引文格式：邓淑丹. 时空大数据全链条管理分析研究与应用[J]. 地理空间信息, 2024, 22(12): 128-132. doi:10. 3969/j.issn. 1672-4623.2024.12.029

时空大数据全链条管理分析研究与应用

邓淑丹[1]
（1. 福建省自然资源地理信息中心，福建 福州 350003）

摘　要：在数据作为生产要素的新时代，如何组织、管理、挖掘时空大数据价值具有重要的研究意义。从测绘地理信息视角，聚焦时空基础地理信息、时空公共专题、空间规划、智能感知数据、互联网萃取数据等，研究基于RTDS的物联网数据管理技术、互联网数据近实时数据抓取与清洗技术、可视化并行计算的空间数据挖掘分析技术、超量信息可视化分析与应用技术，提出时空大数据全链条的汇、治、管、挖方法。成果应用于智慧南平时空信息云平台建设项目，自然资源舆情监测、环境质量监测、地价分析等方面，取得显著成效。

关键词：时空大数据；时空数据组织模型；全链条管理分析；时空数据挖掘分析
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2024）12-0128-05

Research and Application of Spatio-temporal Big Data Full-chain Management Analysis

DENG Shudan[1]
(1. Fujian Provincial Natural Resources and Geomatics Center, Fuzhou 350003, China)

Abstract: In the new era of data as a factor of production, how to organize, manage, and mine the value of spatio-temporal big data has important research significance. From the perspective of surveying and mapping geographical information, focus on spatio-temporal basic geographical information, spatio-temporal public topics, spatial planning, intelligent perception data, Internet extraction data, we studied the Internet of Things data management technology based on RTDS, the Internet data near real-time data capture and cleaning technology, the spatial data mining analysis technology of visual parallel computing, visual analysis and application technology of excess information, and proposed the methods of aggregation, government, management and mining of whole chain of spatio-temporal big data. The research results have been applied to the project of the Smart Nanping spatio-temporal information cloud platform, natural resource public opinion monitoring, environmental quality monitoring, and land price analysis, and have made outstanding achievements.

Key words: spatio-temporal big data, spatio-temporal data organization model, full-chain management analysis, spatio-temporal data mining analysis

地理信息数据快速转变为蕴含时间和空间的时空大数据，促使地理信息产业迈进全新发展阶段。国内众多学者开展了时空大数据的研究[1-5]，但未体系性地研究动静结合的时空组织模型，全链条的时空大数据汇、治、管、用技术。本文从测绘地理信息视角，聚焦城市时空基底数据、人类互联网活动数据、物联网实时数据以及人类开发利用依据等数据，以问题为导向，系统性研究城市动态多源的时空大数据全链条管理分析的方法与技术，为激活数据潜能、释放数据价值提供支撑。

1　时空大数据构成

国家政务管理四大基础库包括人口、法人、宏观经济、自然资源与空间地理数据库，因此，时空大数据构成须在四大基础数据的基础上拓展，具体包括时空基础地理信息、时空公共专题、智能感知、空间规划以及互联网萃取数据。时空基础地理信息数据构建起城市地表、室内外、虚实一体化的地表全空间，在此基础上集成人口、法人、宏观经济等时空公共专题数据形成完备的基础数据。拓展空间规划数据，为国土空间开发利用提供统一的事实依据。接入物联网智能感知数据，形成标准统一的智能感知数据，为政府管理者和决策者掌握城市脉搏、感知城市运行的每个角落提供支撑。通过抓取人类互联网活动数据，经清洗、时空化处理、挖掘分析后，形成政府部门重点感知的社会信息及变化趋势数据，为更科学、更精准地政策调整、资源配置和政务服务提供支撑。

收稿日期：2024-04-14。
项目来源：智慧南平时空信息云平台国家建设试点项目（自然资办函〔2018〕550号）。
作者简介：邓淑丹（1982—），硕士，高级工程师，从事自然资源政务信息化、在线地理信息服务等工作，E-mail：43730947@qq.com。

2　全链条管理分析方法

时空大数据全链条管理分析涵盖数据汇聚、治理、组织、管理与分析等。

2.1　时空数据汇聚

时空数据来源和产生方式不同，其汇聚方式也不同。①时空基础地理信息数据。该数据已严格按照测绘地理信息数据标准生产，但存在涉密信息，须离线拷贝至涉密区。经脱密脱敏后的数据，从涉密区离线拷贝至非密环境的汇聚区。②时空公共专题数据。该数据来源于行业部门，采用前置数据库或在线接口方式按周期在线推送至汇聚区。③空间规划数据。包括批复规划版本和动态调整的规划数据，批复规划版本采用离线拷贝方式汇聚，而动态调整的规划数据由业务系统推送至相应环境的汇聚区。④智能感知的实时数据。该数据由交通、水利、环保等部门或铁塔、移动等公司汇聚而成。对于已具备监测点位空间数据服务和实时数据调用接口的数据，采用在线调用方式汇聚。对于无法在线调用的数据，由实时感知建设端动态推送监测点位和实时数据包至同网络的汇聚区，采用多节点负载方式接收智能感知数据。⑤互联网萃取数据。从微博、微信、博客、论坛等社交平台抓取可用信息，再经过挖掘处理后，形成用于市场动态监测、社会舆情监测的专题数据后，再存入汇聚区。

2.2　时空数据治理

按照时空数据组织要求，对所有汇聚数据注入“三域”标识。对于非空间数据，先采用适宜方式空间化处理，对含地名地址属性的数据，采用地址匹配方式处理；对含准确位置坐标的数据，利用坐标值转换为空间数据；对于蕴含可识别的位置或坐标描述信息的数据，采用语义识别、智能匹配方式生成空间数据。所有空间数据均需完成统一时空基准，统一数据格式，属性和要素关系的一致性处理，再利用ETL技术抽取、清洗、转换、集成处理，形成规范、有效的数据。数据更新则采用增量更新或版本更新的方式，记录历史时态信息，实现客观世界的连续快照。

针对公开使用的数据，还需进行脱密脱敏处理，其中涉密基础地理信息数据按照《公开地图内容表示规范》规定，经脱密处理后可应用于非密应用环境，其余数据按照政务管理要求，去除敏感数据方可使用。

2.3　时空数据组织

传统地理信息数据的组织与存储已无法满足时空大数据管理需求，因此需研究建立全空间信息组织模型，构建面向对象和多源数据混合的存储模式，做到图属性关联、实体对象串联、多时序追溯。

本文将研究的数据按其组织方式划分为10类，如图1所示，A类为图属一体、实时动态更新的空间数据，B类为既有版本又有动态小调整的数据如空间规划数据，C类为不断累加数据如宏观经济数据，D类为各类比例尺DLG数据，E类为图像文件和镶嵌数据集组合数据，F类为数据文件和矢量数据组合数据，G类为数据文件和结合图矢量数据组合数据，H类为抓取数据、清洗数据、矢量数据组合的海量数据，I类为矢量数据和流式数据组合数据，J类为元数据信息、数据标签信息、资源分类信息等数据。针对10类数据特点，开展数据时空组织模型和存储组织模型研究。

2.3.1　数据时空组织模型

由于数据的时间特性各异，其数据时空组织模型也不同，本文采用3种数据时空组织模型，对于E、F、G类中的文件数据无需建立时空模型，无时态属性的J类数据无需建立时空模型。

1）生命周期标识的时空模型。具有产生、消亡

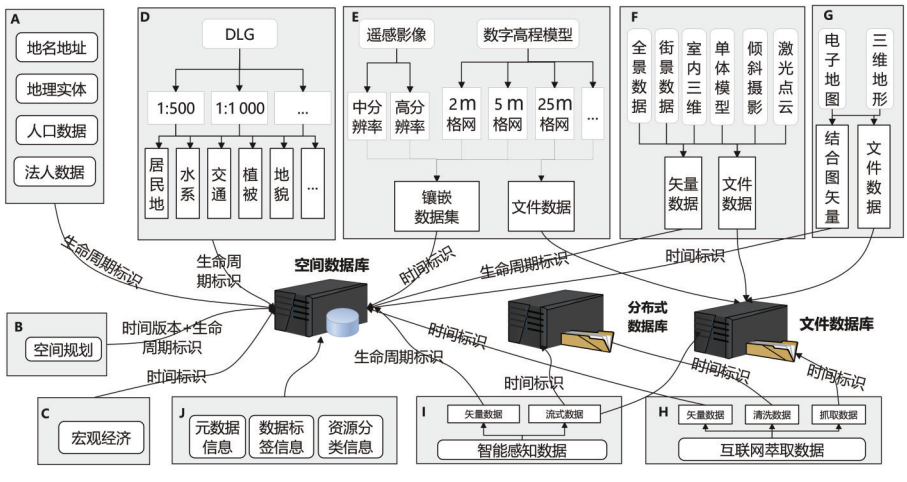


图1　时空数据组织模型

特性的A、D类数据，以及F类、I类的矢量数据，采用该时空模型组织。对每条空间数据赋予属性标识、状态标识、现势性时间、消亡时间、空间标识等信息，形成时空数据的“三域”标识，其中属性标识标明隶属的领域、行业、主题的类别代码，状态标识值域为{0,1}（0标识无效，1标识有效），空间标识为该数据覆盖的城市空间格网编码。当数据更新或删除时，将变更的记录赋予消亡时间和无效状态标识，更新后或新增的数据追录至图层中，并注入其数据产生时点的“三域”标识。

2）时间标识的时空模型。具有产生、无消亡特性的C、H类数据，以及E类的镶嵌数据集、G类的结合图矢量、I类的流式数据，采用该时空模型组织。H类清洗前的数据是将“三域标识”添加至文件数据中，以追加数据文件方式累加更新。I类的流式数据分为视频流与监测流数据，视频流仅对其视频流文件的索引数据赋予时间标识，监测流数据则是为每条数据赋予时间标识，采用不断累加方式更新。其余数据则是对每条数据赋予现势性时间、属性标识、空间标识等信息，以不断追加数据记录方式累加更新，不做数据删除操作。

3）时间版本+生命周期标识的时空模型。B类数据因其特殊性，采用该时空模型组织。以国土空间规划为例，按批复规划数据建立时间版本图层，如2023年福建省双评价图层。对于规划有效时期内，对于微调的数据按生命周期标识组织，更新后或新增的数据追录至图层中，并注入有效的“三域”标识。

2.3.2 数据存储组织模型

大数据时代要实现结构化、非结构化的数据一体化存储，以满足高并发、高时效的管理要求[6]。为此，本文将各类数据按4种数据库组织，由应用端统一存储管理。

空间数据库组织。除J类数据外，其余类别中的空间数据均按图层或数据集方式存入空间数据库中。为提高数据存储效率，对于海量的空间数据按时间、按区域分块存储，并建立各存储的映射关系。

分布式数据库组织。将I类数据的视频感知数据索引文件及监测数据流，以及H类中清洗而成的数据，通过分布式数据库存储，以提高数据检索效率。

文件数据库组织。E、F、G类数据的文件数据，以及H类的抓取数据包，I类的视频文件段，均按照文件数据库组织，将访问量大、访问要求高的文件存入固态存储，其余数据存入机械存储。

2.4 时空数据管理与挖掘

研发时空基础数据管理、智能感知数据管理、互联网萃取数据管理、空间数据挖掘分析、可视化展示、服务发布等时空大数据管理功能，满足时空数据汇聚、处理、服务发布、挖掘分析与应用等需求。

1）时空基础数据管理。提供时空基础地理信息、时空公共专题、空间规划等数据的导入、接边、转换、批量或增量更新，数据质检、数据编辑等功能，相比数字城市基础地理信息数据管理，增加立体的新型测绘地理信息数据，拓展时序化数据管理与追溯能力。

2）智能感知数据管理。提供实时数据的接入、存储、查询、分析、预警等功能，实现传输协议各异的感知数据流统一化、可视化、服务化。

3）互联网萃取数据管理。提供互联网数据的采集、清洗、分析处理、专题加工与加载等功能，发掘互联网大数据价值。

4）空间数据挖掘分析。从海量空间数据中可挖掘有价值的信息[6-8]，为此，提供数据源计算模型构建、计算任务调度、计算结果存储，运行情况实时监测预警等功能，以分析出数据潜在价值。

5）数据可视化。提供数据接入、专题地图定制、统计图表定制等可视化表达功能，以及配置地图查询、时序浏览、空间分析等操作功能，形成可查询、浏览和分析的专题应用。

6）服务发布。实现将时空大数据以及数据挖掘结果发布成数据服务、功能服务以及知识服务等能力，支持建立、删除与修改服务元数据、标签、类别目录等关联信息，以便第三方使用。

7）元数据信息管理。对元数据、标签、分类等信息管理。元数据管理实现时空数据元数据注册、编辑、修改及查询；标签管理与分类管理实现标签及类别定义、修改、删除，支持添加或修改时空数据的标签与分类信息。

8）系统管理与维护。实现用户管理、权限控制、系统数据库参数、数据库的备份与恢复等功能，确保整个系统持续、有效、稳定运行。

3 研究的关键技术

3.1 基于RTDS的物联网数据管理技术

实时数据服务（RTDS，real-time data service）是一种数据采集、传输、处理、分析技术，其最大优势在于实时接收和传输不同来源数据至目标数据库。实时感知数据的产生方式、传输协议、数据结构、输出频率差异大，为此，本文基于RTDS技术实现物联网

实时数据接入、应用与预警。①实时数据接入服务。依据传感器实时数据、JSON等格式数据、传输接口等特征，配置不同的适配器，通过分析输入数据、创建存储数据集、分析目标格式、读取数据、数据清洗、数据存储等流程，实现实时数据的接入。其中，数据清洗包括属性和空间过滤2种规则。②实时数据应用服务。将来源各异的实时数据，按照统一规范的查询、分析接口对外服务，提供涵盖GetCapabilities、DescribeFeatureDataSet、QueryFeature、GetObservation、GetObservationHistory、FeatureAnalytics等接口。③实时数据预警。支持检测周期、检测条件描述、通知格式、通知对象等预警模版定制，按照预警模版定时自动执行监测与发送告警通知。

3.2　互联网数据近实时数据抓取与清洗技术

互联网数据具备来源多、体量大、时效强、低价值密度等特点，为此本文通过构建采集模型设计器、动态调整采集周期、采集数据清洗等组合技术，来提高数据的采集效率和清洗质量。①构建采集模拟设计器。由于数据来源各异，采集规则也不同，为此按网站类别分别设计数据采集规则。②动态调整采集周期。虽然采集模版设置了初始数据采集周期，而每个网站更新频次各异，为减少资源浪费和提高采集效率，综合网站的等级、采集量、最近更新频次等要素制定采集算法，动态调整采集周期，更新频繁越高的网站采集间隔时间越短。③采集数据清洗。对抓取的数据进行垃圾过滤、数据排重、提取关键词、中文分词、语义指纹和摘要等清洗后，形成有效数据，解决数据价值密度低的问题。

3.3　可视化并行计算的空间数据挖掘分析技术

为支持灵活、可定制的数据汇聚治理与挖掘应用，本文研究可视化定制和并行任务计算技术。①可视化计算模型定制技术。采用拖拽方式，选定计算算子，其算子可为数据抽取、数据转换、数据加载、空间数据处理、空间分析、时序预测、深度学习等，再将多个算子单向连接，实现数据接入、处理分析、数据存储等运算作业过程。②任务调度技术。本文通过搭载多个计算节点，实现多任务自动化并行计算，并基于Quartz框架任务触发器按周期动态执行任务。具体实现上，在运算服务器上分配多个计算节点，每个节点配置节点状态、心跳机制、所属的节点组。通过任务调度中心将定制的计算模型任务存放至任务列表，再将任务负载均衡地分配至节点组，节点组将任务平均分配至节点队列中的计算节点，同时把该计算节点从节点队列中移除。计算节点启动线程执行接收任务，按心跳间隔发送消息至任务调度中心，若为任务完成消息，任务调度中心再将其存入对应节点组的节点队列中，以此循环调度。

3.4　超量信息可视化分析与应用技术

时空大数据量大、维度多、关系复杂，是超量信息。超量信息可视化分析与应用是立足将庞大复杂的数据变成直观可交互的图表或分布图，以帮助决策者快速获取关键有价值的数据。为此，本文研究以空间数据可视化、信息图形化技术直观表达抽象信息，再采用插件技术，实现数据配置、数据表达、信息关联、界面布局配置、操作插件配置等功能，以达到零编码可视化应用创建目标。空间数据可视化表达将配置的空间数据，以空间分布图、饼图、热力图、分段设色图、动态气泡、蜂窝图、流向图、轨迹图、路网图、风向图等地图方式表达。信息图形化表达将配置数据按趋势、比例、类别的统计图表等方式表达。空间可视化表达与信息图形化表达并非单独存在，必须建立动态联动关系，如选择图表的某类型数据，其空间可视化表达中也仅显示该类数据。由于数据量较大，为提高地图显示效率，对于统计数值按设定频次固化至数据库中，对于地图数据，按需采用矢量瓦片形式动态存储。

4　研究成果应用

4.1　智慧南平时空信息云平台应用

时空大数据是智慧城市时空信息云平台的基础[9-10]，为此，本项目以智慧南平国家试点项目为例，验证时空大数据全链条管理分析能力。智慧南平时空信息云平台采用研究技术，建成了12类时空基础地理信息数据、6类时空公共专题数据、3类空间规划数据、4类智能感知数据、5类互联网萃取时空数据，共计1 035个图层的时空数据资源池，建成了百余个数据计算模型，具备流程化、定制化的数据挖掘分析能力，构建起智慧南平坚实的时空基底，为南平市智能化规划、建设、管理、服务提供支撑。

4.2　自然资源舆情监测应用

舆情主要发酵于互联网，以自然资源行业舆情为例，验证时空大数据萃取和数据挖掘分析能力的有效性。自然资源的重大网络舆情聚焦不动产登记、耕地、地灾灾情、违法用地、非法占用、征地补偿、违规征地等信息，抓取主流传媒、微博微信、社区网站等媒体相关信息，经分类、推送、跟踪处理后，形成舆情预警报告，以辅助决策者做出科学的应对处理。以南平地区为例，经过3个月的应用，共抓取1万余条相关数据，主要来源于微信和网络媒体，舆情安全

指数95%，热门词语为生态、拆迁、干部等，经清洗分析，2条论点提示为一般舆情预警信息，其中1条论点被15家媒体转载，1条论点被12家媒体转载。通过溯源分析，数据采集时间较发布时间延迟3 h，数据清洗结果与实际一致。

4.3 环境质量监测应用

以南平地区环境监测站智能感知数据接入应用为例，验证实时数据接入、分析、超量信息可视化表达专题功能。已成功接入实时雨量、实时河道水量、实时水库水量以及监测站详细水文信息等数据，总氮、总磷等水质数据，O_3、$PM_{2.5}$等空气质量数据，含氧量、SO_2等废气排放数据，经核对，接入信息有效。同时，采用地图可视化展示与图表展示技术，直观展示各类监测数据分布及实时数值信息。通过设定预警模版进行异常报警，如特大暴雨、大暴雨、超正常等雨量预警，排放物超标等异常预警。

4.4 地价分析应用

以构建南平地区地价分析模型为例，验证数据挖掘分析能力的可用性。以城市基准地价、地价监测数据、土地出让数据为基础，融合房产交易数据等形成分析数据源，利用挖掘分析功能，配置满足条件地段的历年监测价格数据，构建基准地价累积增长率回归预测模型，将选出的不同年份出让价格，结合各年基准地价变化趋势，通过线性回归函数，预测估计得出当前年份的理论出让价格。再栅格化处理理论地块地价，构建克吕格插值模型，利用模型插值计算预测地块。最后，形成指定区域的平均地价、地价总额等报表，为土地储备前地价评估、土地市场交易前地价评估提供辅助决策支持。

5 结 语

本文从测绘地理信息角度研究了时空大数据全链条的汇、治、管、用等方法。成果应用于智慧南平时空信息云平台，形成了“全”、“活”、“动”的时空数据资源池，开展了自然资源舆情监测、环境质量监测、地价分析应用等挖掘分析，取得显著成效。

参考文献

[1] 袁存忠，邓淑丹. 地理信息大数据探讨[J]. 测绘通报，2016(12):105-107

[2] 王家耀，武 芳，郭建忠，等. 时空大数据面临的挑战与机遇[J]. 测绘科学，2017，42(7):1-7

[3] 王家耀. 时空大数据时代的地图学[J]. 测绘学报，2017，46(10):1 226-1 233

[4] 杜江毅，边馥苓. 面向大数据的空间数据挖掘综述[J]. 地理空间信息，2017，15(1):8-11

[5] 秦昆，许凯，吴涛，等. 智能空间信息处理与时空大数据分析探索[J]，地理空间信息，2022，20(12):1-9

[6] 李德仁，张过，蒋永华，等. 论大数据视角下的地球空间信息学的机遇与挑战[J]. 大数据，2022(12):1-12

[7] 王树良，丁刚毅，钟鸣. 大数据下的空间数据挖掘思考[J]. 中国电子科学研究院学报，2013，7(1):8-17

[8] 李德仁，王树良，李德毅. 空间数据挖掘理论与应用(第二版)[M]. 北京:科学出版社，2013

[9] 凌琦，陈志刚，雷帆. 测绘监理在智慧城市时空大数据生产中的应用[J]. 城市勘测，2023(5):106-114

[10] 国家测绘地理信息局. 关于印发《智慧城市时空大数据与云平台建设技术大纲》(2017版)的通知[EB/OL]. (2018-06-07)[2024-03-14]. https://gs. mnr. gov. cn/201806/t20180607_1792791. html

(上接第15页)

与实现[J]. 地理空间信息，2021，36(5):78-80

[10] 孙雪婧，卢向伟，余鑫洋，等. 二三维一体化GIS技术在国土空间基础信息平台建设中的应用[J]. 现代测绘，2021，44(5):1-5

[11] 王红雷，谢义娟，刘诗琦，等. 贵州乱占耕地建房动态监测监管技术平台研究与实现[J]. 地理信息世界，2022，29(6):16-20

[12] 宋禹亭，冯广京，董文君. 国土空间用途管制治权与事权的协调机制研究[J]. 管理现代化，2022，42(4):132-139

[13] 陈慧玲. 依托信息化推动自然资源管理“两统一”的思考:以江苏省自然资源信息化建设为例[J]. 自然资源通讯，2021(15):2-6

[14] 中共中央国务院. 中共中央国务院关于建立国土空间规划体系并监督实施的若干意见[EB/OL]. (2019-05-23) [2023-01-19]. https://www.gov.cn/zhengce/2019-05/23/content_5394187.htm.

[15] 尚嫣然，张悦，薛昊天. 市县联动的国土空间规划传导体系研究:以湖北省黄冈市为例[J]. 规划师，2021，37(19):6-11

[16] 张衍毓，张晓玲，邓红蒂. 新时代国土空间治理科技创新体系研究[J]. 中国土地科学，2021，35(4):9-16

[17] 张新长，廖曦，阮永俭. 智慧城市建设中的数字孪生与元宇宙探讨[J]. 测绘通报，2023(1):1-7

[18] 范合君，吴婷. 新型数字基础设施、数字化能力与全要素生产率[J]. 经济与管理研究，2022，43(1):3-22

[19] 高翔. 建立适应数字时代的政府治理新形态[J]. 探索与争鸣，2021(4):141-146