

**CECS ×××:20××**

中国工程建设标准化协会标准

**城市供水信息系统 基础信息加工处理技术指南**

**Guidelines for basic information processing of urban water supply information system**

（征求意见稿）

**中国工程建设标准化协会标准**

城市供水信息系统 基础信息加工处理技术指南**Guidelines for basic information processing of urban water supply information system**

**CECS ×××:20××**

主编单位： 中国城市规划设计研究院、山东省城市供排水水质监测中心

批准单位： 中国工程建设标准化协会

施行日期：

**目 录**

[前 言 1](#_Toc23497282)

[第一章 总则 2](#_Toc23497283)

[第二章 基础信息分类编码 3](#_Toc23497284)

[第三章 基础信息加工处理基本流程 7](#_Toc23497285)

[第一节 国内外城市供水信息系统发展现状 7](#_Toc23497286)

[第二节 基础信息加工处理基本流程 10](#_Toc23497287)

[第三节 相关行业信息系统标准规范 10](#_Toc23497288)

[第四章 基础信息加工处理的单元技术 12](#_Toc23497289)

[第一节 数据采集 12](#_Toc23497290)

[第二节 数据清洗、转换及装载 14](#_Toc23497291)

[第三节 数据存储与备份 17](#_Toc23497292)

[第四节 数据分析及展示 18](#_Toc23497293)

[第五节 信息安全 21](#_Toc23497294)

[第五章 质量保障 23](#_Toc23497295)

[第六章 系统维护与管理 27](#_Toc23497296)

[附录 29](#_Toc23497297)

[编制依据 29](#_Toc23497298)

# 前 言

近年来，随着国内城市供水监管信息化建设进程的不断加快，很多供水监管机构及企业等都构建很多供水管理信息系统，为城市的供水管理工作提供了有效支撑，但是，很多城市的供水信息系统存在信息孤岛、数据质量不佳、加工处理方面缺少标准规范等问题，给数据的有效分析、价值挖掘和提取带来很多困难，进而影响到了信息化建设对供水监管的支撑作用。

为规范城市供水信息系统基础信息的加工处理，提高城市供水信息系统中基础信息的真实性、准确性、完整性和有效性，编制组经过广泛调研，认真总结实践经验，参考国家和行业有关标准，并广泛征求意见的基础上，特编制了本指南。

本指南的主要技术内容有：1总则；2基础信息的分类编码；3基础信息加工处理的基本流程；4基础信息加工处理的单元技术；5质量保障；6系统维护与管理。

本指南由中国工程建设标准化协会标准负责管理，由中国城市规划设计研究院、山东省城市供排水水质监测中心负责具体技术内容的解释。执行过程中如有意见或建议，请将意见或建议寄送至中国城市规划设计研究院（地址：北京车公庄西路5号，邮政编码：10004），山东省城市供排水水质监测中心（地址：济南市奥体中路5111号，邮编250000）。

主编单位：中国城市规划设计研究院、山东省城市供排水水质监测中心。

参编单位：青岛积成电子股份有限公司、济南水务集团有限公司、杭州绿洁水务科技有限公司、力合科技（湖南）股份有限公司

主要起草人员：

主要审查人员：



# 第一章 总则

### 1 编制目的

为建立基础信息加工处理技术规范体系，提高城市供水信息系统中基础信息的真实性、准确性、完整性和有效性，指导供水信息系统的设计与开发，制定本指南。

### 2 主要内容

本指南提出了城市供水信息系统中基础信息分类编码、数据采集、数据清洗、转换和装载、数据存储与备份、数据分析与展示、质量保障等方面的基本方法和要求。

### 3 适用范围

本指南适用于城市供水监管及供水企业信息系统基础信息的加工处理。

# 第二章 基础信息分类编码

### 1 基础信息类型

城镇供水信息系统基础信息包括结构化数据、半结构化数据和非结构化数据，参照CJ/T 541的规定对其进行分类。

#### 1.1 结构化数据

1. 城镇供水管理基础信息主要包括以下内容：
2. 城镇供水行政主管部门基本情况；
3. 城镇供水单位水厂数量、综合生产能力等汇总信息；
4. 城镇年供水量；
5. 城镇年用水量；
6. 城镇二次供水管理情况；
7. 城镇供水水源地基本情况；
8. 城镇应急水源基本信息；
9. 城镇取水、供水管道信息。
10. 城镇供水单位基础信息主要包括以下内容：
11. 城镇供水单位名称、地址、类型、企业性质、服务类别等基本情况；
12. 城镇供水设施基础情况；
13. 城镇供水单位运营情况：包括水厂数量、服务人口、用户水表数、客户服务、生产运营信息情况等；
14. 水质检测部门基本信息，包括名称、联系方式、资质状况和级别、部门人员情况、检测能力、检测设备等；
15. 水质在线监测基础信息：包括监测点布局和建设信息、监测设备信息等；
16. 水质检测基础信息：包括水源、水厂、管网（包括管网末梢）、二次供水的采样点数量和布设情况。
17. 城镇供水水厂基础信息：包括水厂名称、水厂类型、所属供水单位、水源类型、水源地基本信息、水厂规模、净水工艺、供水总量、消毒方式、水质检测项目、在线监测项目、生产能力、水厂安全管理信息和应急水源等。
18. 城镇供水设施在建、规划拟建项目基础信息：包括项目基本情况、建设内容、投资情况、设计情况、审批情况、水厂净水工艺、管道材质及长度等。
19. 城镇供水单位月供水水量与水压动态信息
    1. 月供水流量监测信息；
    2. 月供水压力监测信息；
    3. 供水压力在线监测信息；
    4. 供水流量在线监测信息；
20. 城镇供水水质动态监测信息
    1. 水源水水质在线监测信息；
    2. 出厂水水质在线监测信息；
    3. 供水管网水质在线监测信息。
21. 城镇供水水质基础信息
    1. 城镇供水水质检测数据：包括水源水、出厂水、管网（管网末梢）水、二次供水等日检、周检、月检、季检、半年检、年检的实验室水质检测数据、水质汇总数据及质量控制等相关信息。
    2. 实验室水质检测数据应包括样品编号、采样点名称、采样点地址或位置、水样类型、采样时间、水厂名称、水样检测类型、检验时间、检测机构、指标代码、指标名称、检测值、指标计量单位等内容。
22. 城镇二次供水基础信息：包括二次供水设施地址、管理单位、设施分类、用水人口、建设时间、建设单位、水质检测单位和时间、清洗单位、清洗时间、日常水质检测数据等。
23. 城镇供水设施在建项目季报信息：包括本季度投入资金金额、资金来源、项目完成情况等。
24. 城镇供水突发水质事件快报信息
    1. 事件基本情况：发生时间、发生地点、事故类型、特征污染物等；
    2. 水质跟踪情况。

#### 1.2 半结构化数据

主要包括城镇供水单位和供水厂的值班日志、电子邮件、维修记录、投诉工单等记录信息类。

#### 1.3 非结构化数据

主要包括以下内容：

(1) 城镇供水各类报表、国家/行业公开的标准文本等文档与图片类信息；

(2) 通知公告类：包括水质督查等通知、供水水质信息公开结果等；

(3) 城镇供水突发水质事件新闻报道类信息；

(4) 城镇供水安防监控类的图像、音频和视频类信息；

(5) 城镇供水规范化考核、二次供水专项检查等专项业务的图像和视频类信息；

(6) 城镇供水客户投诉和处理等客户服务、设备检修系统的图像、音频和视频类信息；

(7) 城镇供水质量控制考核的检测结果上报、考核结果的通报等图片类信息。

### 2 基础信息分类与编码

#### 2.1编码管理

在建立供水信息系统时，应依据CJ/T 474-2015和CJ/T 541-2019的规定，建立城镇供水水质指标及城镇供水管理系统基础信息的编码体系和管理制度，作为供水信息系统开发建设的基础。

#### 2.2分类与编码

为保证城市供水基础信息在同一系统内或不同系统间的交互与加工处理，应对城镇供水信息系统基础信息和供水水质指标进行分类与编码。分类与编码应符合下列规定：

* + 1. 城镇供水信息系统基础信息的分类与编码符合CJ/T 541的规定；
    2. 城镇供水水质指标分类与编码应符合CJ/T 474-2015的规定；
    3. 非结构化数据可参考CJ/T 541中第4章的规定进行分类与编码；
    4. 为保证编码系统的完整性，对于CJ/T 541中未涉及的新增基础信息，应依据CJ/T 541中第4章的规定进行分类与编码；对于CJ/T 474-2015未涉及的新增水质指标，应依据CJ/T 474-2015中的规定进行分类与编码。

# 第三章 基础信息加工处理基本流程

## 第一节 国内外城市供水信息系统发展现状

### 1 国外城市供水信息系统发展现状

国外城市供水信息系统起步早，经过几十年的发展，达到了建立信息系统基础、完善信息系统数据、细化信息系统模块、整合信息系统平台的整体目标。在这个过程中，不仅保证了城市供水信息系统功能的可靠性、完整性、实用性，而且积累了大量的生产运营数据以及使用经验，为其进行数据分析，建立水质、水利模型提供了基础。

大多数国外供水信息系统由计算机辅助(IT)和业务技术支持(OT) 两块技术组成，加强IT与OT之间的融合是国外供水信息系统的主要工作，因此，国外供水信息系统相较于国内，尤其重视生产运营数据的分析与应用。国外供水信息系统不仅在水系统范围内得到大量应用，而且延伸到用户用水量预测、分析相关服务，如对不同类型用水的实时数据进行分析归纳、获得为提高水费发票精准度和灵活性的信息、基于历史数据对水的消费趋势进行分析等。事实上，越来越多发达国家对未来城市供水信息系统所提倡的概念是：建立一个有弹性理念的供水信息系统，也就是通过对系统可能产生的风险进行评估，对系统接纳风险的能力进行评估，进而提出系统弹性恢复的建议措施。

国外城市供水信息系统在发展过程中也存在一些问题。例如，缺少可以参考的经验先例。由于国外起步早，进度快，在寻找有价值的参考案例方面也存在困难，因此只能一边摸索一边发展。

### 国内城市供水信息系统发展现状

受限于国内早期信息化能力的不足，国内城市供水信息系统相比国外起步较晚。大体来看，我国城市供水信息系统的发展主要分为自动化、数字化和智慧化三个阶段：

（1）自动化阶段

在这一阶段，我国城市供水信息系统的建设集中在基础信息的自动化采集上，通过对阀门、泵站、生产工艺过程等的自动化操控，不仅提高了水质、水压和流量等涉水数据的测量水平，而且很大程度上代替了繁重的人工操作，降低了工人的劳动强度。

（2）数字化阶段

在这一阶段，国内城市各供水相关单位利用局域网络、3G/4G网络和数据库技术等，逐步建立起各自的业务系统和数据存储系统，极大提高了信息存储、查询和回溯的效率，初步实现了业务管理和行政办公的信息化。目前，我国绝大部分城市供水信息系统正处于该阶段。

（3）智慧化阶段

这一阶段我国城市供水相关单位成熟运用物联网、云计算、大数据和机器学习等新一代信息技术，在实现对数据全方位采集的同时，对数据进行深度处理，实现信息化建设和管理水平提升的充分结合。目前，我国大多数城市供水信息系统的建设正处于从数字化阶段向智慧化阶段迈进的过程中。

国内城市供水信息系统在建设时期大多缺少整体规划，“重建设，轻管理”的现象较为普遍，系统碎片化问题突出。总体来说，主要存在以下问题：

（1）系统间整合程度低，功能重叠。由于不同单位、企业在依照各自需求进行信息系统建设时缺乏统一的规划，各类业务系统大多按照不同部门各自开展的业务进行职能划分，因此不同系统建设时所采用标准不统一，系统之间交互性差，且存在重复建设问题。

（2）系统内业务流程不统一、不规范。由于业务流程不规范，相对原始的管理模式仍然存在，城市供水行业信息化仅仅实现了现有管理模式的计算机化。与业务流程相配套的国家统一法规、标准相对滞后，因此各地区的单位或企业在建设信息系统时往往根据自身需求，自行制定工作规范和标准，导致信息共享、交换不畅，各系统间无法通力合作，信息系统应有的优势得不到充分发挥。

（3）安全管理与运行维护问题。主要体现在四个方面，一是缺少统一的安全策略，安全防护水平参差不齐；二是网络安全基础设施建设不完善；三是没有全面的IT资源管理平台；四是没有明确的机房建设标准。

### 3基础信息管理存在的问题

#### 3.1信息孤岛问题

目前，国内供水机构大都构建有供水业务管理信息系统，也花费了大量的人力和财力，经过一段时间运行后也积累了大量历史数据，如何准确、高效地对这些数据进行处理，充分挖掘数据中的价值，是当前面临的一个重要挑战。而且在实际应用中也存在一些问题：供水机构的SCADA系统、GIS系统、管维系统、水力模型等各个业务子系统各自为政，信息孤岛现象严重，没有实现共享；因孤岛、分散问题，造成了数据流转周期长，出错率高；还会造成技术、生产部门不能及时获得信息数据，信息整合、处理滞后，即使有“数据中心”也只是将各类来源的数据从一个模块“机械”地拷贝或导入到另一个模块数据库中，当历史数据过于庞大时，就会导致系统运行缓慢、数据丢失等各类问题，从而难以对供水生产全过程进行监控并有效指导公司的生产和经营。

各地智慧水务建设正如火如荼的开展，结合供水机构供水管理现状，开发一套面向服务的耦合城市供水水质综合管理系统很有必要，数据综合分析、深度挖掘不能只停留在口号上。

#### 3.2数据质量问题

目前，数据质量的定义还没有统一的形式。很多文献从不同的角度和应用范围对数据质量进行了定义，包括如下三种:1）数据质量指信息系统满足模式和数据实例的一致性、正确性、完整性和最小性四个指标的程度。2）数据质量是数据适合使用的程度。3）数据质量是数据满足特定用户期望的程度。数据质量维度是数据质量的特征，它们为度量和管理数据的质量提供了一种途径和标准，一般包括数据的一致性、准确性、完整性、统一性等四个方面。

随着水务信息化进程的不断深入，供水信息系统积累了海量数据，但当用户根据需要使用这些数据时，却往往会因各种数据质量问题影响到数据的深层分析和价值挖掘，无法实现真正的“智慧”。常见的问题包括完整性、拼写度、矛盾性，命名冲突以及不一致等问题。数据质量问题来源是多个方面的，可以由人员、流程和系统等问题造成，主要问题包括：1）可能由于采集过程中收集了不符合要求的数据；2）在信息传播过程中数据失真、误码；3）在信息处理过程中数据压缩、转换的执行失败；4）在信息存储过程中某些原因导致的加载失败等。

## 第二节 基础信息加工处理基本流程

### 1 必要性

因城市供水信息系统建设存在信息孤岛、数据质量不高等问题，数据的有效分析与价值挖掘大打折扣，为使系统中各类基础信息发挥出巨大的经济效益和社会效益，应对其进行筛选、加工、浓缩、整理、去粗取精、去伪存真等加工处理，使得零散、无序、彼此独立的信息具有条理性和系统性。信息的加工处理过程，是把大量的这种信息变成能够对企业的决策有指导意义的信息。

### 2 加工处理基本流程

信息加工处理基本流程可分为采集-清洗、转换、装载-存储-检索分析-展示等五个阶段。涉及到具体项目应用时，根据数据采集方式、应用的目的等不同，加工处理流程略有区别。如数据采集方式为人工录入时，因未涉及到数据的跨信息系统抽取问题，其加工处理流程为采集-存储-检索分析-展示等。

加工处理后的信息变成易于观察、传递、分析、利用的形式，实现了信息的效用。如在信息加工处理的关键环节—检索分析中，利用数据分析挖掘技术，从大量的、不完全的、随机的数据中，找寻隐藏在数据背后的信息，揭示数据整体的特征，并预测其发展趋势，进而指导企业的决策活动。

## 第三节 相关行业信息系统标准规范

标准规范对我国各地正在兴起的供水系统建设起着重要的指导和规范作用。目前，在国家与行业层面已形成了不少信息化建设标准，现有标准按内容大致分为信息化标准体系标准、数据库建设类标准、数据采集传输交换类标准、信息安全类标准等。生态环境部在信息化建设工作上起步较早，其标准体系相对完善，但仍缺少基础信息质量控制方面的标准规范。而供水行业信息化标准建设正处于探索性阶段，标准规范的数量还存在不少差距，目前仅在指标编码标准、编制数据交换标准、数据库建设规范、系统建设规范、系统安全规范等方面出台了相应的标准规范。在供水基础信息加工方面，也仅制定了《城镇供水管理信息系统 基础信息分类与编码规则》（CJ/T 541），规定了供水管理信息系统的基础数据分类方法、编码原则与方法、基础信息分类与构成、数据字典分类与代码等内容，但是尚缺乏分类编码后的供水基础信息在采集、清洗、入库、分析展示等具体加工处理质量控制方面的规范文件。供水监管信息化平台建设与管理对基础信息日益增长的质量要求与当前较为落后的标准规范体系之间存在显著矛盾。

# 第四章 基础信息加工处理的单元技术

## 第一节 数据采集

### 1 结构化数据采集

#### 1.1 抽取

选择socket、web service、web socket、http、数据库共享等技术，通过系统中受控的接口定期从供水调度系统、实验室管理系统、管网监测系统等子业务系统或跨平台系统中，对如城市水质动态信息、实验室水质检测数据、压力监测数据、流量监测数据、供水企业设备工况、材料库存、售水情况、维护检修记录等进行跨系统的全量或增量抽取，并完成数据库对应临时表的数据插入。针对以上数据可以从原始业务记录中根据定义的数据抽取规则，将已录入的数据抽取出来并形成标准格式的数据缓存，供进一步的数据转换处理或备份使用。

#### 1.2 在线采集传输

选择socket、光纤、NB-IoT、4G等通讯方式，通过网络端口从城市供水信息系统外部的在线采集设备中，对如压力、流量、水质等监测数据进行采集。数据上传至系统后进行存储，供进一步的数据转换处理或备份使用。被采集数据可以是模拟量，也可以是数字量。

#### 1.3 人工录入

对于以纸质文档或电子文档形式保存的系统无法通过跨系统抽取的结构化数据，如城市供水基础信息、供水单位（公司级）基本信息、水厂基本信息、在建（规划、拟建）项目信息、城市供水单位水质水量月报、在建项目季报、突发水质快报等，可以采用人工录入的方式将数据采集到系统中。负责人工录入的系统合法用户既可以采用http、ftp等技术将数据上传至系统中存储，也可以登录系统中相应的数据录入界面，通过前端嵌入式电子表格工具人工输入相关数据，保存后上传至系统中存储。

#### 1.4 网络爬虫

对于广泛分布于互联网中的结构化的城镇供水信息，可以采用网络爬虫技术，按照预先定义好的与城镇供水信息有关的主题有选择地进行网页爬取，然后根据一定的采集策略从抓取的网页中采集相关数据并逐条存储到数据库中。

### 2 半结构化及非结构化数据采集

#### 2.1 抽取

可以选择socket、web service、web socket、http等技术，通过系统中受控的接口定期从电子邮件系统、设备检修系统、客户服务系统、安保监控系统等业务系统中，对电子邮件、维修记录、投诉工单、视频信息等进行跨系统的数据抽取，将半结构化数据转换为结构化数据进行存储，将非结构化数据以文件的形式存储到系统中。

#### 2.2 人工录入

对于以纸质文档或电子文档形式保存的系统无法通过跨系统抽取的半结构化或非结构化数据，如值班日志、办公文档、图片、各类报表、图像等，可以采用人工录入的方式将数据采集到系统中。负责人工录入的系统合法用户既可以采用http、ftp等技术将数据以xml、json、文件的格式上传至系统中，并将半结构化数据转换为结构化数据进行存储，将非结构化数据以文件的形式存储到系统中。

#### 2.3 网络爬虫

对于广泛分布于互联网中的结构化的城镇供水信息，如发布的信息、新闻、通知公告等，可以采用聚焦网络爬虫技术，按照预先定义好的与城镇供水信息有关的主题有选择地进行网页爬取，然后根据一定的采集策略从抓取的网页中采集相关数据，并将半结构化数据转换为结构化数据进行存储，将非结构化数据以文件的形式存储到系统中。

## 第二节 数据清洗、转换及装载

### 1 数据清洗

#### 1.1 结构化数据清洗

1. 数据清洗原则

结构化数据针对数据重复性、有效性、时效性、合法性进行清洗。

结构化数据清洗前应定义结构化数据编码规则及数据字典，数据编码用于数据身份的唯一识别。数据清洗时根据定义的合理性规则对所抽取的数据进行合法性校验。只有符合过滤条件且具有合法标识的数据才能用于后续的数据转换处理及装载。

1. 库表清洗

库表清洗主要针对结构化数据，清洗过程应包括对原始数据编码字典进行转换、无效数据的处理、特殊符号的替换、指标或监测值是否在合格范围内的判断等，将整理后的数据更新到原有的数据表中。

清洗方法包括非空检核、主键重复清洗、非法代码清洗、非法值清洗、数据格式检核、记录数检核，详见表4-1库表清洗方法。

**表4-1库表清洗方法**

|  |  |  |
| --- | --- | --- |
| **序号** | **清洗方法** | **主要内容** |
| 1 | 非空检核 | 要求实时库表字段为非空的情况下，需要对该字段数据进行检核。 |
| 2 | 主键重复清洗 | 多个业务系统中同类数据经过清洗后，在统一保存时，为保证主键唯一性，需进行检核工作。 |
| 3 | 非法代码清洗 | 非法代码、代码与数据标准不一致等，主要针对实时监测数据中的异常数据。 |
| 4 | 非法值清洗 | 取值错误、格式错误、多余字符、乱码等，需根据具体情况进行校核及修正。 |
| 5 | 数据格式检核 | 通过检查表中属性值的格式是否正确来衡量其准确性，如时间格式、多余字符、乱码。 |
| 6 | 记录数检核 | 指各个系统相关数据之间的数据总数检核。 |

1. 缺失值清洗

在线监测数据具有体量大、连续性等特点，为不影响分析结果，在跨系统抽取过程中，应对缺失的数据进行清洗。

确定缺失值范围，按照缺失比例和字段重要性分别制定策略：

重要性高，缺失率低：通过计算进行填充；通过经验或业务知识估计；

重要性高，缺失率高：尝试从其他渠道取数补全；使用其他字段通过计算获取；

重要性低，缺失率低：不做处理或简单填充；

重要性低，缺失率高：删除该字段。

填充缺失值方法有以下三种：以监测数据经验推测填充缺失值；以同一监测类型的计算结果（均值、中位数、众数等）填充缺失值；以不同监测类型的计算结果填充缺失值。

1. 逻辑错误清洗

针对结构化数据中时效性要求不高的数据，应采用逻辑错误清洗的方式进行，主要规则如下：1）去重。常见的就是去掉多余的空格，人工录入数据时难免会添加一些肉眼不易辨别的空格，造成数据重复等问题。2）去除不合理值。常见的就是单位未注意造成数量级不合理，例如管网漏损率（%）超过100，当检测到不合理值时，需要去除或者用缺失值代替。3）修正矛盾内容。例如某市的城市供水总量必然是大于其公共供水总量，如果出现小于的情况，就需要修正基础信息。

#### 1.2 半结构化及非结构化数据清洗

针对半结构化及非架构化数据以保证数据的合法性为主要清洗目的，对可能出现的数据二义性、重复、不完整、违反业务规则等问题，将有问题的数据先剔除，根据实际情况调整相应的清洗操作，有效清除无效数据，保证数据质量。数据清洗处理有多种形式：取值范围检查、交叉记录验证、格式校验、空值检查、重复记录清除、数据源纠错等。

### 2 数据转换

#### 2.1 结构化数据转换

结构化数据转换应根据供水信息系统已定义的基础信息数据、监测数据的格式进行逻辑转换，将从其他供水业务系统抽取的数据进行必要的转换处理，形成满足目供水信息系统数据库业务规则要求的数据形式。

结构化数据由其他类型业务系统进行转换时，主要转换方法如下：

（1）共享数据库方式

从其他业务系统接入实时数据、历史数据或对数据一致性要求较高时，可采用共享数据库方式。采用此方式应避免接入的系统出现访问受限情况，不建议在本地业务系统设置共享数据库。

（2）远程过程调用方式

此方式适用于从其他业务系统接入实时数据或对数据同步性要求较高时，常用方式为Web Services协议。此方式较易实现，但要求通信双方必须同时工作。

#### 2.2 半结构化及非结构化数据转换

半结构化及非结构化数据在加工处理时，由于受数据格式影响，不可直接转换成量化信息用于分析，可将其转换为结构化数据，转换方法如下：

（1）文件共享方式

需从其他业务系统进行结构化数据转换时，包含固定格式的图片、统计报表、报告等文件，可采用文件共享方式进行交互，常用方式为FTP方式。

文件共享方式须搭建文件共享服务器，宜使用局域网或虚拟局域网网络，保障文件安全。

（2）非机构化数据转换

投诉工单、值班记录等非架构化数据，可使用分词技术或预先设定的匹配规则，进行数据分类转换，按照内容属性、时间属性、空间属性、来源属性、格式属性、使用属性等进行统一分类命名。

### 3 数据装载

#### 3.1 结构化数据装载

结构化数据装载宜使用关系型数据库进行装载和管理，满足对数据的检索、分析、统计需求，同时装载应考虑应用需求、数据备份需求、数据共享需求及数据容灾需求。

#### 3.2 半结构化及非结构化数据装载

半结构化及非结构化数据可按照树形文件形式进行数据装载，树形文件可按照不同文件类型设计，可对单个数据压缩装载，可对多个数据打包压缩。可使用工具软件对已装载数据进行索引。

## 第三节 数据存储与备份

### 1 结构化及半结构化数据存储

结构化数据应针对不同业务数据类型，按照不同的数据库表格进行存储。数据库表格主要包括城镇供水管理基础信息、城镇供水单位基础信息、城镇供水水厂基础信息、城镇供水设施在建、规划拟建项目基础信息、城镇供水单位月供水水量与水压动态信息、城镇供水水质基础信息等表格。

针对监测数据，依据样品来源不同可分为水样数据和质控数据，质控数据又包括标样数据、加标回收数据、手工比对数据。

依据数据数值大小对可以数据分类为超仪器检出限数据、零值数据、负值数据等。

依据数据变化趋势可对数据分类为连续恒值数据、离群偏小数据、离群偏大数据。

对于水质监测数据的分类标识设置可参考表4-2数据分类标识。

**表4-2 数据分类标识**

|  |  |  |
| --- | --- | --- |
| **数据标识** | **标识定义** | **说 明** |
| N | 正常 | 测量数据正常有效 |
| T | 超上限 | 监测浓度超仪器测量上限 |
| L | 超下限 | 监测浓度超仪器下限或小于检出限 |
| P | 电源故障 | 系统电源故障 |
| F | 仪器通信故障 | 仪器数据采集失败 |
| Z | 取水点无水样 | 取水点没有水样或采水泵未正常上水 |
| S | 手工输入数据 | 手工输入的补测值 |
| M | 维护调试数据 | 仪器仪表处于维护期间产生的数据) |
| ZF | 零值异常 | 测量数据为零 |
| WQW | 水质明显变差 | 当前参数水质类别变差 |
| WQB | 水质明显好转 | 当前参数水质类别变好 |
| QCF | 质控失败 | 质控数据不合格，该监测数据无效 |
| OVS | 离群偏小数据 | 小于上一次监测值的1/3倍 |
| OVL | 离群偏大数据 | 大于上一次监测值的3倍 |
| KPF | 关键参数异常 | 仪器上传的关键参数不合格标准 |
| CF | 恒值不变 | 数据连续3条及以上不变 |
| D | 仪器故障 | 仪器故障 |
| B | 仪器离线 | 仪器离线(数据通讯正常) |
| HD | 现场启动测试 | 现场人员通过基站监测系统以手工即时执行的方式发出的命令，并让仪 器自动完成操作，包括水样测试、标样核查测试、加标回收测试、零点 核查、跨度核查等 |

对于工作日志、电子邮件、运行维护、客户服务等半结构化数据，需先转换成结构化数据再进行存储。

### 2 非结构化数据存储

非结构化数据的存储应遵循集中管理、高效利用和支持管理决策的原则，在存储过程中，可以按照文档、图片、音频、视频等类别进行文件存储。为便于数据检索宜对非结构化数据按照一定的规则进行重命名并建立索引。非结构化数据的存储时间原则上不低于半年。

### 3 数据备份

为保证运行安全和数据安全，应建立数据备份、信息冗余和灾害恢复机制，当系统出现故障、数据传输中断等突发事故时，应能在24小时内恢复，数据缺失间隔不能大于1小时。

对数据库进行定期备份，每天自动备份一次，备份数据与原数据宜异地存储，并进行加密保护。

## 第四节 数据分析及展示

数据分析与展示可采用GIS技术、BI技术、组合报表、数字媒体、人工智能等技术手段，实现入库基础信息的条件检索查询、统计、以及监测和检测数据的综合分析。根据不同的分析结果，采用丰富的图表形式直观展示。

### 1 结构化数据分析

#### 1.1 在线监测数据分析

1. 监测指标自动排序，可实现不同站点的同一指标按监测数值大小进行排序。
2. 站点监测类型分类筛选，可按照水源、水厂、管网、二次供水等不同监测点类型进行筛选。
3. 监测指标超标自动排序，可实现数据超标站点自动排序至最前。
4. 监测指标历史曲线分析，可实现每个指标月度、年度曲线分析，提供柱状图、曲线图，统计显示该指标最大值、最小值、平均值。
5. 监测指标对比曲线分析，可实现不同站点同一指标的变化曲线对比。
6. 监测指标站点合格率统计分析，可实现不同站点任意指标、任意时间段内的合格率统计，包含监测次数、合格次数、最大值、最小值、平均值等。
7. 监测站点实时上数率分析，可实现不同站点的月度上数率、有效上数率分析，包含应上数、实际上数、有效上数、采集指标数、运维公司等。
8. 监测站点指标合格率对比，可实现同一年份不同站点同一指标的合格率对比。
9. 监测站点历年合格率对比，可实现历年不同站点相同指标的合格率对比。
10. 监测数据时空渲染分析，可实现历年在线监测指标热力渲染分析、等值线变化分析，分析时间、指标可选择。
11. 报警分析，可实现不同站点的不同指标在任意时间段内的报警信息汇总，提供仪器量程、国标限值对比。

#### 1.2 实验室检测数据分析

1. 检测报告属性检索应用：检索属性包括样品类型、检测性质、采样计划、采样单位、受检单位、评价标准、采样点、采样时间等。
2. 检测报告指标统计应用：可实现不同采样地点任意检测指标在任意时间段内的统计分析，包括样品个数、指标个数、合格率等，可根据采样计划、数据来源、检测类型进行筛选，可根据指标数值进行排序。
3. 检测报告样品统计应用：可实现不同采样地点的样品在任意时间段内的统计分析，包括样品个数、合格率等，可根据采样计划、数据来源、评价标准进行合格情况筛选。
4. 检测指标历史曲线分析应用：可实现历年实验室检测指标变化分析，包含该指标历年最大值、最小值、平均值等。
5. 检测指标历史曲线对比分析应用，可实现不同实验室采样点的同一指标变化曲线对比分析，对比指标包括各采样点指标最大值、最小值、平均值。
6. 供水单位上报数据统计应用，实现同一供水单位内不同检测指标检测次数、超标次数、最大值、平均值的统计，可调阅不同供水单位不同水厂检测指标明细。
7. 供水单位上报数据分析应用，可实现不同供水单位同一检测指标合格率排名分析，可根据采样类型、时间筛选。可实现同一供水单位内不同检测指标合格率排名分析，可根据采样类型、时间筛选。可实现不同供水单位不同采样类型的敏感指标分析。

#### 1.3 专项业务数据分析

专项业务包括水质督察、供水规范化考核、检测业务培训，数据分析以报表形式进行汇总，可进行时间、空间维度查询、统计、排序。

### 2 半结构化数据分析

对水质投诉历史数据进行统计分析，基于投诉信息对应的水厂和水质检测项目进行对比分析。

1. 投诉工单时空数据分析应用，实现任意时间段内不同水质问题、不同供水区域的工单时空分布检索，实现工单热力渲染识别敏感区。
2. 投诉工单解构数据分析应用，实现任意时间段内不同水质问题、不同供水区域的工单构成比例分析，可调阅投诉工单明细。
3. 投诉工单时间轴数据分析应用，实现不同年度不同水质问题、不同供水区域的工单月度数量曲线分析，可调阅投诉工单明细。

### 3 数据展示

#### 3.1 结构化数据展示

1. 数据地图模式

可按照所属供水单位、所属水厂、监测点类型、监测点名称等属性进行分层显示，并可进行浏览模式切换。

1. 数据列表模式

可按照所属供水单位、所属水厂、监测点类型、监测点名称等属性进行查询。可进行实时监测值自动排序，并提示超标报警信息。。

1. 数据曲线模式

可按照所属供水单位、所属水厂、监测点类型、监测点名称等属性进行数据曲线查询，并显示对应的站点以及站点的各项监测指标曲线。

1. 数据卡片模式

可按照所属供水单位、所属水厂、监测点类型、监测点名称等属性进行查询。站点监测数据实时循环更新播放。

1. 历史数据报表导出

可按照采样点、监测项目、监测时间、合格状态等属性字段查询历史数据并导出报表。

#### 3.2 半结构化及非结构化数据展示

半结构化数据宜转换为结构化数据进行数据展示，具体展示方式参照结构化数据展示章节。

非结构化数据可以树形文件索引形式或文件存储工具展示，按照内容属性、时间属性、空间属性、来源属性、格式属性、使用属性等分类展示。

## 第五节 信息安全

信息安全是基础信息数据采集、传输、处理、存储的基本保障，在基础信息加工处理过程中，通常应采用的基础性安全保障技术包括以下方面。

（1）身份认证技术：供水信息系统须使用身份认证技术确定用户身份的合法性，具体技术手段可使用户名口令、身份识别、PKI证书和指纹认证等。

（2）加解密技术：在数据传输、存储过程中进行数据的加解密，可使用的典型的加密算法有对称加密和非对称加密。

（3）边界防护技术：应防止外部网络用户以非法手段进入内部网络，保护内部网络环境、内部数据存储资源，可安装防火墙和入侵检测设备。

（4）访问控制技术：数据库访问与读取，应设置合理的网络安全防范和保护策略，设置数据库资源访问的限制，并在身份识别的基础上，根据身份对数据资源访问的请求加以权限控制。

（5）主机加固技术：应使用主机加固技术，对供水信息系统相关的操作系统、数据库等进行漏洞加固和保护，提高系统的抗攻击能力。

（6）安全审计技术：供水信息系统的访问记录应包含日志审计和行为审计，通过日志审计协助系统管理员在受到攻击后察看网络日志，从而评估网络配置的合理性、安全策略的有效性，追溯分析安全攻击轨迹，并能为实时防御提供手段。通过对用户的网络行为审计，确认行为的合规性，确保数据访问、处理过程的安全。

# 第五章 质量保障

数据采集作为城市供水信息系统基础信息加工处理的首要环节，采集数据的质量好坏决定了信息系统能否有效发挥作用。

### 1 基本要求

#### 1.1 水质实验室数据质量

水质检测实验室在机构、人员、场所环境、设备设施、管理体系等方面要符合《检验检测机构资质认定能力评价检验检测机构通用要求》（RB/T 214—2017）。

#### 1.2 水质在线监测数据质量

在线监测数据质量保证与控制需符合《城镇供水水质在线监测技术标准》（CJJ/T271-2017）相关要求。

### 2 结构化数据质量保障

#### 2.1 人工录入数据

(1) 预审核

数据录入人员应首先对数据信息进行审核，并保留电子版或纸质版原始记录信息，检查无误后将数据录入系统并对原始记录信息进行存档。审核录入完毕后，数据进入下一级审核。

(2) 正式审核

对初步提交系统的预审核数据进行审核，对存在疑问数据进行标记并加批注写明原因，将存疑数据及相关佐证材料发送给数据录入人员协商认定，再次检查原始记录有关信息的准确性，并由预审核人员负责更改数据并重新录入。正式审核人员对重新录入数据进行再次审核，确保无误后将数据进行入库。

#### 2.2 在线采集传输数据

(1) 审核内容

审核内容主要包括在线监测设备的运行信息，水源、水厂、管网水质在线监测设备获取的实时水质数据。

(2) 自助预审核

1) 存疑数据判定

当监测数据出现以下情况时，系统标记为存疑数据，便于人工复核。

1. 监测数据发生突变或连续不变。
2. 监测数据为0值。
3. 监测数据低于仪器检出限。
4. 监测项目的关键状态值（消解温度、消解时长、显色温度等）不在合理范围。
5. 数值间逻辑关系不符合要求。
6. 无效数据判定

监测数据出现以下情况时，系统标记为无效数据，便于人工复核。

1. 水站维护测试时间段内产生的数据。
2. 水质自动分析仪出现故障产生的数据。

(3) 一级审核

对系统自动预审核、自动标记结果进行初审，对系统自动预审的结果进行确认，针对存疑数据和无效数据进行标记，并加批注写明原因，在规定时间内提交存疑数据及相关佐证材料。因仪器设备故障导致的数据无效，须详细说明原因（如泵故障、采水故障等）。

1. 审核时限
   1. 日审核

每天完成对上一天数据的审核，如果没有在规定时间内审核完毕，数据则自动进入下一级审核。

* 1. 月审核

每月初完成对上月数据的一级审核。

1. 审核依据
2. 数据规范性：查看系统过程日志，监测全过程是否运行正常。
3. 质控符合性：质控过程及手段是否符合相关质控要求，质控数据是否合格。
4. 逻辑合理性：相关项目数据之间逻辑关系是否合理，上下游之间数据逻辑关系是否合理。
5. 数据可比性：水体藻类较多、上游断流、冰封期冰层下水深较浅、采样期间水体中有突降雨影响、水体藻类较多、上游断流、冰封期冰层下水深较浅、采样期间水体中有突发性污染团过境等原因导致样品代表性存疑。

(4) 二级审核

主要针对一级审核中标记的存疑数据进行复核。必要时可组织专家进行讨论，判断数据是否有效，最终形成认定结果，并将数据进行入库。

1）审核时段

* 1. 日审核

每天完成对上一天一级审核结果的标记。

* 1. 月审核

每月初完成对上月数据的审核。

2）审核依据

对存疑数据复审时，应综合考虑以下因素：

* 1. 一级审核提交的存疑数据相关佐证材料。
  2. 存疑数据的监测全过程，包括从开始采样到分析结束的全部过程日志和影像资料。
  3. 采样点现场水体及周边状况、气象条件影响。
  4. 断面近一个月以来历史数据及变化趋势。
  5. 河流上下游、湖库各区域各监测项目浓度水平。
  6. 最近一次实验室检测数据。
  7. 质控数据结果。

#### 2.3 跨系统抽取数据

对于跨系统抽取的数据，在抽取过程中及抽取后，其清洗、转换规则和方法参照本指南第4章执行。

### 3 半结构化及非结构化数据质量保障

半结构化及非结构化数据应保障数据的准确性、时效性等，可通过建立数据命名规则、分类规则，使得存储数据可检索；还可通过关键信息与系统中其他的结构化数据进行关联，提高半结构化及非结构化数据的入库质量和应用价值。

# 第六章 系统维护与管理

### 1 维护管理目标

利用各种技术手段，检测、监控城市供水信息系统的运行状态，及时发现故障、处理故障，保证城市供水信息系统的持续、稳定运行，包括：

1. 异常情况下尽快恢复系统的正常运行；
2. 最小化系统运行中发生的故障对系统业务的影响；
3. 确保各类异常问题都得到有效处理而不会有遗漏；
4. 持续优化维护维护管理措施，建立异常情况处理预案；
5. 执行预先制定的措施降低网络安全事件对系统正常运行的影响。

### 2 人员配置

1. 系统运维管理人员至少应该包括系统管理人员、监测设备设施管理人员与网络管理人员，原则上管理人员必须配置备份。
2. 系统运维管理人员上岗前必须经过相应的资质审查与运维管理技术培训。

### 3 维护与管理

1. 系统管理人员应做好系统应用及数据库的操作系统日常管理、权限管理、密钥管理、接口管理、数据备份与恢复、日志记录、信息系统功能维护等工作；当信息系统发生功能故障或数据安全问题时，应及时对其进行维护，必要时系统开发方技术人员应与系统管理人员共同完成系统故障排除工作，保证系统的正常运行。
2. 监测设备设施管理人员应该严格按照监测设备设施的管理规定做好设备的日常巡检、日常维护保养、定期校准与校验等工作；发现设备故障应及时维修或更换，保证监测工作的正常运行。
3. 网络管理人员首先需要做好对系统运行所依赖的物理服务器、核心交换、汇聚路由、安全设备、数据备份设备等硬件与网络设备的日常功能、安全防护、运行日志记录等工作；并且需要对在线监测设备设施是否正常联网上传数据等链路情况进行定期检测；若发现以上设备或链路故障应及时维修或更换，保证硬件与网络设备的正常运行。

# 附录

# 编制依据

1. 《城市排水防涝设施数据采集与维护技术规范》 GB/T 51187
2. 《城镇供水管理信息系统 供水水质指标分类与编码》 CJ/T 474
3. 《城镇供水管理信息系统 基础信息分类与编码规则》 CJ/T 541
4. 《城镇供水水质在线监测技术标准》 CJJ/T271
5. 《水污染源在线监测系统运行与考核技术规范(试行)》 HJ/T355
6. 《水污染源在线监测系统数据有效性判别技术规范(试行)》 HJ/T356
7. 《固定污染源监测质量保证与质量控制技术规范(试行)》 HJ/T373
8. 《环境基础空间数据加工处理技术规范》 HJ 724
9. 《环境信息网络建设规范》 HJ 460