

团 体 标 准

T/XXX XXXX—XXXX

地址数据治理技术规范

Code of practice for address data governance

（征求意见稿）

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

广东省数字政务协会 发 布

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 基本要求 2

 4.1 时空基准要求 2

 4.2 治理原则 2

5 地址名称标准化技术要求 2

 5.1 原数据要求 2

 5.2 数据成果要求 2

 5.3 数据成果统计 3

6 治理成本度量 3

 6.1 人力成本 3

 6.2 非人力成本 3

7 治理流程与内容 4

 7.1 地址评估分流 4

 7.2 方案设计 5

 7.3 地址预处理 5

 7.4 地址层级规范性治理 5

 7.5 地址元素规范性治理 5

 7.6 地址空间性治理 5

 7.7 地址完备性治理 5

 7.8 整体运营迭代 6

 7.9 质量检查 6

8 评价标准 6

 8.1 合规性评价 6

 8.2 正确性评价 6

 8.3 现势性评价 6

 8.4 完整性评价 6

附录 A（资料性） 地址治理过程记录示例 7

附录 B（规范性） 地址治理成本度量参考标准 9

参考文献 11

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。

本文件由广东省数字政务协会归口。

本文件起草单位：

本文件主要起草人：

地址数据治理技术规范

1 范围

本文件规定了地址数据治理的基本要求、技术要求、成本度量、操作流程与治理内容和评价标准的证实方法。

本文件适用于地址数据的生产、维护、更新以及应用服务。

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 23705—2009 数字城市地理信息公共平台地名/地址编码规则

GB/T 32627—2016 信息技术 地址数据描述要求

GB/T 35639—2017 地址模型

3 术语和定义

下列术语和定义适用于本文件。

3.1

小区 zone

具有一定范围、共同属性的相对独立的区域性地理实体, 包含地片名、标识性地物名、居民小区名、自然村名、村民小组和专署区域等。

[来源: GB/T 35639—2017, 5.4.2, 有修改]

3.2

街路巷 road

供各种车辆和行人通行的城市道路。

[来源: GB/T 23705—2009, 3.3, 有修改]

3.3

门(楼)牌 road number

院落、独立门户或建筑物表示地址编号的描述。

[来源: GB/T 23705—2009, 3.5, 有修改]

3.4

院落 area of interest

单位、住宅小区、建筑物群等具有地理边界的一个相对独立的区域范围。

3.5

地址元素 address element

构成地址的具有地理意义的最小语义单元, 地址元素通常是某个地理实体的名称, 是组成地址的基本单元。

[来源: GB/T 32627—2016, 3.7, 有修改]

3.6

地址层次关系 address hierarchical relationship

地址中各地址元素之间存在隶属的组合关系, 包括单条地址横向层次关系和一组地址纵向层次关系。

[来源: GB/T 32627—2016, 3.8, 有修改]

3.7

地址数据 address data

描述标识和定位人们生产、生活和社会活动所处的空间位置的结构化数字信息。

注：地址数据包括且不限于地址、地址名称、地址别称、空间位置、地理编码、地理坐标、标识码等数字信息。

4 基本要求

4.1 时空基准要求

时空基准要求应满足以下要求：

- a) 空间基准应采用 2000 国家大地坐标系（CGCS2000），坐标单位为度（EPSG:4490）；
- b) 高程基准应采用 1985 国家高程基准；
- c) 时间基准应采用公历纪元和北京时间（UTC/GMT +8.00）。

4.2 治理原则

地址数据治理应遵循以下原则：

- a) 完整性：治理前后地址信息不能有丢失、多余及重复，保证数据完整；
- b) 一致性：治理前后地址信息指向应保持一致不能变更地址指向含义，并且保留治理前后地址信息的对应关系；
- c) 安全性：应满足数据安全性原则，在规定的环境下运行及按照规定进行脱敏处理。

5 地址名称标准化技术要求

5.1 原数据要求

待治理原地址数据分为单来源地址信息和多来源地址信息两种情况，要求分别如下：

- a) 单来源地址信息：指待治理地址信息均为同一来源，需对该来源地址信息进行数字化处理。该来源地址数据应具备有效的地址信息字符，必要时应包含地址信息的在用性、有效性以及其他会影响数字化处理要求的说明；
- b) 多来源地址信息：指待治理地址信息有不同来源，需通过数字化融合形成唯一的地址数据成果。每个来源地址数据均应具备有效的地址字符，必要时应包含地址信息的在用性、有效性以及其他会影响数字化处理要求的说明。

5.2 数据成果要求

原地址信息经标准化、数字化成果应当满足统一的格式和结构化标识要求，如表1所示：

表 1 标准化数字化地址数据的标识字段格式要求

序号	标识符	数据项名称	字段类型	是否必须	要求与示例
1	uid	唯一ID	字符型	是	原始地址自带唯一ID/补充地址的唯一ID
2	norm_address	治理前原始地址	字符型	是	示例：广东省深圳市宝安区西乡街道福中福社区悦和路10号富通城五期15-A-301
3	std_address	治理后标准地址	字符型	是	示例：广东省深圳市宝安区西乡街道福中福社区悦和路10号富通城五期15栋A单元3层301室
4	adcode	行政区划代码	字符型	是	示例：广东省深圳市宝安区的编码为440306
5	x	经度	数值型	是	示例：113.849275
6	y	纬度	数值型	是	示例：22.571937
7	province	省级区划名	字符型	是	GB/T 2260-2007规定的省级行政区划名称，包括省、自治区、直辖市、特别行政区等。示例：广东省
8	city	地级市区划名	字符型	是	GB/T 2260-2007规定的地级行政区划名称，包括地级市、地区、自治州、盟等。示例：深圳市
9	county	县级区划名	字符型	是	GB/T 2260-2007规定的县级行政区划名称，包括县、自治县、县级市、旗、自治旗、市辖区、林

序号	标识符	数据项名称	字段类型	是否必须	要求与示例
					区、特区等。示例：宝安区
10	town	乡镇/街道级区划名	字符型	是	GB/T 2260-2007规定的乡镇/街道级行政区划名称，包括镇、乡、民族乡、街道等。示例：西乡街道
11	community	社区/村（居）委名	字符型	否	社区/行政村（居）委会等基层群众性自治组织名称等。示例：福中福社区
12	road	街路巷/小区名	字符型	否	包括街路巷名、片区名、地片名、标识性地物名、居民小区名、自然村名、村民小组、专属区域名。示例：悦和路
13	road_no	门牌号	字符型	否	示例：10号
14	aoi	院落名	字符型	否	示例：富通城
15	sub_aoi	子院落名	字符型	否	示例：五期
16	building	楼栋号	字符型	否	示例：15栋
17	unit	单元	字符型	否	示例：A单元
18	floor	楼层	字符型	否	示例：3层
19	room	户室号	字符型	否	示例：301室
20	result_mark	治理成果说明	字符型	是	描述该地址治理的结果，取值为新增、修改、删除、不变。
21	address_source	地址来源	字符型	是	地址成果对应的原始地址来源
22	type	地址类型	字符型	是	地址分类结果，取值为院落/片区类、楼栋类、户室类、门（楼）牌类、其他类。
23	create_time	创建日期	日期型	是	格式YYYY-MM-DD
24	update_time	更新维护日期	日期型	否	格式YYYY-MM-DD

5.3 数据成果统计

数据成果应包含治理前后变化内容统计结果，包括：

- 治理前后地址元素变化情况，包括各元素补充、修改、无变化的数量和占比。统计示例如附录 A.1 所示。
- 治理前后地址变化情况，包括地址新增、地址修改、地址删除、坐标补充、坐标修改、地址无变化的数量和占比。统计示例如附录 A.2 所示。

6 治理成本度量

6.1 人力成本

人力成本指用于项目实施和交付所需要的人力资源，人力资源包括：

- 人员工资：指人员的工资、奖金、津贴和补贴等；
- 社保和公积金：指五险一金等；
- 管理成本：指项目实施过程中分摊的管理费用。

人力成本按人力资源所对应的人力单价与投入时长综合计算，其中人力单价根据运营难度系数分档确定，单位为元/人月；投入时长根据项目需求、项目规模和数据治理量级等因素确定，单位为人月。

运营难度系数由技术难度系数与业务难度系数的乘积计算，技术难度系数详见B.1，业务难度系数详见B.2。人力单价档位详见B.3。

6.2 非人力成本

非人力成本包括：

- 数据产品费，包括项目中所需购买的外部数据产品（如数据集、数据算法、数据服务等）所发生的费用；
- 设备材料费，包括项目实施过程中所需要采购或租赁的专用资产、软硬件设备、办公耗材和印刷等费用；
- 业务费，包括差旅、会议费和其他所需辅助活动产生的费用。

7 治理流程与内容

由地址评估分流、方案设计、地址预处理、地址层级规范性治理、地址元素规范性治理、地址空间性治理、地址完备性治理、整体运营迭代、质量检查九个环节组成地址治理的流程，流程图见图1。

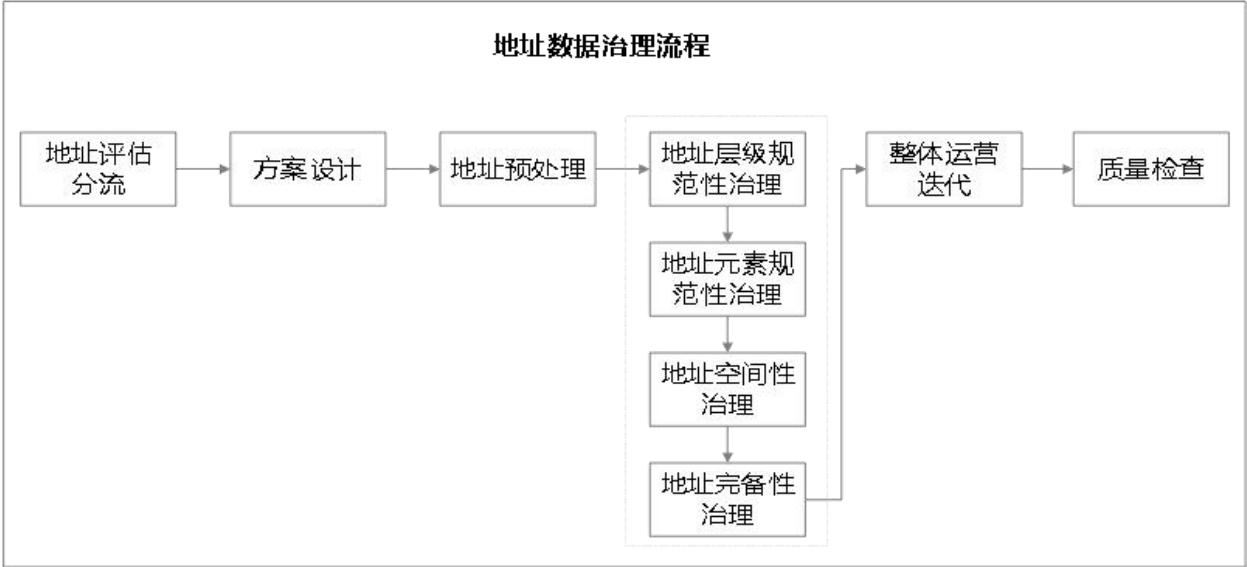


图 1 地址数据数字化治理标准

地址治理成本度量参考附录B。

7.1 地址评估分流

通过地址分类、地址质量评价两个方面对原始地址数据综合评估分流，为方案设计提供数据分析支撑。

7.1.1 地址分类

通过地址分类判断地址详细程度的分布情况并且为后续采取不同的治理侧重点提供分流支撑。地址分类主要包括：

- a) 门（楼）牌类：地址的详细程度仅描述到门（楼）牌级别；
- b) 院落/片区类：地址的详细程度仅描述到院落/片区级别；
- c) 楼栋类：地址的详细程度仅描述到楼栋级别；
- d) 户室类：地址的详细程度描述到户室级别；
- e) 其他类：地址未描述到具体的门（楼）牌、院落、楼栋、户室的任何一种，属于存疑无法判断的地址。

地址分类示例如表2所示。

表 2 地址分类示例

地址分类	数量	占比	类型样例
门（楼）牌类	—	—	广东省深圳市宝安区福海街道新和社区宝安大道6095号
院落/片区类	—	—	广东省深圳市宝安区福永街道白石厦社区凤凰花苑 广东省深圳市龙华区龙华街道向联社区河背新村
楼栋类	—	—	广东省深圳市福田区梅林街道梅丰社区梅华路 183 号碧华庭居翠园 2 栋 广东省深圳市龙华区观澜街道牛湖社区老一村35号
户室类	—	—	广东省深圳市宝安区福海街道桥头社区福盈工业区 C1 栋 301 广东省深圳市宝安区福海街道和平社区和顺路92号201
其他类	—	—	航空路与顺昌路交叉口

7.1.2 地址质量评估

地址质量评估则从地址有效性、地址规范性、空间正确性、地址完备性四个质量维度展开。通过多个维度的综合评估，全面评估地址中的各类潜在地址质量问题，为后续地址治理提供治理方向。具体包括：

- a) 地址有效性：评估地址中存在地址无具体的地理实体，或非地址信息等无法判断其实际位置的地址信息；
- b) 地址规范性：
 - 1) 评估各地址元素的取值是否符合范围、精度、类型等属性取值规范性问题；
 - 2) 评估地址中存在的地址元素重复、缺失、错乱等层级规范性问题；
 - 3) 评估地址中存在的层次关系是否符合地址模型规定的逻辑规范性问题；
- c) 空间正确性：评估地址中坐标的质量和精度是否满足正确性要求；
- d) 地址完备性：评估地址中存在地址重复、地址缺失等重复和未覆盖情况。

质量评估示例如附录A.3所示。

7.2 方案设计

地址数据治理的方案设计，主要包括：

- a) 根据地址数据分析现状和治理效果要求输出地址治理方案；
- b) 根据地址数据治理效果要求和治理方案输出地址治理质检方案。

7.3 地址预处理

地址预处理是地址数据治理的基本环节，是对地址中存在的非规范信息进行统一的预处理，主要包括：

- a) 格式统一，包括繁简体、大小写、全半角等格式统一；
- b) 格式规范化，包括空格、扩号、特殊符号等内容调整和删除；
- c) 冗余信息剔除，包括特殊后缀、人名、无意义词等非地址结构。

7.4 地址层级规范性治理

针对待治理地址，按照地址结构化分词模型规则识别并提取出省级区划名、地级区划名、乡镇/街道级区划名、社区/村（居）委名、街路巷/片区名、门牌号、院落名、子院落名、楼牌号、单元、楼层、户室号等结构化的地址元素层级。该环节应采用成熟的自动化分词算法并辅助以人工修正完成，为后续按地址元素维度进行治理提供支撑。

7.5 地址元素规范性治理

对地址元素进行核实、修改和补充，具体操作包括：

- a) 元素修正：当核实发现地址元素存在错误时，则修正地址元素，包括元素重复、元素名称不规范、元素层次关系错乱等问题；
- b) 元素补充：当核实发现地址元素存在缺失时，则补充地址元素，包括行政区划、街路巷、门牌号、院落等地址元素补充。

治理过程中应使用快递物流、外卖、运营商等行业行为所产生的地址以及地址处理能力，为待治理地址数据中的元素错误和元素缺失的识别提供有效依据，并且为元素修正和元素补充提供有效来源。

7.6 地址空间性治理

采用内业核实和实地外业采核的手段相结合的方式纠正、补充地址的坐标，使得地址坐标质量和精度达到要求，具体操作包括：

- a) 坐标修正：当核实发现地址坐标存在偏差和错误时，则修正坐标到要求的质量和精度；
- b) 坐标补缺：当地址坐标存在缺失时，则按照坐标的质量和精度要求补充坐标。

治理过程中应使用快递物流、外卖、运营商等行业行为所产生的地址空间坐标，为待治理地址数据中的坐标错误和坐标缺失的识别提供有效依据，并且为坐标修正和坐标补充提供有效来源。

7.7 地址完备性治理

通过内业多源地址校验和实地外业采核相结合的方式，识别缺失、多余、无效的地址，具体操作包括：

- a) 缺失补充：当识别到新的地址时，则将新的地址增加到全量地址库中并标记新增状态；
- b) 多余删除：当识别到地址数据所描述的地理实体重复、拆除或其他原因不在使用时，则将地址标记为删除状态。

多源地址应使用政府、企业等各单位的应用所产生的地址进行整合校验，为地址缺失和存在性识别提供有效依据，并且为地址补充、删除标记提供有效来源。

7.8 整体运营迭代

按照质量要求对地址进行查漏补缺及精细修复，主要包括：

- a) 制定地址综合检查项，对治理完成的地址进行整体检查；
- b) 检查发现的地址、地址元素、地址坐标问题，制定针对性的解决方案并完成修复工作；
- c) 核验经过修复后的地址或地址元素，是否修复成功。

7.9 质量检查

根据地址治理质检方案对治理后的成果进行质量检查，并输出质检报告。质检报告包括：

- a) 数据检查内容：时空基准、数据结构、规范性、空间性、完备性；
- b) 质检指标：说明质量检查各个维度的正确性指标，及是否达到合格标准；
- c) 质检问题：说明质量检查发现的主要问题，及每类问题示例。

质检报告示例如附录A.4所示。

8 评价标准

8.1 合规性评价

评价治理后的地址是否满足时空基准要求、治理成果要求及地址治理原则。

8.2 正确性评价

正确性评价主要包括：

- a) 地址文本：以地址全量计算地址文本表述正确率。正确的定义包括，元素规范、层级正确、同一院落内属性一致、行政区划名与官方提供的行政区划一致等；
- b) 空间坐标：以地址全量计算空间坐标正确率。正确的定义需满足各级别地址的坐标精度要求。

8.3 现势性评价

现势性评价主要包括：

- a) 存在性：评价地址数据是否真实有效的存在；
- b) 覆盖率：评价地址数据对标现实的覆盖情况。

8.4 完整性评价

评价治理成果、治理过程中要求输出的材料、报告、指标是否完整。

附录 A
(资料性)
地址治理过程记录示例

A.1 治理前后地址元素变化统计示例

治理前后地址元素变化统计示例见表A.1。

表 A.1 治理前后地址元素变化统计示例

元素名称	数量			数量占比		
	补充	修改	无变化	补充	修改	无变化
行政区划编码						
坐标						
省/直辖市						
市/地级市						
区县/县级市/开发区						
街道/镇/乡						
社区/行政村						
道路/片区						
门牌号						
院落						
子院落						
楼栋						
单元						
楼层						
房间						

A.2 治理前后地址变化情况统计示例

治理前后地址变化情况统计示例见表A.2。

表 A.2 治理前后地址变化情况统计示例

地址变更情况	数量	占比
地址新增		
地址修改		
地址删除		
坐标补充		
坐标修改		
地址无变化		

A.3 地址质量评估示例

地址质量评估示例见表A.3。

表 A.3 地址质量评估示例

评价项	评价子项	问题地址数量	问题地址占比
地址有效性	存疑地址		
	特殊字符		
	元素异常		

评价项	评价子项	问题地址数量	问题地址占比
	地址元素重复		
	地址元素缺失		
	地址元素重复		
	地址层次关系		
空间正确性	院落级坐标		
	楼栋级坐标		
地址完备性	地址重复		
	地址缺失		

A.4 质量检查报告示例

质量检查报告示例见表A.4。

表 A.4 质量检查报告示例

地址数据治理质检报告							
自查 ☐ 过程检查 ☐ 最终检查 ☒				检查时间:XX 年 XX 月 XX 日			
质检任务名称			送检人员				
质检物数量			检验方式				
问题详表							
检查内容		问题	示例	质检数量	错误个数	错误率	修改意见
基本要求	时空基准						
	地址层级结构						
规范性	地址文本						
空间性	地址坐标						
完备性	地址缺失						
	地址多余						
基本要求正确率: XX%							
文本正确率: XX%							
坐标正确率: XX%							
完备正确率 (覆盖且正确): XX%							
综合正确率: XX%							
检验结论			检查人员		审核人员		
备注: XX 正确率=1- (检查错误地址数/检查地址总数)*100%, 其中单维度正确性中的错误地址数仅以该单维度的错误计入错误数, 综合正确性中的错误地址数以任一维度的错误计入错误数。							

附录 B
(规范性)
地址治理成本度量参考标准

B.1 技术难度级别表

技术难度级别取值见表B.1。

表 B.1 技术难度系数取值表

难度级别	难度系数	备注
T1	1.0	一般难度，比较容易开展的重复性、日常性、辅助性技术工作
T2	1.2	初级难度，需要一定技术水平能力支撑运营相关工作
T3	1.5	中级难度，需要较高技术水平能力支撑运营相关工作
T4	2.0	高级难度，需要高级技术水平能力支撑运营相关工作

[来源：T/DGAG 024-2024 政务信息化项目 服务成本度量规范系统业务运营服务 附录A.1]

B.2 业务难度级别表

业务难度级别取值见表B.2。

表 B.2 业务难度系数取值表

难度级别	难度系数	备注
B1	1.0	一般难度，比较容易开展的重复性、日常性、辅助性业务工作
B2	1.2	初级难度，需要一定业务水平能力支撑运营相关工作
B3	1.5	中级难度，需要较高业务水平能力支撑运营相关工作
B4	2.0	高级难度，需要高级业务水平能力支撑运营相关工作

[来源：T/DGAG 024-2024 政务信息化项目 服务成本度量规范系统业务运营服务 附录A.2]

B.3 人力单价档位取值表

采用工作量计算服务费用的地址治理服务，应按表B.3标准计取人力单价：

表 B.3 人力单价档位取值表

人月单价（元/人月）	运营难度系数区间	备注
第一档单价	1	初级人员
第二档单价	(1, 1.5]	中级人员
第三档单价	(1.5, 2.6]	项目管理人员
第四档单价	(2.6, 4.1]	高级人员
注：人月单价可参照各地区人社部门发布的相应信息传输、软件和信息技术服务业岗位的行业工资价位并结合各投标单位实际情况以及市场现状综合确定基准人月单价。		

B.4 地址治理内容难度系数参考表

地址治理内容难度系数参考表见表B.4。

表 B.4 地址治理内容难度系数参考表

治理内容大类	治理内容小类	技术难度系数	业务难度系数
地址评估分流	地址评估分流	T3	根据实际项目数据来源、数据基础现状、治理效果要求程度等情况综合确定
方案设计	方案设计	T4	
地址预处理	地址预处理	T2	

T/XXX XXXX—XXXX

治理内容大类	治理内容小类	技术难度系数	业务难度系数
地址层级规范性治理	地址层级规范性治理	T2	
地址元素规范性治理	地址元素规范性治理	T2	
地址空间性治理	内业治理	T2	
	外业治理	T3	
地址完备性治理	内业治理	T3	
	外业治理	T4	
整体运营迭代	整体运营迭代	T3	
质量检查	质量检查	T2	

参 考 文 献

- GB/T 2260—2007 中华人民共和国行政区划代码
GB/T 37118—2018 地理实体空间数据规范
GB/T 42528—2023 时空大数据技术规范
T/DGAG 024—2024 政务信息化项目 服务成本度量规范系统 业务运营服务
-