

能源监管服务数字化平台
数据支撑技术要求

Digital platform for energy regulatory services—Technical requirements for data
support

地方标准信息服务平台

2023 - 12 - 26 发布

2024 - 03 - 25 实施

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 2

5 数据支撑处理流程 2

6 数据存储与分析 3

7 数据治理 7

8 数据发布 9

参考文献 10

地方标准信息服务平台

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。

本文件由山西省能源局提出、组织实施和监督检查。

山西省市场监督管理局对标准的组织实施情况进行监督检查。

本文件由山西省能源标准化技术委员会归口。

本文件起草单位：山西大数据产业发展有限公司、华为技术有限公司、山西省能源发展中心。

本文件主要起草人：张永森、孙宇彤、崔国栋、王茂盛、阳军、侯秉让、杜青、邵国荣、兰世忠、杨海鹏、赵芮、王晓宇、杨杰、宋兰兰、贾灏然、吉凤鸣、刘耀宏、刘泱、康杰、王艳、李峰。



能源监管服务数字化平台 数据支撑技术要求

1 范围

本文件规定了能源监管服务数字化平台的数据支撑处理流程以及数据支撑相关的数据存储与分析、数据治理、数据发布等方面的技术要求。

本文件适用于能源监管服务数字化平台的建设及实施工作。

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 35295—2017 信息技术 大数据 术语
GB/T 37721 信息技术 大数据分析系统功能要求
GB/T 37722 信息技术 大数据存储与处理系统功能要求
GB/T 40685—2021 信息技术服务 数据资产 管理要求
DB14/T 2936—2023 能源监管服务数字化平台 通用要求
DB14/T 2939—2023 能源监管服务数字化平台 数据采集要求
DB14/T 2940—2023 能源监管服务数字化平台 数据发布规范

3 术语和定义

GB/T 35295—2017、GB/T 40685—2021、DB14/T 2936—2023 确立的以及界定的术语和定义适用于本文件。

3.1

能源监管服务数字化平台 Digital platform for energy regulatory services

在自动化、信息化、智能化等相关技术的支撑下，对能源领域相关信息进行采集、整合、分析，并进行充分应用，有效实施全过程动态监管，并能为能源企业、产业链上下游企业等提供智能化服务的数字化智慧平台。

[来源：DB14/T 2936—2023， 3.1]

3.2

数据资产 data asset

合法拥有或者控制的，能够进行计量的，为组织带来经济和社会价值的数据资源。

[来源：GB/T 40685—2021， 3.1]

3.2

数据仓库 data warehouse

在数据准备之后用于永久性存储数据的数据库。

[来源：GB/T 35295—2017， 2.1.35]

3.3

数据湖 data lake

一种在系统或存储库中以自然格式存储数据的方法。

4 缩略语

下列缩略语适用于本文件。

ACID: 原子性、一致性、隔离性、持久性 (Atomicity、Consistency、Isolation、Durability)

API: 应用程序编程接口 (Application Programming Interface)

CPU: 中央处理器 (Central Processing Unit)

DDL: 数据定义语言 (Data Definition Language)

IoT: 物联网 (Internet of Things)

Jar: Java平台上的归档文件格式 (Java Archive)

JSON: Java脚本对象描述 (JavaScript Object Notation)

MD5: 信息摘要算法 (Message-Digest Algorithm)

MPI: 消息传递窗口 (Message Passing Interface)

MR: 分布式运算框架 (Map Reduce)

SLA: 服务等级协议 (Service Level Agreement)

SQL: 结构化查询语言 (Structured Query Language)

XML: 可扩展标记语言 (Extensible Markup Language)

5 数据支撑处理流程

5.1 应基于能源监管服务需求，按照 DB14/T 2936—2023 对数据支撑的要求，将能源监管服务相关数据，统一存储并进行数据分析，然后通过 API 发布或数据资产发布方式，实现数据资产在各组织机构或监管部门的共享，数据支撑处理流程应符合图 1 的要求。

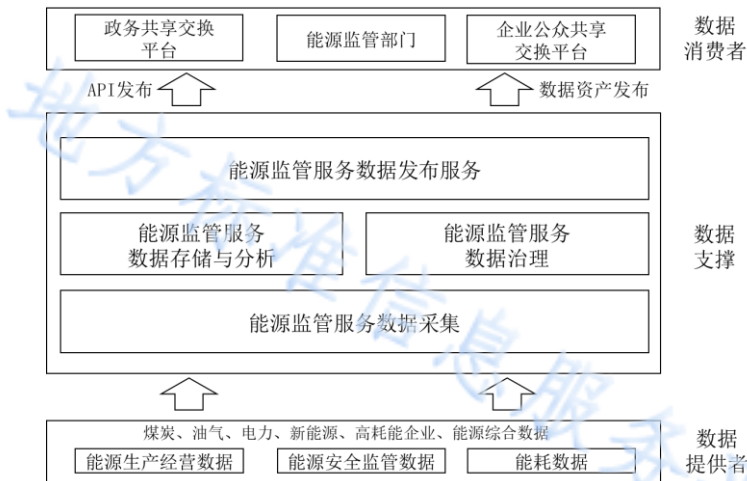


图 1 能源监管服务数字化平台数据支撑处理流程图

5.2 能源监管服务数字化平台应根据监管服务需要，采集各个系统的数据，数据采集内容及数据采集方式应符合 DB14/T 2939—2023 的要求。

5.3 能源监管服务采集的数据进行预处理之后，应传输到云平台集中存储，并建立统一数据湖，实现能源生产经营数据、能源安全监管数据、能耗数据等数据统一存储，需要提供能源监管服务的数据存

储能力，数据存储方式包括但不限于分布式文件存储、数据湖存储、分布式数据仓库、时序数据库等，并符合 GB/T 37722 的要求。

5.4 能源监管服务数字化平台应实现数据治理，提供数据集成、规范设计、数据开发、数据质量、数据资产、数据安全、数据服务等功能，基于能源监管服务需要，整理并生成数据资产，包括但不限于贴源层、整合/明细层、主题链接层、集市层，并通过数据发布，开放给上层不同应用系统或共享交换平台调用。

5.5 能源监管服务数字化平台应提供数据分析能力，支撑能源监管服务各主题库、专题库和指标的生成，数据分析能力包括但不限于批处理分析、流处理分析、流批一体融合分析、交互分析、搜索分析、事件分析、虚拟测点计算分析、时序分析等，并符合 GB/T 37721 和 GB/T 37722 的要求。

5.6 能源监管服务数字化平台应满足能源监管服务的数据发布需求，数据发布的流程及发布的数据内容符合 DB14/T 2940—2023 的要求，数据发布方式应满足本文件的技术要求。

5.7 能源监管服务数字化平台应优先选择符合国家自主可控和安全可信相关要求的全栈自主创新技术。

6 数据存储与分析

6.1 数据存储

6.1.1 分布式文件存储

结合能源监管服务需求，通过分布式文件存储，用于存储能源监管服务文件数据，分布式文件存储要求如下：

- a) 应提供文件的上传、下载、读写、复制、移动、删除、访问控制等功能；
- b) 应提供容错机制和系统高可用机制，包括数据块的备份、系统宕机恢复、灾备恢复等功能；
- c) 应提供文件块的完整性校验和同步功能，保证数据的完整性与一致性；
- d) 应提供存储模块的弹性扩展功能，支持存储节点的扩缩容；
- e) 应提供文件系统级别的数据压缩和解密功能；
- f) 应提供快速检索功能，支持数据资源的统一检索、编目、增加和删除操作；
- g) 应提供文件的搜索、批量操作、回收站、快照等功能；
- h) 宜提供小文件合并成大文件集中存储的功能；
- i) 宜提供存储资源隔离能力，能够通过盘配置目录的存储空间及文件数量实现逻辑隔离；
- j) 宜提供元数据的快速迁移功能，能够复制目录级文件元数据信息实现元数据快速迁移。

6.1.2 数据湖存储

结合能源监管服务需求，通过数据湖存储，提供分布式文件系统之上的ACID数据库能力，支持数据的高效更新、增量查询等能力，数据湖存储要求如下：

- a) 应提供结构化数据的分布式存储机制，实现数据存储的可扩展性；
- b) 应提供多样化的API接口满足各类上层应用的使用需求；
- c) 应提供一种高效的数据组织格式，实现历史数据的高效更新、快速删除能力；
- d) 应提供湖内结构化数据的事务能力，以保证数据的一致性、完整性；
- e) 应提供湖内数据的增量查询能力，支持实时数据分析；
- f) 应提供自动化的小文件合并能力，自动管理文件大小和布局，以优化数据访问速度；
- g) 应提供自动化的历史数据清理能力，以防止数据过度膨胀影响文件系统健壮性；
- h) 应支持多样化的计算引擎和存储引擎，能够平滑的在引擎之间迁移；

- i) 应支持流批融合的计算能力，实现同一个作业内完成批量、增量的数据处理能力。

6.1.3 分布式数据仓库

结合能源监管服务需求，通过分布式数据仓库，提供数据集市能力，用于存储能源监管服务经过数据治理之后的主题联接层和指标等数据，分布式数据仓库要求如下：

- a) 应提供基于无共享的分布式大规模全并行处理的关系型数据库；
- b) 应提供满足SQL语法，包括存储过程、自定义函数、序列、触发器等高级语法功能；
- c) 应提供保障全局强一致性事务能力，可以针对单表和多表并发插入、更新、删除等操作
- d) 应提供以表为单位的行、列存储引擎，用户可指定表的存储格式，满足不同场景业务需要；
- e) 应提供自适应压缩算法，支持可调整压缩级别，压缩级别越高，压缩能力越强；
- f) 应提供在线弹性伸缩功能，扩容过程中数据持续可查询；
- g) 应提供多租户管理功能，租户间CPU、内存、IO等资源隔离，相互不干扰；
- h) 应提供图形化安装部署、补丁升级、扩容缩容、备份恢复、巡检告警、事件管理、日志管理、参数设置、集群启停、节点替换等集群管理功能；
- i) 应提供图形化主机监控、数据库监控、性能监控、TOP SQL、会话监控、作业监控等集群监控平台；
- j) 应提供图形化故障诊断（慢主机）、性能优化、SQL熔断、SQL诊断、负荷分析报告功能；
- k) 宜提供范围、列表等分区功能，分区键支持整型、字符型、时间型等数据类型；
- l) 宜提供行存、列存表B-Tree索引功能，支持函数/表达式索引，提供查询性能。

6.1.4 时序数据库

结合能源监管服务需求，针对IoT数据，时序数据库要求如下：

- a) 应支持按存储组方式对数据进行管理；
- b) 应支持数据按照时间维度分级存储管理；
- c) 应支持分钟级弹性伸缩；
- d) 应支持高性能多维聚合查询；
- e) 应支持亿级时间线。

6.2 数据分析

6.2.1 批处理分析

结合能源监管服务需求，通过批处理分析功能，用于实现高性能的离线批处理作业运行，批处理分析要求如下：

- a) 应支持多种数据类型的离线分析，包括结构化和非结构化数据；
- b) 应支持离线计算任务进度与状态的实时上报；
- c) 应支持多节点离线任务的联动执行；
- d) 应提供多种语言的API，以满足不同应用的开发需求；
- e) 应支持作业调度；
- f) 应支持分散-聚集的处理方式；
- g) 应支持批处理计算框架运行在分布式资源管理之上；
- h) 应支持SQL、函数式编程、图计算等方式处理数据；
- i) 应支持在线查看任务运行日志；
- j) 应支持任务内部子任务的失败重试，以及当运行节点出现异常时能够恢复计算结果。

6.2.2 流处理分析

结合能源监管服务需求，通过流处理分析功能，实现能源监管服务IT数据实时处理能力，流处理分析要求如下：

- a) 应支持多种数据源中获得实时流数据，完成高吞吐、低延时的实时计算，并将结果输出到消息队列或者进行持久化；
- b) 应提供用户级别的访问控制功能，支持对消息处理任务进行创建、浏览、终止、激活、去激活等操作，并对用户级别操作记录审计日志；
- c) 应支持采用滑动窗口的方式实时分析任务，时间窗口大小可任意调节；
- d) 应支持容错机制，在出现故障的情况下，系统具备恢复中间计算结果重试失败任务的能力；
- e) 宜支持直接读取前置任务计算结果作为数据源的方式来编排作业，实现复杂作业拆分的同时保证计算的实时性。

6.2.3 流批一体融合分析

结合能源监管服务需求，通过流批一体融合分析功能，可执行批处理作业和流处理作业，实现能源监管服务IT、OT融合数据流批一体融合处理能力，流批一体融合分析要求如下：

- a) 应支持流批融合一体查询SQL语言；
- b) 应支持多种场景下的流式SQL，如位置信息分析等；
- c) 应支持常用时间窗口，包括跳跃窗口、滑动窗口等；
- d) 宜支持基于SQL语言的批、流数据的模式识别；
- e) 宜支持流批融合的计算引擎；
- f) 宜支持事件驱动的流处理，降低处理延迟；
- g) 宜支持处理乱序、超期数据结果旁路输出；
- h) 宜支持对复杂任务的调度，如支持深度学习的训练、MPI任务。

6.2.4 交互分析

结合能源监管服务需求，通过交互分析功能，适用于海量数据、中高并发，多个租户共享查询引擎的场景，交互分析要求如下：

- a) 应支持对于数据湖中结构化数据的准实时查询；
- b) 应支持统一标准SQL查询语言；
- c) 应支持基本的数据库、模式、表、视图、函数等对象的创建、删除、描述等SQL语句；
- d) 应支持读取各种类型数据格式文件；
- e) 应支持管理节点的高可用保障能力；
- f) 应支持基于分布式调度引擎部署作业的能力；
- g) 宜支持动态调整资源占用情况，在查询较少时自动释放过剩资源；
- h) 宜支持基于视图建立一个实体化的表副本；
- i) 宜支持对于不同的数据源关联查询的能力；
- j) 宜支持多个数据湖之间关联查询的能力。

6.2.5 搜索分析

结合能源监管服务需求，通过检索分析功能，实现能源监管服务全文检索分析能力，搜索分析要求如下：

- a) 应提供基于索引的海量数据全文检索能力；
- b) 应支持独立的索引结构数据存储格式，满足分布式存储的可靠性、一致性要求；

- c) 应支持索引数据压缩存储，对索引数据结构进行特定优化的压缩方式实现高压缩比；
- d) 应支持文档的嵌套查询能力；
- e) 应支持向量化检索能力。

6.2.6 事件分析

结合能源监管服务需求，通过用户自定义事件，可以设置能源监管服务触发条件、结束条件以及各种复杂的计算逻辑，实现从数据流中捕获满足规则的数据记录，生成一个事件，事件分析要求如下：

- a) 应支持用户自定义事件类型、事件规则；
- b) 应支持复杂事件规则由多个算子构成，各个算子通过连接线串联起来，适用于比较复杂的业务场景；
- c) 应支持预集成有状态聚合计算、过滤算子、聚合计算、字段计算、数据过滤、数据关联等工业主流算子；
- d) 应支持预集成表达式函数，包括不限于数学计算、逻辑计算、类型转换等函数
- e) 应支持事件规则图形化、低代码配置；
- f) 应支持查看历史事件记录；
- g) 应支持事件结果实时开放。

6.2.7 虚拟测点计算分析

结合能源监管服务需求，针对 IoT 海量设备数据的特点，提供属性级实时准确的计算能力，支持以物实例的属性数据作为输入，同时将输出结果写入到物实例的属性，让物实例在承载设备上报的测量数据的基础上，同时能承载分析计算后的指标数据，虚拟测点计算分析要求如下：

- a) 应支持转换计算，对单个或多个参数的值进行常规数学计算，表达式中不能使用系统的时序聚合函数；
- b) 应支持聚合计算，对单个或多个参数的值进行周期性的时序聚合计算，表达式中可以使用系统的时序聚合函数；
- c) 应支持流计算，引用流作业管理中定义的作业，将输入数据传递给作业进行计算；
- d) 应支持工业通用基础算子预集成，包含不限于输入算子、输出算子、过滤算子、扁平算子、嵌套算子、快照算子等；
- e) 应支持通用基础算子预集成，包含不限于输入算子、输出算子、过滤算子、扁平算子、嵌套算子、快照算子等；
- f) 应支持高级算子预集成，包含不限于去噪算子、张量拆解算子、张量组装算子等；
- g) 应支持SQL语句编写函数；
- h) 应支持函数、任务按照图形化、低代码方式配置。

6.2.8 时序分析

结合能源信息监管要求，提供针对IoT数据的时序洞察/分析能力，帮助用户便捷对历史数据进行时间维度上的分析，时序分析要求如下：

- a) 应支持时序数据多种聚合计算，包括不限于平均值、最小值、最大值、总数、记录数量等5种；
- b) 应支持对所选时序数据按时间做偏移，在预览列表新增一条记录，常用于同一测点不同时间的对比分析；
- c) 应支持时序数据按照阶梯图，最小/最大范围，显示数据点；
- d) 应支持图形化方式查看资产数据、或者资产数据的历史曲线，包括不限于折线图、热力图、散点图等图表；

e) 应支持类型、堆积、共享等等折线图显示方式。

7 数据治理

7.1 数据集成

结合能源监管服务数据集成要求，应支持批量数据迁移和实时数据集成，数据集成要求如下：

- a) 批量数据迁移应提供同构/异构数据源之间批量数据迁移的功能；
- b) 应支持批量迁移表或者文件，支持同构/异构数据库之间整库迁移，一个作业可迁移几百张表；
- c) 应支持文件增量迁移、关系型数据库增量迁移、面向列的分布式存储数据库增量迁移，以及使用筛选条件配合时间变量函数实现增量数据迁移；
- d) 应支持当迁移作业执行失败时，将数据回滚到作业开始之前的状态，自动清理目的表中的数据；
- e) 应支持去隐私、字符串操作、日期操作等常用字段的数据转换功能；
- f) 在迁移文件到文件系统时，批量数据迁移应支持对写入云端的文件进行加密；
- g) 应支持使用MD5校验，检查端到端文件的一致性，并输出校验结果；
- h) 应支持将迁移过程中处理失败的、被清洗过滤掉的、不符合字段转换或者不符合清洗规则的数据自动归档到脏数据日志中，方便用户分析异常数据。并支持设置脏数据比例阈值，来决定任务是否成功。

7.2 规范设计

应支持能源监管服务数据加工并业务化的功能，应提供主题设计、数据标准、数据建模、指标体系等功能，规范设计要求如下：

- a) 通过主题设计构建统一的数据分类体系，用于目录化管理所有业务数据，便于数据的归类、查找、评价、使用；通过分层架构对数据分类和定义，可帮助用户厘清数据资产，明确业务领域和业务对象的关联关系；
- b) 通过数据标准管理构建统一的数据标准体系，数据标准流程化、系统化，对每一行数据、每一个字段的具体取值进行标准化，从而提升数据质量和易用性；
- c) 通过数据建模构建统一的数据模型体系，自顶向下构建企业数据分层体系，沉淀企业数据公共层和生产经营主题库、双碳管理主题库、安全监管主题库、企业服务主题库、产业赋能主题库等5大主题库，便于数据的流通、共享、创造、创新，提升数据使用效率，极大的减少数据冗余、混乱、隔离、不一致以及谬误等；
- d) 通过指标体系建设构建能源监管服务统一的数据指标体系，定义业务指标要素，统一指标计算规则，拆解关联技术指标，形成指标资产。

7.3 数据开发

结合能源监管服务数据治理需求，应提供一站式敏捷数据开发平台，帮助能源监管服务快速构建大数据处理中心，数据开发要求如下：

- a) 应支持可视化的图形开发界面、丰富的数据开发类型（脚本开发和作业开发）、全托管的作业调度和运维监控能力、内置行业数据处理管道和一键式开发和全流程可视化；
- b) 应支持多人在线协同开发和管理多种大数据云服务，降低能源监管服务用户使用大数据的门槛；
- c) 应有数据管理功能，支持管理多种数据仓库，支持可视化和DDL方式管理数据库表；
- d) 应与批量数据迁移无缝集成，支持批量数据迁移，支持20多种异构数据源之间可靠高效的数据传输，实现多数据源集成到数据仓库；

- e) 应提供在线脚本编辑器，支持多人协作进行SQL、系统命令行脚本在线代码开发和调测。支持使用变量和函数；
- f) 应提供图形化设计器，支持拖拽式工作流开发，快速构建数据处理业务流水线；
- g) 应支持预设数据集成、SQL、MR、大数据处理引擎、系统命令行、机器学习等多种任务类型，通过任务间依赖完成复杂数据分析处理；
- h) 应支持统一管理在脚本开发和作业开发使用到的文件、jar、归档文件类型的资源；
- i) 应支持单次调度、周期调度和事件驱动调度，周期调度支持分钟、小时、天、周、月等多种调度周期。

7.4 数据质量

结合能源监管服务数据治理需求，应提供对能源监管服务数据全生命周期管控、数据处理全流程质量监控、异常事件实时通知、生成数据质量质量报告的能力，数据质量要求如下：

- a) 应支持对业务指标和数据质量进行监控，数据质量可检验，及时发现数据质量问题；
- b) 应提供对业务指标数据进行质量管理的有效工具，可以灵活的创建业务指标、业务规则和业务场景，实时、周期性进行调度，满足业务的数据质量监控需求；
- c) 应提供配置数据质量检查规则，在线监控数据准确性的能力，数据质量可以从完整性、有效性、及时性、一致性、准确性、唯一性六个维度进行单列、跨列、跨行和跨表的分析，也支持数据的标准化，能够根据数据标准自动生成标准化的质量规则，支持周期性的监控。

7.5 数据资产

结合能源监管服务数据治理需求，应提供企业级的元数据管理功能，厘清能源监管服务信息资产；数据资产管理可视，支持钻取、溯源等；通过数据地图，实现能源监管服务数据资产的数据血缘和数据全景可视，提供数据智能搜索和运营监控，数据资产要求如下：

- a) 应提供元数据管理能力，支持创建自定义策略的采集任务，可采集数据源中的技术元数据。支持自定义业务元模型，批量导入业务元数据，支持关联业务和技术元数据、全链路的血缘管理和应用；
- b) 应提供数据搜索能力，服务于数据分析、数据开发、数据挖掘、数据运营等数据表的使用者和拥有者，提供方便快捷的数据搜索服务，提供功能强大的血缘信息及影响分析；可通过关键词搜索数据资产，支持模糊搜索、快速检索、定位数据。

7.6 数据安全

结合能源监管服务数据治理需求，数据安全应为能源信息监管服务提供数据生命周期内统一的数据使用保护能力；通过敏感数据识别、分级分类、隐私保护、资源权限控制、数据加密传输、加密存储、数据风险识别以及合规审计等措施，建立安全预警机制，增强整体安全防护能力，让数据可用不可得和安全合规，数据安全要求如下：

- a) 应提供访问权限管理，通过创建权限策略实现对资源的访问控制；
- b) 应提供敏感数据识别功能，通过用户创建或内置的数据识别规则和规则组自动发现敏感数据并进行分级分类；
- c) 应提供隐私保护功能，通过数据静态脱敏、数据水印两种方式实现敏感数据保护；
- d) 应提供详细的客户端访问资源的鉴权日志记录以及授权日志记录，给用户授权/鉴权审计所需要的信息，方便更好的做到安全管控。

7.7 数据服务

结合能源监管服务数据治理需求，数据服务应支持搭建统一的数据服务总线，帮助能源监管服务统一管理对内对外的API服务，支撑业务主题/画像/指标的访问、查询和检索，提升数据消费体验和效率，数据服务要求如下：

- a) 应提供快速将数据表生成数据API的能力；
- b) 应支持现有的API快速注册到数据服务平台以统一管理和发布。

8 数据发布

- 8.1 能源监管服务数字化平台数据发布的基本要求、数据发布流程、数据发布方式等应遵循 DB14/T 2940—2023 的要求。
- 8.2 能源监管服务数字化平台应支持 API 发布方式。
- 8.3 能源监管服务数字化平台应支持数据资产发布方式。



参 考 文 献

- [1] 《中华人民共和国网络安全法》
 - [2] 《中华人民共和国数据安全法》
 - [3] 《数据安全管理办法》
 - [4] GB/T 22239-2019 信息安全技术网络安全等级保护基本要求
 - [5] KSSJZL12-2023 智能化矿山数据融合共享规范 数据存储技术规范
 - [6] KSSJZL13-2023 智能化矿山数据融合共享规范 数据质量管理规范
-

地方标准信息服务平台