



中华人民共和国国家标准

GB/T 43707—2025

科学数据溯源元数据

Scientific data provenance metadata

2025-01-24 发布

2025-01-24 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 总体要求	2
4.1 元数据信息	2
4.2 元数据的描述方法	2
4.2.1 概述	2
4.2.2 定义	2
4.2.3 英文名称	2
4.2.4 字段类型	2
4.2.5 值域	2
4.2.6 短名	2
4.2.7 注解	3
5 数据集元数据	3
5.1 标识符	3
5.2 名称	3
5.3 最近提交日期	3
5.4 创建日期	3
5.5 描述	4
5.6 关键词	4
5.7 访问限制	4
5.8 链接地址	4
5.9 提交机构名称	4
5.10 提交机构通信地址	5
5.11 提交机构邮政编码	5
5.12 提交机构联系电话	5
5.13 提交机构电子邮箱	5
5.14 数据类型	5
5.15 语种	5
5.16 数据来源	6
5.17 更新频率	6
5.18 记录数	6
5.19 存储量	6
5.20 安全分类	6
5.21 安全分级	7

5.22	质量报告	7
5.23	权限声明	7
5.24	时间参数	7
5.25	空间参数	7
6	活动元数据	8
6.1	活动名称	8
6.2	活动时间	8
6.3	活动描述	8
7	执行实体元数据	8
7.1	参与人	8
7.1.1	参与人姓名	8
7.1.2	参与人贡献	8
7.1.3	参与人单位名称	9
7.1.4	参与人通信地址	9
7.1.5	参与人联系电话	9
7.1.6	参与人电子邮箱	9
7.2	工具	9
7.2.1	工具名称	9
7.2.2	工具描述	10
8	元数据示例	10
附录 A (规范性)	代码表	11
附录 B (资料性)	科学数据溯源元数据示例	13
参考文献		18

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中华人民共和国科学技术部提出。

本文件由全国科技平台标准化技术委员会(SAC/TC 486)归口。

本文件起草单位：中国科学院计算机网络信息中心、国家科技基础条件平台中心、广州物联网研究院、中国标准化研究院、国家海洋信息中心、国家气象信息中心、中国农业科学院农业信息研究所、中国科学院地理科学与资源研究所、中国林业科学研究院资源信息研究所、中国科学院过程工程研究所、中国科学院微生物研究所、中路高科交通科技集团有限公司、中国科学院青藏高原研究所、自然资源部第一海洋研究所、南方电网互联网服务有限公司、北京信息科技大学。

本文件主要起草人：朱艳华、胡良霖、廖方宇、赫运涛、陈鲁鑫、王志强、高瑜蔚、赵月红、李新、纪平、胡泊、樊景超、王漪、郭学兵、吴林寰、刘冬梅、梁国霞、李俊瑶、杨青海、王妍、宋转玲、赵琳、王丽娜、高尚、刘英迪、桂媛、游新冬。

引 言

随着我国科学数据开放共享的力度不断加大,提升数据质量的需求也日益强烈。然而,如何评价数据质量,不同的研究者持有不同的观点。

科学数据溯源信息记录数据流转过程,可以客观反映数据质量。科学数据溯源信息可以通过元数据形式进行描述。然而现有元数据标准大多聚焦在数据内容描述信息,尚不能覆盖溯源模型的其他方面(如数据活动、执行实体等),科学数据溯源元数据有待进一步补充和完善。本文件根据 GB/T 30522—2014,采用摘要描述方法,界定了科学数据通用领域溯源元数据的格式和内容,供各学科领域根据实际进行扩展和补充。

本文件是对 GB/T 34945—2017 的推进和落实,将进一步明晰溯源模型语义描述、提高互操作水平,有助于改善科技资源平台的数据质量,提升科技平台数据安全。

科学数据溯源元数据

1 范围

本文件规定了科学数据溯源元数据的总体要求、数据集元数据、活动元数据、执行实体元数据,并给出了元数据示例。

本文件适用于科学数据溯源过程的跟踪和记录。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 7408 数据元和交换格式 信息交换 日期和时间表示法

GB/T 32843 科技资源标识

3 术语和定义

下列术语和定义适用于本文件。

3.1

科学数据 scientific data

在自然科学、工程技术科学等领域,科学研究活动中形成的以及通过观测监测、考察调查、检验检测等方式获取的原始及其衍生信息的记录,或可用于科学研究活动的其他数据。

[来源:GB/T 43708—2025,3.1]

3.2

科学数据溯源 scientific data provenance

科学数据在研究活动中的演变信息和演变处理内容的记录。

[来源:GB/T 34945—2017,2.1,有修改]

3.3

科学数据生存周期 scientific data lifecycle

科学数据从采集加工、存储备份、传输交换、开放共享、使用服务、安全处置,最终实现再利用的一个循环过程。

[来源:GB/T 43708—2025,3.3]

3.4

元数据 metadata

关于数据的数据。

[来源:GB/T 30522—2014,3.3]

3.5

数据集 dataset

可以标识的数据集合。