引文格式：王宏,于雪鸥,曾涛,等. 基于数据立方体的郑州城市地质数据分类组织[J]. 地理空间信息, 2023, 21(12): 93-96. doi:10. 3969/j.issn. 1672-4623.2023.12.022

基于数据立方体的郑州城市地质数据分类组织

王　宏[1,3,4]，于雪鸥[2]，曾　涛[1,3,4]，乔东玉[1,3,4]，朱学立[1,3,4]，田　隆[1,3,4]
（1. 河南省地质研究院，河南 郑州 450016；2. 河南省地质科学研究所，河南 郑州 450001；
3. 河南省自然资源科技创新中心（智慧城市时空大数据），河南 郑州 450016；
4. 河南省地质矿产勘查开发局河南省地质大数据应用中心，河南 郑州 450016）

摘　要：地质数据具有多维度属性，采用单一的数据分类方法对数据进行组织，难以满足地质数据多元化应用和服务的需求。针对地质数据难以有效分类这一问题，以郑州市为例，建立了该市多要素城市地质调查数据分类立方体模型，实现了郑州市多要素城市地质调查数据可根据业务需求动态生成数据组织目录，为海量、多源异构地质数据的快速检索与发现奠定坚实的基础。

关键词：郑州城市地质数据；数据分类；数据立方体；地质数据分类立方体
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2023）12-0093-04

Classification and Organization of Zhengzhou Urban Geological Data Based on Data Cube

WANG Hong[1,3,4], YU Xueou[2], ZENG Tao[1,3,4], QIAO Dongyu[1,3,4], ZHU Xueli[1,3,4], TIAN Long[1,3,4]
(1. Henan Institute of Geology Research, Zhengzhou 450016, China; 2. Henan Institute of Geological Sciences, Zhengzhou 450001, China; 3. Science and Technology Innovation Centre of Nature Resources of Henan Provinces(Smart City Space-time Big Data), Zhengzhou 450016, China; 4. Henan Geological Big Data Application Center, Henan Bureau of Geology and Mineral Exploration, Zhengzhou 450016, China)

Abstract: Geological data has multi-dimensional attributes. It is difficult to meet the demands of diversified applications and services of geological data by using a single data classification method to organize the data. To solve the difficulty in effective classification of geological data, we established the Zhengzhou multi-factor urban geological survey data classification cube by using the geological data classification cube model according to the data cube theory. Then,we achieved the Zhengzhou multi-factor urban geological survey data todynamicallyorganize data contents based on business demands, which laid a solid foundation for the rapid retrieval and discovery of massive, multi-source heterogeneous geological data.

Key words: Zhengzhou urban geological data, data classification, data cube, geological data classification cube

郑州市是全国首批多要素城市地质调查示范城市之一。针对城市发展对地质工作的需求，开展了城市三维地质结构与地下空间调查、地质资源勘查与评价、城市地质安全评价、城市地质环境调查与监测和城市地质大数据云平台建设5个方面的工作，为解决制约城市发展的空间、资源、安全、环境等问题提供地质解决方案。伴随着多要素城市地质调查工作的开展，郑州市多要素城市地质数据资源也将逐渐形成。郑州市多要素城市地质数据资源建设过程中，地质数据资源的分类是数据管理体系建设的基础。受专业认知、数据应用等因素的影响，对地质数据有多种分类方式[1-4]。面向不同的数据使用人员、应用场景，仅使用一种数据分类方式是不合理的。传统的数据分类往往仅考虑领域专业性问题，导致分类只能由地质学家进行，而不同的专家分类各异，难以达成一个统一的分类标准。与此同时，在倡导地质数据社会化服务的背景下，地质数据的使用者往往不一定是地质学家，甚至不具备地质专业知识，而是更关心数据所能解决的实际问题的人。地质数据是有多个维度属性的[5-6]，地质数据的多维度属性使得分类系统不应成为固定的体系而应随着不同的领域或应用进行调整。

收稿日期：2022-06-22。
项目来源：国家自然科学基金资助项目（42071384）；郑州市自然资源与规划局资助项目（郑财招标采购-2020-111）；河南省地质研究院基本科研项目（2023-331-XM027-KT02）。
第一作者简介：王宏（1987—），硕士，工程师，研究方向为地质大数据挖掘与地学信息系统建设，E-mail：mickeydad@126.com。

1 郑州市多要素城市地质数据资源体系建设

郑州市多要素城市地质数据资源体系建设以提供全面、权威、及时、便捷的地质信息服务为目标，以形成有效的地质数据汇聚体系为基础[7]，围绕郑州市城市发展和生态文明建设遇到的资源、环境、空间、安全问题，基于大数据技术，开展多源异构地质数据的集成、抽取、挖掘，并建立多专题的数据库，为各类地质应用提供支撑。

地质数据来源广泛，传感监测网、智能手机等丰富了数据采集方式。地质数据类型复杂，既有地质工作产生的数据，也有其他行业的数据；既有以往存档的地质资料，也有新的实时监测数据；既有结构化数据，又有非结构化数据。针对不同的数据其存储策略也会不同。考虑数据更新频率不同，地质调查成果档案、自然资源部门相关工作成果等按要求汇交入库；各类收集数据要及时进行入库；各类地质监测数据要实时汇聚入库；提供与相关业务系统的共享交换。考虑到数据结构不同，数据存储方式也有区别，目前通用的地质数据存储方式为：结构化数据能够使用关系数据库技术来处理，海量非结构化数据适合采用分布式文件系统HDFS，无模式的半结构化数据适合采用分布式Key/Value存储引擎Hbase等[8-9]。采集汇总后的地质数据经过数据清洗、处理、集成、融合、入库，按照物理分布、逻辑统一的技术路线，按照统一的地质资源数据目录和元数据规范进行存储和管理[10]。不同的地质数据其针对数据的业务应用场景或地质问题不同，在认真梳理地质数据应用场景的基础上，建立各类地质数据库，为相关的地质应用提供数据支撑。

针对城市发展对地质工作的需求，郑州市多要素城市地质调查开展了城市三维地质结构与地下空间调查、地质资源勘查与评价、城市地质安全评价、城市地质环境调查与监测4个方面调查工作，形成了相关的地质数据资源。

2 地质调查数据分类

2.1 地质数据分类立方体

1996年Jim Gray[11]提出数据立方体（Data Cube），围绕主题，数据立方体可以从不同的维度展现数据的特征，用户可以根据实际需求从不同的视角对数据进行分析[12-13]。数据立方体由维和事实度量定义[14]，将地质数据分类立方体表达为：

$$\text{GeologyDataCube}=(\text{Dimension},\ \text{Measure},\ \text{Hierarchy},\ \text{Entity})$$

式中，GeologyDataCube为一个分类体系；Dimension为地质数据分类立方体的维，地质数据立方体有多个维度$\text{Dimension}=\{\text{Dimension}_1,\text{Dimension}_2,\ldots,\text{Dimension}_n\}$，$\text{Dimension}_k(1\leqslant k\leqslant n)$为地质数据的第$k$个分类特征；$\text{Measure}=\{\text{Measure}_1,\text{Measure}_2,\ldots,\text{Measure}_n\}$为地质数据分类立方体对应的分类特征值，其中$\text{Measure}_k(1\leqslant k\leqslant n)$对应的是$\text{Dimension}_k(1\leqslant k\leqslant n)$的分类特征值集合，包含多个分类特征值Tag；地质数据分类立方体支持构建多个分类维度间的层级关系，$\text{Hierarchy}=\{\text{Hierarch}y_{1\to di},\text{Hierarch}y_{2\to dj},\ldots,\text{Hierarch}y_{n\to dk}\}$为地质数据分类立方体将会使用$n$维对数据进行分类组织，并且$\text{Hierarch}y_{x\to\text{Dimension}_y}(1\leqslant x\leqslant n,1\leqslant y\leqslant n)$为地质数据分类立方体的$x$个层级是按照$\text{Dimension}_y$进行组织的；$\text{Hierarch}y_1$为分类体系的最高级，$\text{Hierarch}y_2$为$\text{Hierarch}y_1$的下一级，以此类推，地质数据分类立方体至少有一个分类层级。建立地质数据分类立方体的目的是对地质数据进行多维度的组织，需要建立地质数据与具体维度特征值间的映射关系。Entity为具体的整个地质数据集合，地质数据之间没有层次等级，每个地质数据实体关联一个或多个特征值标签，建立数据与特征维度（Dimension）之间的对应关系。

2.2 分类立方体实现

郑州市多要素城市地质调查数据的应用对象为政府部门、地质专业人员、相关科研机构和企事业单位、社会公众，从数据应用、数据处理方式、数据获取方式、数据公开要求等方面考虑，将地质数据分类立方体按专业类别、数据来源、比例尺、行政区、密级、数据类型、服务类型、数据维度、加工程度、工作手段共10个维度进行划分。

地质数据与分类体系间的映射通过标签技术实现，即根据数据维度分类特征值为每个地质数据打上相应的标签，可以为数据打上多个维度的分类特征值，并且在一个数据维度下，可以对应多个分类特征值。例如，某个地质图覆盖范围为郑州市二七区和惠济区，此时，维度行政区会有2个标签，即键值对<行政区>:<二七区>、<行政区>:<惠济区>。

3 地质调查数据组织

3.1 目录树的建立

郑州市多要素城市地质调查数据依据所构建的数据分类立方体进行组织管理，即根据数据立方体建立

数据目录。需要先确定哪些维度及其分类特征参与数据目录树的建立，并需要确定各维度的层级关系，目录树的深度等于参与构建目录树的维度。

假设地质大数据分类体系的确定了地质专业、区域、比例尺3个维度。那么用户可以对此3个维度进行任意组合，形成目录树的层次结构。

当用户选择地质专业、比例尺的维度顺序时，目录树排布如图1所示。

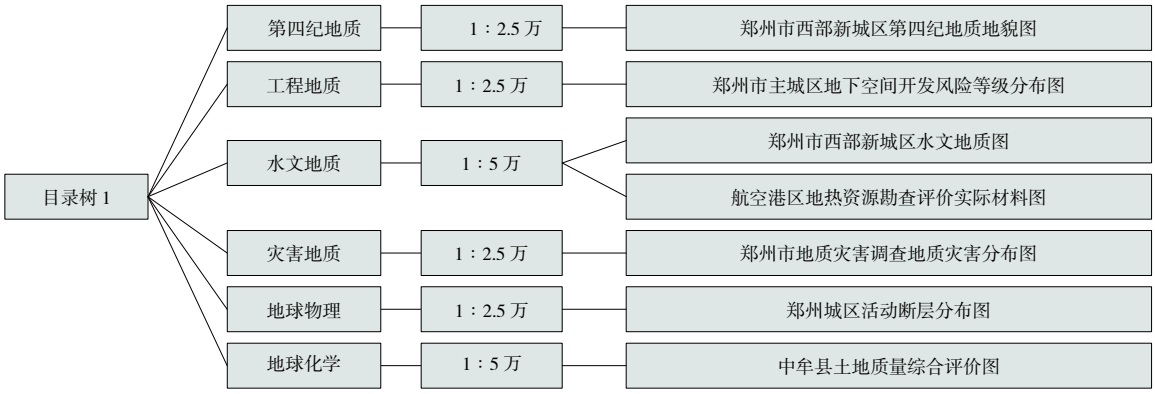


图1　按地质专业、比例尺维度顺序组织的目录树

3.2　目录塌陷

当某个数据未填选取的数据标签时，则无法参与目录排布，这种情况叫做目录塌陷。例如表2中，当“郑州市西部新城区第四纪地质地貌图（1：2.5万）”的比例尺标签值缺失时，则按地质专业、比例尺的维度顺序时排布目录树时，会发生目录塌陷，即“郑州市西部新城区第四纪地质地貌图（1：2.5万）”数据实体无法排入目录的任何节点下，此时可有2种解决办法：

1）发生目录塌陷的数据实体不参与目录排布，按照这个方法，目录树组织如图2所示。

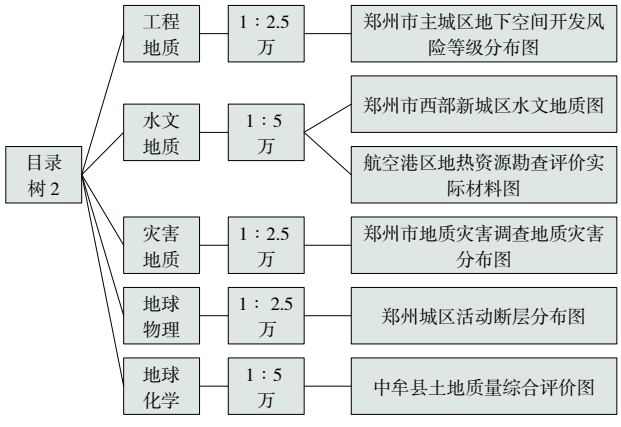


图2　目录坍塌下不参与排布方式的目录树

2）发生目录塌陷的数据实体一律在该级节点以其他的方式形成一个特殊节点，但该数据的其他特征值仍参与目录树的构建。在对郑州市多要素城市地质数据进行组织时，采用第2中方式处理，目录树如图3所示。

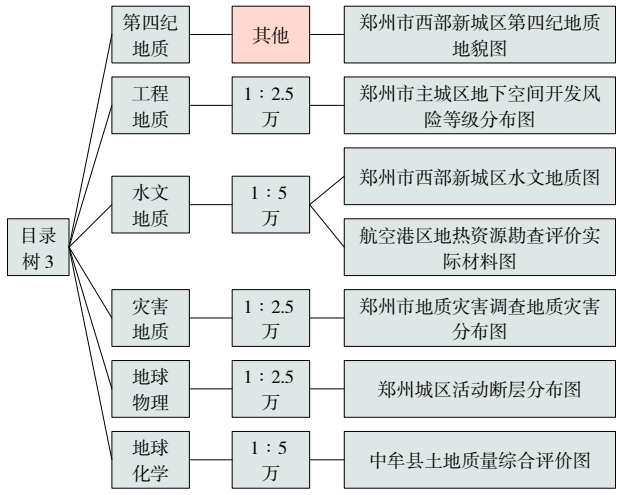


图3　目录坍塌下采用空节点方式组织的目录树

3.3　同一维度下具有多属性时的目录组织

维度对应多标签时，数据实体有可能会被多次排布到不同的维度标签下，在表2中，郑州市主城区第四纪地质地貌图（1：2.5万）和新近系明化镇组地热资源开发利用区划图（1：10万）在区域维度下被标记了多个分类属性值，表明该数据覆盖范围包括中原区、金水区、二七区、管城区，按比例尺、地质专业、区域维度顺序排布的目录树如图4所示。

4　结　论

地质数据具有多维度属性，地质数据的某一维度是在特定专业角度下对数据的“透视”，以发现数据中的特征规律。采用单一的数据分类方法对数据进行组织，难以满足地质数据多元化应用和服务的需求，

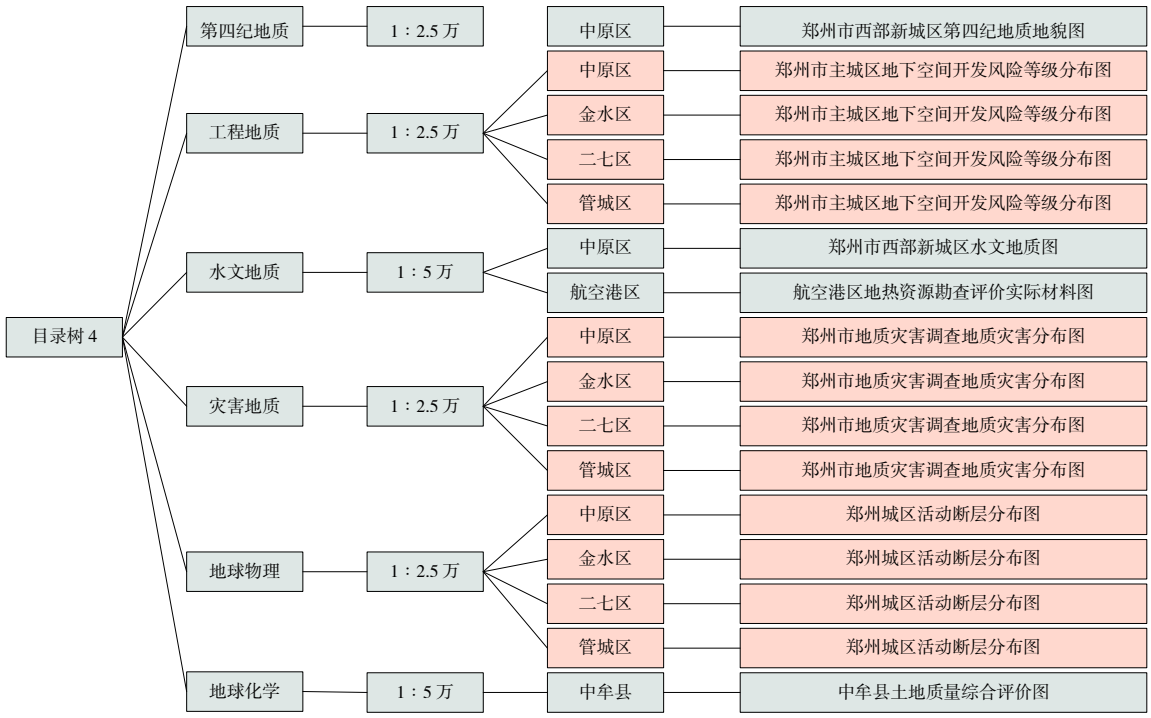


图4 维度对应多标签下的目录排布

因此，本文引入数据立方体理论，抽象提出了地质数据分类立方体模型，根据数据应用场景，建立了包含专业类别、数据来源、比例尺、行政区等10个维度的郑州市多要素城市地质调查数据分类立方体，讨论了使用该分类立方体生成数据目录树过程中的目录塌陷以及数据在同一维度下具有多属性时的目录组织问题。通过郑州市多要素城市地质调查数据分类立方体，实现了郑州市多要素城市地质调查数据可根据业务需求动态生成数据组织目录，并为海量、多源异构地质数据的快速检索与发现奠定坚实的基础。

参考文献

[1] 林良俊,李亚民,葛伟亚,等. 中国城市地质调查总体构想与关键理论技术[J]. 中国地质,2017,44(6):1 086-1 101

[2] 张号,王小军,王炜,等. 西安市城市部件普查成果质量评价研究[J]. 地理空间信息,2022,20(2):159-161

[3] 刘素茹,陈鑫祥,吴锦超. 地理信息空间数据管理的对象存储研究[J]. 地理空间信息,2022,20(3):59-62

[4] 杨锦坤,韩春花,韦广昊,等. 自然资源信息化框架下海洋资源数据管理体系设计有关思考[J]. 海洋信息,2020,35(1):12-17

[5] 邹逸江,李德仁,王任享. 空间数据立方体分析操作原理[J]. 武汉大学学报(信息科学版),2004,29(9):822-826

[6] Zhao J, Collins C, Chevalier F, et al. Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets[J]. IEEE Transactions on Visualization and Computer Graphics, 2013,19(12):2 080-2 089

[7] 谭永杰. 地质大数据与信息服务工程技术框架[J]. 地理信息世界,2016,23(1):1-9

[8] 李敏,傅洁,陈安蜀,等. 地质大数据体系建设初探[J]. 河北地质大学学报,2018,41(2):16-20

[9] 孙红,郝泽明. 大数据处理流程及存储模式的改进[J]. 电子科技,2015,28(12):167-172

[10] 吴冲龙,刘刚,何珍文,等. 城市地质环境信息系统[M]. 北京:科学出版社,2016

[11] Gray J, Bosworth A, Lyaman A, et al. Data Cube: A Relational Aggregation Operator Generalizing Group-by, Cross-tab, and Sub-totals[J]. Data Mining and Knowledge Discovery,1997,1(1): 29-53

[12] 汪浩洋,吴伟,向超,等. 武器装备测试数据立方体模型的构建[J]. 计算机测量与控制,2021,29(9):142-146

[13] 信俊昌,王国仁,李国徽,等. 数据模型及其发展历程[J]. 软件学报,2019,30(1):142-163

[14] 郭国航,李虎,胡钦,等. 量子科学实验卫星的多维数据立方体方法研究[J]. 空间科学学报,2021,41(3):467-474

[15] 闾国年,俞肇元,周良辰,等. 地理实体分类与编码体系的构建[J]. 现代测绘,2019,42(1):1-6

[16] 国土资规〔2017〕1号:国土资源部关于加强地质资料管理的通知[S].