

T/GXDSL

团 体 标 准

T/GXDSL —2026

新材料研发与选型应用材料基因工程数据 规范

Data Specification for Material Genome Engineering in R&D, Selection and Application of New
Materials

(工作组讨论稿)

(本草案完成时间: 2026-01-29)

2026 - - 发布

2026 - - 实施

广西电子商务企业联合会 发布

目 次

前 言	II
1 引言	1
2 范围	1
3 规范性引用文件	1
4 术语和定义	2
4.1 材料基因工程	2
4.2 材料科学数据	2
4.3 数据规范	2
4.4 元数据	2
4.5 数据溯源	2
4.6 数据质量	3
5 数据分类与编码规范	3
5.1 数据分类体系	3
5.2 数据编码规则	3
6 数据采集与描述规范	4
6.1 通用采集要求	4
6.2 核心元数据描述	4
6.3 关键数据描述细则	5
7 数据存储与管理规范	5
7.1 存储架构	5
7.2 数据格式	5
7.3 数据组织	6
7.4 数据质量管理	6
7.5 数据溯源管理	6
8 数据共享与交换规范	6
8.1 共享原则	6
8.2 共享数据内容与级别	6
8.3 数据接口	7
8.4 数据标识与引用	7
9 数据安全、伦理与知识产权	7
9.1 数据安全	7
9.2 数据伦理	7
9.3 知识产权	8
10 实施与监督	8
11 附则	8

前 言

本文件依据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。

本文件由广西产学研科学研究院提出。

本文件由广西电子商务企业联合会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

新材料研发与选型应用材料基因工程数据规范

1 引言

新材料研发与选型应用材料基因工程，是集成高通量计算、高通量实验与专用数据库等前沿技术，加速关键核心新材料发现、迭代与工程化应用的战略性技术体系，对支撑我国制造业高端化、智能化、绿色化转型、保障产业链供应链安全具有重要意义。为规范全国范围内材料基因工程活动中数据的采集、存储、管理、共享与应用全流程，保障数据的准确性、一致性、安全性与可复用性，提升国家新材料研发创新效率与科技成果转化水平，夯实国家新材料产业创新生态的数据基础，特制定本标准。本标准旨在建立一套统一、科学、可操作的国家级数据规范，为全国材料基因工程数据生态建设、科技资源统筹共享提供核心技术支撑。

2 范围

本标准规定了我国新材料研发与选型应用材料基因工程活动中，数据采集、描述、存储、管理、共享、交换与应用的基本要求、通用框架及核心技术规范。适用于全国范围内从事新材料基因工程相关工作的高等院校、科研院所、骨干企业、第三方服务机构及行业管理部门，在材料计算模拟、实验制备、性能检测、选型应用与服役性能评价等全链条环节中开展的数据治理活动，为国家层面科技资源统筹、产业协同创新提供合规依据。

3 规范性引用文件

下列文件对于本标准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标准；凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本标准，同时应符合国家最新产业政策与技术规范要求。

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

GB/T 3058-2021 信息技术数据质量评价指标

GB/T 35295-2017 信息技术大数据术语

GB/T 37973-2019 信息安全技术大数据安全管理指南

GB/T 29811.1-2013 材料科学数据第1部分：数据分类与描述

GB/T 39665-2020 材料科学数据元数据

GB/T 40005-2021 材料科学数据共享数据接口通用要求

GB/T 36344-2018 信息技术数据溯源描述模型

GB/T 34078.1-2017 基于云计算的电子政务公共平台技术规范第1部分：系统架构

国家数据安全法、个人信息保护法及新材料产业发展相关国家级政策文件

4 术语和定义

GB/T 35295-2017、GB/T 29811.1-2013界定的以及下列术语和定义适用于本标准，同时契合国家新材料产业创新发展语境。

4.1 材料基因工程

通过融合高通量计算、高通量实验、材料数据库与数据挖掘等前沿技术，系统集成材料组成、工艺、结构与性能全要素数据，实现关键核心新材料快速研发、精准设计、优化迭代与工程化选用的战略性技术体系，是支撑国家科技自立自强的重要技术方向。

4.2 材料科学数据

在材料研究、开发、生产、测试与应用全过程中产生的，反映材料成分、工艺、结构、性能及环境影响等核心信息的原始数据、衍生数据、分析数据和关联数据的总称，是国家战略性科技资源的重要组成部分。

4.3 数据规范

对数据的类型、格式、精度、单位、编码、语义及采集与记录方法等进行的统一界定与强制性或推荐性规定，是保障数据跨领域、跨机构、跨区域互联互通与复用的基础。

4.4 元数据

描述数据的数据，用于定义和说明数据的内容、覆盖范围、质量、管理方式、数据所有者及权属信息等核心要素，是实现数据可发现、可追溯、可管理的关键支撑。

4.5 数据溯源

描述数据来源、产生过程、流转历史、演变关系及责任主体的全链条追踪信息，是保障数据真实性、可靠性与合规性的核心手段。

4.6 数据质量

数据满足国家科技研发、产业应用、行业管理等特定使用要求的能力特性，核心包括准确性、完整性、一致性、时效性、可访问性与安全性，是数据价值转化的前提。

5 数据分类与编码规范

5.1 数据分类体系

为适配国家新材料全产业链研发与应用需求，材料基因工程数据按产生阶段与核心内容分为以下五类，覆盖材料全生命周期关键数据要素：

5.1.1 成分与工艺数据：包括材料化学组成、元素配比、原料技术参数、合成方法、制备工艺关键参数、热处理制度、加工路径及过程控制数据等，是保障材料制备一致性与可重复性的核心基础数据。

5.1.2 结构表征数据：包括晶体结构、相组成、微观组织、缺陷信息、表面形貌、界面特性等，来源于X射线衍射、电子显微镜、光谱分析等标准化表征手段，是关联材料结构与性能的关键桥梁数据。

5.1.3 性能测试数据：包括力学性能、物理性能、化学性能、热学性能、电学性能、磁学性能及服役性能等实验测量结果，需严格遵循国家或国际标准测试方法，是材料选型与工程化应用的核心依据数据。

5.1.4 计算模拟数据：包括第一性原理计算、分子动力学模拟、相场模拟、有限元分析等产生的能量、电子结构、相图、性能预测等数据，是缩短新材料研发周期、降低研发成本的重要支撑数据。

5.1.5 环境与选型数据：包括材料使用环境条件、工况参数、选型决策依据、成本分析、生命周期评价及失效案例数据等，是支撑新材料精准选型、保障装备安全可靠运行与产业绿色低碳发展的重要应用数据。

5.2 数据编码规则

为实现全国范围内数据的统一识别与管理，采用分级分类编码体系，编码结构统一规范为：MGE-[数据大类代码]-[数据子类代码]-[序列号]，编码规则全国通用，便于跨机构数据对接与共享。示例：MGE-CP-02-0001，表示材料基因工程-成分与工艺数据-合成方法类-第1条记录；其中数据大类代码、子类代码由国家相关主管部门统筹统一编制与更新。

6 数据采集与描述规范

6.1 通用采集要求

所有数据采集工作应遵循可追溯、可重复、标准化、合规化的核心原则，严格落实国家数据安全与科技资源管理相关要求。采集所用设备与仪器需定期按国家计量标准校准，并完整记录校准信息与溯源凭证；原始数据需全程留痕、不可篡改，任何数据处理与变换操作须详细记录算法、步骤、参数及责任主体，确保数据可复现。

6.2 核心元数据描述

依据GB/T 39665-2020及国家科技资源共享相关要求，每条数据记录必须包含以下核心元数据，确保数据的可发现、可管理与可复用：

6.2.1 标识符：数据的全国唯一永久性标识，适配国家科技资源标识体系。

6.2.2 标题：简明、准确描述数据核心内容的标准化标题。

6.2.3 贡献者：明确载明数据生产者、提供者、管理者及其所属机构信息，并标注核心责任人员身份及职责。

6.2.4 详细含唯一含号参数（含文献标注格式），确保数据来源全程来源：详细说明数据产生所依托的实验设备（含型号、唯一编号）、计算软件（含名称、版本号）、仿真模型参数或文献出处（按规范格式标注），确保数据来源全程可追溯、可核验。

6.2.5 界定限，建立动态管控机制，确保数据应用的有效性日期：明确界定数据产生日期、入库日期、最后修改日期及有效期限，建立数据时效性动态管理机制。

6.2.6 精准界定类别参数范围阈值边界，明确数据适用范畴覆盖范围：精准界定数据对应的材料体系类别、工艺参数区间、性能指标阈值、应用场景边界等核心信息，明确数据适用范畴与限制条件。

6.2.7 数据格式与量纲：统一采用国际单位制（SI）量纲，明确数据存储的标准化格式，保障数据具备跨平台互操作性。数据格式与量纲：统一采用国际单位制（SI）量纲，明确数据存储的标准化格式，保障数据具备跨平台、跨系统互操作性（interoperability）。

6.2.8 数据质量描述：详细说明数据准确性验证依据、测量不确定度计算方法、置信区间范围，及质量验证的流程与结果判定标准。数据质量描述：详细说明数据准确性验证依据、测量不确定度计算方法与数值、置信区间范围，及质量验证的流程、标准与结果判定结论。

6.2.9 权限信息：明确数据访问、使用、共享的权限边界边界与合规要求，严格衔接本标准第九章关于严格衔接本标准第九章关于知识产权与数据安全的相关的相关管理规定，确保权责清晰。

6.3 关键数据描述细则

6.3.1 成分数据：应明确元素、化合物的标准名称及规范及规范化学式，含量优先采用质量百分比（wt%）（wt%）或原子百分比（at%）（at%）表示，精度需符合国家相关测试标准规定，明确标注规定，明确标注有效位数与测量不确定度数值数值。

6.3.2 工艺数据：须完整记录全流程工艺参数涵盖信息，涵盖温度、压力、时间、气氛、升降温速率、加料顺序等关键参数，明确参数测量方法、精度等级等级及过程控制标准标准，确保工艺过程可精准过程可精准复现。

6.3.3 结构数据：应完整载明表征方法标准名称、仪器型号与唯一唯一编号、测试条件参数参数、分辨率指标指标、标样信息及数据分析方法与软件（含版本）（含版本号），附件须包含原始表征图谱与完整完整数据曲线，确保结构信息可验证、可追溯、可追溯。

6.3.4 性能数据：必须严格遵循国家或国际相关测试标准开展实验工作，完整记录测试标准编号、样品制备工艺与状态参数参数、测试环境条件（温度、湿度等）（温度、湿度等）、测试设备信息及原始曲线数据（如应力-应变曲线、热分析曲线等），确保性能数据具备具备权威性、可比性与可复现性与可复现性。

7 数据存储与管理规范

7.1 存储架构

为支撑国家层面数据统筹管理与共享利用利用，鼓励采用“国家-区域-机构”三级分布式存储架构，实现原始数据、过程数据与与结果数据的分类存储、关联管理与分级管控与分级管控。支持基于云计算、大数据技术的集约化存储方案，严格严格符合GB/T 34078.1-2017相关技术技术要求，确存储系统具备高具备高安全性、高高可靠性、高高扩展性与完善的容灾备份完善的容灾备份能力，核心数据必须必须落实异地备份与冗余存储与冗余存储要求。

7.2 数据格式

7.2.1 鼓励优先采用开放、非专有、标准化的数据格式，其中晶体结构数据采用CIF标准标准格式，分子结构数据采用MOL或SDF格式，通用表格数据采用CSV或JSON格式，原始图谱数据采用国家推荐的标准化格式，保障数据保障数据跨平台、跨机构交互顺畅顺畅、可解读。

7.2.2 对于确需使用专用格式的数据，须同步步提供完整的格式说明文档、数据解析规则及可读取该格式的公开软件（含版本（含版本信息）），确保数据长期可访问、可解读、可解读、可复用。

7.3 数据组织

数据应按照“国家科技项目/产业项目-材料体系-数据类型-批次/版本”的层级逻辑进行标准化目录组织，实现数据的层级化管理、快速检索与精准定位与精准定位。每个数据包必须包含数据文件、元数据文件、数据溯源日志及及README说明文件，其中README文件需明确载明数据使用范围、操作操作注意事项及合规使用使用要求。

7.4 数据质量管理

7.4.1 建立国家层面统一的数据质量评估与管控流程，严格严格依据GB/T 3058-2021设定完整性、准确性、一致性、时效性、安全性等核心校验指标与判定指标与判定规则，构建构建“采集-审核-入库-运维-更新-更新”全流程、闭环式数据质量管控机制。

7.4.2 入库数据必须经过形式审查与技术审核双重校验，审核流程全程留痕、责任可追溯，审核流程全程留痕、责任可追溯：形式审查重点核查元数据完备性、格式规范性与手续合规性；技术审核由领域权威专家对数据的科学性、合理性、准确性及应用价值进行综合综合判定，审核意见作为数据入库的核心依据意见作为数据入库的核心依据。

7.5 数据溯源管理

严格依据GB/T 36344-2018建立数据数据全生命周期溯源信息记录体系，至少完整记录完整记录数据产生源头、所有权变更轨迹轨迹、数据处理与变换全全步骤、关键衍生关系及各环节责任主体信息，确保数据从产生、处理、存储、处理、存储到应用的全程可追溯、可追责。

8 数据共享与交换规范

8.1 共享原则

严格严格遵循“FAIR”（可发现、可访问、可互操作、可重用）核心原则，立足国家新材料产业创新发展大局，在严守守知识产权、数据安全底线与保障底线与保障国家利益的前提下，鼓励各类创新主体有序有序开放共享材料基因工程数据资源，构建全国一体化、协同高效的、协同高效的数据共享生态体系体系。

8.2 共享数据内容与级别

律法要求规定实行三级分类、动态更新与发布结合国家保密法律法规、产业安全要求与知识产权保护规定，共享数据实行三级分类管理，分级标准由国家相关主管部门统筹制定、动态更新与发布实施。

8.2.1 公开级：向社会公众无条件开放的非涉密、非敏感、非商业秘密数据，优先保障基础共性、公益性、公益性数据的公开共享，全面全面支撑全行业创新研发与产业升级研发与产业升级。

8.2.2 受限级：需经数据提供方书面书面申请审批，，或签订正式正式数据使用协议后方可获取的数据，涉及关键技术细节、阶段性研发成果或商业敏感信息，共享范围严格严格限定于合规的科研与产业应用场景。

8.2.3 保密级：涉及国家秘密、核心商业秘密或重大安全的数据，严格遵照国家相关保密法律法规与管理制​​度执行，严禁擅自共享、传播与使用，切实与使用，切实保障国家科技安全、产业安全与信息安全与信息安全。

8.3 数据接口

数据共享交换接口必须严格严格符合GB/T 40005-2021技术技术要求，适配国家数据共享交换平台统一统一技术规范，提供基于API的标准化数据查询、检索与获取服务。接口应支持常见查询语言与数据格式，内置内置身份认证、权限管控、操作审计与异常预警与异常预警功能，确保数据共享交换过程安全、规范、可追溯过程安全、规范、可追溯。

8.4 数据标识与引用

所有共享数据必须赋予赋予永久性唯一标识符，鼓励采用国家认可的数字对象标识符（DOI），纳入国家科技资源标识体系统一管理与追溯与追溯。数据引用格式需严格严格符合国家标准规范，完整包含标识符、数据数据标题、贡献者、发布机构、发布年份、版本号号等核心信息，确保数据成果规范引用、权益可追溯。

9 数据安全、伦理与知识产权

9.1 数据安全

数据全生命周期安全管理严格遵循GB/T 37973-2019及《中华人民共和国数据安全法》《中华人民共和国《中华人民共和国数据安全法》《中华人民共和国网络安全法》》等相关法律法规规定，建立与数据分级分类相匹配的差异化安全防护体系。严格严格落实访问控制、加密传输、数据数据脱敏、操作审计、容灾备份等技术与管理措施，核心数据、敏感数据必须存储于安全可信的存储的存储环境，严防数据泄露、篡改、丢失、丢失与滥用，全面全面保障国家科技安全、产业安全与信息安全与信息安全。

9.2 数据伦理

涉及人体实验、生物材料、生态环境影响或可能引发重大伦理风险的研究数据，其采集、使用与共享必须经国家级或省级伦理审查委员会审查通过并备案并备案，严格保护相关主体的隐私权、知情权与合法权益，恪守科研诚信与伦理准则，严禁任何严禁任何违规采集、、滥用与传播伦理敏感与传播伦理敏感数据的的行为。

9.3 知识产权

数据提供方依法享有其数据的原始知识产权与合法权益，数据共享与使用需严格严格遵循国家知识产权战略相关要求与法律法规规定与法律法规规定。数据使用方应在约定的许可范围与使用场景与使用场景内规范使用数据，引用数据成果时必须注明数据来源与贡献者；数据数据来源与贡献者；数据共享协议应明确约定约定数据使用权限、成果归属、利益分配机制及侵权责任追究条款，保障数据要素合法有序流通与价值转化。

10 实施与监督

由国家工业和信息化部、科学技术部等相关主管部门统筹指导、监督实施、监督实施，广西产学研科学研究院负责具体解释，联合全国新材料领域产学研用协同创新平台、行业协会、行业协会推动标准在全国范围内全面全面落地应用，优先在国家重大科技项目、重点产业工程中推广执行执行。各实施单位（高校、科研院所、企业等）应设立专职数据管理员岗位，明确岗位职责与工作流程，明确岗位职责与工作流程，负责本单位数据规范的落地执行、数据审核、质量管控、安全管理与信息、安全管理与信息上报工作，主动主动接受国家与地方行业管理部门的监督指导。鼓励建立国家级第三方数据质量评价与认证机制，由权威机构定期对本标准实施情况开展开展全国范围内的评估与反馈，及时收集行业诉求、、技术发展新需求与标准执行过程中的问题与标准执行过程中的问题，为标准的修订完善提供科学依据，确保标准的时效性、、适用性与权威性与权威性。

11 附则

本标准由广西电子商务企业联合会负责解释。本标准自发布之日起试行，试行期为一年。试行期满后，根据实施反馈情况进行修订和完善。各相关单位可依据本标准制定具体的实施细则。若本标准与国家新颁布的法律法规或强制性标准有不一致之处，应以国家法律法规和强制性标准为准。本标准所引用的规范性引用文件如有更新，其最新版本适用于本标准。广西电子商务企业联合会将根据技术发展和应

用需求，适时组织对本标准的复审与修订工作，以保障其持续的先进性和适用性。本标准的有效实施，有赖于各级医疗机构、主管部门、技术服务商和各相关方的共同努力，通过规范智慧医院数据互联互通共享技术，推动医疗健康数据资源有效整合与安全共享，提升医疗服务质量和效率，促进智慧医院建设规范化发展，为推进健康中国建设提供技术支撑。
