

ICS 点击此处添加 ICS 号

CCS 点击此处添加 CCS 号

ZJSEE

浙江省电力学会标准

T/ZJSEE XXXX—XXXX

地市级能源大数据中心能源数据质量管理 规范

Quality Management Specification for Energy Data in City level Energy Big Data
Centers

草案版次选择

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX - XX - XX 发布

XXXX - XX - XX 实施

浙江省电力学会 发布

目 次

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 数据质量管理框架	2
5.1 概述	2
6 感知层	2
6.1 采集协议标准化	2
6.2 数据校验标准化	3
6.3 数据格式标准化	3
7 网络层	3
7.1 安全标准	3
7.2 通信协议标准	3
7.3 数据接入设备	4
8 平台层	4
8.1 质量指标	4
8.2 数据质量评价方法	8
8.3 数据质量报告	10
8.4 数据质量事件	11
8.5 数据存储版本控制	11
8.6 数据审计	12
8.7 数据备份	12
8.8 数据恢复	12
9 应用层	12
9.1 收集用户反馈	12
9.2 持续改进	12
附 录 A	13
电力能源数据类型指标一览表	13

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。

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。

本文件由浙江省电力学会提出。

本文件由浙江省电力学会技术归口和解释。

本文件起草单位：

本文件主要起草人：

本文件首次发布。

本文件在执行过程中的意见或建议反馈至浙江省电力学会标准工作委员会（地址：浙江省杭州市南复路1号，邮编：310008，网址：<http://www.zjsee.org/>，邮箱：zjseeorg_bz@163.com）。

地市级能源大数据中心能源数据质量管理规范

1 范围

本文件规定了地市级能源大数据中心数据质量管理规范中感知层、网络层、平台层、应用层的要求。本文适用于地市级能源大数据中心电、热、气三类能源数据的质量管理。

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 2589 综合能耗计算通则
GB/T 29873 能源计量数据采集系统数据传输协议
GB/T 30149 电网通用模型描述规范
GB/T 36073-2018 数据管理能力成熟度评估
GB/T 38673-2020 信息技术大数据大数据系统基本要求
GB/T 40685—2021 信息技术服务数据资产管理要求
ISO/IEC 27001 信息安全管理标准

3 术语和定义

下列术语和定义适用于本文件。

3.1

能源大数据 Energy big data

电力、煤炭、石油、天然气、水务、热力、新能源和可再生能源等能源生产、传输、存储、消费、交易过程中产生的海量大数据。

3.2

能源数据质量管理 Energy data quality management

指导和控制能源机构数据质量的协调活动。

3.3

能源大数据中心 Energy big data center

能源大数据的数据汇聚中心、价值创造中心、开放共享中心。一般具有实体或虚拟的组织管理形式和机构，包含基础设施、数据管理、产品服务、运营管理、安全防护部分。

3.4

数据架构 Data architecture

通过组织级数据模型定义数据需求，指导对数据资产的分布控制和整合，部署数据的共享和应用环境，以及元数据管理的一套整体规范。

[GB/T 36073-2018, 3.4]

3.5

数据标准 Data standards

数据的命名、定义、结构和取值规范方面的规则和基准。

[GB/T 36073-2018, 3.5]

3.6

能源大数据分类 Energy big data categorization

根据能源大数据的属性或特征，将其按一定的原则和方法进行区分和归类，并建立一定的分类体系和排列顺序。

4 缩略语

下列缩略语适用于本文件。

CoAP: 受限应用协议 (Constrained Application Protocol)

CSV: 逗号分隔值 (Comma-Separated Values)

JSON: 对象表示法 (JavaScript Object Notation)

MQTT: 消息队列遥测传输 (Message Queuing Telemetry Transport)

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

5 数据质量管理框架

5.1 概述

地市级能源大数据中心能源数据质量管理平台框架见图1，包括下列主要部分：

- 感知层：实现采集协议标准化、数据校验标准化和数据格式标准化；
- 网络层：在网络层采用安全标准、通信协议标准、数据格式标准、数据包校验规则、接入设备进行管理控制；
- 平台层：通过质量指标标准化、数据质量评价方法标准化、数据质量报告标准化、数据质量事件标准化、数据存储版本控制、数据审计、定期备份数据、建立数据恢复流程保证平台层可控性；
- 应用层：采集用户反馈并持续改进数据质量。

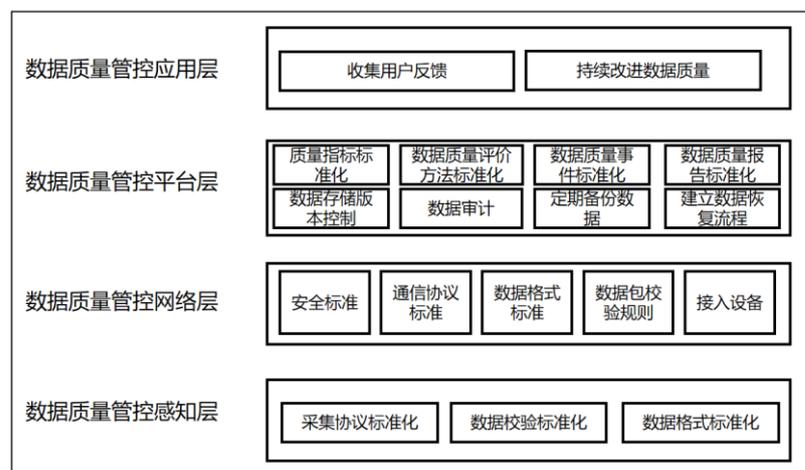


图1 数据质量管理框架

6 感知层

6.1 采集协议标准化

应对不同的能源设施，制定统一的数据采集协议，确保不同来源和不同类型的数据按照统一标准进行采集（包括记录数据采集设备、数据能源类型、数据采集频率等）。

6.1.1 数据采集设备

数据采集设备应满足下列要求：

- a) 现场数据采集设备支持对能耗数据 7~10 天的存储需求；
- b) 数据网关支持根据上位机命令启动数据采集和根据预设周期或时刻启动数据采集两种命令数据收集模式；
- c) 数据采集渠道、规范数据格式以及相关的流程和方式满足表 1 要求，并满足数据采集的合规性、正当性、一致性要求；

表 1 规范性能源数据采集质量要素表

采集渠道	规范数据格式	统一数据单位
智能计量设备、API 接口集成、数据交换平台	包括但不限于 JSON 或 XML 格式	如电力数据需包含电流（单位：安培）、电压（单位：伏特）、功率（单位：千瓦）等关键字段

- d) 当网络或设备发生故障时，现场数据采集设备能够存储未能正常上报的能耗数据，待恢复正常后进行断点续传，保证能耗数据的一致性和完整性。

6.1.2 数据采集频率

数据采集频率应满足下列要求：

- a) 能源数据采集设备从能源计量终端中采集能源信息的周期为 5 分钟到 24 小时。
- b) 能源数据采集设备上上传能源信息到能源监测应用系统的周期为 5 分钟到 24 小时。

6.2 数据校验标准化

对不同的能源设施，制订统一的数据校验方法（包括范围检查、格式验证、完整性检查）并记录。校验方法应满足下列要求：

- a) 针对不同能源类型（电力、热能、天然气）制定专门的校验规则。
- b) 根据校验规则对读入数据进行提取和计算，检查是否为有效的文件，检查模型中的实体是否与校验模板对应，检查模型中实体的属性是否缺失以及属性值是否超出范围；
- c) 使用数据结构校验文件的有效性，基本架构不应发生变更。实体校验规则和属性校验规则按照用户的需求制定，编辑在校验系统外部的配置文件中，校验系统运行时调用配置文件对实体和属性进行校验。系统外部的配置文件可根据用户的不同需求发生变更；
- d) 各编码应遵循企业内部标准，且针对不同能源设施制定专属编码体系，确保设备识别的唯一性和可追溯性。

6.3 数据格式标准化

应对所有能源设施制订统一的数据格式标准（包括数据文件格式、数据类型名称），宜满足下列要求：

- a) 电力数据：采用统一的文件格式，包括但不限于 CSV 或 JSON，确保数据类型名称明确，包括但不限于电流（单位：安培）、电压（单位：伏特）、功率（单位：千瓦时）；
- b) 热能数据：规定数据记录格式，包括热供应量（单位：吉焦耳）、供回水温差（单位：摄氏度）、热效率等关键参数，文件格式推荐为 XML，便于数据交换；
- c) 天然气数据：明确数据结构，涉及气体流量（单位：立方米）、压力（单位：帕斯卡）、热值（单位：兆焦耳/立方米），采用 Parquet 或 Avro 格式以支持大数据分析处理。

7 网络层

7.1 安全标准

应采用国际通用的数据安全和加密标准，包括但不限于 TLS/SSL 加密协议、ISO/IEC 27001 信息安全管理，应按 GB/T 40685—2021 中 6.4 和 GB/T 37973—2019 中 8.3 的规定，包括数据资产确权、数据分类分级、数据加密存储、数据备份恢复等。

7.2 通信协议标准

应遵循行业内通信协议标准，包括但不限于 MQTT、CoAP，确保不同设备和系统间兼容性和互操作性。

7.3 数据接入设备

应记录接入的设备及其接入权限。

7.3.1 数据接入要求

应满足下列要求：

- a) 按照 GB/T 38673—2020 中 6.1 的规定支持各类型能源的结构化数据、非结构化数据全量、增量、实时接入，支持采用数据抽取、数据同步、数据接口、消息队列、网页抓取和数据填报等方式；
- b) 数据量较大、单批量接入可能会造成系统故障的，采用分批接入。

7.3.2 接入时效要求

数据接入应采用实时接入或定时批量接入的方式，并满足下列要求：

- a) 实时接入定义为数据从源头产生至到达大数据中心的延迟时间不超过 5 分钟可视为实时，具体可根据业务进行调整；
- b) 定时批量接入作为补充方式，适用于数据量大且实时性要求相对宽松的场景，批量接入应设定周期性执行，具体频率依据业务需求和系统处理能力确定。

8 平台层

8.1 质量指标

8.1.1 规范性

能源数据质量描述要素子元素定义见表2。

表 2 规范性能源数据质量描述要素子元素

子元素名称	子元素概述	能源数据类型	子元素描述	指标
数据标准	数据符合数据标准的度量。数据标准为数据在命名、创建、定义、更新和归档时遵循的标准，包括国际标准、国家标准、行业标准、地方标准或相关规定等。	电力能源	包括电压、电流、频率等标准化度量单位和质量参数	应满足 GB/T 30149、GB/T 2589、GB/T 29873 要求
		热能源	涉及温度、流量、热效率的标准化度量	
		天然气能源	包括气体成分、压力、体积等标准化度量单位	
数据模型	数据符合数据模型的度量。数据模型是一种直观描述数据组织结构的手段，是数据表达的规范。	电力能源	电力能源数据模型包括电网结构描述、电力系统操作数据、电力市场交易数据等	应满足 GB/T 30149 要求
		热能源	热能源数据模型包括热网分布模型，热能产生和传输模型等	
		天然气能源	气能源数据模型包括气体能源的生产、存储、输送和使用的整个流程描述、气体成分分析、压力和流量数据，以及与安全相关的监测数据、气体能源的市场交易数据等	
元数据	数据符合元数据定义的度量。元数据为用于提供有关资源的结构数据，用于刻画其他数据，以供检索或使用信息。	电力能源	数据来源（如传感器类型和位置）、采集时间、数据精度	应满足用户使用的采集周期、精度的标准，同类不符合采集周期和精度标准的数据数量不应超过 5%
		热能源	数据来源、采集时间、精度等信息	
		天然气能源	气体类型、来源、采集时间、存储条件	

表2 规范性能源数据质量描述要素子元素（续）

子元素名称	子元素概述	能源数据类型	子元素描述	指标
业务规则	数据符合业务规则的度量。业务规则是一种权威性原则或指导方针，用于描述业务交互，并建立行动和数据行为结果及完整性的规则。	电力能源	负荷分配规则、紧急情况下的供电优先级	数据应符合相关时间、顺序、优先等级等规则，同一类型不满足业务规则的数据量不超过各数据能源类型的5%。
		热能源	热能分配和温度控制规则	
		天然气能源	气体分配规则、安全存储和运输规则	
安全规范	安全规范是安全和隐私方面的规则，包括数据权限管理，数据脱敏处理等。	电力能源	电气安全标准、设备维护和检测规范	数据应满足 GB/T 43697-2024 要求，存储应有权限约束、外部传输应加密，避免监听。
		热能源	热网系统安装和运维的安全标准	
		天然气能源	气体泄漏检测标准、防火防爆标准	

8.1.2 完整性

完整性的能源数据质量描述要素子元素定义见表3。

表3 完整性能源数据质量描述要素子元素

子元素名称	子元素概述	能源数据类型	子元素描述	指标
数据元素完整性	按照能源业务规则要求，数据集中应被赋值的数据元素的赋值程度。	电力能源	电流、电压、功率等基本电力参数的完整记录；时序数据的完整性，例如每分钟或每小时的电力消耗数据；设备和组件的状态信息，如变压器、电缆的完整性数据	数据应根据存储频率进行存储、刷新，保持完整性，因通讯故障、延迟、填报不及时造成空数据，同类数据中满足数据完整性的比例应大于95%。
		热能源	温度、热流量、热效率等关键测量参数的完整性	
		天然气能源	气体的质量参数，如成分、热值的完整记录；输送系统的完整性数据，包括压力、流量、温度等；存储设施的监测数据，例如气体储量、容器完整性	
数据记录完整性	按照业务规则要求，数据集中应被赋值的数据记录的赋值程度	电力能源	完整的电网运行日志，包括所有事故、维护和检修记录；用户用电历史记录完整性，包括峰值和谷值用电量；完整的能源交易记录，包括购买、销售和定价信息	数据应根据相关规则进行存储和更新，保持完整性，同类数据中满足数据完整性的比例应大于95%。
		热能源	热供应和使用的完整记录，包括所有操作和维护活动	
		天然气能源	完整的气体生产和输送记录，包括每日产量和输送量；安全检查和泄漏检测的完整记录；气体买卖合同和交易记录的完整性	

8.1.3 准确性

准确性的能源数据质量描述要素子元素定义见表4。

表 4 准确性能源数据质量描述要素子元素

子元素名称	子元素概述	能源数据类型	子元素描述	指标
数据内容正确性	数据内容是否是预期数据。	电力能源	电力消耗、产量、设备状态和性能数据的准确性	数据在传输过程中应与现场采集器的数据保持100%一致；
		热能源	供热量、温度、流量数据的准确性	
		天然气能源	气体质量和流量的准确度，安全监测数据的正确性等	
数据格式合规性	数据格式(包括数据类型、范围、长度、精度等)是否满足预期要求	电力能源	数据记录是否遵循行业标准格式	数据传输过程中应遵循约定的标准通讯协议，按照标准协议进行解析，并按照约定的数据类型、量纲、系数、偏移量进行读写存储，应与现场采集器保存100%一致。
		热能源	热供应数据是否标准化表达，如检查功率单位和时间戳	
		天然气能源	气体参数(如体积、压力)是否统一度量单位，安全报告和环境监测数据是否标准化格式	
数据重复率	特定字段、记录、文件或数据集意外重复的度量	电力能源	同一电力数据在不同系统中的重复率、重复的用电记录和设备日志等	除数据备份外，特定唯一性数据应保存唯一性，重复性数据应自动筛查，并给予删除。
		热能源	热供应记录的重复性、环境监测数据的重复记录等	
		天然气能源	流量和压力数据的重复录入、交易和合同数据的重复性等	
数据唯一性	特定字段、记录、文件或数据集唯一性的度量	电力能源	每个用户或设备是否唯一标识符，事故和故障报告的唯一性等	除数据备份外，特定唯一性数据应设置标识，避免重复输入，没有设置唯一性标识，应判定为BUG，应及时修正。
		热能源	热网系统组件的唯一标识，每个监测点的数据唯一性等	
		天然气能源	气体源和终端用户的唯一标识、安全事件的唯一记录等	
脏数据出现率	正确字段、记录、文件或数据集之外无效数据的度量	电力能源	不准确的读数和错误的设备信息的出现比率、无效或错误的用户数据出现比率等	无效数据应自动进行标识，经过修正过的数据应消除标识，脏数据占比不应大于同类型数据数量的5%。
		热能源	错误的热供应记录比率等	
		天然气能源	错误或失效的安全监测数据比率等	

8.1.4 一致性

一致性的能源数据质量描述要素子元素定义见表5。

表 5 一致性能源数据质量描述要素子元素

子元素名称	子元素概述	能源数据类型	子元素描述	指标
相同数据一致性	同一数据在不同位置存储或被不同应用或用户使用时，数据的一致性；数据发生变化时，存储在不同位置的同一数据被同步修改	电力能源	例如在不同系统（如发电、输电、分配）中相同参数（如电压、电流）的一致性	a) 同一个历史数据被调用应100%一致； b) 同一个实时数据在被不同应用调用，允许产生偏差，偏差和读取数据的频率和元数据波动相关，在5倍最大轮询周期内，不应超过2次偏差大于元数据最大波动。
		热能源	例如在不同热网系统测量的相同环境参数（如光照强度、温度）的一致性	
		天然气能源	例如在不同地点或时间的气体质量参数（如热值、成分）的一致性	
关联数据一致性	数据关联性用于度量存在关系的数据，即关联关系是否缺失或错误。数据关联关系包括函数关系，相关系数，主外键关系，索引关系等。数据之间存在关联性问题会影响分析的结果。根据数据关联性检查关联数据的一致性。	电力能源	例如电力消耗数据与相应的用户信息和账单数据的一致性	a) 数据与台账的关联约束应100%一致； b) 数据与关联数据的偏差应在物理规律的约束范围20%以内。
		热能源	例如供热量与环境因素的关联数据一致性	
		天然气能源	例如气体销售量与生产量、存储量的关联数据一致性	

8.1.5 时效性

时效性的能源数据质量描述要素子元素定义见表6。

表 6 时效性能源数据质量描述要素子元素

子元素名称	子元素概述	能源数据类型	子元素描述	指标
基于时间段的正确性	基于日期范围的记录数或频率分布符合能源业务需求的程度	电力能源	长期电力消耗和发电数据的准确性，如月度、季度报告；维护和运营数据的长期记录准确性，如设备性能变化等	基于时段的元数据修正后，所关联的结果数据应进行自动或提示手动重算修正，以保证数据准确性，根据相关数据校核筛查，同一类型的数据正确性应大于95%。
		热能源	供热量的长期准确性，包括季节变化和年度产量；环境影响对长期发电效率的影响数据准确性等	
		天然气能源	长期气体生产、输送和消费量的准确记录；存储设施的长期监测数据准确性，如气体存储量变化	
基于时间点及时性	基于时间戳的记录数、频率分布或延迟时间符合业务需求的程度	电力能源	实时电力需求和供应数据的更新频率和及时性；事故和故障响应数据的实时记录和报告等	基于时间戳的数据因更新频率、通讯延时、通讯故障等原因造成统计和计算数据的不准确，应自动进行标记，同一类型的数据正确性应大于95%。
		热能源	环境条件变化对供热量的即时影响等	
		天然气能源	实时气体流量和压力数据的更新和记录；安全监测（如泄漏检测）数据的实时更新等	
时序性	数据集中同一实体的数据元素之间的相对时序关系	电力能源	电力数据的时间序列分析，如负荷预测、消费趋势；设备性能和故障率的时间序列分析等	同一实体的数据应符合一定的物理规律，违背物理规律的数据应判定为无效数据，同一类型的数据正确性应大于95%。
		热能源	供热量的时间序列分析，如日变化、季节性影响；系统性能的时间序列监测，如长期效率衰减等	
		天然气能源	气体生产和消费的时间序列分析，包括季节性和趋势变化；管网运营的时间序列分析，如压力变	

			化趋势等	
--	--	--	------	--

8.1.6 可访问性

可访问性的能源数据质量描述要素子元素定义见表7。

表7 可访问性能源数据质量描述要素子元素

子元素名称	子元素概述	能源数据类型	子元素描述	指标
可访问	数据在需要时的可获取性	电力能源	电网数据的访问权限；用户用电历史和账单信息的访问控制等	数据应设置访问权限管理，有权限的用户应100%可访问，无权限的用户应100%拒绝访问。
		热能源	供热数据的访问权限、系统维护和环境监测数据的访问管理等	
		天然气能源	气体流量和质量数据的访问控制、安全监控和事故报告的访问权限等	
可用性	数据在设定有效生存周期内的可使用性	电力能源	电力消耗和发电数据的实时更新和持续可用性、电网状态和故障响应数据的持续监控和即时更新等	数据应持续有效，同一类型的无效数据不应超过5%。
		热能源	热网系统输出和环境数据的实时可用性、系统性能和故障诊断数据的持续监测等	
		天然气能源	气体生产和输送数据的实时更新和可用性、安全监测系统（如泄露检测）的数据实时更新和持续监控等	

8.2 数据质量评价方法

能源数据质量评价流程是产生和报告能源数据质量结果的一系列步骤，在进行能源数据质量识别之后，进行数据质量评价。能源数据质量识别见表8，评价流程步骤见表9。

表8 能源数据质量识别

能源数据类型	数据质量识别要求
电力能源	(1) 检查数据采集方法是否能准确测量电流、电压、功率等。 (2) 确保存储格式可以有效记录时间序列数据，如每分钟或每小时的能耗。 (3) 分析工具应能处理大规模电网数据，并支持实时监控和预测。
热能源	(1) 数据采集方法应能精确测量温度、热流量等。 (2) 存储格式应适合记录分布式热网数据，并能处理多源数据。 (3) 分析工具需能评估热效率和优化供热系统。
天然气能源	(1) 采集方法需准确测量气体压力、流量、成分等。 (2) 存储格式应适合高压气体输送系统，并能承受大量数据的存储需求。 (3) 分析工具应支持安全监控和流量优化。

表9 评价流程步骤

步骤	活动	概述	能源数据类型	描述
----	----	----	--------	----

1	识别可用能源数据质量描述要素、能源数据质量描述要素子元素及数据质量范围。	根据 5.2 识别能源数据质量描述要素、能源数据质量描述要素子元素以及数据质量范围。	电力能源	识别电流、电压、功率、测量精度、时间戳准确性、数据记录完整性等能源数据质量描述要素子元素，确定覆盖整个电网的发电、输电和分配等数据覆盖范围
			热能源	识别温度、热流量、供热效率、环境影响因素、设备效能、能源分配等能源数据质量描述要素子元素，确定集中供热系统和分布式热能使用等数据覆盖范围
			天然气能源	识别气体压力、流量、成分、测量稳定性、数据一致性、安全监测等能源数据质量描述要素子元素，确定涵盖从气源到终端用户的整个供应链等数据覆盖范围。
2	识别数据质量测量	对每个测试，识别数据质量策略、数据质量值类型及数值质量值单位。	电力能源	数据质量策略包括确保事件序列准确性，如负载和产能；值类型和单位包括电压（伏特）、电流（安培）、功率（瓦特）等。
			热能源	数据质量策略包括测量温度和流量的一致性和准确性；值类型和单位包括温度（摄氏度）、热流量（焦耳/秒）。
			天然气能源	数据质量策略包括监控气体压力和流量的稳定性；值类型和单位包括压力（帕斯卡）、体积流量（立方米/小时）。
3	选择并运用数据质量评价方法	对每个被识别的数据质量测量，选择数据质量评价方法	电力能源	关注电网的稳定性和能源分配效率，使用时间序列分析来评估负载和产能数据的一致性和预测准确性
			热能源	关注热能供应的可靠性和系统效率，运用热力学模型和效率分析来评估热能的传输和利用效率
			天然气能源	关注系统的安全性和气体的质量控制，利用流体动力学分析和压力/流量监控来评估气体输送系统的性能
4	确定数据质量结果	结果为：能源数据质量描述要素结果、数据质量值或数据质量值集合、数据质量值单位及数据质量测量日期	电力能源	数据质量描述要素结果：电网稳定性、负载响应时间等。 数据质量值集合：如电压波动率、频率稳定性指标等。 值单位：伏特、赫兹等。 测量日期：指示数据收集的具体时间点或时间段等。
			热能源	数据质量描述要素结果：供热效率、温度一致性等。 数据质量值：供热量、温度变化率等。 值单位：焦耳、摄氏度等。 测量日期：收集数据的具体日期等。
			天然气能源	数据质量描述要素结果：气体纯度、压力稳定性等。 数据质量值：气体流量、压力变化等。 值单位：立方米/小时、帕斯卡等。 测量日期：数据采集的准确日期等。
5	确定一致性	若数据规范或用户需求中已指定一致性质量等级，将其与数据质量结果相比可确定一致性。一致性数据质量结果（“通过”或“不通过”）是能源数据质量描述要素数据质量结果与一致性质量等级比较后的结果	电力能源	确定一致性标准，如电网数据的稳定性标准，比较实际数据质量结果与一致性标准，输出一致性结果（通过或不通过）。
			热能源	确定一致性标准，如供热系统效率和温度控制的一致性标准，比较实际数据质量结果与一致性标准，输出一致性结果（通过或不通过）。
			天然气能源	确定一致性标准，如气体流量和压力数据与一致性标准，比较实际数据质量结果与一致性标准，输出一致性结果（通过或不通过）。

8.2.1 准确性

应采用交叉验证、多源数据校验、异常检测保证准确性，并满足下列技术要求：

- a) 交叉验证：通过将数据集分成若干个互斥的子集，轮流用其中若干个子集作为训练集，其余子集作为验证集，对模型进行评估；

- b) 多源数据校验：根据获取校验内容和预先制定的校验规则生成校验任务和与所述校验任务对应的配置文件；当需要执行校验任务时，执行所述校验任务对应的配置文件；
- c) 异常检测：对原始数据进行清洗和处理，包括数据去噪、缺失值填充和特征选择等；根据问题的特点，通过特征提取、转换和构造等方法，确定适合异常检测的特征；根据数据的特点和问题的需求，选择合适的异常检测模型；可以同时尝试多种模型，比较它们的性能和效果；使用标记的训练数据集对异常检测模型进行训练，优化模型的参数和参数设置；使用训练好的模型对新数据进行异常检测；通过计算异常分数或距离度量来判断数据是否异常，对异常检测的结果进行评估和验证。

8.2.2 完整性

应进行缺失值分析、必要字段检查、完整性约束检查（数据库层面）保证完整性，并满足下列要求：

- a) 缺失值分析：采用完全随机缺失、随机缺失、非随机缺失方法进行分析；
- b) 必要字段检查：进行完整性检查、正确性检查、正确性检查；
- c) 完整性约束检查：提供定义完整性约束条件的机制，保证数据的正确性和相容性，例如检查主码值是否唯一，如果不唯一则拒绝插入或修改；检查主码的各个属性是否为空，只要有一个为空就拒绝插入或修改。

8.2.3 及时性

应进行数据更新频率检查、时间戳分析保证及时性，并满足下列要求：

- a) 数据更新：归集数据应真实可用，已归集的能源数据应按时进行更新。当已更新数据不满足应用需求时，数据提供者应对自身源数据进行更新检查，配合数据更新工作；
- b) 数据更新频率：当原始数据发生更新时，应按照数据采集、数据整合、数据分析的步骤对更新数据进行再次归集，保证更新频率处于可控范围；
- c) 时间戳分析：当执行备份操作时，确定哪些文件自上次备份以来已更改的关键指标。

8.2.4 一致性

应进行跨系统对比、数据格式和单位检查、规则和逻辑校验保证一致性，并满足下列要求：

- a) 跨系统对比：应支持向用能企业提供能效综合评价、行业能效对比、区域能效对比等。
- b) 规则和逻辑校验：根据各指标特点增加针对性校验规则。

8.3 数据质量报告

数据质量报告作为能源数据质量的评价结果，能源数据质量报告主要内容见表10，内容要求如下：

- a) 编号：给表中每个条款编号。
- b) 名称：报告条款名称。
- c) 说明：定义或描述条款内容。
- d) 条件：描述报告改条款的必要条件，或需要该条款的条件。其中，必备代表应有该条款；
- e) 条件可选代表规定条件被满足时应有该条款；可选代表该条款是可选的。

表 10 能源数据质量报告主要内容

编号	名称	说明	条件
1	质量报告	报告章节	必备
1.1	报告名称	报告的名称	必备
1.2	报告范围	该报告所评价的数据集的范围	必备
2	数据质量测量	报告章节	必备
2.1	数学描述	数据质量测量的数学描述	必备
2.2	数据质量值	数据质量测量的结果值	必备
2.3	数据质量值单位	数据质量测量结果值的单位或值类型	必备
2.4	可信度	计算或估计的数据质量策略的可信度	必备

表10 能源数据质量报告主要内容（续）

编号	名称	说明	条件
----	----	----	----

2.5	可信度单位	可信的的单位或值类型	必备
3	一致性可信度	报告章节	必备
3.1	一致性结果可信度	一致性结果的可信度	必备
3.2	一致性结果可信度单位	一致性结果可信度的单位或值类型	必备
3.3	参考文档	一致性评价所参考的文档信息	可选
4	质量评价方法信息	报告章节	必备
4.1	检查策略	所用检查策略信息	必备
5	数据质量评价方法	报告章节	必备
5.1	假定	开发和应用该数据质量评价方法的隐含假定信息	必备
5.2	处理算法	为确定数据质量结果, 处理数据的算法	必备
5.3	参数信息	数据质量评价方法所用参数信息	可选
5.3.1	参数定义	所用参数定义	必备
5.3.2	参数值	所用参数值	必备
5.3.3	参数单位	所用参数值的单位	必备
5.4	完全检查方法	完全检查方法的信息	条件可选
5.4.1	完全检查过程	完全检查过程的详细描述	必备
5.4.2	数据项描述	数据项的内容定义	必备
5.4.3	参考文档	完全检查所参考的文档	可选
5.5	抽样检查方法	抽样检查方法的信息	条件可选
5.5.1	抽样检查方法类型	抽样检查方法的类型信息	必备
5.5.2	抽样过程	抽样过程详细描述	必备
5.5.3	抽样比率	样本占群体的比率	必备
6	综合质量	报告章节	可选
6.1	综合质量值	综合质量结果值	必备
6.2	综合质量值单位	综合治理结果值的单位或值类型	必备
6.3	综合方法	综合方法详细描述	必备
6.4	时间	综合时间	可选
6.5	综合质量报告	针对综合质量的报告	可选
7	其他	报告章节	可选

8.4 数据质量事件

能源数据质量事件的分级通常基于对业务影响、修复难度和问题的广泛性的评估。能源数据质量事件的定级和处理措施见表10。

表 11 能源数据质量事件的定级和处理措施

能源数据质量事件定级	事件描述	处理措施
低级事件	轻微的数据不一致性或格式错误	通过自动化工具或简单的手动过程进行纠正
中级事件	数据不完整或轻微的数据损坏	需要数据团队更深入地分析并采取修复措施
高级事件	严重的数据损坏、数据丢失或安全问题	需要进行全面的数据恢复、系统重构或安全加固, 并进行事后审计

8.5 数据存储版本控制

应对关键数据实施版本控制, 跟踪数据变更历史, 便于追溯和恢复, 并满足下列要求:

- 允许跟踪数据的变更和回滚, 以确保数据的可重复性和可解释性;
- 每当记录发生变更时, 更新时间戳会被自动更新;
- 每当记录发生更改时, 版本号相应增加;
- 使用触发器自动化版本控制;
- 定期生成数据库的快照, 将数据库中某一时刻的数据状态完整保存, 以便将来可以恢复到该状态。

8.6 数据审计

应记录数据的添加、修改、删除等操作，以便于数据质量的审计和追踪；

8.7 数据备份

应定期备份重要数据，以降低数据丢失风险，并根据能源类型及其采集频率的差异，实行差异化备份策略，确保数据安全性和时效性，并满足下列要求：

- a) 鉴于电、热、气等能源数据的采集频率不同，备份周期应与其相匹配。对于高频采集的实时数据（如电力数据每5分钟采集一次），每日进行增量备份，每周进行一次全量备份，以保持数据的连续性和完整性。对于采集频率较低的能源数据（如月度结算数据），每月进行一次全量备份。
- b) 备份策略综合考虑数据的采集频率、重要性及存储资源，确保在有限资源下达到最佳的数据保护效果。

8.8 数据恢复

8.8.1 应建立数据恢复流程，制定详细的数据恢复流程和计划，确保在数据损坏或丢失时可以迅速恢复。

8.8.2 数据恢复流程和计划应明确从发现问题、报告流程、初步评估、选择恢复策略、执行恢复操作、验证数据完整性到恢复后总结的多个步骤。流程中应包括时间线管理、责任分配、所需资源和预期完成时间关键要素。

9 应用层

9.1 收集用户反馈

9.1.1 应基于数据使用的能源数据质量监控要求能源数据使用节点提供一个明确的、易于访问的平台（如内部网站、电子邮件系统）供用户报告数据质量问题，并应建立一个数据质量团队，负责接收、记录和分类用户的反馈。

9.1.2 应鼓励用户反馈数据问题，并将用户反馈纳入数据质量改进过程，反馈内容应满足下列要求：

- a) 进行统一化流程管理；
- b) 建立统一格式（如表格等）形式，对用户意见定期回访；
- c) 数据质量团队持续跟进反馈内容。

9.2 持续改进

应建立机制收集使用过程中的反馈，不断优化和调整规范内容，定期更新规范，以适应新的技术发展和市场需求，应根据数据质量监控结果和用户反馈，持续改进数据处理、存储和管理流程。

附 录 A
(资料性)
电力能源数据类型指标一览表

A.1 指标详见下表

表 A.1 电力能源数据类型指标

序号	数据类型	一级类型	二级类型	量纲	备注	
1	模拟量	电压	A相电压 (Ua)	V		
			B相电压 (Ub)	V		
			C相电压 (Uc)	V		
			平均相电压 (U_lnavg)	V		
			相电压不平衡率 (U_Inub)	%		
			AB线电压 (Uab)	V		
			BC线电压 (Ubc)	V		
			CA线电压 (Uca)	V		
			平均线电压 (U_llavg)	V		
			线电压不平衡率 (U_lub)	%		
			零序电压 (3Uo)	V		
			A相n次谐波电压 (nTHD_Ua)	V		
			B相n次谐波电压 (nTHD_Ub)	V		
			C相n次谐波电压 (nTHD_Uc)	V		
			直流电压 (Udc)	V		
			第n组串电压 (U_nPV)	V		
			电流	A相电流 (Ia)	A	
				B相电流 (Ib)	A	
		C相电流 (Ic)		A		
		平均相电流 (I_lnavg)		A		
		相电流不平衡率 (I_Inub)		%		
		AB线电流 (Iab)		A		
		BC线电流 (Ibc)		A		
		CA线电流 (Ica)		A		
		平均线电流 (I_llavg)		A		
		线电流不平衡率 (I_lub)		%		
		零序电流 (3Io)		A		
		A相n次谐波电流 (nTHD_Ia)		A		
		B相n次谐波电流 (nTHD_Ib)		A		
		C相n次谐波电流 (nTHD_Ic)		A		
		直流电流 (Idc)		A		
		第n组串电流 (I_nPV)		A		
		功率		总有功功率 (Psum)	kW	
				A相有功功率 (Pa)	kW	
			B相有功功率 (Pb)	kW		
			C相有功功率 (Pc)	kW		
			总无功功率 (Qsum)	kVar		
			A相无功功率 (Qa)	kVar		
			B相无功功率 (Qb)	kVar		
			C相无功功率 (Qc)	kVar		
			总视在功率 (Ssum)	kVA		
			A相视在功率 (Sa)	kVA		
			B相视在功率 (Sb)	kVA		
			C相视在功率 (Sc)	kVA		
			直流功率 (P)	kW		

序号	数据类型	一级类型	二级类型	量纲	备注
			虚拟电厂基线负荷(P_BaseLine)	kW	
			虚拟电厂响应负荷(P_RespondValue)	kW	
			充电功率(P_Charge)	kW	
			放电功率(P_Discharge)	kW	
			设置功率(P_Setting)	kW	
		功率因数	系统功率因数(PF)	COS ϕ	
			A相功率因数(PF_a)	COS ϕ	
			B相功率因数(PF_b)	COS ϕ	
			C相功率因数(PF_c)	COS ϕ	
		相位角	Ub相滞后Ua相位角(Ub ϕ)	°	
			Uc相滞后Ua相位角(Uc ϕ)	°	
			Ia相滞后Ua相位角(Ia ϕ)	°	
			Ib相滞后Ua相位角(Ib ϕ)	°	
			Ic相滞后Ua相位角(Ic ϕ)	°	
通信故障计数	通信故障计数(Counting_CommFault_)	times			
2	状态量	通信故障	0-Online, 1-Offline	ON/OFF	
		开关位置	0-OFF, 1-ON	ON/OFF	
3	累积量	电度量	正向有功电度(Ep_imp)	kWh	
			正向无功电度(Eq_imp)	kVarh	
			反向有功电度(Ep_exp)	kWh	
			反向无功电度(Eq_exp)	kVarh	