Sustainable Development 可持续发展, 2025, 15(7), 116-124
Published Online July 2025 in Hans. https://www.hanspub.org/journal/sd
https://doi.org/10.12677/sd.2025.157192

Hans汉斯

基于位置－语义模式构建多源数据融合的县级基础地理实体数据库

孙晓霞，周秀华

江苏省测绘工程院，江苏 南京

收稿日期：2025年6月9日；录用日期：2025年7月10日；发布日期：2025年7月22日

摘　要

自然资源部出台多项政策和技术文件，部署推进新型基础测绘和实景三维中国建设。县级新型基础测绘体系实体数据库建设是其重要任务之一。本文以江苏省出台的相关技术标准为前提，探索了一套县级基础地理实体转换的方法：以大比例尺地形要素为本底库，基于位置－语义模式，融合多源矢量数据，充分利用各类存量数据的优势，避免了重复采集。实践中借助python提高转换的自动化程度，从而提升生产效率。

关键词

多源数据，融合，地理实体

A County-Level Fundamental Geographic Entity Database Constructed Based on a Hierarchical Position-Semantic Model

Xiaoxia Sun, Xiuhua Zhou

Jiangsu Survey and Mapping Engineering Institute, Nanjing Jiangsu

Received: Jun. 9th, 2025; accepted: Jul. 10th, 2025; published: Jul. 22nd, 2025

Abstract

A series of policies and technical documents were released by the National Resources Bureau, calling for significant progress in advanced basic surveying (novel) and Three-Dimensional China construction in reality. Constructing the entity database for a county-level novel basic surveying system

文章引用: 孙晓霞, 周秀华. 基于位置-语义模式构建多源数据融合的县级基础地理实体数据库[J]. 可持续发展, 2025, 15(7): 116-124. DOI: 10.12677/sd.2025.157192

is one of its important tasks. This paper explores a method for converting base geographic entities using Jiangsu Province's relevant technical standards as a premise: taking high-resolution terrain elements as the fundamental repository, utilizing layer relationships based on position-semantic modeling, fusing multi-source vector data, optimally utilizing the advantages of heterogeneous data sources to avoid redundancy in data collection. In practice, Python was utilized to enhance automation levels, thereby improving work efficiency for data conversion.

Keywords

Multi-Source Spatial Data, Fusion, Geographical Entity

Copyright © 2025 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

实景三维中国作为真实、立体、时序化反映人类生产、生活、生态空间的时空信息，是数字中国整体框架构建的核心要素和重要内容。地理实体是现实世界中客观存在、具有空间位置、可以相互区分的地理对象，是实景三维数据的重要组成部分[1]-[3]。

虽然《实景三维中国建设技术大纲》(下文简称《技术大纲》)中规定了基于大比例尺地形要素转换生产基础地理实体数据的技术流程，但在实际生产中有许多实体无法直接转换，比如院落实体，道路实体，构筑物实体等，而且大比例尺地形要素的时效性较差，需要融合其他数据进行补充，以扩展属性，并提高时效性。比如，杨婷等[4]通过分析比较 1:10,000、1:50,000 基础地理信息数据和第三次全国国土变更调查数据，按照实景三维中国建设地理实体构建要求，提出了一种多源数据融合的自然资源地理实体构建方法。李新双等[5]提出一种基于地名地址匹配的数据融合方法，分别对含空间坐标和仅包含位置描述的两种不同类型政务数据进行匹配，提高了地名地址匹配的效率和准确度。叶芬等[6]提出融合 1:10,000 基础地理信息和第三次全国国土调查年度变更等多源数据，以地理实体编码为统一标识符，构建基于知识－规则的省级地理实体融合转换方法。因此，如何利用大量已有数据生产地理实体有待深入研究。

本文以江苏省某县级市为试验区，在《技术大纲》的基础上，结合获取的数据情况，探索融合多源数据的县级大比例尺地形数据转换生产地理实体的方法，为实现数字中国、数字政府、数字经济等国家战略提供有力支撑。

2. 融合多源数据的县级地理实体转换流程

融合多源数据的县级地理实体转化流程如图 1 所示。

1) 源数据收集分析：收集、整理大比例尺地形数据和各类专题数据，分析各数据的用途，用于生产基础地理实体数据。

2) 映射转换：制作大比例尺地形要素与基础地理实体的几何图形映射表和属性映射表，规范大比例尺地形要素与基础地理实体映射关系。依据大比例尺地形要素与基础地理实体映射表进行几何图形和属性映射，开展源数据的映射转换与编辑。

3) 源数据预处理：将收集到的源数据进行处理，包括数据清洗，格式转换，基准统一，数据结构一致化，线构面，数据提取等。

4) 数据融合：多源数据的几何融合和属性融合。

5) 语义化处理：按照《江苏省基础地理实体数据库规范》进行属性完善，并构建实体之间的关系。
6) 数据建库：按照《江苏省基础地理实体数据库规范》进行地理实体的检查入库。

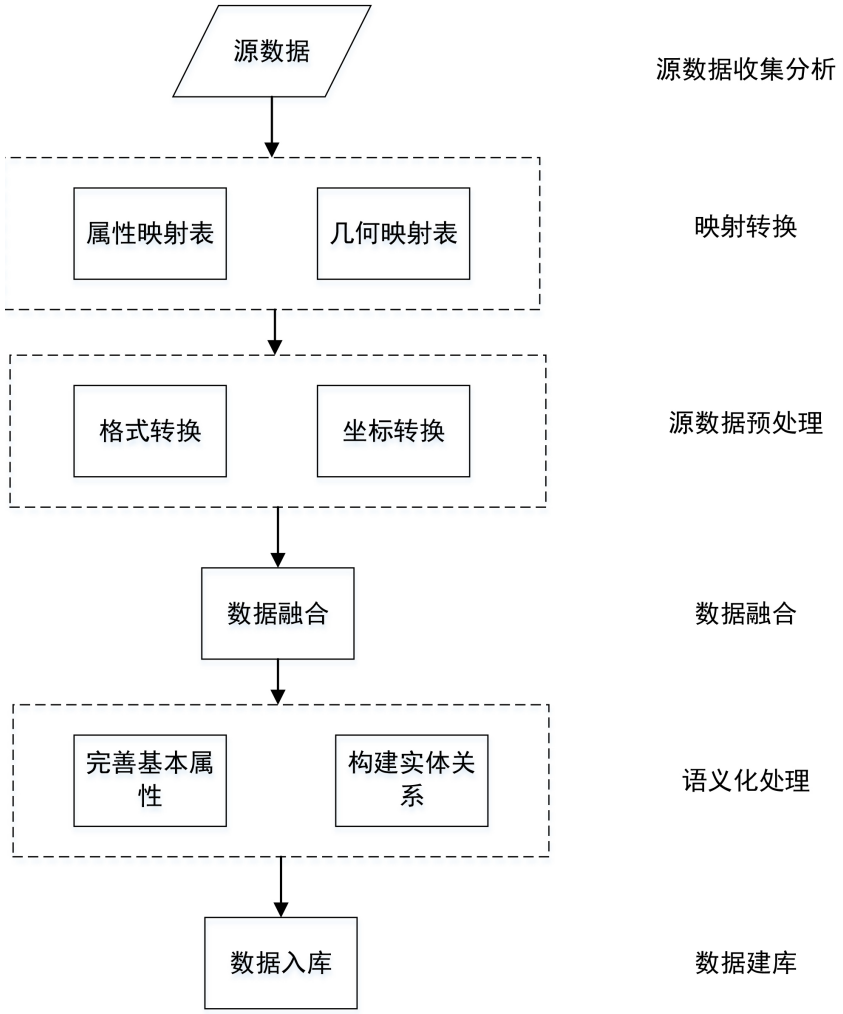

Figure 1. Conversion flowchart
图 1. 转换流程图

3. 关键技术研究

数据融合的技术流程如图 2 所示。

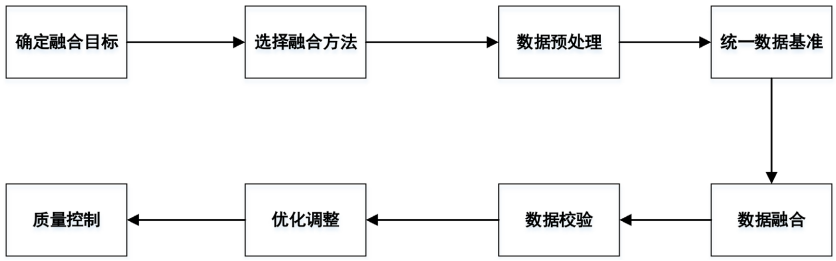

Figure 2. Fusion flowchart
图 2. 融合流程图

3.1. 点数据融合的方法

点状数据的融合采用基于语义和位置信息进行匹配的方法[7]。语义相似度采用 Jaccard distance 方法，距离相似度采用欧式距离计算，基于语义和位置加权判断是否为新增点，若为已有点，则用于补充属性信息，填充缺失值。计算方法如下：

根据样本的特征设置距离相似度权重 ω_D 和语义相似度权重 ω_L，两者权重值满足 $\omega_D + \omega_L = 1$，最终计算公式为 $sim(A, B) = \omega_D \cdot D_{A,B} + \omega_L \cdot L_{A,B}$。

3.2. 线数据融合的方法

线状数据的融合基于 Hausdorff 距离开展[8] [9]。Hausdorff 距离是衡量两个集合之间相似程度的一种方法。计算方法如下：

有集合 A 和集合 B，H_upper 表示从集合 A 到集合 B 的最大最小距离，即对于任意点 a 属于集合 A，找到与之最近的点 b 在集合 B 中的距离，并记录所有这样的距离的最大值。H_lower 表示从集合 B 到集合 A 的最大最小距离，两个集合之间的 Hausdorff 距离是其上半部和下半部中的最大值。计算公式为 Hausdorff 距离 $D = \max\{H_upper, H_lower\}$。

3.3. 面数据融合的方法

面状数据的融合与线状数据相似，先借助缓冲区分析筛选一定范围内的面要素，从而缩小查找范围，提高计算效率。再用 Hausdorff 距离对比相似性，方法与线状数据融合的方法相同[10] [11]。

3.4. 融合原则

针对每类数据，分析不同的数据源，确定每类数据的融合源目标，按点、线、面的几何类型分别选择融合方法，对数据进行一致化处理，以大比例尺地形数据为基础，其他数据源按照可靠度和重要度进行权重划分，融合中标记为相似的数据用于补充大比例尺地形数据的属性，对于冲突的属性字段选择可靠度高的数据源进行补充，对于唯一有值的属性字段则直接更新赋值；标记为不相似的数据则用于新增[12]-[16]。几何和属性特征都融合后，结合影像进行数据的校验，对于不合理的部分进行调整优化，保证数据融合后的正确性。

4. 具体实践

江苏省某县级市以大比例尺地形数据为主要数据源，融合多源数据进行转换生产基础地理实体，流程如下。

4.1. 源数据收集分析

本次实践收集到的数据有大比例尺地形数据，专题要素数据，地形图结合表数据，不动产登记数据，天地图成果数据，地名地址和兴趣点数据，第三次全国国土调查数据，土地利用年度变更调查数据，“十四五”省级基础测绘数据，自然村数据，三区三线数据，行政区划数据，2024 年影像数据。

Table 1. Utilization of multi-source data
表 1. 多源数据利用情况

序号	数据名称	图层名称	利用情况分析
1	政区数据	县、乡镇范围	生产县级行政区、乡(镇、街道)行政区实体
2	天地图数据	POI	地名实体补充数据源

续表

3	不动产数据	自然幢	生产自然幢数据
4	用地分类地块	LAND	提取交通运输用地数据，作为道路范围面数据
5	建筑物	BUILDING	补充大比例尺地形图中房屋面属性，包括用途类型、建筑结构、建筑层数、建筑高度
6	控保建筑和文保单位	PROTECT_BUILDING	采集院落现状范围
7	公共厕所	TOILET	生产厕所实体
8	固定垃圾	DUMP	生产垃圾收集点实体
9	高压铁塔及其连线	HV_TOWER, HV_TOWER_LINE	生产长输输电管线实体
10	桥梁	BRIDGE	桥梁中心线，作为桥梁实体生产补充数据源，同时利用此数据补充大比例尺地形图中桥梁面属性

基于以上多种数据源，构建地名、国土空间规划、管线及设施、建(构)筑物及设施、交通、农林用地与其他土地、其他管理区域、其他管理实体、水利、山体、水体、行政区划单元、院落等 13 类实体。表 1 列出了部分数据的利用情况。

4.2. 映射转换

表 2 列出的是部分字段属性映射表。表 3 列出的是部分大比例尺地形要素与基础地理实体的几何图形映射和属性映射表。多种源数据按照映射表进行数据的归一化处理和提取。

Table 2. Field mapping table
表 2. 字段映射表

映射字段	目标字段	目标字段类型	目标字段长度
ANGLE	JD	DOUBLE	18.2
DWMC	STMC/TYMC	TEXT	100
JZCS	CS	SHORT	
FWZH	ZH	TEXT	20
JZJG	FWJG	TEXT	20
LB	LX	TEXT	50
LFKD	KD	DOUBLE	18.2
MC	STMC/TYMC	TEXT	100

Table 3. Mapping table of large-scale topographic elements and basic geographic entities
表 3. 大比例尺地形要素与基础地理实体映射表

序号	大比例尺地形要素				基础地理实体要素		
	地物编码	符号名称	GIS 图层名称	GB 码	实体分类代码	几何特征	实体图层
1	141300	在建房屋	DLGJ_JMD_PY	310302	230101	RG	STJBB_FW
2	141400	破坏房屋	DLGJ_JMD_PY	310700	230110	RG	STJBB_FW
3	141500	棚房	DLGJ_JMD_PY	310600	230102	RG	STJBB_FW
4	163210	等级公路路基线	DLGJ_JT_LN	420000	220208	ML	STJBB_DL
5	163220	等级公路中心线	DLGJ_JT_LN_C	420000	220208	ML	STJBB_DL
6	163300	等外公路路边线	DLGJ_JT_LN	420400	220204	ML	STJBB_DL
7	163920	城市主干道边线	DLGJ_JT_LN	430501	220302	ML	STJBB_DL

4.3. 源数据预处理

将所有数据统一为2000国家大地坐标系，进行数据清洗和提取，并按照《技术大纲》对数据进行一致性处理，所有数据层按照表4添加公有属性字段，部分数据层需要添加专有属性，比如道路层如表5，房屋层如表6。

Table 4. New public attributes of data layers
表4. 数据层新增公有属性

序号	字段名称	字段别名	字段类型	约束条件
1	TYID	图元标识码	TEXT (26)	M
2	STID	实体标识码	TEXT (50)	M
3	JHLY	几何来源	TEXT (50)	O
4	SXLY	属性来源	TEXT (100)	O
5	TYLX	图元类型	TEXT (50)	M
6	XXLY	信息来源	TEXT (100)	O
7	CJSJ	采集时间	TEXT (8)	M
8	XWSJ	消亡时间	TEXT (8)	M
9	XZQH	行政区划	TEXT (100)	M
10	LX	类型	TEXT (50)	M
11	GB	国标分类码	LONG	M
12	TYMC	图元名称	TEXT (100)	O
13	STMC	实体名称	TEXT (100)	O

Table 5. New attributes of road layer
表5. 道路层新增属性

序号	字段名称	字段别名	字段类型	约束条件
1	QDMC	起点名称	TEXT (50)	O
2	ZDMC	终点名称	TEXT (50)	O
3	SXXFX	上下行方向	TEXT (4)	O

Table 6. New attributes of building layer
表6. 房屋层新增属性

序号	字段名称	字段别名	字段类型	约束条件
1	CS	层数	SHORT	O
2	ZH	幢号	TEXT (20)	O
3	FWJG	房屋结构	TEXT (20)	O
4	DZ	地址	TEXT (255)	O

4.4. 数据融合

按照数据融合的流程进行多源数据融合，包括几何信息和属性信息。本实践中认为大比例尺地形数

据是可靠度最高，重要度最高的数据源，补充新增数据后要进行标记，参照影像进行核对，若后期大比例尺数据此区域有更新时，便于快速查找定位并替换，以保证数据整体的精度。

4.5. 语义化处理

(1) 实体数据编码

本实践中实体标识码采用“6 位行政区划码 + 6 位实体分类代码 + 8 位顺序码”的方式编码，同一个实体有一个唯一实体标识码，比如“北京西路”在几何图形上虽然分为 3 段，但每段的实体标识码都是相同的，且“北京西路”的面状图元和线状图元都共用同一个实体标识码。河流、桥梁、湖泊等有多种图元表达形式的实体都采用此方式编码。

(2) 图元数据编码

本实践中图元标识码采用“6 位行政区划码 +6 位实体分类代码 +2 位几何类型代码 +8 位顺序码”的方式编码，每一个图元有一个唯一图元标识码，一个实体标识码可以对应一个图元标识码，也可以对应多个图元标识码。

(3) 实体关系构建

本实践中构建了如表 7 的实体关系，用于反映不同类型实体间的空间位置关系，并通过知识图谱的方式进行可视化表达，如图 3。

Table 7. List of entity spatial relationships
表 7. 实体空间关系列表

基础地理实体类别			相关实体类型	关系名称
一级类	二级类	三级类		
院落	/	/	房屋	整部关系
交通	城际公路	/	交通附属设施(高速出入口、公交站点、收费站、加油气站、充电桩)	依赖关系
			交通连通交叉口	依赖关系
		/	其他通道(桥梁、涵洞、隧道、匝道)	连通关系
	城市道路	/	交通附属设施(快速路出入口、公交站点)	依赖关系
			交通连通交叉口	依赖关系
		/	其他通道(桥梁、涵洞、隧道、匝道)	连通关系
	乡村道路	/	其他通道(桥梁、涵洞)	连通关系
	轨道交通	/	交通附属设施(火车站入口)	依赖关系
			其他通道(桥梁、涵洞、隧道)	连通关系
水系水利	湖泊	/	河、湖岛	依赖关系
			交通附属设施(桥梁)	依赖关系
			有名称河流	连通关系
	河流	/	水利附属设施(闸、坝)	依赖关系
			其他过水通道(涵洞、输水隧道)	依赖关系
			交通附属设施(桥梁)	依赖关系
	水库	/	闸、坝、溢洪道、出水口	依赖关系
	沟渠	/	水利附属设施(泵站)	依赖关系
政区	区县级行政区	/	乡镇级行政区	包含关系

Figure 3. Knowledge graph example of courtyard-house relationship
图 3. 院落与房屋关系知识图谱示例

5. 结语

本文以县级市大比例尺地形数据为基础，结合多源专题数据，提出了一种基于位置和语义的多源数据融合方法，并结合 PYTHON 开发处理工具，实现数据融合转换的自动化操作。然而，当前研究仍面临数据异质性、计算资源需求和语义理解能力不足等问题。未来，会继续研究融合影像数据的方法，在数据更新中借助人工智能模型训练样本，用于识别变化区域，辅助检测已灭失或者变更的区域，从而提高矢量数据的时效性，进一步提升实际应用价值。

参考文献

[1] 自然资源部. 实景三维中国建设技术大纲, 自然资测绘函[2021]68 号[S]. 北京, 2021.
[2] 尹向军, 黄国平, 孟军, 等. 一种地理实体构建及应用方法[J]. 测绘科学, 2024, 49(3): 168-173.
[3] 刘纪平, 刘坡, 翟亮, 等. 基础地理实体的特征分析与表达初探[J]. 测绘科学, 2023, 48(8): 1-9.
[4] 杨婷, 次仁旺堆, 徐许雄, 符钟壬, 范先锋. 多源数据融合的自然资源地理实体构建方法[J]. 北京测绘, 2025, 39(5): 670-675.
[5] 李新双, 陈军胜, 刘化冰, 赵丽丽. 基于地名地址的多源异构数据融合方法研究[J]. 城市勘测, 2024(3): 50-53.
[6] 叶芬, 胡燕, 杨琪琳, 胡晓丹. 基于知识-规则的省级基础地理实体多源数据融合转换方法[J]. 测绘通报, 2023(7): 160-164.
[7] 孙晓霞. 融合语义和位置信息的兴趣点实体匹配方法[J]. 地理空间信息, 2023, 21(5): 22-24.
[8] 邓敏, 赵彬彬, 徐震, 徐凯. GIS 空间目标间距离表达方法及分析[J]. 计算机工程与应用, 2011, 47(1): 35-39+45.
[9] 林木棵. 城市存量道路要素自动构建地理实体研究[J]. 测绘科学, 2023, 48(3): 57-63.
[10] 马京振, 孙群, 李少梅, 等. 多尺度矢量空间数据融合的基本问题研究[J]. 测绘科学, 2022, 47(9): 208-216.

[11] 王化娟, 张新长, 贠法长. 不动产测量成果构建地理实体的方法研究[J]. 测绘科学, 2023, 48(3): 16-24.

[12] 高培超, 谢一茹, 叶思菁. 基于 PYTHON 的 GIS 空间分析[M]. 北京: 北京师范大学出版集团, 2019.

[13] 王贝贝, 董静. 多源数据支持下的全要素地形图生产实践[J]. 测绘科学, 2023, 48(3): 120-127.

[14] 张祖涛, 刘坡, 翟亮, 等. 黄山市地理实体统一标识体系设计与实现[J/OL]. 测绘科学, 2025, 50(6): 157-167. http://kns.cnki.net/kcms/detail/11.4415.P.20250425.1324.002.html, 2025-06-03.

[15] 刘伟, 顾和和. 基于语义的地理信息空间关系检索[J]. 测绘科学, 2013, 38(5): 16-18+27.

[16] 王磊, 李长辉, 邢汉发. 地理实体化数据处理关键技术研究[C]//广州市城市规划勘测设计研究院. 第四届“测绘科学前沿技术论坛”论文精选. 北京: 测绘出版社, 2012: 557-559+569.