International Standard
FINAL DRAFT

ISO/IEC FDIS 5212

Information technology — Data usage — Guidance for data usage

ISO/IEC JTC **1**/SC **32**

Secretariat: **ANSI**

Voting begins on:
2024-01-23

Voting terminates on:
2024-03-19

Reference number
ISO/IEC FDIS 5212:2024(en)

COPYRIGHT PROTECTED DOCUMENT

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.

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.

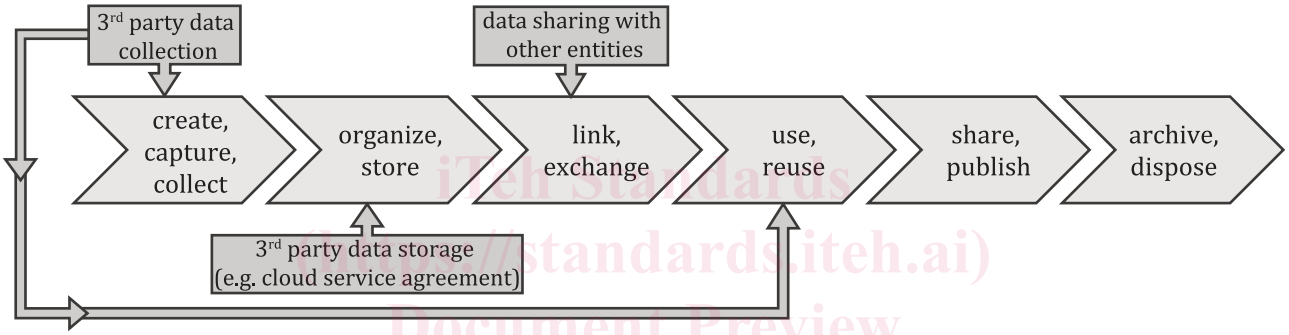

Figure 2 — Third parties and the data lifecycle

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.

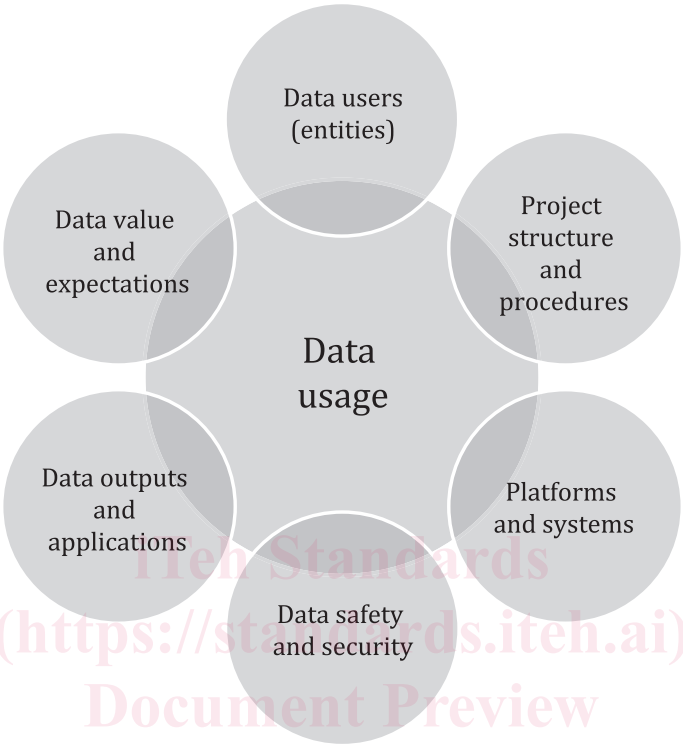

Figure 3 — The data environment

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	