

ICS XX.XXX

CCS XX

团 体 标 准

T/CSSTI XXXX-YYYY

特色数据目录建设 技术与管理规范

Technical and management specifications for

construction of characteristic data catalog

(征求意见稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

中国科学技术情报学会

发 布

目 次

目 次	1
前 言	2
引 言	3
特色数据目录建设 技术与管理规范	1
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 特色数据目录建设原则	2
5 特色数据目录分类筛选要求	3
6 特色数据目录编制流程	5
7 特色数据目录管理要求	6
附 录 A （规范性附录） 特色数据目录元数据	9
参 考 文 献	12

前 言

本标准按照GB/T 1.1-2020给出的规则起草。

本标准由中国科学技术情报学会提出并归口。

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这些专利的责任。

本标准起草单位：中国科学技术信息研究所、安徽省科学技术情报研究所、吉林省科学技术信息研究所、山东省科学技术情报研究院、浙江省科技信息研究院、厦门市科学技术信息研究院、菏泽市科学技术信息研究所、苏州市科技服务中心、新疆维吾尔自治区科技发展战略研究院、北京数智盈动科技有限公司、海又知信息科技（南京）有限公司、北京嘉韵楷达气象科技有限公司。

本标准主要起草人：张英杰、赵伟、王莉、张越、张玄玄、曹悦、吴绮睿、沈贺文、杨晓辉、张文君、毛刚、陈晓玲、李潇、史筱飞、王衍、钟丽丹、陈卫、李丽婷、郑琦琳、董传民、李士杰、于连生、陈洪梅、陈玲、马燕、张唯、宋刘陈、胡芳槐、魏永超。

本标准首次发布。

引 言

本文件定义了特色数据目录的技术与管理规范，旨在进一步加强和规范特色数据资源建设和管理，提高特色数据资源开放共享水平，保障特色数据资源安全，更好支撑国家科技创新、情报服务和数据管理工作。特色数据目录数据资产化活动中的一项重要功能，它可以有效识别数据资产，为数据要素化和数据交易流通提供关键支持。

特色数据目录建设 技术与管理规范

1 范围

本文件规定了特色数据目录建设的基本定义、分类筛选要求、编制流程和服务管理要求。本文件提供了建立特色数据集、围绕构建专题数据形成特色化应用及开展高质量数据建设等活动的指导。

本文件适用于有特色数据目录建设需求的科技管理和研究机构、企业等创新主体、科技和产业数据供应商、政府管理部门、智库等。

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 4894 信息与文献 术语

GB/T 5271.1 信息技术 词汇 第1部分：基本术语

GB/T 18391.1 信息技术 元数据注册系统 (MDR) 第1部分：框架

GB/T 42108 信息与文献 组织机构元数据

3 术语和定义

下列术语和定义适用于本文件。

3.1

特色 characteristic

数据目录在收集、处理、利用、服务等过程中形成的差异化特征或优势。包括专业化特色、精细化特色、领域特色、热点特色、场景特色等某个特色或多个特色。

3.2

数据目录 data catalog

以核心元数据为主要描述方式，按照数据分类体系对数据进行有序排列的一组信息。能够将所有的数据资产整合成清单，帮助数据专业人员为任何分析和业务目的找到相关数据，并根据描述数据的特征，对数据进行快速检索、定位与获取。

3.3

数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

[来源: GB/T 5271.1-2000, 定义1.1.2]

3.4

元数据 metadata

定义和描述其他数据的数据。

[来源: GB/T 18391.1-2009, 定义3.2.16]

3.5

主题领域 subject field

专门知识的领域。

[来源: GB/T 4894-2009, 定义4.1.1.1.7]

3.6

数据目录管理主体 data catalog management principal

被授权对数据目录进行操作的组织。

4 特色数据目录建设原则

4.1 真实性

为确保特色数据目录的真实性，建设应遵循以下原则：

- a) 数据源验证：应对数据源进行验证，确保数据源可信，包括核实数据提供者的信誉和声誉、检查数据的来源和采集方法、调查数据产出方式是否符合研究、管理和生产等活动实际以及评估数据的准确性和一致性。
- b) 数据质量管理：应施行严格的数据质量管理措施，包括数据清洗、去重、标准化和验证等步骤，建立数据质量检查机制和指标，定期监测数据的质量，并及时进行修复和更新。
- c) 安全和隐私保护：应考虑安全和隐私保护，确保数据在采集、存储、传输和使用过程中得到妥善保护，符合相关法规和标准。应采取必要的安全措施，如数据加密、访问控制和身份验证，以防止未经授权的访问和数据泄露。
- d) 独立验证和审计：应进行独立验证和审计，包括通过独立的第三方机构或内部审计团队对数据目录进行评估，以验证数据的真实性、发现潜在的问题和风险并提供改进建议。

4.2 及时性

为确保特色数据目录的及时性，建设应遵循以下原则：

- a) 明确目标 and 需求：应定义特色数据目录的目标和需求，包括确定所需的数据类型、数据源、数据格式、数据更新频率等方面的要求。应在数据目录管理主体规定的程序和期限内，完整地提供特色数据目录元数据并及时更新；
- b) 自动化数据收集和更新：应使用抓取工具、API集成或实时数据同步等自动化工具和技术，确保数据能够自动从源头收集并定期更新到目录中。减少人工操作干预，以提高数据更新的速度和准确性；
- c) 设计合理的数据处理流程：应确保数据处理流程中的清洗、转换和集成高效进行，包括数据质量检查、标准化、去重等步骤；
- d) 建立实时监控和报警机制：应及时发现数据更新失败、数据异常或数据质量问题，并及时采取纠正措施，帮助保持数据目录减少潜在错误；

- e) 定期审查和更新：应通过定期审查评估目录的性能和及时性并识别改进机会。根据反馈和用户需求，对目录进行更新和优化，以确保其与不断变化的数据环境保持同步。

4.3 科学性

为确保特色数据目录的科学性，建设应遵循以下原则：

- a) 数据收集科学：应确保数据的收集过程采用科学方法，包括明确定义数据收集的目的、采样方法和样本规模，并确保数据收集过程具有可重复性和可比性；
- b) 数据验证与验证过程透明：应确保数据的验证过程是科学的，并且透明可追溯，包括明确验证方法、标准和流程，以及数据验证的结果和可信度评估；
- c) 数据处理方式科学：数据处理应按照标准加工处理规范进行，包括采用合适的数据分析方法和模型，确保可发现性、可获得性、互操作性和可重复利用性；
- d) 数据存储方式科学：数据应该以科学合理的格式存储，便于发布和使用；
- e) 专业团队与专家咨询：应组建具备专业知识和技能的团队并充分利用专家咨询，以应用科学方法和最佳实践来指导数据目录的建设过程，为特色数据目录提供领域知识和科学见解。

5 特色数据目录分类筛选要求

5.1 专业化特色

专业化特色是指在特定领域或行业中，数据目录具备独特的专业知识经验以及对该领域或行业相关数据的深入理解和专业处理能力，以形成深度竞争优势。专业化特色的筛选特征包括以下四个维度：具备长期专注并深耕于产业链某一环节相关数据能力、对特定领域或行业的广泛知识和专业经验、对特定领域或行业的数据分类和标准化能力、注重数据质量能力。例如，在社会舆情分析领域，某些机构可能具有独特的数据来源和分析能力，可以提供更准确、更实用的预警舆情数据。在人才资源领域，某些机构可能有更丰富、更深入的人才数据库，可以提供更多元、更精准的高层次专家人才数据。

表 1 专业化特色的筛选特征

维度	释义
长期专注深耕	积累大量与产业链环节相关数据，并具备对该领域或产品的深刻理解
	提供更全面、详尽和深入的数据洞察，满足用户的特定领域或行业需求
具备领域专业知识和经验	了解该领域的核心概念、行业趋势、市场规模和竞争环境
	理解数据的背景和含义
	为用户提供对数据的深入解读、领域见解和定制化的分析
	提供实用数据专家合作支持
数据分类和标准化	能够根据领域需求将数据进行合理的分类和组织，具备一定的可比性
	制定适合特定领域的数据标准和规范
	进行严格的数据验证和清洗
数据质量保障	数据来源可靠
	采用科学的数据采集、整理和验证流程
	实施数据质量监控和反馈机制，及时纠正和改进数据质量问题

5.2 精细化特色

精细化特色是指特色数据目录在技术、产品、管理上实现精细化，能够提供高质量数据，并根据用户的具体需求提供个性化、深入和易于使用的数据服务，帮助用户做出更明智的决策和战略规划，实现组织竞争和运营效率提升。精细化特色的筛选特征包括以下四个维度：具备数据及时供给能力、满足用户个性化需求能力、深入挖掘数据的能力、数据更易于理解和分析应用能力。例如，在金融行业，精细化特色的数据目录需要及时提供细分市场数据、个股数据分析、宏观经济数据、高频交易数据等，以图表和报告等直观、清晰的方式呈现各市场的动态和特点，支持交易策略的制定和执行。

表 2 精细化特色的筛选特征

维度	释义
数据时效性	数据及时收集、整理和发布
	数据反映快速变化的市场和行业情况
	保持数据及时更新
数据细分和个性化	支持对数据进行更细致的划分和个性化定制
	支持导出特定的数据视图、指标和报告
	提供用户容易访问和利用相关数据的方式
数据深度和广度	提供更多维度和更详细的数据，帮助用户深入分析
数据易用性	利用数据可视化技术等，以用户友好的方式呈现数据

5.3 领域特色

领域特色是指数据目录具备与某领域核心关注点和业务需求紧密相关的特定数据，可以满足用户在行业、地区和主题领域做出更具针对性和有价值的决策。领域特色的筛选特征包括以下三个维度：针对特定行业、特定地理位置特点、特定主题提供独特的数据服务能力。例如，医疗健康领域关注医疗机构数据、疾病统计和医疗技术发展数据；能源行业关注能源消耗、发电量和可再生能源数据等；在人工智能、新能源、数字经济等战略性新兴产业领域，提供相应的数据资源和分析服务，为相关行业和企业提供参考和支持；对粤港澳、长三角地区类脑智能、量子信息等行业提供最新数据，利用数据进行提前预测、布局 and 规划。

表 3 领域特色的筛选特征

维度	释义
行业领域	在某些传统产业、具有战略意义或发展前景的产业、地方鼓励的支柱和优势产业未来产业领域提供特色数据服务
地区领域	针对不同地区具有的不同经济、社会和环境特点而提供特色数据服务
主题领域	特色数据目录这些主题可能是当前热门的话题，社会关注的议题或特定用户群体感兴趣的领域

5.4 热点特色

热点特色是指特色数据目录与当前社会、经济或科技领域关注度较高的热门议题、趋势或事件相关，并提供相关的数据支持和洞察。热点特色的筛选特征包括以下四个维度：在政策热点领域、新兴技术领域、前沿研究领域及当前社会热点话题领域提供特色数据服务能力。例如，为一带一路、RCEP、公共卫生、社会事件等提供相应的数据资源和分析服务，帮助用户了解最新动态和趋势。

表 4 热点特色的筛选特征

维度	释义
政策热点	特色数据目录提供政策制定和评估所需的相关数据资源和分析服务,为政策制定者和研究人员提供参考和支持
技术热点	特色数据目录提供新兴技术领域相关的数据支持。
研究热点	特色数据目录能够提供与前沿研究领域动向、趋势相关的数据支持。
话题热点	特色数据目录提供社会热议的话题分析所需的数据服务,有效提高事件响应度。

5.5 场景特色

场景特色是指构建符合应用场景的数据目录使服务过程中各类数据与场景有效衔接,完成基于特定场景、用户角色的服务匹配、组合和推荐,提升数据服务的用户效能。场景特色的筛选特征包括以下三个维度:数据服务赋能决策场景、赋能产业场景、赋能智库场景。赋能决策场景的典型场景如,政策制定、区域规划、招才引智、招商引资等。赋能产业场景典型场景如,产业监测、产业规划、技术对接、投融资服务、企知识产权评议或预警等。赋能智库场景的典型场景如,研究辅助、专家协同等。

表 5 场景特色的筛选特征

维度	释义
赋能决策场景	特色数据目录为政府提供国家政策和相关数据资源和数据支持,通过产业政策制定,调节产业资源分配,有效支撑业务、战略和投资决策过程。
赋能产业场景	特色数据目录为产业用户提供高质量数据服务,帮助产业参与者了解行业动态、市场趋势、竞争情况等关键信息,快速建立服务提供商、资源、渠道间的协作关系,促进产业高质量发展。
赋能智库场景	特色数据目录为智库专家供数据支持,帮助进行深入研究分析,通过专家专业性知识的数字化加工整合有效提高知识的传播和利用效率,促进智库研究和分析的数字化转型。

6 特色数据目录编制流程

特色数据目录编制体现多方协同参与数据编目视角下流程,包含数据资源调查、数据资源编目、数据应用等内容。其中涉及的主要角色与职责如下。

- a) 数据提供主体: 进行数据目录内容规划、数据资源梳理、数据目录注册;
- b) 数据目录管理主体: 进行数据目录元数据接收审核、数据目录收录凭证出具、数据目录发布、数据目录管理、数据目录服务提供;
- c) 数据使用主体: 进行数据目录服务获取。

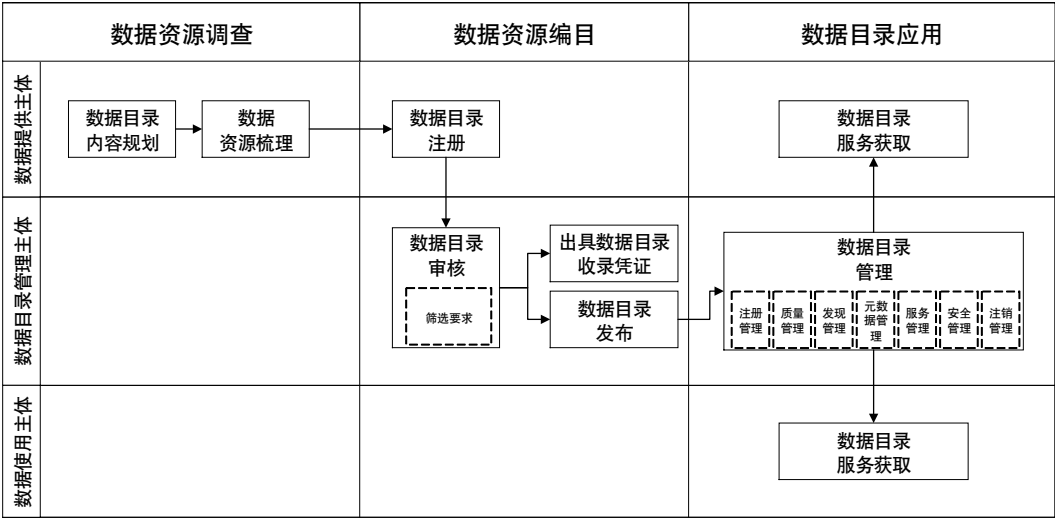


图 1 特色数据目录编制流程

7 特色数据目录管理要求

7.1 特色数据目录注册管理

特色数据目录提供对目录元数据进行注册、存储、发布等，以便用户可以方便地搜索、浏览和访问特色数据的相关信息。

- a) 元数据注册：允许数据提供者注册其数据集的元数据信息。元数据包括关于数据集的详细描述和特征；
- b) 元数据存储：将注册的元数据存储存储在统一的数据目录数据库中，以便进行统一管理和检索；
- c) 元数据发布：通过数据目录的网站、应用程序或接口将元数据信息发布给用户，用户可浏览目录中的数据集列表，并查看每个数据集的元数据信息。

7.2 特色数据目录质量管理

特色数据目录提供数据目录质量规范化管理，以提高目录使用便捷性和反馈数据准确性，更好地为用户服务。

- a) 数据目录筛选管理：帮助快速找出所需数据，在特色数据目录中建立通用筛选条件，筛选出代表性结果，有效融合不同数据源信息，提高数据可用性及后续利用便捷性；
- b) 数据规范化管理：规范化处理复杂数据格式，避免因数量级和量纲差异而造成数据准确度产生偏差，在进行数据分析时可快速获取标准化数据，有效提高数据资源利用率；
- c) 目录层级管理：细化各目录层级并对各层级目录下的内容质量进行评估和维护，对类目下的具体划分标准进行管理，使得用户能够快速寻找所需领域、产业、分工等不同维度的信息；
- d) 目录更新管理：定期进行维护和更新数据目录元数据，在系统中设置相关的更新频率，有效提醒管理者更新数据。

7.3 特色数据目录发现管理

数据目录发现管理提供对高价值数据发现的导航功能。

- a) 搜索结果管理：提供数据搜索、按相关度进行结果排序并提供相应信息，通过建立标引等方式，使用户能够快速访问检索结果以及历史记录，实现高效检索；
- b) 信息检索管理：提供专业化检索工具和检索方式，可通过简单检索、专业检索等复合策略使得系统反馈，使用先进技术手段更新检索技术，以提高用户体验满意度。

7.4 特色数据目录元数据管理

特色数据目录元数据管理对元数据的收集、组织和维护以保证元数据质量，元数据应符合附录A的规定。

- a) 数据对象管理：对数据对象进行统一化管理，通过对数据对象按标准或规则进行分类存储并在相关数据对象间建立联系，支持后续进行数据分析时快速调用、跳转和查询数据，提高数据利用率。支持主题专家通过业务术语表、标签、关联、用户自定义注释、分类、评级等形式贡献业务知识，对元数据进行及时更新；
- b) 数据交互管理：数据交互管理过程中通过数据信号匹配、多方验证等方式进行核验，完成数据交互过程。

7.5 特色数据目录服务管理

数据目录服务管理提供检索、查询、共享、交换、分析、可视化等数据服务的管理，结合多种先进技术，实现高性能数据服务能力。

- a) 协同服务：
 - 1) 数据协同：应形成数据协同网络，包括不同部门、单位、地区间的数据共享与协同建设，实现数据资源共享、互通互联；
 - 2) 产业协同：应协同产业发展，推动各要素合理配置和产业有机融合，整合产业链上下游相关企业、生产要素、资源数据，实现产业协同发展；
 - 3) 专家协同：应对专家分领域进行数据整合，对专家优势研究方向和数据进行融合，有利于数据和专家多维度的交互分析。
- b) 智能服务：
 - 1) 数据智能：利用大数据、人工智能技术对数据进行挖掘、分析和利用等智能化处理，有效提取知识，并通过建立模型、推理等方式对现有问题提供解决方案和实现预测，提高数据目录建设的智能化程度；
 - 2) 决策智能：将数据目录建设与决策支持系统结合，通过对数据进行测算和预测提供智能化决策支持；
 - 3) 服务智能：利用人工智能技术实现数据目录建设自动化、管理系统化、服务智能化，在用户检索时自动分析识别用户隐性需求，主动地提供智能化服务。
- c) 运营服务：
 - 1) 策略规划：根据实际需要和发展规划，制定建设方案和实施计划，明确建设目标任务。根据特色数据目录的特点和受众群体，选择适合的推广渠道；
 - 2) 宣传管理：加强特色数据目录的相关信息传播，提高公众对数据开放的认知理解，增加使用的信心和积极性，鼓励企业、学术机构等优质内容提供者入驻特色数据目录，提高用户使用率；
 - 3) 品牌维护：以用户体验为中心进行内容更新、质量保持、服务质量和水平提升，提高使用和分享目录的意愿，有效维持品牌生命力。

7.6 特色数据目录安全管理

特色数据目录安全管理提供数据目录的安全保障。

- a) 访问控制：根据用户属性进行数据查看、修改等权限管控，限制特色数据目录的访问权限。敏感数据的访问设定身份验证和授权机制，避免恶意修改带来的不利影响；
- b) 数据加密：采用强密码和加密算法对特色数据目录中的敏感数据进行加密，包括数据存储和数据传输过程中的加密；
- c) 审计日志：记录特色数据目录的操作日志和访问日志，追踪和检测异常行为，及时发现和应对安全威胁；
- d) 数据备份和迁移：定期备份特色数据目录的数据，确保数据的可恢复性。

7.7 特色数据目录注销管理

特色数据目录注销管理对数据目录进行合理的注销处理。

- a) 制定注销政策：制定明确的数据目录注销政策和程序，明确注销的条件、流程和责任方，并确保政策符合相关法规和合规要求。
- b) 数据清除和销毁：在进行注销后，彻底清除和销毁存储在数据目录中的数据，使用安全的数据清除方法，确保数据无法恢复。
- c) 系统关闭和访问控制：在注销数据目录后，关闭相应的系统和服务，以防止未经授权的访问，更新访问控制策略，撤销已授权用户的访问权限。
- d) 留存合规记录：根据相关法规和合规要求，保留必要的注销记录和日志。记录注销的时间、原因和相关的操作细节，以备审计和合规检查。

附录 A
(规范性附录)
特色数据目录元数据

特色数据目录元数据应保证数据目录要素完整、内容规范准确，参考《政务信息资源目录编制指南》中政务信息资源核心元数据相关内容，结合DB33/T 2349-2021《数字化改革 公共数据目录编制规范》中提及到的数据目录编制内容，对特色数据目录编制过程中的数据目录要素进行梳理，并对其属性进行补充，以服务于特色数据目录内容展示、浏览与检索。数据目录元数据建议包含以下要素：

- a) 数据资源标识符；
- b) 数据名称；
- c) 数据特色分类；
- d) 科技要素分类；
- e) 数源单位；
- f) 数据量；
- g) 数据存储容量；
- h) 数据摘要；
- i) 数据格式；
- j) 所属信息系统名称；
- k) 更新频率；
- l) 数据版本
- m) 创建日期；
- n) 修改日期；
- o) 数据标签；
- p) 数据有效期限；
- q) 联系邮箱；
- r) 数据项描述：
 - 1) 数据项名称；
 - 2) 数据类型；
 - 3) 数据长度；
 - 4) 字段描述；
 - 5) 共享属性；
 - 6) 开放属性；
 - 7) 是否主键。

表 A.1 特色数据目录元数据规范表

要素名称	说明	
1.数据资源标识符	定义	数据目录的唯一不变的标识字符。
	数据类型	字符型。
2.数据名称	定义	缩略描述数据目录内容的标题。
	数据类型	字符型。
3.数据特色分类	定义	说明特色数据目录的特色分类信息。
	数据类型	字符型。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/366040103145010215>