引文格式：闻达,吴夏倩,潘慧婕,等. 浙江省在线标准地址库自动化质检研究与设计[J]. 地理空间信息, 2024, 22(1): 117-119. doi:10. 3969/j. issn. 1672-4623.2024.01.028

浙江省在线标准地址库自动化质检研究与设计

闻　达[1]，吴夏倩[1]，潘慧婕[1]，周　煜[1]，沈钜龙[1]，卢春阳[1]
（1. 浙江省测绘科学技术研究院，浙江 杭州 311100）

摘　要：依据浙江省标准地址库日常更新运维情况，为了满足多源数据质检要求，解决人工质检效率低、错误率高、缺乏自动化程序跟踪处理等问题，为离线和在线地址库数据设计了质检规则，归纳和分析了地址库的空间关系和属性关系，并开发了对应的质检模块，形成了质检流程闭环，以满足质量评定和质检规范要求，实现标准地址库数据自动化、智能化、规范化质检，进而提高浙江省标准地址库整体质检效率，节约了人力和时间成本，大幅提高了数据质量。

关键词：地址库；在线质检；数据质量
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2024）01-0117-03

Research and Design of Automatic Quality Inspection for Online Standard Address Database in Zhejiang Province

WEN Da[1], WU Xiaqian[1], PAN Huijie[1], ZHOU Yu[1], SHEN Julong[1], LU Chunyang[1]
(1. Zhejiang Academy of Surveying and Mapping, Hangzhou 311100, China)

Abstract: According to the daily update operation and maintenance of standard address database in Zhejiang Province, in order to meet the quality inspection requirements of multi-source data and solve the problems of low inspection efficiency, high error rate and lacking of automatic program tracking processing, we designed quality inspection rules for the offline and online address database data, summarized and analyzed the spatial relationship and attribute relationship of address database, and developed corresponding quality inspection modules to form a closed loop of the quality inspection process and meet the quality assessment requirements and quality inspection specification requirements, which could realize the automatic, intelligent and standardized quality inspection of standard address database data, improve the overall quality inspection efficiency of standard address database in Zhejiang Province, save the labor and time costs, and greatly improve the data quality.

Key words: address database, online quality inspection, data quality

浙江省“基层治理四平台”数字化建设，提出了建立完善日常工作中涉及的基础信息维护与更新、网格绘制等标准制度，明确将地址数据采集更新纳入网格员工作内容，并要求各类重点场所信息与建筑物地址编码进行关联。地址是最常用的社会公共信息，不仅与人们的日常生活息息相关，而且是政府行政、经济建设、社会治理重要的基础信息资源[1]。在信息化体系中，地址是不可或缺的重要节点和桥梁，但因其普遍性和复杂性，导致标准地址的建设与应用面临诸多难题[2]。目前，浙江省地址数据存在来源部门多、标准不一、覆盖区域不全等问题，从而导致社会治理信息落图和分析研判困难。本文基于标准地址库质量评定方法，将质检工作量大、缺少自动化检查工具等难题作为切入点，利用FME技术和Python的空间地理信息接口设计了相关的质检模块，可根据地址库数据质检要求，自动检查地址数据的规范性、完整性、唯一性、准确性等，进而提升数据质量、提高质检效率。

1　质检内容概述

标准地址库质检包括概要检查和抽样详查，概要检查是对全域数据进行统筹检查，包括坐标系、字段、格式等数据的逻辑一致性、规范性；抽样检查是对指定范围内的数据进行详细检查，包括数据采集全面性、数据唯一性、定位准确性、类型正确性、属性完整性等。

1.1　标准地址概述

地址是现代社会经济发展过程中所必需的战略性基础信息资源，对于国家和地区发展具有重要意义[3]。地址数据是基础地理数据重要的组成部分，是

收稿日期：2022-06-01。
项目来源：浙江省测绘科学技术研究院院长科研资金资助项目（DFP2021D0506）。
第一作者简介：闻达（1991—），工程师，主要从事测绘信息化和天地图技术服务工作，E-mail：552494278@qq.com。

人们最常用的地理信息数据之一[4]，可为社会公众的工作和生活提供详实的检索、定位信息，为公众的出行、旅游、购物等活动提供方便[5]。标准地址是指具有统一规范表达形式和统一编码的位置信息描述，用于标识和定位人们生产、生活所处的位置。根据粒度标准地址可分为门址、楼址和户室地址，门址是指具有独立门牌号码的小区、院落和沿街商铺等地址；楼址是指独立的建筑物地址，一个楼址对应一个相对独立的建筑物；户室地址是建筑物内带独立编号的一套或一间户室的地址。标准地址由行政区划、基本区域限定物、院门、院内区域限定物、建筑物、单元户室6部分组成。标准地址库是组织、管理、融合和分析地表海量地址时空信息的数据库基础，也是描述、构建和认知地址空间，进而解决标准地址编码问题的时空平台[6]，承担着整合信息资源、支撑技术应用的职能，可为政府各职能部门和社会公众提供可靠、高效、准确的地理位置定位服务[7]。

1.2 质检规则设计

根据标准地址库质量评定方法，结合实际地址库质检工作要求，质检规则规定了地址数据的逻辑一致性、采集全面性、数据唯一性、定位准确性、属性正确性、属性完整性、属性规范性和数据提交的规范性等。详细质检规则设计见表1。

1.3 技术路线设计

标准地址库质检的技术路线：①基于FME工具搭建质检转换器模型，分别采用直连和本地读取的方式读取在线云数据库和离线数据库，并利用空间分析工具判断门楼址、户室地址与建筑物的定位、距离、压盖等关系；②利用属性判断工具检查数据是否完整、准确、遗漏；③输出检查成果报告和成果定位数据，作业人员及时进行修改，修改完成后再次提交质检，直至质检通过，得到最终的地址库成果。

2 质检研究设计

本文根据技术路线，从数据源的读取设计入手，针对空间和属性的质检要求设计每个功能模块，完成地址库数据质检研究设计。

2.1 多源数据读取设计

地址库的数据主要包括政法委提交的离线数据和在线云数据库，离线数据为GDB格式库体数据，在线云数据库为MySQL地址库，二者类型、格式、字段、数据访问限制各不相同。对于相同的质检规则，需满足不同用户对数据访问和抽取的要求。离线GDB地址库是标准格式版地址库数据，属于线下提交成果，用于在线数据库的增量更新和质检中的数据详查（表2）。

表1 质检规则设计

质检项	质检规则
逻辑一致性检查	坐标系是否为CGCS2000坐标系
	字段是否有缺失
采集全面性	依据空间位置统计建筑物面内地址点数量
	检查楼址与户室地址的关联情况，并提取关联与未关联户室的楼址
数据唯一性	检查建筑物面上是否有多个地址点
	检查门址点与楼址点全称是否重复
定位准确性	检查楼址与建筑物面的关联情况，即建筑物面上楼址点与建筑物面中心点距离超过5 m的楼址点
	检查地址网格代码与网格数据代码一致时，地址点是否定位在网格面内
	检查门、楼址的自重叠
	检查楼址与关联户室地址的重叠
	检查楼址与门址的重叠
定位正确性	检查与建筑物面重叠的门址点
	检查与建筑物面不重叠的楼址点
属性完整性	根据地址数据字段属性表，检查地址的字段（GBD库体：MYTYE，GRID_CODE，ADDRESS_CODE，ADDRESS，CREATETIME，LEGAL_STATUS，STATE；MySQL库体：MTYPE，GRID_CODE，SZS，SZQX，SZZ，LEGAL_STATUS，DZ）是否为空
	根据楼址的UUID字段匹配户室地址的LID字段，检查相应的楼址和户室地址，对比名称相同的字段，检查字段值不一致的楼址和其对应的户室地址
属性正确性	检查门、楼、户室地址唯一码是否重复
	对比门、楼址与网格面中行政区划字段（市、县、乡、社区）的一致性
	检查地址网格代码与网格数据中代码的一致性
	检查地址名称异常字符（不规则字符、空格、回车键等）
	检查地址类型正确性
属性规范性	检查分节内容不规范的地址

在线MySQL云数据库是存储在浙江省政务云上的数据库，其表结构更加复杂，部分数据库表结构见表3。在数据源读取上，通过配置数据库IP地址、端口以及数据库的用户名和密码，实现对数据库链接和要素图层的获取（图1），主要用于质检中的数据整体抽样概查。

2.2 空间关系质检设计

地址库的空间关系质检主要是检查地址数据的定位准确性，包括地址点与建筑物面的空间关系、与统

表2　标准格式GDB数据库表结构（以户室地址为例）

字段名	字段中文名	字段类型（长度）
HOME_ADDRESS_UUID	户室地址永久标识码	Varchar（32）
HOME_ADDRESS_CODE	户室地址代码	Varchar（27）
AFFILIATED_BUILDING_ADDRESS_UUID	所属楼址永久标识码	Varchar（32）
UNIT	单元门号	Varchar（20）
FLOOR_NUM	楼层数	Varchar（20）
FLOOR	楼层	Varchar（20）
HOME_NUM	户室号	Varchar（30）
REMARK	描述	Varchar（50）
POI	兴趣点名称	Varchar（256）
PHOTO	照片	Varchar（256）
LON	经度	Double
LAT	纬度	Double
Z	高程	Double
LEGAL_STATUS	法定状态	Int（4）
CREATETIME	创建时间	Date
UPDATETIME	更新时间	Date
DYINGTIME	消亡时间	Date
STATE	使用状态	Int（4）

表3　MySQL数据库表结构（部分）

字段名	字段类型（长度）	字段说明
ID	Int（11）	主键
FLOOR_ID	Varchar（20）	楼层ID号
UNIT_ID	Varchar（20）	单元ID
POI	Varchar（200）	兴趣点名
GEO	Geometry（0）	空间字段
STATE	Int（11）	使用状态
LEGAL_STATUS	Int（11）	法定状态
BZ	Varchar（400）	备注
LID	Varchar（100）	楼址主键UUID
NAME	Varchar（100）	名称
UUID	Varchar（100）	户室地址永久唯一码
CREATE_DATE	Timestamp	创建时间
CREATOR_ID	Int（11）	创建用户ID
EDITOR_ID	Int（11）	修改过用户ID
EDIT_DATE	Timestamp	修改时间
EDIT_MEMO	Varchar（500）	修改结果说明

一网格面的空间关系以及地址点相互之间的重叠关系。

地址点与建筑物面的空间关系，一方面检查落在建筑物面的门、楼址点，规定门址不能落在建筑物面内，楼址必须落在建筑物面内，通过PointOnAreaOverlayer工具可判断点与面的压盖关系，属于地下



图1　MySQL数据库链接界面

停车场的楼址点可不在面内，因此进一步过滤名称，将不在面内且名称中不包含“地下”的楼址点输出；另一方面，检查落在建筑物面的楼址点位置与建筑物面中心点的距离是否符合要求，设置阈值进行检查，仍采用点面压盖工具筛选落在面内的点，并通过Matcher匹配工具过滤“一面多点”的情况，利用CenterPointReplacer提取面的中心点，利用构线工具LineBuilder连接地址点和面的中心点，计算投影长度，设置长度阈值，将与面中心点距离大于阈值的点输出。

地址与统一网格面的空间关系，包括地址点是否落在对应的网格面内、地址中的网格代码（GRID_CODE）与网格数据的代码（ADCODE）是否一致以及行政区划字段是否一致。通过点面压盖工具判断地址点与网格面的关系，并设置属性聚合，将网格面的代码赋值到地址点的列表中（图2），最后判断其代码的一致性。行政区划字段一致性判断同理。

地址点相互之间的重叠关系包括门楼址的自重叠、楼址与户室地址的重叠以及门址与楼址的重叠，其中门楼址的自重叠通过PointOnPointOverlayer工具判断点与点之间是否压盖，判断Overlaps压盖值是否为1，输出压盖点；楼址与户室地址、门址的重叠检查均通过地理缓冲工具GeographicBufferer对楼址点作一个细微的缓冲面，将户室地址和门址点分别与楼址缓冲面进行压盖分析，相交的地址点则与楼址点重叠。

2.3　属性关系质检设计

地址库的属性关系质检主要是针对数据的唯一性以及属性的完整性、正确性和规范性，包括地址的名称和唯一码是否唯一、户室匹配楼址字段是否相同、是否存在异常字符、关键字段是否为空等。地址名称

（下转第123页）