

ICS 35.040

CCS A 00



团体标准

T/CSTM 00837—2022

材料基因工程数据 元数据标准化基本原则与方法

Materials genome engineering data—
Metadata standardization principle and method

2022-08-31 发布

2022-11-30 实施

中关村材料试验技术联盟

发布

前 言

本文件参照 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。

本文件由中国材料与试验团体标准委员会材料基因工程领域委员会（CSTM/FC97）提出。

本文件由中国材料与试验团体标准委员会材料基因工程领域委员会通则技术委员会（CSTM/FC97/TC01）归口。

全 国 标 准 发 布 使 用

引 言

材料基因工程通过数据驱动模式来加速材料研究，而当前材料数据离散分布、多源异构的现状特点使得数据的访问、获取、集成和再利用极为不便，严重制约了材料基因工程的高效推进。为此，中国材料与试验团体标准委员材料基因工程领域委员会（CSTM/FC97）制定了 T/CSTM 00120《材料基因工程数据通则》（以下简称《通则》），以 FAIR（Findable, Accessible, Interoperable, Reusable, 可发现、可获取、可互操作、可再利用）原则为目标来对材料数据进行标准化。《通则》将数据分为样品信息、原始数据（未经处理的数据）与衍生数据（经分析处理得到的数据）三类，以操作（样品制备/表征/计算/数据处理）为条目单位，在每条数据中需包含数据唯一标识、丰富的元数据记录及原始结果数据来保证其符合 FAIR 原则要求。这里，样品可以是实验产生的实物，也可以是经计算产生的虚拟物。同理，原始数据可以来自于表征或是直接的测量，也可以通过模拟计算产生。《通则》明确了材料基因工程数据的标准化内容和基本方向，仍需明确具体的标准化原则与方法作为标准编制的指导依据。

元数据是用于描述数据的数据，可对数据的产生背景、内容、过程及其它特征进行全面性描述，构成了数据访问和利用的实际承载基础，因此，材料数据标准化实现的关键在于其元数据的标准化。为促进材料数据标准化目标的切实落地和全面性开展，本文件基于《通则》要求，对材料基因工程数据（样品信息/原始数据/衍生数据）的元数据标准制定原则与方法进行明确，确保各元数据标准细则内容完整、形式一致，使其有效保障数据符合 FAIR 原则要求。