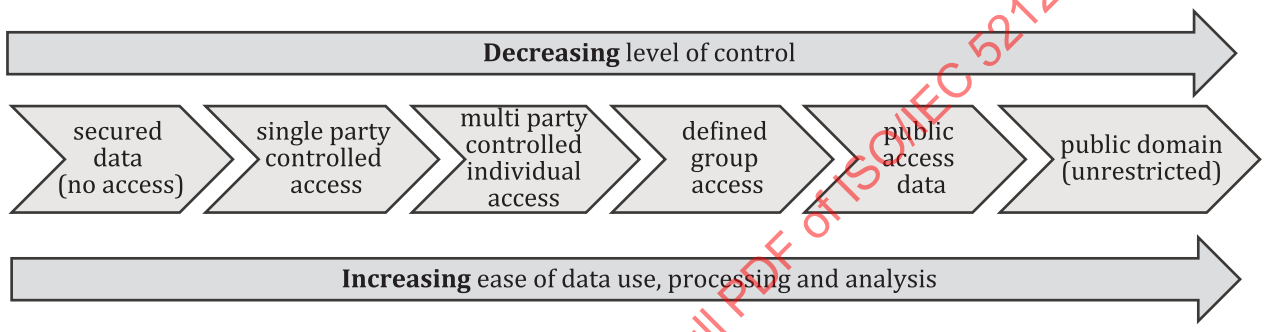

Figure 5 — Data access domains

Establishing a framework for data sharing

In all data sharing activities, organizations should assess and document:

— the justification for sharing data including:
  — the purpose for sharing data;
  — the expected outputs of sharing data;
  — the knowledge and insights expected;
  — the value of the data sharing project.

— stakeholders;

— authorized participants or users in the data sharing project including organizations, individuals, platforms, and systems, for example electronic data interchanges (EDI);

— measures to ensure the authorization of data sharing participants;

— data subjects particularly in relation to individuals and personal information;

— changes in data security, privacy, and sensitivity levels of the data;

— risk assessment insights developed in examining data stages (for example using the data lifecycle) or the data environment;

— risk mitigation measures for each data sharing party;

— the steps in the data sharing process;

— post sharing risk assessment including any mitigation measures;

— the review and audit pathways including as a minimum:

  — pre and post sharing data set retention;

  — metadata registry updates;

  — version control;

— concluding steps.

8.4.2 Data sharing within an organization

Data sharing within an organization can be considered low risk, however privacy, security and financial dissimilarities can require boundaries or risk management measures to preserve compliance obligations. Internally, this can require the development of a data sharing process or platform to maintain delineation between different levels of authorization. Importantly, organizations should include any technical service providers involved in the development of these processes or platforms as part of the risk assessment as they are likely to have visibility of all security and authorization levels and therefore can require separate control measures.

The high frequency of data sharing within organizations in today's enterprise environments indicates that a fundamental understanding of safe data management principles should be made available to all individuals. In all situations, all actions, analysis and changes to the data should be recorded in the *metadata*.

8.4.3 Data sharing between organizations

Data sharing between or among organizations can offer significant societal benefits in realizing:

— increased digitization;

— improved service efficiencies particularly in government services;

— increasing response time in relation to material issues.

Organizations seeking to enter a shared data arrangement should consider the following:

— the level of readiness to share data;

— the internal processes to receive and manage data from another entity;

— additional resources required to manage the shared data;

— benefits and risks to the organization in sharing data;

— benefits and risks to customers or data subjects in sharing data;

— benefits and risks to society in sharing data (particularly for public service organizations);

— management measures for new issues arising from data sharing including:

  — changes in security, privacy and sensitivity;

  — new insights;

  — commercial considerations resulting from new data sets and insights.

Data sharing projects should be appropriately scoped prior to the sharing of any data with all participating organizations operating from a shared agreement of the project. Organizations should consider the use of a data sharing agreement which should incorporate the following:

a) the data sharing purpose;