

ICS 35.040
CCS L72



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

GB/T 43707-2025

Scientific data provenance metadata

科学数据溯源元数据

Issued on: January 24, 2025

Implemented on: January 24, 2025

**Issued by: State Administration for Market Regulation;
Standardization Administration of the PRC**

Table of Contents

1 Scope	1
2 Normative references	1
3 Terms and Definitions	1
4 General Requirements	2
4.1 Metadata Information	2
4.2 Metadata Description Method	2
4.2.1 Overview	2
4.2.2 Definition	2
4.2.3 English name	2
4.2.4 Field Type	2
4.2.5 Range	2
4.2.6 Short Name	2
4.2.7 Note	3
5 Dataset Metadata	3
5.1 Identifier	3
5.2 Name	3
5.3 Latest submission date	3
5.4 Creation Date	3
5.5 Description	4
5.6 Keywords	4
5.7 Access Restrictions	4
5.8 Link address	4
5.9 Name of submitting organization	4
5.10 Correspondence address of submitting institution	5
5.11 Submitting institution postal code	5
5.12 Contact number of submitting institution	5
5.13 Submitting institution email address	5
5.14 Data Types	5
5.15 Language	5
5.16 Data Source	6
5.17 Update frequency	6
5.18 Number of records.	6

5.19 Storage	6
5.20 Security Category	6
5.21 Security Class	7
5.22 Quality Report	7
5.23 Permission Statement	7
5.24 Time parameters	7
5.25 Spatial parameters	7
6 Activity Metadata	8
6.1 Activity Name	8
6.2 Activity Time	8
6.3 Activity Description	8
7 Execution Entity Metadata	8
7.1 Participants	8
7.1.1 Name of Participant	8
7.1.2 Contribution of Participants	8
7.1.3 Name of Participant	9
7.1.4 Participant Communication Address	9
7.1.5 Participant contact number	9
7.1.6 Participant Email	9
7.2 Tools	9
7.2.1 Tool Name	9
7.2.2 Tool Description	10
13 Reference	18

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. Please note that some of the contents of this document may involve patents. The issuing organization of this document does not assume the responsibility for identifying patents. This document was proposed by the Ministry of Science and Technology of the People's Republic of China. This document is under the jurisdiction of the National Science and Technology Platform Standardization Technical Committee (SAC/TC486). This document was drafted by: Computer Network Information Center of the Chinese Academy of Sciences, National Science and Technology Infrastructure Platform Center, Guangzhou Internet of Things

Research Center Institute of Agricultural Sciences, China National Institute of Standardization, National Marine Information Center, National Meteorological Information Center, Agricultural Information Institute of Chinese Academy of Agricultural Sciences, China Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Institute of Resource Information, Chinese Academy of Forestry, Institute of Process Engineering, Chinese Academy of Sciences, Institute of Microbiology, Chinese Academy of Sciences, Zhonglu High-Tech Transportation Technology Group Co., Ltd., Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Ministry of Natural Resources First Institute of Oceanography, China Southern Power Grid Internet Services Co., Ltd., and Beijing Information Science and Technology University. The main drafters of this document are. Zhu Yanhua, Hu Lianglin, Liao Fangyu, He Yuntao, Chen Luxin, Wang Zhiqiang, Gao Yuwei, Zhao Yuehong, Li Xin, Ji Ping, Hu Bo, Fan Jingchao, Wang Yi, Guo Xuebing, Wu Linhuan, Liu Dongmei, Liang Guoxia, Li Junyao, Yang Qinghai, Wang Yan, Song Zhuanling, Zhao Lin, Wang Lina, Gao Shang, Liu Yingdi, Gui Yuan, You Xindong.

As my country continues to increase its efforts to open up and share scientific data, the need to improve data quality is also growing. Different researchers hold different views on quality. Scientific data traceability information records the data flow process and can objectively reflect the data quality. However, most existing metadata standards focus on data content description information and cannot cover other aspects of the traceability model. The scientific data traceability metadata needs to be further supplemented and improved. GB/T 30522-2014, using the abstract description method, defines the format and content of scientific data general domain traceability metadata for various disciplines. The scope will be expanded and supplemented according to actual conditions. This document is the promotion and implementation of GB/T 34945-2017, which will further clarify the semantic description of the traceability model and improve the level of interoperability. This platform will help improve the data quality of the scientific and technological resource platform and enhance the data security of the scientific and technological platform.

Scientific Data Provenance Metadata

1 Scope

GB/T 43707-2025 defines the provenance metadata of scientific data - the record of where a dataset came from, what was done to it, by whom and with what. Provenance is what separates a reusable dataset from a file: a result that cannot be traced back through its processing steps to the instrument or the survey that produced it cannot be reproduced, corrected or credited, and China's national scientific data centres now hold enough data that the question is a practical one rather than a philosophical one. The standard sets the general requirements, including the metadata description method by which each element is defined - its definition, English name, field type, value range, short name and notes - and then the metadata elements themselves at the dataset level, covering the identifier, the

name, the submission dates, the responsible parties, the sources, the processing steps and the derivation relations between datasets. It took effect on 24 January 2025.

This document specifies the general requirements for scientific data provenance metadata, dataset metadata, activity metadata, and execution entity metadata, and provides A metadata example is given. This document is applicable to the tracking and recording of the scientific data traceability process.

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.

GB/T 7408 Data elements and exchange formats Information exchange Date and time representation

GB/T 32843 Identification of scientific and technological resources

3 Terms and definitions

The following terms and definitions apply to this document.

3.1 scientific data In the fields of natural sciences, engineering sciences, etc., the Records of original and derived information obtained by other means, or other data that can be used for scientific research activities. [Source: GB/T 43708-2025, 3.1]

3.2 Records of the evolution of scientific data during research activities and the content of its evolution processing. [Source: GB/T 34945-2017, 2.1, modified]

3.3 Scientific data is collected and processed, stored and backed up, transmitted and exchanged, shared, used and served, safely disposed, and finally reused. Circular process. [Source: GB/T 43708-2025, 3.3]

3.4 metadata Data about data. [Source: GB/T 30522-2014, 3.3]

3.5 dataset A collection of data that can be identified.