

ICS 07.080  
CCS A 40



**NATIONAL STANDARD OF THE  
PEOPLE'S REPUBLIC OF CHINA**

**GB/T 45782-2025**

---

**Biotechnology - Requirements for data formatting and  
description in the life sciences**

生物技术 生命科学中数据格式和描述的要求

**(English Translation)**

**Issued on: May 30, 2025**

**Implemented on: September 1, 2025**

---

**Issued by: State Administration for Market Regulation, China National  
Standardization Administration**

## Table of Contents

|                                                                                                     |
|-----------------------------------------------------------------------------------------------------|
| Preface                                                                                             |
| Introduction                                                                                        |
| 1 Scope                                                                                             |
| 2 Normative references                                                                              |
| 3 Terms and Definitions                                                                             |
| 4 Recommendations and requirements for entity and concept descriptions in life sciences data        |
| 4.1 General                                                                                         |
| 4.2 Recommended universal identifier scheme for biological and conceptual entities                  |
| 4.3 Data formatting and contextual descriptive data (metadata) for biological entities and concepts |
| 5 Technical and organizational recommendations and requirements for data formats                    |
| 5.1 Overview                                                                                        |
| 5.2 Organizational Responsibility                                                                   |
| 5.3 Documentation                                                                                   |
| 5.4 Version Control and Change Log                                                                  |
| 5.5 Compatibility                                                                                   |
| 5.6 Scalability                                                                                     |
| 5.7 Compression                                                                                     |
| 5.8 Structure and control elements                                                                  |
| 5.9 Data Type Requirements in Data Formats                                                          |
| 5.10 Conformance and Compatibility                                                                  |
| 5.11 Data Integrity                                                                                 |
| 5.12 Format Verification                                                                            |
| 5.13 Data Tracing                                                                                   |
| 6 Semantic recommendations and data format requirements                                             |
| 6.1 General                                                                                         |
| 6.2 Minimal consensus for biological data annotation                                                |
| 6.3 Syntax and Substantiation                                                                       |
| 7 Requirements for terminology and ontologies applicable to biological data annotation              |
| 7.1 General                                                                                         |
| 7.2 Requirements of biological ontology                                                             |
| 8 Requirements for domain-specific data standards                                                   |

## 8.1 Overview

## 8.2 Specific requirements for domain-specific data standards

## 9 Requirements for biological data repositories

### 9.1 General

### 9.2 Requirements for biological data repositories

## Appendix A (Informative) Examples of Common Data Formats in Life Sciences

### A.1 Overview

### A.2 Data formats for OMICS (omics), biochemistry and molecular biology methods

### A.3 Formats of Bioimaging Data

### A.4 Data formats for computer models of biological systems

### A.5 Data formats for life science model simulations and their results

### A.6 Descriptors for data and model quality measurements

## Appendix B (informative) Minimum reporting standards for data, models and metadata

### B.1 Overview

### B.2 Minimum reporting standards

### B.3 Domain-specific ontologies, taxonomies, and controlled vocabularies

## Reference

## Foreword

This document is in accordance with the provisions of GB/T 1.1-2020 "Guidelines for standardization work Part 1: Structure and drafting rules for standardization documents" Drafting.

This document is equivalent to ISO 20691:2022 "Requirements for data formats and descriptions in biotechnology life sciences".

## Introduction

Life science research and its application in biotechnology, diagnostics and pharmaceuticals rely on a wide range of detection, biology, functional research, this includes the use of derived data for biological, biotechnological and physiological the computation, reconstruction, modeling and simulation of biotechnology processes and their application in biotechnology workflows.

Biotechnology research spans a broad range of biological sciences and biotechnology fields and applications (e.g., human health, genetically engineered organisms, environmental science, agricultural industry, bioremediation, DNA sequencing, chromatography, microscopy, etc.). Only by using computational methods can the data be

consistently described, Data-driven, data-intensive, and big data analysis methods in life sciences are only possible if they are structured and integrated.

This document provides information on the storage, representation, meaning, interpretation, exchange and reuse of data in the life sciences.

The basic requirements are set to meet the key needs of establishing an interoperable and unambiguous data recording, description and transmission framework, thereby enhancing the benefits of Utilization rate and maximization of data value.

Data in the life sciences from different sources and recorded at different times must be findable, accessible, interoperable and reusable (FAIR) Data sets are valuable and useful only if they are accessible and stored in a well-structured, consistent format.

The platform-restricted format ensures data version control, data archiving, and data traceability. Complete and updatable metadata (i.e., data describing the data) this document provides requirements and recommendations for standardized interoperable data formats in the life sciences.

The Biotechnology and Biotechnology communities define data formats and description standards that provide a conceptual framework for the many different subfield-specific data formats and description standards defined by the biotechnology and biotechnology communities.

This document describes a set of standards to facilitate the use of the referenced field-specific formatting and description standards and their interoperability.

Therefore, this document provides the overall data format and popular description rules of the relevant sub-fields.

In addition, it provides specific standards for creating special (sub)domains, ensuring interoperability and implementation.

rules and guidelines for implementing these standards.

Data formats and

Requirements for description

## 1 Scope

This document defines the requirements for data and corresponding metadata in the life sciences (including biotechnology and biomedicine and non-human biological research and development).

It provides a consistent format and documentation for data (i.e., data that describes the data and its context).

findable, accessible, interoperable and reusable guidance.

This document applies to manual or other activities in the life sciences that systematically collect, record or integrate data and corresponding metadata for other purposes.

Computational workflow.

This document specifies the format requirements for raw experimental or procedural data obtained manually and machine-generated data.

Requirements for the storage, sharing, access, interoperability and reuse of data and corresponding metadata in the field of life sciences.

This document specifies requirements for systematically acquiring large amounts of data from automated high-throughput workflows in the life sciences, and for He also has requirements for large and small data sets for life science technology and artificial data acquisition.

This document is applicable to many areas in biotechnology and life sciences, including but not limited to. Basic/applied in all areas of life sciences Research, and industrial, medical, agricultural, or environmental biotechnology (excluding those used for diagnostic or therapeutic purposes) and its methodological drivers, such as Genomics (including massively parallel sequencing, metagenomics, epigenomics, and functional genomics), transcriptomics, translomics, protein Proteomics, metabolomics, lipidomics, glycomics, enzymology, immunochemistry, synthetic biology, systems biology, systems medicine and related field.

## 2 Normative references

The contents of the following documents constitute the essential clauses of this document through normative references in this document.

For referenced documents without a date, only the version corresponding to that date applies to this document; for referenced documents without a date, the latest version (including all amendments) applies to this document.

ISO 8601-1 Date and time-Representations Part 1: Basic principles Note. GB/T 7408.1-2023 Date and time information exchange representation Part 1: Basic principles (ISO 8601-1:2019, IDT) ISO 8601-2 Date and time information interchange representation Part 2: Extensions [ISO 8601-2 Dateandtime-Rep-

## 3 Terms and definitions

The following terms and definitions apply to this document.

A standard for character encoding in electronic communications.

**Note 1:** ASCII codes represent text in computers, telecommunications equipment and other equipment.