CNAS 技术报告

科研数据特征划分及其分类体系表

中国合格评定国家认可委员会

发布日期：2019 年 01 月 01 日

前 言

科研实验室是实施科技创新的重要组成部分和基础技术保障，出具的数据是科学成果的依据，其质量直接影响科学成果的可靠性和科学性。但科研实验室存在科研数据难以重复、可靠性难以评估的问题。如何对科研实验室进行规范管理，实现科研数据的可靠和可重复，是科学界面临的重大课题。为了更好的规范科研实验室的管理和运行，依托“十三五”国家重点研发计划 “科研实验室认可关键技术研究”项目中的“科研数据不确定性表征方法和评估技术研究”课题（编号：2016YFF0203801）， CNAS 组织编制了《科研数据特征划分及其分类体系表》。

本文件由中国合格评定国家认可委员会提出并归口。

本文件主要起草单位：中国合格评定国家认可中心、北京理工大学。

本文件主要起草人：周桃庚、吕京、傅华栋、宋桂兰。

发布日期：2019 年 01 月 01 日

科研数据特征划分及其分类体系表

1 范围

本标准定义了科研数据的特征划分和分类方法，规定了科研数据分类体系表组成、结构及编码规则和编制要求，用以保证科研数据组织、存储及交换的一致性。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是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 13016 标准体系表编制原则和要求

3 术语和定义

下列术语和定义适用于本文件。

3.1 科研数据 Research Data

在科技活动（实验 、观测 、探测 、调查等)或通过其他方式所获取的反映客观世界的本质、特征 、变化规律等的原始性、基础性数据，以及根据不同科技活动需要,进行系统加工整理的数据产品和相关信息。

3.2 定量数据 Quantitative Data

可以量化和验证的数据，适用于统计操作。

3.3 定性数据 Qualitative Data

近似或表征但无法测量的某物或现象的属性，特征，性质等的数据。

3.4 科研数据分类 Research Data Classification

根据科研数据的属性或特征，将其按照一定的原则和方法进行区分和归类，并建立起一定的分类体系和排列顺序，以便更好地管理和使用科研数据的过程。

3.5 线分类法 method of linear classification

将分类对象按选定的若干属性(或特征)逐次分为若干层级，每个层级又分为若干类目。同一分支的同层级类目之间构成并列关系，不同层级类目之间构成隶属关系。

3.6 面分类法 method of area classification

选定分类对象的若干属性(或特征)，将分类对象按每一属性（或特征）划分成一组独立的类目，每一组类目构成一个“面”。再按一定顺序将各个“面”平行排列。使用时根据需要将有关“面”中的相应类目按“面”的指定排列顺序组培在一起，形成一个新的复合类目。

3.7 混合分类法 method of composite classification

将线分类法和面分类法组合使用，以其中的一种为主，另一种做补充的分类方法。

4 分类原则

按照科研数据的特征和属性及其相互间客观存在的逻辑关联以及信息组织的要求进行分类。分类名称应尽量沿用各专业习惯名称，但是要保证规范化，避免发生概念混淆或二义性，尽量使分类方案简便、实用和便于记忆。

4.1 完整性和扩展性结合

分类体系在总体上应具有很大的概括性和包容性，能够容纳各种各样的科研数据和满足将来可能产生的信息需求。分类在反映数据的属性和数据间的相互关系上保持相对的完整性。SDS/T 2122—2004 4.3

4.2 实用性

确保每个类目下要有科研数据，不设没有意义的类目，类目划分要符合用户对科研数据分类的普遍认识。SDS/T 2122—2004 4.4

4.3 揭示性原则

分类应尽可能反映科研数据集的内容、对象和属性特点，以便于检索使用，为深入分析科研数据集的关联和影射关系提供便利。TR-REC-018 5.5

4.4 规范性原则

所使用的语词或短语能确切表达类目的实际内容范围，内涵、外延清楚；类名采用科学、规范、通用的术语或译名；在表达相同的概念时，做到语词的一致性；在不影响类目涵义表达的情况下，保证用语的简洁；每个类目都要有专指的检索意义。此外，类目涵义以及与他类目的关系，必要时还需通过类目体系和类目注释加以说明或限定。TR-REC-018 5.6

4.5 系统性原则

将选定的分类对象的特征（或特性），按其内在规律进行系统化排列，确保类目唯一、结构合理、层次清晰，减少冗余。

5 科研数据分类编码的方法

5.1 分类方法

按照GB 7027 ，采用混合分类法进行。

科研数据依其特性属性，分为两大类（一级分类）。各大类科研数据依其数据特征，选取内容属性、观测属性、来源属性 、格式属性、处理属性、状态属性、格式属性、表现属性等不同组合进行分类（二级分类）。

5.2 编码方法

分类代码由大类代码和二级分类的各属性代码组成，各代码之间用下横线“_”分割。

科研数据各大类和属性代码用两种方法表示，一是简码（为阿拉伯数字代码），二是标识符（为英文字母和阿拉伯数字组成，通常第一位应为英文字母）。

大类简码用一位英文大写字母表示，其标识符用四位英文大写字母组成。二级分类的属性简码用三位阿拉伯数字表示。

标识符可以体现大类和属性的基本意义，应便于人工识别、符合英文缩写习惯。

6 科研数据分类体系表结构

6.1 层次结构

科研数据分类体系表**层次结构图**见图1。

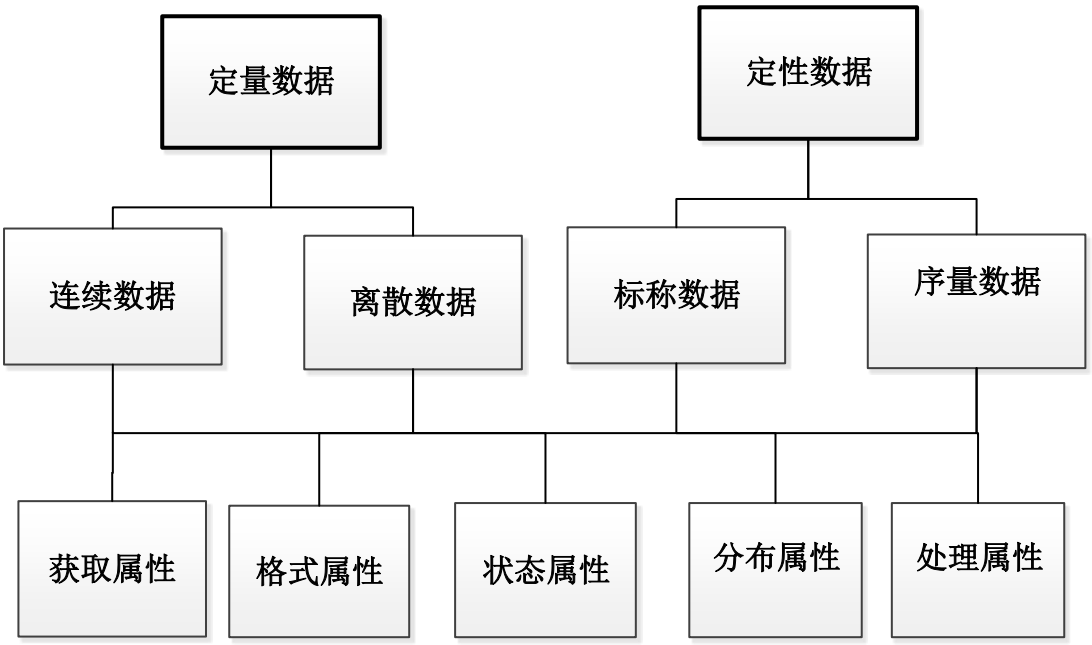

图1 科研数据分类体系表层次结构图

6.2 数据获取属性分类

数据获取属性指科研数据的获取方式和获取来源，表1中列出了依据数据获取属性进行的小类划分和代码。

表 1 数据获取属性分类

数据获取主题类型	代码	说明
实验/试验		通过实验、试验获取的数据
仿真		基于原始数据，采用模型、方法通过模拟、计算，获得的信息、数据。
派生或编译		利用通常来自不同数据源的现有数据点，通过某种转换（如算术公式或聚合）来创建的新数据。
参考文献或规范		对较小（同行评议的）数据集进行（静态或有机）组合聚集或收集，最有可能出版和策划的数据。
……		

6.3 数据格式属性分类

数据格式属性指科研数据的表达形式与数据格式，表2中列出了依据数据格式属性进行的小类划分和代码。

表 2 数据格式属性分类

数据格式类型	代码	说明
文档		文本格式或其他相应格式存储的数据。
影像		遥感影像、图片等常用格式存储的数据。
地图		数字化地图、电子地图等常用格式存储的数据。
模型		数学方法的数学表达式及其说明，或工具化的模拟软件或构件。
表格		数据表常用格式存储的数据。
视频		观测过程记录、实验过程记录等常用格式存储的视频数据。
音频		观测过程记录、实验过程记录等常用格式存储的视频数据。
服务		基于 Web 发布的、外部用户可调用的各类数据处理与加工资源。
工具		可离线运行的数据处理与加工资源。
……		

6.4 数据状态属性分类

数据状态属性指科研数据所处的状态，表3中列出了依据数据状态属性进行的小类划分和代码。

表 3 数据状态分类

数据状态类型	代码	说明
动态		
静态		

6.5 数据分布属性分类

数据分布属性指科研数据的统计特征，表4中列出了依据数据分布属性进行的小类划分和代码。

表 4 数据分布属性分类

数据分布类型	代码	说明
正态分布		
矩形分布		
三角分布		
梯形分布		
反正弦分布		

Weibull 分布		
……		

6.6 数据处理属性分类

数据处理属性指科研数据的使用方便程度，表5中列出了依据数据处理属性进行的小类划分和代码。

表 5 数据处理属性分类

数据处理类型	代码	说明
原始数据		
辅助数据		
未污染数据		
已校准数据		
汇总数据		
可视化数据		
……		