引文格式：张向军,李茜楠,马强,等. 地理实体数据建设关键技术研究及应用探索[J]. 地理空间信息, 2024, 22(8): 64-67. doi:10. 3969/j.issn. 1672-4623.2024.08.014

地理实体数据建设关键技术研究及应用探索

张向军[1]，李茜楠[1]，马　强[2]，丁　磊[3]，钟　斌[1*]
（1. 河南省遥感院，河南 郑州 450003；2. 光山县自然资源局，河南 信阳 465450；
3. 中国人民解放军战略支援部队信息工程大学 数据与目标工程学院，河南 郑州 450001）

摘　要：通过地理实体在新型基础测绘项目中的建设，分析实体数据的组织结构和多源异构空间数据的融合方式，设计了一种实体扩展码编码规则，开发了实体标识码编码工具。解决了多源数据因空间结构差异、属性信息繁杂造成的应用不便、共享困难等问题；设计的扩展码规范了实体码结构，并基于编码工具，实现了实体码快速编码。采用该方法建设了试点区域地理实体数据库，构建了成果展示系统和应用场景，为地理实体数据建设及应用提供了有效的解决方案。

关键词：地理实体；多源数据融合；扩展码；快速编码；实体应用
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2024）08-0064-04

Research on the Key Technologies and Application Exploration of Geo-entity Data Construction

ZHANG Xiangjun[1], LI Qiannan[1], MA Qiang[2], DING Lei[3], ZHONG Bin[1]
(1. Henan Institute of Remote Sensing, Zhengzhou 450003, China;
2. Guangshan County Bureau of Natural Resources, Xinyang 465450, China;
3. School of Data and Target Engineering, PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China)

Abstract: Through the construction of geo-entity in new fundamental surveying and mapping project, we analyzed the organization structure of entity data and the fusion mode of multi-source heterogeneous spatial data, designed an entity extension coding rule, and developed a coding tool based on this system, which could solve the applying and sharing problems of multi-source data caused by spatial structure difference and complex attribute information. The coding system standardizes the entity code structure, and based on the coding tool, the entity code can be fast coded. Using this method, we constructed the geo-entity database in the pilot area, and built the display system and application scenario, which could provide an effective solution for the construction and application of geo-entity.

Key words: geo-entity, multi-source data fusion, extension code, fast coding, entity application

当前，新型基础测绘体系建设在各试点陆续开展实验，地理实体相关研究逐渐增多，但在多尺度数据集成应用和联动更新[1-3]、实体编码等方面都尚未形成一套标准化、规范化的技术体系。文献[4]评价了二维矢量空间数据特征融合等相关理论、算法和技术的研究现状；文献[5]研究了地理实体的多态特征，提出了解决多种层属特征的“地理实体组织法”；文献[6]研究了政务专题数据与地理实体快速融合的技术流程；文献[7]设计了多源数据融合和实体分类的技术方法；文献[8-9]提出了地理实体属性语义相似性的度量方法和基于本体属性特征学习的地理要素分类语义映射方法；文献[10]提出了一种通用的地理实体数据模型；文献[11]阐述了面向城市级应用的地理实体构建范围和规则；文献[12-16]在地理实体标识码体系建设方面，提出了不同的设计方案。现有的研究成果为地理实体建设奠定了研究基础。基于这些研究，本文从实体数据组织展开分析，研究多源异构空间数据融合方式、实体扩展码结构和编码工具，并基于智慧城市试点项目探索地理实体成果应用，为以地理实体为视角和对象的新型基础测绘体系建设提供技术借鉴。

1　地理实体数据建设关键技术

地理实体数据建设质量决定着实体成果的优劣。多源数据成果作为地理实体建设的数据基础，具有格式与坐标基准不统一、几何形态和属性信息不一致、

收稿日期：2023-04-07。
项目来源：国家自然科学基金资助项目（42201443）。
第一作者简介：张向军（1968—），高级工程师，研究方向为摄影测量与遥感，E-mail：2946494668@qq.com。
通信作者：钟斌（1967—），本科，正高级工程师，研究方向为摄影测量及地理信息应用，E-mail：2946494668@qq.com。

生产周期长、联动更新困难、现势性差等特点，不利于实体建设与更新。另一方面，为保证实体码的唯一性，码位一般较长，且一个区域的实体量往往数百万条以上，人工编码工作量巨大，不便更新与维护，容易造成人为错误[16]。因此，研究多源异构空间数据融合方式，构建一套规范化的编码标准，开发自动化的实体编码程序，实现快速编码，是提高实体建设效率的关键和难点。

1.1　地理实体数据组织

地理实体是现实世界中占据一定且连续的空间位置、单独具有同一属性或完整功能的自然地物、人工设施或地理单元，采用标识码来标识每个对象[17]。地理实体模型设计阶段一般包括空间信息、属性信息和实体关系。

实体数据最基本的特征是空间特征，具有特定的地理意义与空间唯一性。一般采用点、线、面、体的形式来描述和表达。一个地理实体可采用多种表现形式。建筑物实体在不同的应用场景中，可用二维面或三维体进行表达。无论其表现形式如何，不同场景的同个实体要素，标识码保持一致。

1）地理实体的属性信息一般分为三类，即基本属性、扩展属性和专业属性。基本属性是实体要素的必备属性，可确定实体唯一性，包括名称、地址、类别、实体标识码、生命周期等；扩展属性是指通过地理实体标识码关联到的经济、社会、法人等方面信息，用以扩展实体应用范围，挖掘数据效能；专业属性一般是专题数据具备的属性信息，由相关职能部门根据管理和业务需求提供的行业性、专题性的属性信息。通过扩展实体的属性信息，挂接相关社会经济信息，实现信息交换与共享。

2）地理实体间具备地物要素间基本的空间关系、类属关系、时间关联关系和逻辑关系。综合考虑实体间的关系特征，建立合适的实体数据库结构，是地理实体建设的关键。实体关系信息一般以应用为导向，按需构建，按类定义。

1.2　多源异构空间数据融合

多源空间数据的融合处理基本思路是先进行数据清洗，即梳理已收集的多源数据，针对数据的时空多态特征进行分析和评定，提取目标数据；其次进行格式转换、投影变换及属性整合等集成处理；再进行空间位置和拓扑关系匹配，分析各尺度数据几何和属性对应关系，剔除重复数据，更新现势性，完成相关数据空间结构与属性信息的一致性处理。经质检，最终得到空间、属性和对应关系协调一致的数据融合成果。具体流程如图1所示。多源数据融合的关键在于几何特征与属性特征融合，其中基于几何特征的融合是其核心内容。

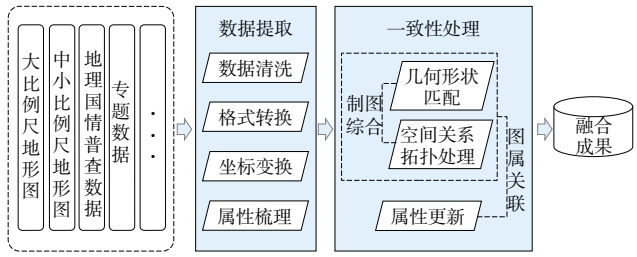


图1　多源数据融合流程图

1.2.1　几何特征融合

多源空间数据通过选择提取、制图综合、特征分析等方式进行要素融合。主要采用的技术方式有：

1）基于位置特征和拓扑特征的同名要素匹配。通过统一坐标基准，同名地理要素的位置、几何形状和走向具有高相似性，同时参考要素间的拓扑关系特征，对不同来源的同名地理要素进行匹配和一致化处理，最终经综合取舍，得到融合后的要素。综合取舍时一般用现势性高的数据更新现势性低的数据，相同时态下精度高的大比例尺数据更新中小比例尺数据。

2）基于属性特征的同名要素匹配。首先梳理不同数据源的各空间要素基本属性信息，包括要素代码、要素类别、名称、性质、空间位置信息等内容；建立多尺度数据代码转换表，表1列出了1∶10 000与1∶1 000基础地理信息要素之间部分映射关系；查找并分析同类属性信息，将拥有相同或高相似性属性信息的地理要素进行匹配。

表1　1∶10 000与1∶1 000基础地理信息要素代码映射示例

1∶10 000比例尺数据			1∶1 000比例尺数据		
地面河流	210101	有向线 范围结构线	地面河流	2101012 2101013	有向线 范围线构面
地下河段	210102	有向线 范围结构线	地下河段	2101022 2101023	有向线 范围线构面
时令河	210200	有向线 范围结构线	时令河	2102002 2102003	有向线 范围线构面

1.2.2　属性信息融合

不同时态的空间数据，同名地物间的属性信息不尽相同。属性信息的融合关键在于对比分析多源数据各要素不同属性字段的信息内容，将同名要素的同义信息建立关联映射，评估现势性更强，语义描述更加

清晰、正确的属性内容，将这些内容进行规范化整理，并更新到融合后的空间数据上。以面状水系数据为例，如图2所示，对表达同一空间实体的不同来源的数据A、B，进行评估分析，A中包含的属性信息有“河流名称代码、名称、库容量、通行月份、类型”等，B中的属性字段为“STBM、HLLX、YSYF、KRL、TYPE”等，提取出有效的属性信息，建立A、B数据属性对应关系，如“河流名称代码–STBM”、“库容量–KRL”、“类型–TYPE”等，再经映射聚合为该空间实体的基本属性信息，通过挂接一些非空间数据信息对其属性内容进行扩展，最终得到较完善的实体成果属性。

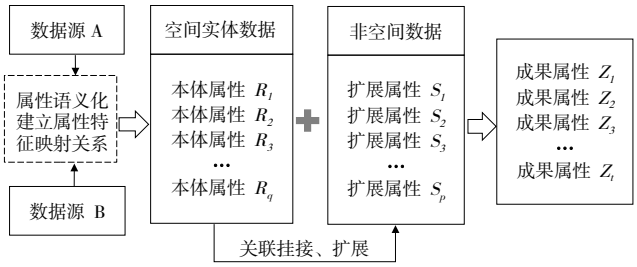


图2 属性融合流程

1.3 地理实体编码

根据自然资源部文件要求，基础地理实体的空间身份编码结构采用“专有标识域+标准域+扩展域”三段划分的编码范式，本文在此基础上进行编码规则设计。

1.3.1 扩展码结构设计

地理实体融合了各行业、多类别空间数据，包含许多新地物和管理对象，进行扩展码结构设计时，为了明确多源数据标识，同时有利于实体数据更新，本文将“扩展域”设计为定长10位的数字码，采用“4位数据源标识+4位时间标识+2位空间表达形式”的组织形式。

1）数据来源标识设计为“2位数据源大类+2位要素类”的组合形式，综合分析已收集的数据，划分数据源大类，使用“01~99”分类表示；要素类按照水系、交通、境界、房屋、植被等类别进行划分，同时区分点、线、面的空间形式，使用“01~99”分类表示；

2）时间标识为源数据现势性，保留到年份；

3）基础地理实体一般采用二维或三维的空间形式进行表达，根据具体表达形式代码使用“20”、“30”。

按照地理实体标识码的设计标准，采用C#作为开发语言，基于ArcGIS Add-In开发框架与Visual Studio软件开发工具，本文研发了适用于ArcGIS Desktop10.1及以上软件使用的编码插件，实现地理实体数据的快速编码，提高编码效率。

研发的编码工具具有自动编码、批量编码、选择编码、编码合并4个功能，编码时可进行图层选择，工具界面如图3所示。

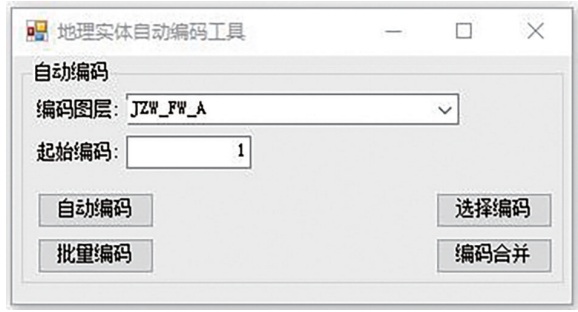


图3 自动编码工具

2 地理实体成果应用

依托于智慧城市试点项目，建设面向新型基础测绘的地理实体数据库，基于实体成果构建展示系统，并设计一些应用场景。

2.1 实体成果

本文基于Oracle数据库平台构建地理实体数据库，汇聚试点城市全空间、多要素、时序化的时空信息，采用“一码多态”，融合多源异构数据，构建基础地理实体数据库，实体库采用集中式存储在政务云环境。实体属性信息包括身份编码、图元信息、实体类别等。

2.2 成果应用

2.2.1 模型浏览

以二维实体成果数据为基础，制作三维单体化实体，构建属性一致的二三维一体化地理实体模型。如图4所示，该实体通过唯一的标识码，关联社会经济信息，挂接附件内容，依托地理实体数据库升级电子地图服务，为自然资源管理及多样化社会应用需求提供空间位置关联服务，及在线地图综合表达与展示服务。

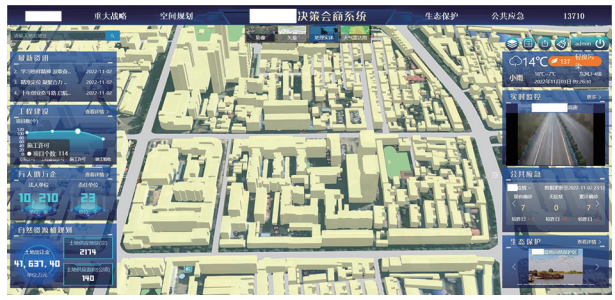


图4 二三维一体化地理实体数据

2.2.2 场景应用

地理实体的应用可以扩展到各行各业、方方面面。

1）将人口大数据、法人大数据和其他部门专题资源与地理实体进行空间匹配关联，形成地理实体

码、人口身份证号和统一社会信用代码“三码合一”，展现城市大数据地图，形成各类分析图表，助力数字政府建设。

2）基于住房楼栋、医院、紧急避险场所和其他公共设施实体的信息匹配，统计小区人口数、紧急出口距离最近避难点等，分析最优避险路线，制定最佳应急处置方案，提高公共应急保障能力。

3）以产业园区地理实体、载体地理实体和楼栋地理实体为基础，关联招商信息，面向用户，集中监控重点洽谈项目、空置载体情况、园区招商引资考核指标，满足以图说话、用图决策的管理需求，评价产业集聚区是否存在低产值、低效用地的情况，分析企业法人、社会经济在该区域内的分布情况及产业结构的调整方向等。

3 结　语

新型基础测绘体系的建设目前还处于试点探索阶段，以地理实体为核心的空间地理信息数据库是加快基础测绘转型升级的关键。本文研究了多源数据的融合方式，探索了地理实体的构建方法，设计了符合实验区特色的实体身份编码，开发了批量、快速编码工具，为其他地理实体研究提供了可借鉴的技术方案。在今后的实践中，还需形成一套系统的实体数据生产、整合建库、联动更新、应用服务等方面的综合性解决方案，制定规范化的标准体系和政策体系，扩展应用范围，从而更好地推动新型基础测绘体系建设。

参考文献

[1] 李成名,印洁,刘晓丽,等.“实体中国”与新型基础测绘的思考[J].测绘科学,2019,44(6):230-234

[2] 彭清山,曹文涛,李海亭,等.新型基础测绘产品体系及技术实现研究[J].城市勘测,2019(6):5-9

[3] 贾丹,王维.基于地理实体的基础测绘生产组织模式探索[J].测绘通报,2017(4):125-129

[4] 孙群.多源矢量空间数据融合处理技术研究进展[J].测绘学报,2017,46(10):1 627-1 636

[5] 季晓林.地理实体的多态特征研究[J].测绘科学技术学报,2014,31(3):310-319

[6] 张建英,刘高.地理实体与政务专题数据关联融合方式研究[J].城市勘测,2018(4):25-28

[7] 尹鹏程,王守东,宋玉兵,等.新型地理实体数据关键技术研究及应用[J].北京测绘,2020,34(7):881-885

[8] 赵云鹏,孙群,刘新贵,等.面向地理实体的语义相似性度量方法及其在道路匹配中的应用[J].武汉大学学报(信息科学版),2020,45(5):728-735

[9] 赵云鹏,孙群,温威伯,等.基于本体属性特征学习的地理要素分类语义映射方法[J].测绘科学技术学报,2020,37(1):84-89

[10] 林木棵,郭功举,刘一宁.面向新型基础测绘的地理实体数据模型实现与应用[J].测绘通报,2021(增刊):75-78

[11] 王琳,郭功举,刘一宁.面向智慧城市建设的地理实体构建方法[J].测绘通报,2022(2):20-24

[12] 王云帆.一种基于空间特征的地理实体编码方法[J].测绘通报,2018(10):66-70

[13] 张亮,周志诚,厉芳婷,等.基于地理实体的数据库建设探讨[J].地理空间信息,2021,19(1):122-124

[14] 李先波.基础地理实体数据整合技术研究[J].北京测绘,2019,33(8):903-906

[15] 上海市测绘地理信息学会.基于地理实体的全息要素采集与建库:T/SHCH 001-2020[S].

[16] 姚驰,孙毅中.一种矢量数据空间特征的编码方法:中国,2011102003072[P].2011-12-21

[17] 国家基础地理信息中心.地理实体空间数据规范:GB/T 37118-2018[S].北京:中国标准出版社,2018

（上接第11页）

515-532

[17] 陈洁,陆锋,程昌秀.可达性度量方法及应用研究进展评述[J].地理科学进展,2007,25(5):100-110

[18] 李亚婷,潘少奇,苗长虹.中原经济区县际经济联系网络结构及其演化特征[J].地理研究,2014,33(7):1 239-1 250

[19] 孟德友,魏凌,樊新生,等.河南“米”字形高铁网构建对可达性及城市空间格局影响[J].地理科学,2017,37(6):850-858

[20] 李思维.基于城市流强度空间权重矩阵的构建与应用[J].统计与决策,2017,32(24):70-73

[21] 张可云,王裕瑾,王婧.空间权重矩阵的设定方法研究[J].区域经济评论,2017,4(1):19-25

[22] Zhang C P. An Analysis of Urban Spatial Structure Using Comprehensive Prominence of Irregular Areas[J]. International Journal of Geographical Information Science, 2008, 22(6): 675-686

[23] Israel E, Frenkel A. Social Justice and Spatial Inequality [J]. Progress in Human Geography, 2018, 42(5): 647-665

[24] 李博闻,黄正东,蒯希,等.基于空间公平理论的公共交通服务评价：以深圳市为例[J].地理科学进展,2021,40(6):958-966