

doi: 10.12029/gc20240919002

左群超, 张志辉, 宋越. 2025. 大数据与数智化时代: 全过程全要素质量控制新范式[J]. 中国地质, 52(3): 890–914.

Zuo Qunchao, Zhang Zhihui, Song Yue. 2025. Era of big data, digitalization and intelligentization: New paradigm of the whole process and total factor quality control[J]. Geology in China, 52(3): 890–914(in Chinese with English abstract).

大数据与数智化时代: 全过程全要素质量控制新范式

左群超^{1,2}, 张志辉¹, 宋越²

(1. 中国地质调查局发展研究中心, 北京 100037; 2. 中国地质调查局自然资源综合调查指挥中心, 北京 100055)

摘要:【研究目的】当前大数据热点几乎集中在如何发现或揭示大数据隐藏的价值或秘密, 极度缺乏深入系统且具实战性的理论研究或最佳实践, 用以关注如何提升大数据建设质量。然而, 大数据建设质量对成功发现或揭示大数据隐藏的价值或秘密以及开展或实施科学而精准的决策至关重要。尤其对我国新一轮找矿突破战略行动矿产资源大数据建设工作实施全过程全要素的数智化质量控制并高水平实现战略目标, 具有重要的现实意义。【研究方法】首先, 通过分析总结现有相关研究成果, 包括全国矿产资源潜力评价大数据建设(2006—2013 年)质量控制经验、矿集区找矿预测大数据实体数据模型和地球科学领域质量控制模型基础框架。然后, 借助大数据思维和数智化技术, 建立矿集区找矿预测及大数据实体建设全过程全要素质量检查与评价数智化体系。最后, 提出一种全过程全要素质量控制新范式, 并可扩展且支持地球科学及其他领域的大数据资产建设质量控制。【研究结果】提出了数据划分和数据粒度质量控制理论与方法, 建立了矿集区找矿预测及大数据实体建设全过程全要素质量检查与评价数智化体系, 开发了矿集区找矿预测质量控制软件, 制定了矿集区找矿预测质量检查与评价技术系列规范, 高效支持并完成了全国矿产资源潜力评价(2006 年至 2013 年)、全国矿集区调查及深部找矿预测(2016 年至 2021 年)大数据实体建设质量控制工作。【结论】建立的矿集区找矿预测及大数据实体建设全过程全要素质量控制数智化体系具有原创性、实战性、高效性和普适性, 可直接扩展到并支持地球科学及以外领域大数据资产建设质量控制; 提出的全过程全要素质量控制新范式具有科学性、有效性和普适性; 提出的“数据模型定义符合信息本体约束”的概念具有普适性和实际意义, 适用于二维或高维图件类数据实体和简单数据表或复杂数据库等表格类数据实体。特别地, 本文构建的质量控制新范式及其方法技术已直接应用并支撑我国新一轮找矿突破战略行动矿产资源大数据建设工作。

关键词: 数据质量控制; 大数据思维; 数字化转型; 智能化转型; 数智化技术; 地质调查工程; 矿产勘查工程

创新点: (1) 建立了基于数据划分、数据粒度、质量检查与评价模型的质量控制理论与方法以及矿集区找矿预测及大数据实体建设全过程全要素质量控制数智化体系, 创建了适用于地球科学广阔领域数据资产建设质量控制新范式, 展示了数智赋能质量控制颠覆传统思维及做法, 具有无可比拟的优势。(2) 研制了基于概念化、共享性、明确性和形式化的矿集区找矿预测大数据实体数据模型、质量检查与评价模型、地球科学知识库核心框架以及支撑矿集区找矿预测及大数据实体建设全过程全要素质量控制的计算机软件系统, 显著提高了质量控制的规范性、客观性和工作效率。

中图分类号: P628 文献标志码: A 文章编号: 1000-3657(2025)03-0890-25

Era of big data, digitalization and intelligentization: New paradigm of the whole process and total factor quality control

收稿日期: 2024-09-19; 改回日期: 2024-10-21

基金项目: 中国地质调查局项目(DD20190570、DD20160052)、全国矿产资源潜力评价计划项目(2006–2013 年)、老矿山深部和外围找矿计划项目(2012–2014 年)、全国危机矿山接替资源找矿国家专项(2004–2010 年)联合资助。

作者简介: 左群超, 男, 1964 年生, 教授级高级工程师, 主要从事矿产地质调查与深部找矿预测、地质大数据建设与应用、计算机地质智能方法技术研究; E-mail: zuoqc163@163.com。

ZUO Qunchao^{1,2}, ZHANG Zhihui¹, SONG Yue²

(1. *Development and Research Center of China Geological Survey, Beijing 100037, China*; 2. *Command Center of Natural Resource Comprehensive Survey of China Geological Survey, Beijing 100055, China*)

Abstract: This paper is the result of mineral exploration engineering.

[Objective] The current hotspots on big data are almost focused on how to discover or reveal the hidden value or secrets of big data, and there is an extreme lack of systematic, in-depth, and practical research or solutions on how to improve the quality of big data construction. However, the quality of big data construction is crucial for successfully discovering or revealing the hidden value or secrets of big data, and for making or implementing scientific and accurate decisions. **[Methods]** Firstly, by analyzing and summarizing the existing relevant research results, including the quality control experiences of the big data construction of China's National Mineral Resource Potential Evaluation (2006–2013), the entity data models for ore-searching prognosis in mineralization concentrating areas, and the basic framework of quality control models in the field of earth science. Then, with the help of big data thinking and digital and intelligent technologies, a digital and intelligent system of the whole process and total factor quality control for ore-searching prognosis and big data entity construction in mineralization concentrating areas is established. Finally, a new paradigm of the whole process and total factor quality control is proposed, which can be extended and supported to the quality control of big data asset construction in the field of earth science and beyond. **[Results]** This study has proposed the quality control theory and method on data division and graininess, established a digital and intelligent system of the whole process and total factor quality control for ore-searching prognosis and big data entity construction in mineralization concentrating areas, developed quality control software, formulated standards and specifications for data quality check and evaluation, and efficiently supported and fulfilled the quality control works such as self check, mutual check, special check, supervision and check, field acceptance, initial review, final review, re-examination, and confirmation of acceptance for the construction of big data entities for ore-searching prognosis in mineralization concentrating areas. **[Conclusions]** The digital and intelligent system of the whole process and total factor quality control for ore-searching prognosis and big data entity construction in mineralization concentrating areas has originality, practicality, efficiency and universality, which can directly be expanded to and support the quality control of big data asset construction in the field of earth science and beyond. The proposed new paradigm of the whole process and total factor quality control also is scientific, effective, and universally applicable. The proposed concept of ontology constraint for data model has universality and practical significance, that is, the definition of data model should meet the requirements of conceptualization, sharing, explicitness, and formalization, applicable to two-dimensional or high-dimensional map-type data entities, as well as simple (e.g. data sheet) or complex (e.g. relation database) table-type data entities.

Key words: data quality control; big data thinking; digital transformation; intelligent transformation; digital and intelligent technologies; geological survey engineering; mineral exploration engineering

Highlights: (1) The quality control theory and method based on data division, data granularity, quality inspection and evaluation model, as well as the digital and intelligent system of the whole process and total factor quality check and evaluation for ore-searching prognosis and big data entity construction in mineralization concentrating areas are established. A new paradigm for the quality control of data assets construction in the broad field of Geoscience is created, which shows that digitalization and intelligentization enable quality control to subvert traditional thinking and practice, and has unparalleled advantages. (2) Based on conceptualization, sharing, definiteness and formalization, the big data entity data model, quality inspection and evaluation model and the core framework of Geoscience Knowledge Base, as well as the computer software system supporting the whole process and total factor quality check and evaluation for ore-searching prognosis and big data entity construction in mineralization concentrating areas have been developed, which have significantly improved the standardization, objectivity and work efficiency of quality control.

About the first author: Zuo Qunchao, male, born in 1964, professor-level senior engineer, engaged in mineral geological survey and deep prospecting prediction, construction and application of geological big data, and computer geological intelligent methods and technologies; E-mail: zuoqc163@163.com.

Fund support: Supported by the projects of China Geological Survey (No.DD20190570, No.DD20160052), National Mineral Resources Potential Evaluation Plan project (2006–2013), Deep and Peripheral Prospecting Plan for Old Mines (2012–2014),

National Special Project for Prospecting Replacement Resources in Crisis Mines (2004–2010).

1 引 言

数学地球科学是一门融合了地球科学和定量方法的交叉学科,数学地球科学的发展不仅有助于理解地球的演化历史和地质结构,还能更准确地预测和模拟地质灾害发生的可能性。因此,数学地球科学在探索地球奥秘和保护人类安全方面发挥着重要的作用(赵鹏大, 2019; 成秋明, 2021)。地球科学大数据具有传统大数据“4V”特征(即数量大(Volume)、种类多(Variety)、速度快(Velocity)和真实性(Veracity)),地质大数据也具备科学大数据“三高”特征(即高维度、高计算复杂性和高不确定性),这些特点使得地质大数据的挖掘和分析更加复杂和更具挑战性,因此知识驱动的地球科学大数据分析成为当前地学研究的热点(Kennedy and Hagan, 2001; 陈建平等, 2017; 翟明国等, 2018)。

人类社会正在进入一个全新的大数据时代,大数据的出现已经深刻改变了人类的生活、生产和思维方式,随着大数据不断增长和发展,科学研究也步入了以大数据为驱动力的第四科学范式(Jonathan et al., 2011; Tolle et al., 2011; Guo et al., 2014; Carranza and Laborte, 2015; Bianco et al., 2017)。大数据研究已成为社会发展和技术进步的迫切需求,对各行各业产生了深远的影响(李国杰和程学旗, 2012; Chen and Zhang, 2014; Marković, 2015; 郭华东, 2018)。科学家通过广泛的数据参与科学研究,通过实时监测和分析解决各种科学问题。大数据在地球科学研究中具有巨大的潜力和优势,知识图谱和演化模型可以帮助我们更好地了解地球科学领域的知识结构和发展趋势(翟明国等, 2018; 周永章等, 2020; 周成虎等, 2021)。数字找矿需要大量的信息技术和定量方法,对矿集区成矿要素进行数字化、模型化、定量化、可视化和网络化处理,实现更准确和高效的找矿预测(赵鹏大和孟宪国, 1992; 吴冲龙等, 2016)。通过利用大数据和数据密集型计算分析,可以揭示地质要素之间的关系,使地质现象能够得到更准确的分析,推动了地球学科的不断进步和发展(吴冲龙和刘刚, 2019; 左

仁广等, 2021)。利用定量数据对地质体进行分类和判断,标志着地质学依赖定量数据进行分析和研究的发展方向(赵鹏大和孟宪国, 1992; 赵鹏大等, 2003)。应对海量数据时,迅速处理数据变得至关重要,而大数据对人工智能发挥着关键作用(Guo, 2018; 李国杰, 2024)。建立一个开放的共享平台和统一的标准,是推动地质大数据研究和发展的关键因素(张旗和周永章, 2017)。科学研究中数据的质量至关重要,确保数据的准确性、完整性和可靠性是基础(Ali et al., 2019)。科学数据评价包括科学数据质量评价、元数据评价、开放评价、管理评价和平台评价等多个方面,尽管已有一些评价指标和方法,但科学数据评价仍处于不成熟阶段(蔡丽华和倪代川, 2021)。对数据客观科学的评估体系是保障并推动数据开放共享的基础,也是开放数据管理体系的核心(王超和张辉, 2020)。面向多维度数据质量的模糊综合评价方法,通过构建合适的数据质量指标体系,并使用德尔菲法和层次分析法确定权重集,进行模糊转换和计算,最终得到数据质量等级(宋俊典和刘丰源, 2018)。科学数据质量的评价通常从多个主要角度进行,包括数据的开放性(撒旭等, 2020)、数据期刊中的同行评议过程(Zhang and Xiao, 2020)、数据恢复和保存(Schmidt et al., 2021)等。现有的评价指标与 FAIR 原则高度契合,即实现数字资源的可查找性、可访问性、互操作性和可重用性(蔡丽华和倪代川, 2021),这表明国内外在提高科学数据开放性、透明性和可重用性方面有一致的看法。科学数据开放评价研究主要集中在构建评价体系和分析比较国外政府开放数据项目上(Zuiderwijk et al., 2021),这些研究关注的指标包括政府效率、透明与问责、环境可持续性、包容性、经济增长和支持企业家精神等方面。前人研究提供了一个科学的数据风险评估框架,为数据管理资源的分配和优先级讨论提供了参考(Mayernik et al., 2020)。在评价科学数据开放方面,由于数据开放与计算机科学密切相关(Lü and Ma, 2019),应将计算机技术融入科学数据开放研究中。评估科学数据平台应该以用户为重心,涵盖专业人士和普通用

户,以更好地为用户提供服务(Máchová et al., 2018)。中国政府《“十四五”数字经济发展规划》(国发[2021]29号,2021)高度重视提升数据质量,认为“数据要素是数字经济深化发展的核心引擎”,强调“把握数字化发展新机遇,充分发挥数据要素作用,强化高质量数据要素供给”。矿集区找矿预测大数据实体(含盖原始、成果和实物资料数据三大部分)具有产出多源、种类多样、规模庞大、结构复杂、时空异构等特点。数据质量问题主要源自采集、传输、存储、处理和分析等环节,可能导致找矿预测结果不准确、不可靠或不稳定等。数据质量表征维度主要包括准确性、完整性、一致性、及时性和可用性等五方面。科研人员应重视数据采集过程,确保数据来源可靠,数据处理合规,从而保障找矿预测结果的可信度和可靠性。大数据思维与数智化变革给矿集区找矿预测研究既带来新机遇,也面临新挑战,大数据质量优劣对矿集区找矿预测成败至关重要。然而,目前有效的科学数据质量控制理论方法以及技术框架能力仍然存在诸多局限,特别是地勘行业,科学大数据质量检查与评价体系尚未完全建立。

本研究在我国矿产资源潜力评价成果数据质量控制方法技术(左群超等, 2013)、矿集区找矿预测理论方法技术体系应用导论(左群超等, 2022)、矿集区找矿预测理论方法技术体系框架(张志辉和左群超, 2024c)、矿集区找矿预测数据库质量检查与评价方法(左群超, 2022)、矿集区找矿预测大数据实体数据模型构建和应用软件研发(张志辉和左群超, 2024a)等基础上,通过开展矿集区找矿预测大数据实体建设质量检查与评价全过程全要素的深入分析和系统研究,进而建立矿集区找矿预测大数据实体建设全过程全要素质量检查与评价数智化体系,提升矿集区找矿预测大数据实体建设质量和效率,促进矿集区找矿预测获得可靠性和准确性结果;同时为地学科学领域数据资产建设质量控制工作,提供一种科学有效且普遍适用的全过程全要素质量控制新范式,从而推动地球科学技术进步和学科领域的全面发展。特别地,本研究对有效支撑我国新一轮找矿突破战略行动并高质量开展矿产资源大数据建设等工作,具有重要的现实意义。

2 质量检查与评价数智化体系

2.1 相关质量控制理论基础

建立数据质量检查与评价数智化体系需要综合考虑多种理论,其中数据划分和数据粒级理论是本次研究提出的关于数据质量检查与评价的核心思想与理论方法。只有这样,才能采取全面统一模式实现数据质量检查与评价。在选择或综合质量理论的应用时,需要充分了解具体需求和数据特点,确保理论的灵活运用能够有效地指导实际工作,从而建立一个有效的数据质量检查与评价数智化体系,确保有效质量控制,提高数据的准确性和可靠性。本次研究主要融合了以下几种基本质量理论。

2.1.1 数据划分(Data Partitioning)和数据粒级(Data Granularity)理论

数据划分是将大型数据集有效分割成多个小型部分的过程,划分的根本目的是把复杂问题简单化,这种方式能够显著提高处理速度和整体效率,主要包括水平划分、垂直划分和混合划分。数据粒级关注处理数据的细节或详细程度,根据细节的不同,可以将数据粒度分为详细粒度和汇总粒度。详细粒度保留了未划分前数据的完整性,适合深入分析和精细计算;汇总粒度则适用于快速检索,有效减少存储需求,提升系统响应速度。在数据库管理中,数据粒度的选择直接影响性能表现、存储需求以及能够解答的业务问题类型。合理的数据划分和粒级选择应相互补充。通过精心设计的数据划分,可以实现对不同粒度数据的处理和管理。同时,针对具体应用需求、数据特性以及系统性能要求,必须综合考虑数据划分策略与数据粒级选择之间的关系,以达到最佳效果。

2.1.2 信息质量理论(Information Quality Theory)

信息质量理论的概念更广泛,不仅关注数据本身的质量,还关注信息的效用、相关性、及时性、易用性等多个维度,强调从用户需求和信息使用场景出发,综合评估信息的多维质量(Strong et al., 1997)。数据质量管理理论(Data Quality Management Theory)是信息质量理论的一个分支,专注于数据本身的质量和管理,主要通过数据的准确性、完整性、一致

性、及时性等质量维度来评估和管理数据。数据质量管理是信息质量理论在数据层面的具体应用,二者在实际应用中是紧密相关的。在实际应用中,数据质量管理是信息质量管理的重要基础。高质量的数据是高质量信息的必要条件,因此,在很多应用场景中,数据质量管理理论被广泛用于确保信息质量。在矿集区找矿预测中,预测数据的质量直接影响预测结果的准确性。

2.1.3 统计质量控制理论 (Statistical Quality Control Theory)

统计质量控制理论源于工业生产领域,主要用于监控和改进产品质量(Bartky, 1942)。在数据质量检查中,可以运用统计方法评估数据的稳定性和一致性,识别异常值并进行数据清洗。统计方法在数据质量检查中发挥重要作用,帮助识别并解决数据质量问题。

2.1.4 数据治理 (Data Governance) 理论

数据治理是管理和监督数据资产的重要方法。通过数据治理,可以确保数据的质量、安全性、有效性和合规性得到有效维护和监督。在矿集区找矿预测中,良好的数据治理框架能够确保数据质量检查和评价的有效执行。因此,数据治理在矿集区找矿预测中扮演着关键角色,有助于保证数据质量的有效管理和评价,从而提高预测和决策的准确性和可靠性。

2.1.5 综合信息集成理论

在矿集区进行找矿预测时,整合多种类型的数据至关重要。综合信息集成理论提供了一种有效的方法,确保不同来源的数据能够被整合在一起,并保持数据的一致性和可比性(Tononi and Sporns, 2003)。这样可以帮助研究人员更全面地分析数据,提高找矿预测的准确性和可靠性。未来整合多源异构数据的能力将是矿业领域研究的重要发展方向,有助于科研人员更好地理解矿集区的地质特征和资源分布规律。

2.1.6 地质数据特有的质量要求理论

地质数据质量要求涉及多个方面,包括地质解释的一致性、数据采集的标准和方法、数据的分辨率和覆盖范围等(王卷乐和陈沈斌, 2006)。地质解释的一致性意味着需要确保在不同地质学家间的观点和解释是一致的,以减少主观误差。数据采集

的标准和方法则直接影响数据的准确性和可靠性,必须严格遵守标准、规范、规程或相关要求。数据的分辨率和覆盖范围也是影响数据质量的重要因素,高分辨率和广覆盖范围的数据可以提供更全面准确的信息。在矿产资源预测过程中,综合考虑地质数据的质量要求尤为重要,只有保证数据的质量,才能更准确地识别潜在矿产资源的分布和赋存情况。

2.2 检查与评价数智化体系

2.2.1 重构并优化矿集区找矿预测流程

矿集区找矿预测流程,事实上包括矿集区找矿预测研究和与之伴随的矿集区找矿预测大数据实体建设两方面同始同终且相辅相成的流程。因此,重构并优化矿集区找矿预测流程,就是重构并优化矿集区找矿预测研究和与之伴随的矿集区找矿预测大数据实体建设的业务流程。具体包括:基于大数据思维与数智化技术、矿集区找矿预测理论方法和矿集区找矿预测数据模型(左群超等, 2022; 张志辉和左群超, 2024b)等研究成果,先重构并优化矿集区找矿预测研究过程,后重构并优化矿集区找矿预测大数据实体建设过程,再将后者与前者从数据实体底层以有效、规范、完备、数字化的数据模型方式有机融合并无缝衔接,最终得到重构并优化后的矿集区找矿预测及大数据实体建设一体化业务流程,为实现“矿集区找矿预测大数据实体建设全过程全要素质量检查与评价数智化体系”奠定基础架构。

矿集区找矿预测研究工作一般包括“找矿预测预研究”、“野外专项调查研究”、“矿产资源潜力评价”和“勘查区大比例尺找矿预测”等基础研究任务。这些基础研究任务同时蕴涵矿集区找矿预测大数据实体的生产与建设的所有建库工作,涵盖了矿集区找矿预测大数据实体所包含的原始资料数据、成果资料数据和实物资料数据的生产与建设。矿集区找矿预测所有研究工作的各环节或各阶段应遵循“矿集区找矿预测理论方法技术及要求”,矿集区找矿预测大数据实体生产建设过程的各环节或各阶段应执行“矿集区找矿预测大数据实体数据模型”,二者有机融合且无缝衔接构成矿集区找矿预测及大数据实体建设一体化业务流程(图 1)。

2.2.2 构建质量检查与评价数智化体系

质量检查与评价数智化体系是以矿集区找矿预测及大数据实体建设一体化流程(图 1)作为体系构建的基础架构,并借助大数据思维和数智能化技术实现的。质量检查与评价数智化体系组成的全部要素涵盖项目组织实施机构、项目质量控制相关人

员、矿集区找矿预测及大数据实体建设过程、矿集区找矿预测大数据实体建设软件(如:矿集区找矿预测大数据实体数据模型应用软件 GeoMapBM)、质量控制活动、质量控制活动输入、质量控制活动输出、质量控制对象、质量控制记录以及质量控制软件(如: GeoDQCBM 软件)等十大类。十大类要素

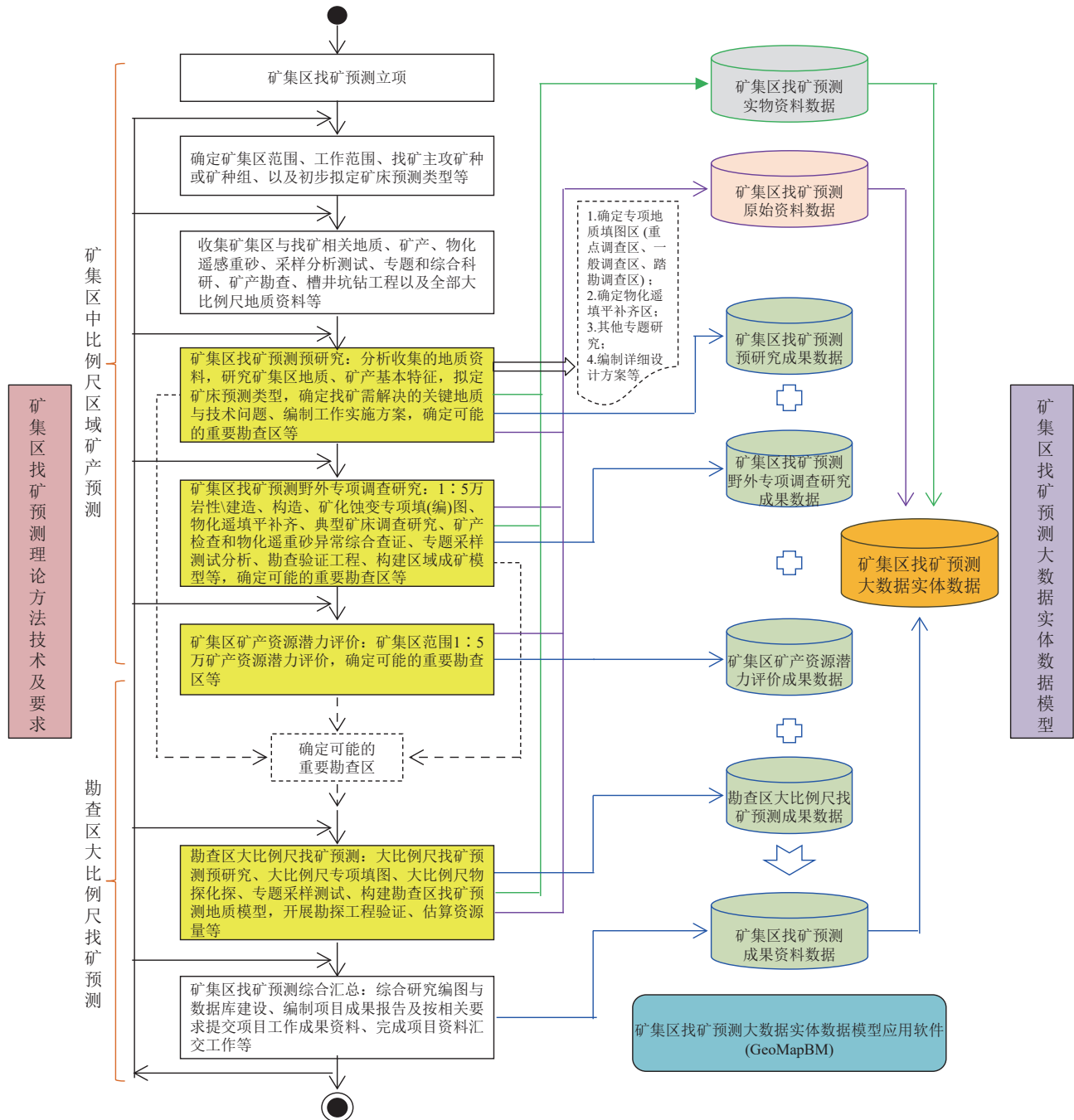


图 1 矿集区找矿预测及大数据实体建设一体化流程

Fig.1 Integrated process of ore-searching prognosis and big data entity construction in mineralization concentrating areas

借助大数据思维和数智化技术构成一个有机整体,即“矿集区找矿预测大数据实体建设全过程全要素质量检查与评价数智化体系”(图 2)。

现通过简要分述图 2 中各类要素内容及职责角色,以解析、说明和展示“矿集区找矿预测大数据实体建设全过程全要素质量检查与评价数智化体系”的组成结构、关联关系以及目标使命与实际效果。

项目组织实施机构要素(简称“体系组织要素”),包括项目投资方及主管部门、项目承担方及主管部门和项目监督机构及主管部门。体系组织要素主要负责激活矿集区找矿预测及大数据实体建设过程,也负责激活质量控制活动等。

项目质量控制相关人员要素(简称“体系人员要素”),包括项目组及成员、项目作业小组及成员、督检专家组及成员、野外工作验收专家组及成员、初审专家组及成员、终审专家组及成员、复核专家组及成员、确认专家组及成员。体系人员要素主要参与矿集区找矿预测及大数据实体建设过程、激活或参与完成质量控制活动等。

矿集区找矿预测及大数据实体建设过程要素(简称“体系过程要素”),即矿集区找矿预测大数据实体建设全过程全要素质量检查与评价数智化体系的“全过程”,包括矿集区找矿预测研究所有过程和矿集区找矿预测大数据实体生产与建设所有过程。体系过程要素涵盖矿集区找矿预测及大数据实体建设一体化业务流程的所有环节或阶段(图 1)。

矿集区找矿预测大数据实体建设软件要素(简称“建库软件要素”),即作者早已实现的矿集区找矿预测大数据实体数据模型应用软件(GeoMapBM)(左群超等, 2022; 张志辉和左群超, 2024a)。GeoMapBM 严格遵行矿集区找矿预测大数据实体数据模型规范,支持矿集区找矿预测大数据实体建设所有过程(图 1)。

质量控制活动要素(简称“体系活动要素”),包括质量自检、互检、专检、野外工作验收、初审、终审、复核、确认接收以及项目期间组织的质量监督检查(即督检)等全部质量控制活动。每个质量控制活动均有相应的激活角色、激活约束、活动输入、活动输出、活动处理本身以及活动结束条件;多个质量控制活动有指定的先后激活逻辑。激活角色

主要是项目组织实施机构、项目质量相关人员等。

质量控制活动输入要素(简称“体系输入要素”或“质量控制模型库”),即从底层逻辑或后台逻辑层面控制每个质量控制活动的质量检查与评价处理本身,以计算机易用的“质量控制模型库”形式存在,便于质量控制软件有效访问或调用。质量控制模型库一般包括基础层和应用层两部分,基础层即质量控制库基础架构,可扩展并用于支撑整个地球科学领域的地质质量控制;应用层即质量控制库应用部分,可基于基础层应用到地球科学领域特定范围或分支的数据质量控制。本文关注的地球科学特定范围或分支,就是矿集区找矿预测,其质量控制模型库应用层主要包括:项目验收规则库,矿集区找矿预测数据模型库(二维与三维图件类数据模型和简单与复杂表格类数据模型),图件类质量模型库、编图规则库、空间拓扑规则库,表格类质量模型库、采集规则库、约束规则库,质检质评记录模版库、质评表单模版库、质量评审意见模版库、质量评述报告模版库、质量附件模版库,以及其他模版、规则或知识库。

质量控制活动输出要素(简称“体系输出要素”),包括质量自检日志、互检日志、专检日志、督检质量报告、野外工作验收(数据资料部分)意见及野外工作验收(数据资料部分)质量评述报告、初审意见及初审质量评述报告、终审意见及终审质量评述报告、复核意见及复核质量评述报告、确认接收报告等。体系输出要素由激活角色借助质量控制软件,激活相应质量控制活动,质量控制活动对指定的矿集区找矿预测大数据实体部分或整体执行相应质量检查与评价处理,质量控制活动结束后产生或输出质量检查与评价记录表单或评述报告以及相应附件等。

质量控制对象类要素(简称“实体数据要素”),即质量控制体系检查或评价的实体数据对象。本文指矿集区找矿预测大数据实体本身,可以是一个或多个矿集区找矿预测项目数据库整体(包括原始资料数据、成果资料数据、实物资料数据三部分),或者是一个矿集区找矿预测项目数据库的原始资料数据、成果资料数据、实物资料数据三部分之一或组合、抑或是矿集区找矿预测项目数据某一部分(如:单个图件、图件子集、图件集;单个表格、表格

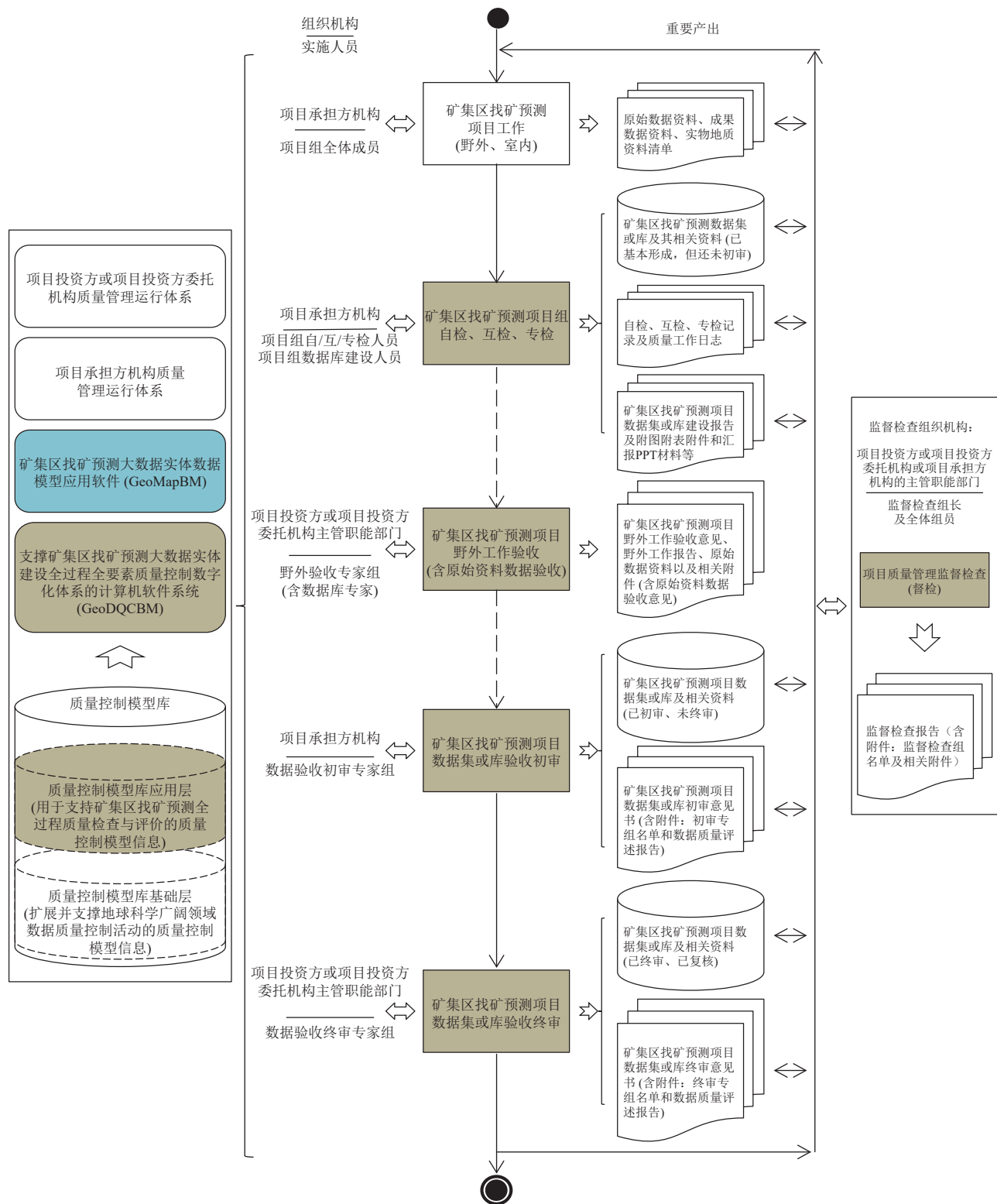


图 2 矿集区找矿预测及大数据实体建设全过程全要素质量检查与评价数智化体系

Fig.2 Digital and intelligent system of the whole process and total factor quality check and evaluation for ore-searching prognosis and big data entity construction in mineralization concentrating areas

子集、表格集;或其组合等)。此类实体数据对象在质量控制软件系统内又称为“实体数据流”。

质量控制记录类要素(简称“质量数据要素”),即质量控制体系对实体数据对象的数据质量实施检查或评价而产生的自检日志、互检日志、专检日志、检查记录、评价记录、质量评述表单、质量评述报告以及相关附件等各类检查或评价记录,统称质量控制记录。此类质量控制记录在质量控制软件内又称为“质量数据流”。

质量控制软件类要素(简称“体系核心要素”),即矿集区找矿预测数据质量控制软件 GeoDQCBM 为质量控制软件要素。GeoDQCBM 作为质量控制数智化体系全要素中的核心要素,一方面关联并协调质量检查与评价数智化体系内所有其他要素,另一方面驱动并协同所有关联质量控制活动。GeoDQCBM 内主要存在两种数据流,即实体数据流和质量数据流。实体数据流即矿集区找矿预测大数据实体部分(如:原始资料数据、成果资料数据或实物资料数据;图件数据、或表格数据等)或整体(如:一个项目数据库或多项目数据库);质量数据流即 GeoDQCBM 检查或评价实体数据流的质量现状而产生。

以最具典型的“终审”质量控制活动,即矿集区找矿预测项目数据库终审验收为例,简单说明并展示“矿集区找矿预测大数据实体建设全过程全要素质量检查与评价数智化体系”完成“终审”质量控制活动关联的体系要素、实施过程以及实际效果(图 3)。“终审”质量控制活动的申请者、激活者、终审专家组、质量终审对象以及质量终审结果,分别为项目承担方或其主管部门、项目投资方或其主管部门、数据库终审专家组、矿集区找矿预测项目数据库(包括原始资料数据、成果资料数据、实物资料数据三部分)、矿集区找矿预测项目数据库终审意见与终审质量评述报告及其相关附件等。为突出说明“终审”质量控制活动的重要内容,图 3 略除了“终审”质量控制活动的申请者、激活者、终审专家组以及质量控制软件等要素。

2.3 质量控制关键要素实现

2.3.1 质量控制软件

质量控制软件作为质量检查与评价数智化体系全部构成要素的核心要素,对关联协同体系其他

各要素并使之发挥有效作用、高效支撑体系各活动并形成客观合规的质量结论至关重要。本研究通过构建“矿集区找矿预测大数据实体建设全过程全要素质量检查与评价数智化体系”,设计并实现了质量控制软件。该软件中文命名为:矿集区找矿预测数据质量控制软件,英文命名为 GeoDQCBM,即 Geology Data Quality Control By Models 的缩写,意即可扩展适用到地球科学广阔领域的的数据质量控制。

GeoDQCBM 采用模型驱动及自动演进理论研制实现(左群超等, 2022; 张志辉和左群超, 2024a),使得 GeoDQCBM 具有传统研制方法无可比拟的优点:前端质量功能界面与后端质量控制模型隔离且相对独立,而且前端质量检查与评价逻辑完全由后端质量控制模型决定。即软件前端质量功能界面专注于处理相对简单且稳定的质量控制活动事务环节,包括自检、互检、专检、督检、初审、终审、复核以及确认接收等,顺应了业内约定俗成的质量控制活动,降低了学习成本且极具实战性;后端质量控制模型则专注于描述异常复杂且多变的质量检查评价算法逻辑或知识规则,适应了实体数据的质量固有特征,控制了复杂多变的地学领域数据质量需求。

GeoDQCBM 软件的登录界面及其实现的功能模块主要包括:图件数据质量检查、批量图件质量检查、图件子集质量评价、图件集质量评价;表格数据质量检查、批量表格质量检查、表格子集质量评价、表格集质量评价;图件表格质量评价;导出质量检查记录、导出质量评述报告等模块等(图 4)。

2.3.2 质量控制模型库

质量控制模型库作为质量检查与评价体系的“智慧大脑”,全面全程控制着前端质量检查与评价操作的处理逻辑,其作用如同全能专家,不仅拥有丰富的领域知识经验,悉晓相关标准要求,深知数据应用质量需求,擅长利用领域知识经验、标准要求以及应用需求,履行检查与评价数据质量职责,给出客观合规的质量评述意见和质量评述报告;而且,借助质量控制软件现代技术手段,新增有别于全能专家的优点,即效率高、成本低、规范且统一、稳健且公正等,同时,可以采用“全检全评”方式而弃用“抽检抽评”方式。

质量控制模型库涵盖领域术语、理论、模型、

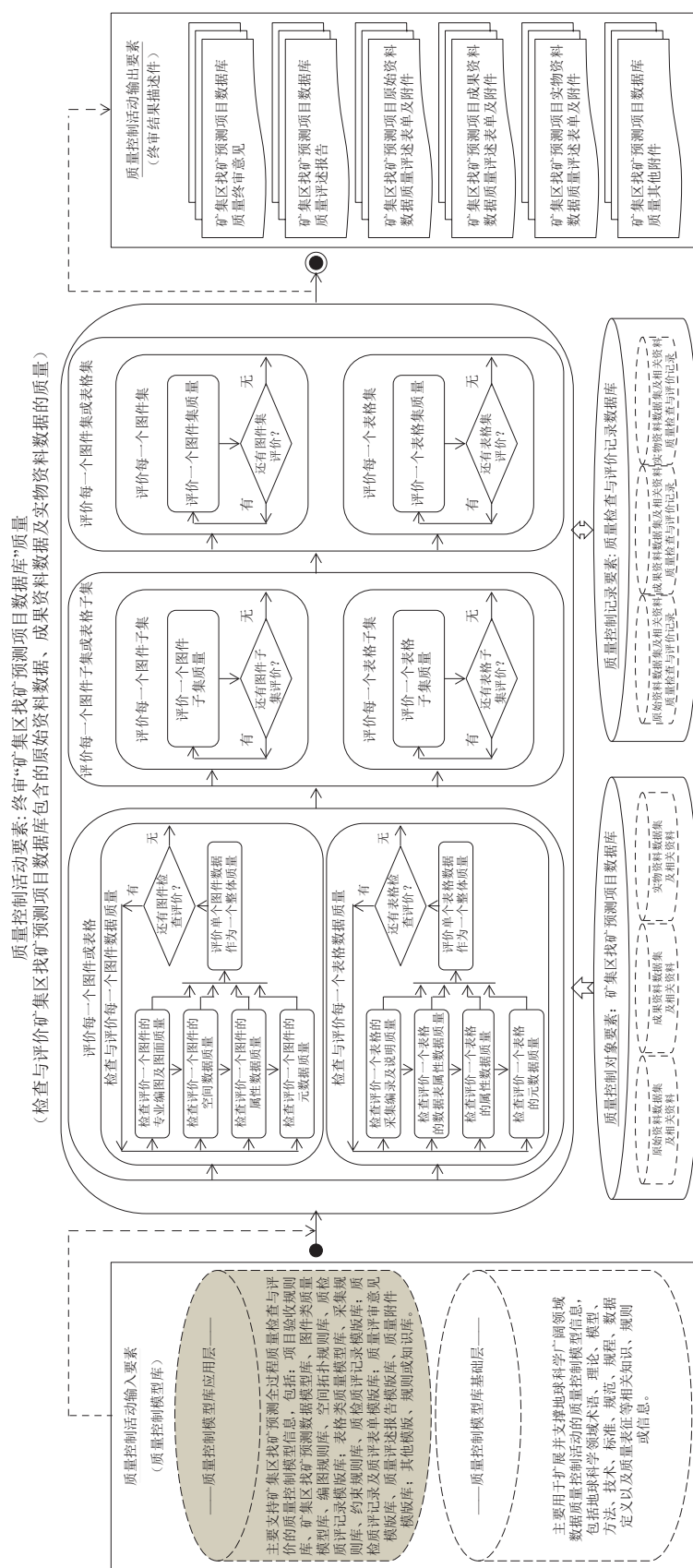


图 3 矿集区找矿预测项目数据库建设终审质量控制活动

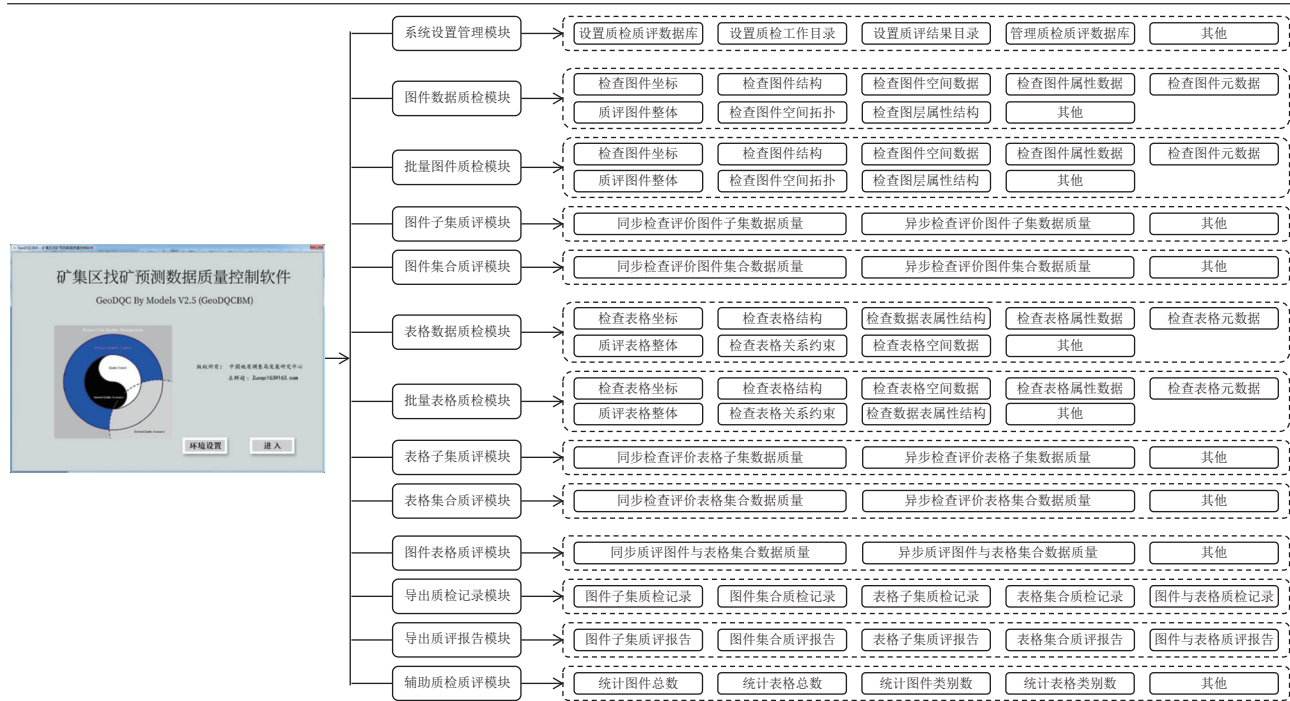


图4 GeoDQCBM 软件登录界面及功能模块框图

Fig.4 Diagram of login interface and functional modules of GeoDQCBM software

方法、技术、标准、规范、规程、数据定义以及质量表征等与质量相关全部内容,采用以方便质量控制软件访问或调用的计算机可有效利用的数据、信息、知识或模版库等形式存在,诸如:领域术语库(如:矿物、岩石、矿床等术语)、领域理论库(如:板块构造学、大陆动力地质学等理论)、领域知识库(如:地史学、地层学、地质构造学等知识)、领域模型库(如:地球圈层结构、矿床形成机理等模型);领域方法库(如:地物化遥钻探调查或测量、采样测试分析、专题综合编图等方法或手段)、领域规则库(如:研究部署、野外工作验收、项目管理等规则);领域数据模型库(如:矿集区找矿预测数据模型、矿产资源潜力评价数据模型、数字地质调查数据模型等);领域质量模型库(如:二维或三维图件类质量模型、简单或复杂表格类质量模型等)、领域检查评价模版库等。

本研究采取规划宏微统筹和阶段跃迁迭代策略构建质量控制模型库,既满足领域应用无限和质量追求不止的宏微需求,又锚定矿集区找矿预测全过程质量控制有限目标。通过建立“矿集区找矿预测及大数据实体建设全过程全要素质量检查与评价数智化体系”,借助包括地质本体(Geological

ontology)、知识图谱(Knowledge Graph)、面向对象(OOA/OOD)、数据建模(Data Modeling)、数据元(Data Element)、数据字典(Data Dictionary)、数据库建设等数智化信息技术,设计并实现了满足矿集区找矿预测全过程质量检查与评价的质量控制模型库,主要内容包括两部分:支持地球科学领域数据质量控制的基础部分和直接应用于矿集区找矿预测全过程质量控制部分。后者主要包括:项目验收规则库、矿集区找矿预测数据模型库、图件类质量模型库、编图规则库、空间拓扑规则库、质检质评记录模版库;表格类质量模型库、采集规则库、约束规则库;质检质评记录模版库、质评表单模版库、质量评审意见模版库、质量评述报告模版库、质量附件模版库以及其他模版、规则或知识库。质量控制模型库的构建方法及质量控制模型库的分层架构(图5)。

3 质量检查与评价核心方法

3.1 数据划分及其粒级

3.1.1 数据划分

经典数学集合理论包涵“集合”和“集合划分”等基本概念。简单地说,“集合”指的是具有相同属

性的事物整体,而“集合划分”则是将具有相同属性的事物整体(即集合)分成若干个互不重叠或重复的小组(即子集),这些互不重叠或重复的小组合并在一起正好是未划分前的事物整体。由于数据库或数据集本身就是数据的集合,本研究在系统构建并实现数据质量检查与评价数智化体系时,将经典数学集合理论中的“集合”和“集合划分”理论与本研究具体实际相结合,直接提出“数据划分”思想方法及矿集区找矿预测数据划分方案。

提出矿集区找矿预测数据划分方法及划分方案的真正目的,一方面是为了将复杂、繁重或重复的传统人工质量控制工作转化为计算机及信息技术应用问题,充分且有效地借助现代数智化信息技术,完成一系列复杂、繁重或重复的质量控制任务,如自检、互检、专检、野外工作验收、监督检查、初审、终审、复核以及确认接收等矿集区找矿预测大数据实体建设全过程质量控制活动。另一方面,为了确保采用的计算机质量控制技术方法和实施手段满足规范统一、客观可信以及实战且高效的需求,实现效率高、成本低、规范且统一、稳健且公正以及数智化质量控制的目标。因此,必须基于矿集区找矿预测大数据实体内在逻辑,对其进行科学、

合理的数据划分。

矿集区找矿预测数据划分方法及划分方案或称“数据划分树”(图 6)简述如下:(1)矿集区找矿预测大数据实体划分为矿集区找矿预测原始资料数据、矿集区找矿预测成果资料数据、矿集区找矿预测实物资料数据。(2)矿集区找矿预测原始资料数据划分为图件类数据集、表格类数据集,矿集区找矿预测成果资料数据划分为图件类数据集、表格类数据集,矿集区找矿预测实物资料数据划分为图件类数据集、表格类数据集。(3)图件类数据集按照“图件类数据集(即图件集)→图件子集→图件→图件的四个部分”四个层次逐级划分。必须明确指出:相同类型的图件能组成一个图件子集,不同类型的图件应分属不同的图件子集且各种图件子集不得重叠或重复。关于图件类型的定义及种类已在矿集区找矿预测数据模型规范(左群超等, 2022)中有明确规定。单个图件是图件类数据集划分的基础,一个图件包括专业编图及图面表达信息集、空间数据数据集、属性数据数据集以及元数据数据集。(4)表格类数据集按照“表格类数据集(即表格集)→表格子集→表格→表格的四个部分”四个层次逐级划分。必须明确指出:相同类型的表格能组

质量控制模型库构建迭代过程及分层结构(可扩展并支撑地球科学领域其他数据资产建设质量控制活动)

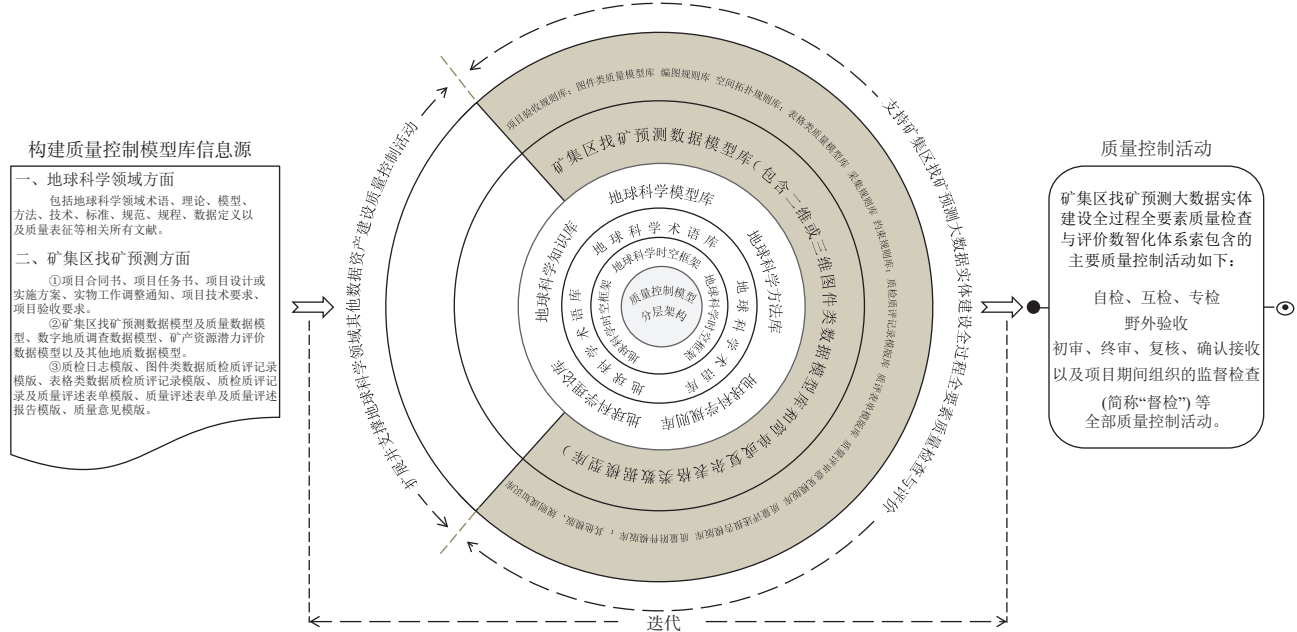


图 5 矿集区找矿预测质量控制模型库构建方法及分层结构

Fig.5 Construction method and hierarchical structure of the quality control model database for ore-searching prognosis in mineralization concentrating areas

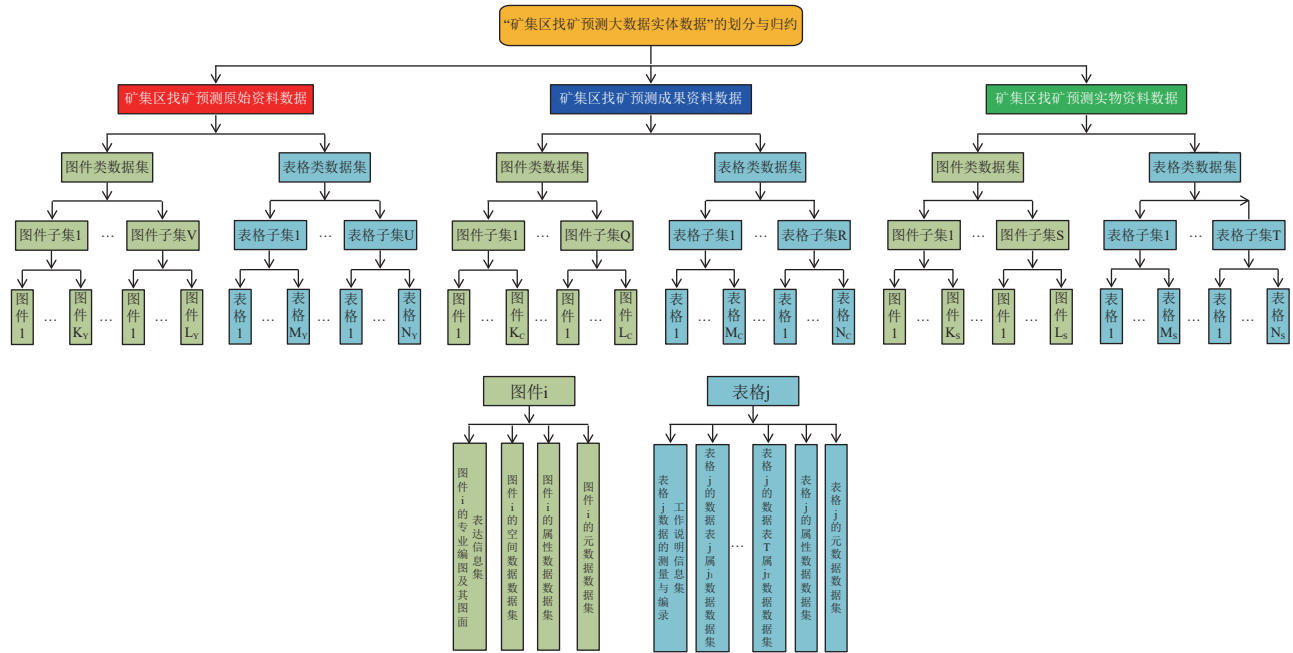


图6 矿集区找矿预测大数据实体数据的划分与归约方案

Fig.6 Division and reduction scheme of big data entities for ore-searching prognosis in mineralization concentrating areas

成一个表格子集,不同类型的表格应分属不同的表格子集且各种表格子集不得重叠或重复。关于表格类型的定义及种类已在矿集区找矿预测数据模型规范(左群超等,2022)中有明确规定。单个表格是表格类数据集划分的基础,一个表格包括表格数据的测量与编录工作说明信息集、表格的各个数据表的属性数据集、表格的全部数据表的属性数据集、表格数据的元数据集。

数据划分方法及划分方案所提及图件种类或表格种类,即图件类数据模型定义的图件种类或表格类数据模型定义的表格种类。图件类数据模型和表格类数据模型应符合“信息本体属性约束”(见本文“讨论”章节“信息本体属性约束”部分)。矿产勘查预测领域或地球科学其他及以外领域的、能用符合“信息本体属性约束”的图件类数据模型描述的所有二维或高维空间对象的图件,均属于数据划分方法及划分方案所提及的图件;矿产勘查预测领域或地球科学其他及以外领域的、能用符合“信息本体属性约束”的表格类数据模型描述的所有结构与关系相对简单的数据表或结构与关系异常复杂的数据库,均属于数据划分方法及划分方案所提及的表格。

从数据表现形式而言,矿集区找矿预测大数据

实体不论是作为整体本身,还是其中一部分(如原始资料数据、成果资料数据或实物资料数据),均可归约为由图件类数据或表格类数据组成。因此,本研究提出的矿集区找矿预测数据划分方法及划分方案,不仅理论严谨,而且方案完备,为质量检查与评价数智化实现奠定数据划分基础。

3.1.2 数据粒度

“数据粒度”概念源自数据库技术领域,用以描述数据对象的细化程度。一般而言,数据对象越细化,其粒度越小;反之,数据对象越粗化,其粒度越大。本研究引入“数据粒度”概念目的在于设计合理的数据对象粒度分级方案,以便构建数据质量检查与评价模型并简化计算和操作,减少人工干预。

本研究设计了五个级次的数据粒度:巨粒、粗粒、中粒、细粒和微粒。在巨粒级别中包括矿集区找矿预测的原始资料数据、成果资料数据、实物资料数据;粗粒级别则包括图件类数据集、表格类数据集;中粒级别包括图件子集、表格子集;细粒级别涵盖具体的图件、表格;微粒级别进一步包括图件或表格的各个部分,如:图件的专业编图及图面表达信息集、图件的空间数据数据集、图件的属性数据数据集、图件的元数据数据集;表格的测量与编录工作说明信息集、表格的每个数据表的属性数据

数据集、表格的属性数据数据集、表格的元数据数据集。这样的数据粒度方案有助于进行矿集区找矿预测数据库的质量检查与评价工作。

3.1.3 数据划分及数据粒度

(1) 实体数据划分及粒度

矿集区找矿预测数据划分方案或称“数据划分树”的根节点为“矿集区找矿预测大数据实体数据”。根的三个直接子节点分别为矿集区找矿预测原始资料数据、矿集区找矿预测成果资料数据、矿集区找矿预测实物资料数据。三个直接子节点则由图件类数据集或表格类数据集子树构成。矿集区找矿预测大数据实体数据的数据划分、数据粒度以及实例说明(表 1)。

(2) 图件类数据划分及粒度

图件类数据被划分为图件类数据集(即图件集)、图件子集、图件和图件的四个部分。图件类数据的数据划分、数据粒度以及实例说明(表 2)。

(3) 表格类数据划分及粒度

表格类数据被划分为表格类数据集(即表格集)、表格子集、表格和表格的四个部分。表格类数据的数据划分、数据粒度以及实例说明(表 3)。

3.2 质量检查评价模型

质量检查与评价模型(简称质量模型)是用于检查矿集区找矿预测大数据实体的部分或整体的各类数据质量指标的实际状况,并以此为依据评价所检查的部分或整体数据质量水平所处等级的数据模型。该质量模型主要建立在矿集区找矿预测大数据实体数据模型基础之上,而矿集区找矿预测大数据实体总可以归结为图件类数据和表格类数据两大类。因此,从用户视角而言,矿集区找矿预测大数据实体的质量检查与评价模型实际上由矿集区找矿预测图件类数据的质量检查与评价模型(即图件类数据质量模型)和矿集区找矿预测表格类数据的质量检查与评价模型(即表格类数据质量模型)两部分构成。

3.2.1 图件类数据质量模型

基于图件类数据的数据划分、数据粒度以及它们的层次与逻辑联系,图件类数据质量模型包括七个子模型,分别涉及单个图件的专业编图和图面表达、单个图件的空间数据、单个图件的属性数据、单个图件的元数据、单个图件、图件子集以及图件

集的质量检查与评价。这些子模型之间的层次与逻辑关系(图 7)揭示了矿集区找矿预测图件类数据质量检查 and 评价的基本原理框架,支持自检、互检、专检、野外工作验收、初审、终审、复核、确认接收以及项目期间组织的质量监督检查等所有质量控制活动。

基于矿集区找矿预测的图件类数据模型、编图处理规则、图件类拓扑规则以及质量控制模型基础层(地球科学领域术语、知识、理论、模型、方法、技术、标准、规范、规程、数据定义以及质量表征等),图件类数据质量模型的七个质量检查与评价子模型各自具有其独特的作用并共同支持图件类数据质量检查与评价实现。单个图件专业编图及图面表达信息质量检查与评价模型用于检查与评价单个图件的专业编图处理及图面表达信息的质量。单个图件空间数据质量检查与评价模型用于检查与评价单个图件的空间数据质量。单个图件属性数据质量检查与评价模型用于检查与评价单个图件的属性数据质量。单个图件元数据质量检查与评价模型用于检查与评价单个图件的元数据质量。单个图件质量检查与评价模型用于质量检查与评价单个图件整体的数据质量。图件子集质量检查与评价模型用于检查与评价图件子集的数据质量。图件集质量检查与评价模型用于检查与评价图件集的数据质量。

3.2.2 表格类数据质量模型

基于表格类数据的数据划分、数据粒度以及它们的层次与逻辑联系,表格类数据质量模型包括七个子模型,分别涉及单个表格的数据测量和编录工作说明、单个表格的每个数据表属性数据、单个表格属性数据、单个表格元数据、单个表格、表格子集以及表格集的质量检查与评价。这些子模型之间的层次与逻辑关系(图 8)揭示了矿集区找矿预测表格类数据质量检查 and 评价的基本原理框架,支持自检、互检、专检、野外工作验收、初审、终审、复核、确认接收以及项目期间组织的质量监督检查等所有质量控制活动。

基于矿集区找矿预测的表格类数据模型、数据测量和编录规则、表格类约束规则以及质量控制模型基础层(地球科学领域术语、知识、理论、模型、方法、技术、标准、规范、规程、数据定义以及质量表征等),表格类数据质量模型的七个质量检查与

表 1 矿集区找矿预测大数据实体数据的数据划分及数据粒级一览

Table 1 Data division and graininess of big data entities for ore-searching prognosis in mineralization concentrating areas			
层次	数据划分	数据粒级	实例
0	矿集区找矿预测大数据实体数据 (由矿集区找矿预测原始资料数据、矿集区找矿预测成果资料数据、矿集区找矿预测实物资料数据构成)	根	河北宣化—丰宁矿集区找矿预测项目建设完成的“河北宣化—丰宁矿集区找矿预测数据库”，由“河北宣化—丰宁矿集区找矿预测原始资料数据”、“河北宣化—丰宁矿集区找矿预测成果资料数据”和“河北宣化—丰宁矿集区找矿预测实物资料数据”构成
1	矿集区找矿预测原始资料数据 (由图件类数据集和表格类数据集构成)	巨粒	河北宣化—丰宁矿集区找矿预测项目建设完成的“河北宣化—丰宁矿集区找矿预测原始资料数据”，由原始资料图件类数据集和原始资料表格类数据集构成
	矿集区找矿预测成果资料数据 (由图件类数据集和表格类数据集构成)		河北宣化—丰宁矿集区找矿预测项目建设完成的“河北宣化—丰宁矿集区找矿预测成果资料数据”，由成果资料图件类数据集和成果资料表格类数据集构成
	矿集区找矿预测实物资料数据 (由图件类数据集和表格类数据集构成)		河北宣化—丰宁矿集区找矿预测项目建设完成的“河北宣化—丰宁矿集区找矿预测实物资料数据”，由实物资料图件类数据集和实物资料表格类数据集构成

表 2 图件类数据的数据划分及数据粒级一览

Table 2 Data division and graininess of map-type data entities			
层次	数据划分	数据粒级	实例
2	图件集 (由若干类图件组成)	粗粒	河北宣化—丰宁矿集区找矿预测项目编制完成的所有成果资料图件 (包含多种图件类, 如5万分幅建造构造图件类、5万分幅构造纲要图件类等), 构成一个图件集
3	图件子集 (由同一类图件组成)	中粒	河北宣化—丰宁矿集区找矿预测项目编制完成的属于5万分幅建造构造图件类的所有图件, 构成一个1:5万分幅建造构造图件类图件子集
4	图件 (单个图件)	细粒	河北宣化—丰宁矿集区找矿预测项目编制完成的崇礼幅1:5万建造构造图件 (包括专业编图及图面表达信息集、图件空间数据数据集、图件属性数据集、图件元数据数据集)
5	一个图件数据的四个部分	微粒	崇礼幅1:5万建造构造图件的专业编图及其图面表达信息集
			崇礼幅1:5万建造构造图件的空间数据数据集
			崇礼幅1:5万建造构造图件的属性数据集
			崇礼幅1:5万建造构造图件的元数据数据集

表 3 表格类数据的数据划分及数据粒级一览

Table 3 Data division and graininess of table-type data entities			
层次	数据划分	数据粒级	实例
2	表格集 (由若干类表格组成)	粗粒	河北宣化—丰宁矿集区找矿预测项目测量完成的所有原始资料表格 (包含多种表格类数据, 如5万分幅1:5万化探测量表格类数据、5万分幅1:5万磁法测量表格类数据等), 构成一个表格集
3	表格子集 (由同一类表格组成)	中粒	河北宣化—丰宁矿集区找矿预测项目测量完成的属于5万分幅1:5万化探测量的所有原始资料表格数据, 构成一个5万分幅1:5万化探测量表格类表格子集
4	表格 (单个表格)	细粒	河北宣化—丰宁矿集区找矿预测项目测量完成的崇礼幅1:5万化探测量数据集 (包括表格的测量与编录工作说明信息集、表格的每个数据表属性数据集、表格的所有数据表属性数据集、表格的元数据集)
5	一个表格数据的四个部分	微粒	崇礼幅1:5万化探测量的专题工作设计、野外工作验收报告、测量及编录工作报告等
		微粒	崇礼幅1:5万化探测量的样品测试分析数据表的属性数据构成的数据集
		微粒	崇礼幅1:5万化探测量的采样点位记录数据表的属性数据构成的数据集
		微粒	崇礼幅1:5万化探测量数据集的所有数据表的属性数据构成的数据集
	表格数据的元数据数据集	微粒	崇礼幅1:5万化探测量数据集的元数据集

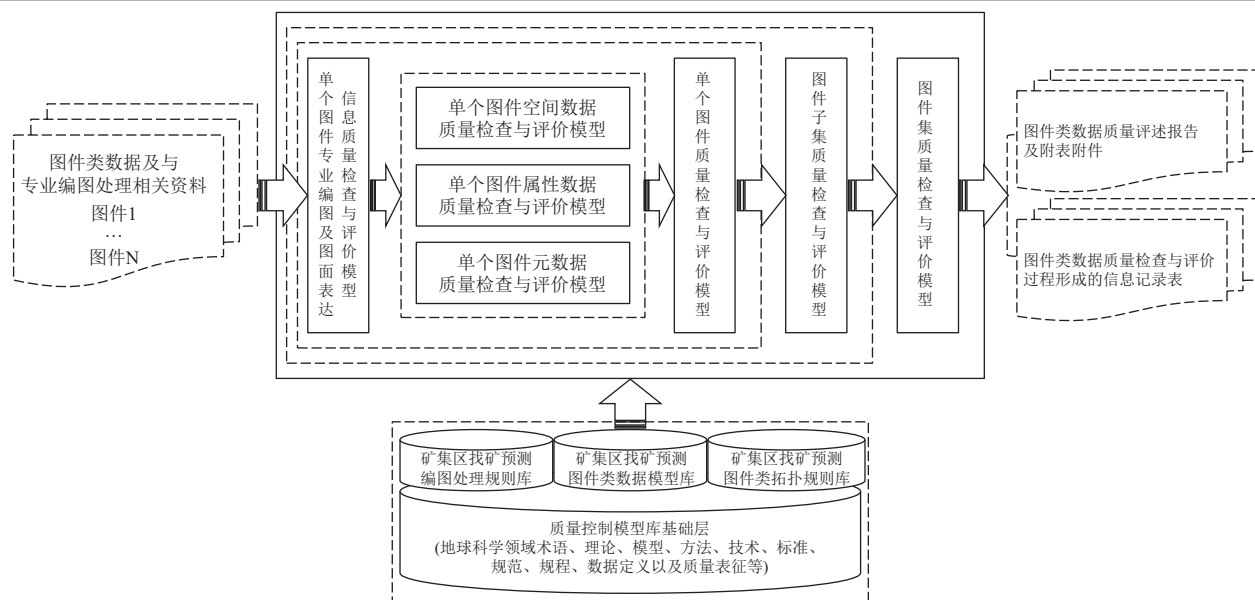


图 7 图件类数据质量模型

Fig.7 Quality models of map-type data

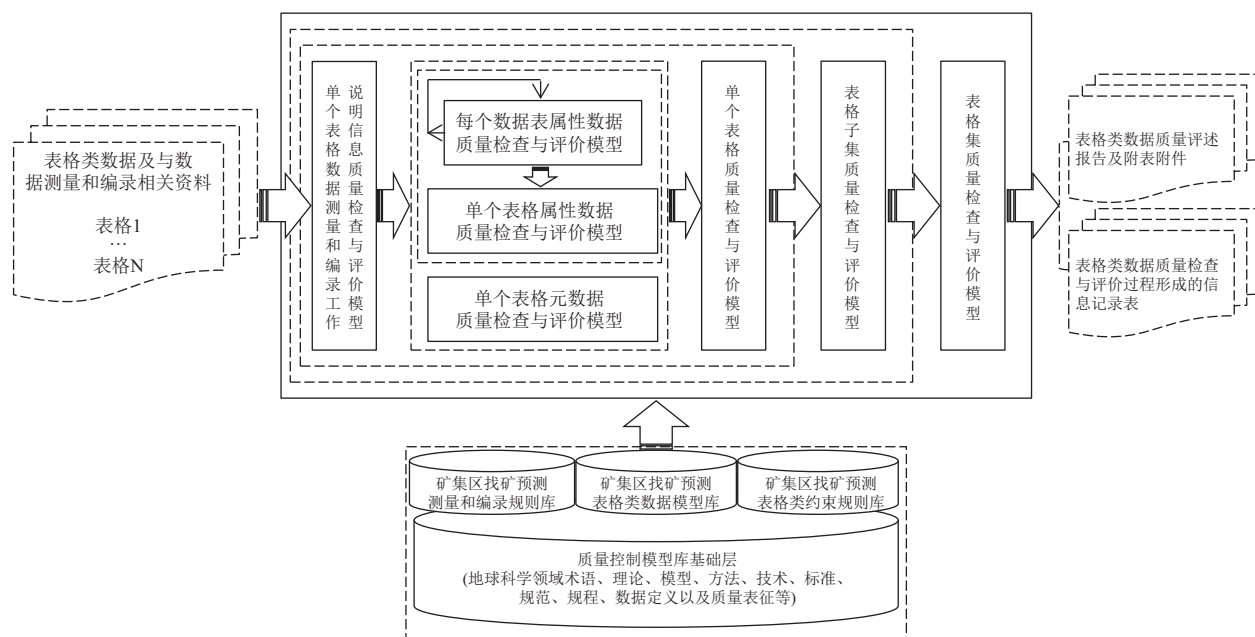


图 8 表格类数据质量模型

Fig.8 Quality models of table-type data

评价子模型各自具有其独特的作用并共同支持表格类数据质量检查与评价实现。单个表格的数据测量和编录工作说明信息质量检查与评价模型用于检查与评价单个表格的数据测量和编录工作的质量。单个表格的每个数据表属性数据质量检查与评价模型用于检查与评价单个表格的每个数据表属性数据质量。单个表格属性数据质量检查与

评价模型用于检查与评价单个表格的属性数据质量。单个表格元数据质量检查与评价模型用于检查与评价单个表格的元数据质量。单个表格质量检查与评价模型用于质量检查与评价单个表格整体的数据质量。表格子集质量检查与评价模型用于检查与评价表格子集的数据质量。表格集质量检查与评价模型用于检查与评价表格集的数据质量。

3.3 缺陷分级及检查项

3.3.1 缺陷与分级

数据质量元素(即质量元素)是指描述数据实体与相应规范或利用需求对数据实体要求符合程度的定性或定量的质量特性。根据不同质量特性及重要程度,质量元素可以分类分级,即描述特定数据实体的质量特性的质量元素集可由若干不同类别的质量元素构成,同时高层级的质量元素还可分为若干低层级的不同类别的质量元素。例如,数据完整性(种类完整性、数量完整性等)、逻辑一致性(与设计方案一致性、与技术标准一致性等)、空间定位准确度、属性数据准确性、图面整饰规范性、数据源头有效性等是常见质量元素,有助于描述数据的质量等级,确保质量检查与评价的可靠性和合理性。

数据质量缺陷(即质量缺陷)是指数据实体的质量特性不满足相应规范或利用需求的情况或事实。在数据质量检查与评价中,用质量缺陷数(即缺陷数)来量化数据质量检查发现的各类不符合相应规范或利用需求的问题。依据质量缺陷和质量元素的概念,质量缺陷是质量元素所描述数据质量特性的负面方面。因此,依据质量元素分类分级和重要程度等特点,质量缺陷可分类分级并进行比较,利用计算机软件并采用全检方式高效检查发现数据实体存在的质量缺陷并统计缺陷数。

矿集区找矿预测大数据实体数据种类多而繁杂且为一次性采集、编录或编研而成,每一图件、表格、样品或数据项均不具备类似工厂生产品的重复特征。因此,为了准确度量、精细检查和客观评价此类数据实体质量,本研究采用缺陷分级系统:极严重缺陷、严重缺陷、重缺陷、次重缺陷、轻缺陷、次轻缺陷。按照中国地质调查局地质调查技术标准《数字地质数据质量检查与评价》(DZ/T 0268—2014)除了极严重缺陷外,其余不同级缺陷间存在一定量算关系,即“一个严重缺陷/一个重缺陷/一个次重缺陷/一个轻缺陷/一个次轻缺陷=16/9/5/2/1”。

极严重缺陷表示数据实体的极重要质量元素不符合相应规范或利用需求的情况或事实,到了极严重不符合相应规范或利用需求的程度,表示该数据实体达不到预期目的,以致不经返工不能提供用

户使用。

3.3.2 质量检查项

质量检查项(即检查项)是指用于数据实体的质量元素方面的最小检查单元。通俗而言,检查项就是作用于数据实体的质量检查与评价的最小数据对象。应根据相应规范、利用需求和数据实体内容特征,界定和描述数据实体的检查项、缺陷与分级。系统、科学而有效地界定和描述检查项、缺陷与分级,对实现矿集区找矿预测大数据实体的质量检查与评价至关重要。简要说明矿集区找矿预测图件类数据和表格类数据的检查项、所有可能存在的缺陷以及缺陷分级情况如下。

3.3.2.1 图件类数据检查项及缺陷与分级

根据矿集区找矿预测图件类数据质量检查与评价基本原理框架(图7),图件类数据检查项及缺陷与分级应涵盖四个层次:单个图件的四个部分、单个图件、图件子集和图件集。单个图件的四个部分的检查项及缺陷与分级的界定和描述,是界定与描述单个图件、图件子集和图件集的检查项及缺陷与分级的基础。单个图件的检查项及缺陷与分级是对单个图件的四个部分检查项及缺陷与分级的简单应用。图件子集的检查项及缺陷与分级,是对相同类型若干图件的检查项及缺陷与分级的汇总应用;而图件集的检查项及缺陷与分级,则是对不同类型若干图件子集的检查项及缺陷与分级的综合应用。

实现图件类数据质量检查与评价,首先需要详细界定和具体描述图件的四个部分的检查项、所有可能存在的缺陷以及缺陷分级,然后基于前者综合界定和描述图件类数据(适合单个图件、图件子集和图件集)的检查项、缺陷与分级。表4列出了图件的四个部分的检查项及缺陷与分级,表5列出了图件类数据的检查项及缺陷与分级。

3.3.2.2 表格类数据检查项及缺陷与分级

根据矿集区找矿预测表格类数据质量检查与评价基本原理框架(图8),表格类数据检查项及缺陷与分级应涵盖四个层次:单个表格的四个部分、单个表格、表格子集和表格集。单个表格的四个部分的检查项及缺陷与分级的界定和描述,是界定与描述单个表格、表格子集和表格集的检查项及缺陷与分级的基础。单个表格的检查项及缺陷与分级

表 4 图件的四部分的质量检查项及缺陷与分级

Table 4 Quality check items, defects, and defect grading for the four parts of one map

图件的四部分	质量检查项	可能存在的缺陷与分级
图件的专业编图及图面表达信息	1) 编图工作流程符合规定程度	1) 编图工作流程完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	2) 编图使用资料数据有效性 (包括来源、时效、比例尺、质量、精度等) 符合规定程度	2) 编图使用资料数据有效性 (包括来源、时效、比例尺、质量、精度等) 完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	3) 编图资料数据处理方法或过程符合规定程度	3) 编图资料数据处理方法或过程完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	4) 图面信息内容符合规定程度	4) 图面信息内容完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	5) 图面表达符合规定程度	5) 图面表达完全不符合要求 (严重)、部分不符合要求 (重、次重、轻、次轻)
	6) 编图比例尺符合规定程度	6) 编图比例尺完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	7) 编图空间范围符合规定程度	7) 编图空间范围完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	8) 编图技术参数符合规定程度	8) 编图技术参数完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	9) 编图说明书内容符合规定程度	9) 编图说明书内容完全不符合要求 (严重)、部分不符合要求 (重、次重、轻、次轻)
	10) 其他相关规定符合程度	10) 其他相关规定符合程度不符合要求 (轻、次轻)
图件的空间数据	1) 图件坐标系统参数符合规定程度	1) 图件坐标系统参数完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	2) 图件地图投影及参数符合规定程度	2) 图件地图投影及参数完全错误 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	3) 图件使用图例图式符合规定程度	3) 图件使用图例库与统一规定不一致 (严重、重、次重、轻、次轻)
	4) 专业图层空间实体符合专业图层内容规定程度	4) 专业图层空间实体完全不符合专业图层内容规定 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	5) 图件的空间拓扑精度符合规定程度	5) 图件的空间拓扑精度不符合规定要求 (重、次重、轻、次轻)
	6) 相关边界图层 (国界、省界、工作区边界、典型矿床研究区边界等) 符合规定程度	6) 相关边界图层 (国界、省界、工作区边界、典型矿床研究区边界等) 不符合要求 (重、次重、轻、次轻)
	7) 图件、图层等结构及代码符合规定程度	7) 图件、图层等结构及代码不符合规定要求 (严重、重、次重、轻、次轻)
	8) 引用专业图层符合规定程度	8) 引用专业图层不符合要求 (严重、重、次重、轻、次轻)
	9) 辅图、辅表符合规定程度	9) 辅图、辅表完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	10) 辅助图层或地理图层符合规定程度	10) 辅助图层或地理图层不符合要求 (重、次重、轻、次轻)
	11) 图层分类 (包括专业图层、引用图层、辅助图层、地理图层、自增图层) 符合规定程度	11) 图层分类 (包括专业图层、引用图层、辅助图层、地理图层、自增图层) 完全不符合要求 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	12) 图件四或六至范围与编图范围一致性符合程度	12) 图件四或六至范围与编图范围不一致 (严重、重、次重、轻、次轻)
	13) 其他相关规定符合程度	13) 其他相关规定符合程度 (轻、次轻)
图件的属性数据	1) 图层属性表结构符合规定程度	1) 图层属性表结构完全不符合规定 (极严重)、部分不符合要求 (严重、重、次重、轻、次轻)
	2) 图层属性数据填写格式符合规定程度	2) 图层属性数据填写完全不符合规定 (极严重)、部分不符合规定 (严重、重、次重、轻、次轻)
	3) 图层属性数据填写内容符合现有地质情况研究程度	3) 图层属性数据填写内容完全不符合现有地质情况研究程度 (极严重)、部分不符合现有地质工作研究程度 (严重、重、次重、轻、次轻)
	4) 图层属性数据填写率符合地质资料拥有程度	4) 图层属性数据填写率不符合地质资料拥有程度 (极严重)、部分不符合地质资料拥有程度 (严重、重、次重、轻、次轻)
	5) 图元编号\特征代码填写符合规定程度	5) 图元编号\特征代码填写不符合规定程度 (轻、次轻)
	6) 其他相关规定符合程度	6) 其他相关规定符合程度 (轻、次轻)
图件的元数据	1) 图件的元数据内容填写符合规定程度	1) 图件的元数据内容完全未填写 (次重)、元数据内容填写不符合规定 (轻、次轻)
	2) 图件的元数据格式 (XML 格式和自由文本格式) 符合规定程度	2) 图件的元数据格式 (XML 格式和自由文本格式) 不符合规定 (轻、次轻)
	3) 图件的元数据填写内容与图件数据 (空间和属性) 一致性符合程度	3) 图件的元数据填写内容与图件数据 (空间和属性) 不一致 (次重、轻、次轻)
	4) 其他相关规定符合程度	4) 其他相关规定符合程度 (轻、次轻)

表 5 图件类数据的质量检查项及缺陷与分级

Table 5 Quality check items, defects, and defect grading for map-type data entities		
图件类数据	质量检查项	可能存在的缺陷与分级
完整性	1) 相关资料附件齐全程度	1) 缺少必要的相关报告、附件、附表等 (极严重、严重)、缺少相关的统计表格、清单、卡片等 (重、次重、轻、次轻)
	2) 图件种类完整性程度	2) 缺少必要的图件类 (极严重、严重)、缺少相关的图件类 (重、次重、轻、次轻)
	3) 图件数量完整性程度	3) 缺少必要的图件 (极严重、严重)、缺少相关的图件 (重、次重、轻、次轻)
	4) 其他相关规定符合程度	4) 其他相关规定符合程度 (轻、次轻)
一致性	1) 与确认设计方案一致性程度	1) 提交内容与任务设计不一致 (极严重、严重)、部分内容与任务设计不一致 (重、次重、轻、次轻)
	2) 与规定技术要求一致性程度	2) 提交内容与规定技术要求不一致 (极严重、严重)、部分内容与与规定技术要求不一致 (重、次重、轻、次轻)
	3) 其他相关规定符合程度	3) 其他相关规定符合程度 (轻、次轻)
规范性	1) 内容规范程度	1) 提交内容与内容规范不一致 (极严重、严重)、部分提交内容与内容规范不一致 (重、次重、轻、次轻)
	2) 形式规范程度	2) 提交内容目录形式与内容目录规范不一致 (极严重、严重)、部分提交内容目录形式与内容目录规范不一致 (重、次重、轻、次轻)
	3) 管理规范程度	3) 提交内容未经过相应初审流程等 (极严重、严重)、初审流程不符合要求 (严重、重、次重、轻、次轻)
	4) 其他相关规定符合程度	4) 其他相关规定符合程度 (轻、次轻)

是对单个表格的四个部分检查项及缺陷与分级的简单应用。表格子集的检查项及缺陷与分级,是对相同类型若干表格的检查项及缺陷与分级的汇总应用;而表格集的检查项及缺陷与分级,则是对不同类型若干表格子集的检查项及缺陷与分级的综合应用。

实现表格类数据质量检查与评价,首先需要详细界定和具体描述表格的四个部分的检查项、所有可能存在的缺陷以及缺陷分级,然后基于前者综合界定和描述表格类数据(适合单个表格、表格子集和表格集)的检查项、缺陷与分级。表 6 列出了表格的四个部分的检查项及缺陷与分级,表 7 列出了表格类数据的检查项及缺陷与分级。

4 讨 论

本文前述已对全过程全要素质量检查与评价数智化体系的组成框架、具体实现、所用方法技术以及数据划分与粒级质量控制理论等重要内容均有详实阐述,还需进一步讨论全过程全要素质量控制数智化体系的几个关键点,便于系统、完整或深入理解,同时简要说明今后研究方向和任务。

(1)本研究建立的全过程全要素质量检查与评价数智化体系高效性、客观性和实战性运行的代表性案例。

案例 1: 动用全国 30 个省(直辖市、自治区)的

地质科研力量,历经 8 年(2006 年至 2013 年)建设形成的全国矿产资源潜力评价大数据资产(总数据量约 10TB)。其空间范围覆盖全国 32 个省(直辖市、自治区),潜力评价矿种或矿组包括煤、铀、铁等 25 种以上;涉及专业包括物化遥自然重砂资料应用、成矿地质背景研究、成矿规律研究和矿产预测研究等;数据内容主要包括物、化、遥、自然重砂、地质背景、成矿规律以及矿产预测等专题图件 11 万个、8 大类全国级规模基础数据(包括地质、矿产地、航磁、重力、化探、重砂、工作程度、地质地理底图等)。借助本研究建立的质量控制体系、质量控制软件以及经统一受训过的同等数量业内权威专家组支持,即:使用“同一把尺子”对该数据资产建设质量实施终审检查与评价,并分别按 32 个省(直辖市、自治区)矿产资源潜力评价项目形成正式质量终审意见、质量评述报告以及检查与评价明细记录附表附件等,实际所用时间为 15 个工作日(按传统方式测算需 1 年以上时间方能完成)。较之传统方式,提高工作效率 25 倍以上。而且,据全国地质资料馆反馈:全国矿产资源潜力评价所汇交数据资料的质量等级为优秀(完整性、一致性、规范性、易用性),10 余年来的借阅利用率一直保持最高水平。

案例 2: 动用全国地矿、有色、冶金和核工业系统共 30 个地勘单位的勘查预测力量,历经 6 年

表 6 表格的四部分的质量检查项及缺陷与分级

Table 6 Quality check items, defects, and defect grading for the four parts of one table

表格的四部分	质量检查项	可能存在的缺陷与分级
表格数据的测量(或采集)与编录工作说明信息	1) 测量(或采集)与编录工作部署及实施采用的技术要求、规范及标准符合规定程度	1) 测量(或采集)与编录工作部署及实施采用的技术要求、规范及标准完全不符合规定(极严重)、部分不符合规定(严重、重、次重、轻、次轻)
	2) 测量(或采集)与编录设备及环境参数符合规定程度	2) 测量(或采集)与编录设备及环境参数完全不符合规定(极严重)、部分不符合规定(严重、重、次重、轻、次轻)
	3) 测量(或采集)与编录获取资料内容要素符合规定要求程度	3) 测量(或采集)与编录获取资料内容要素完全不符合规定要求(极严重)、部分不符合规定要求(严重、重、次重、轻、次轻)
	4) 测量(或采集)与编录完成任务及实物工作量符合设计规定程度	4) 测量(或采集)与编录完成任务及实物工作量完全不符合设计规定(极严重)、部分不符合设计规定(严重、重、次重、轻、次轻)
	5) 其他相关规定符合程度	5) 其他相关规定符合程度(轻、次轻)
表格的数据表的属性数据	1) 数据表结构符合规定程度	1) 数据表结构完全不符合规定(极严重)、部分不符合规定(严重、重、次重、轻、次轻)
	2) 数据表的属性数据填写使用的资料数据有效性(包括来源、时效、比例尺、质量、精度、处理等)符合要求程度	2) 数据表属性数据填写使用的资料数据有效性完全不符合要求(极严重)、部分不符合要求(严重、重、次重、轻、次轻)
	3) 数据表的属性数据填写内容符合现有地质情况研究程度	3) 数据表的属性数据填写内容完全不符合现有地质情况研究程度(极严重)、部分不符合现有地质情况研究程度(严重、重、次重、轻、次轻)
	4) 数据表的属性数据填写率符合地质资料拥有程度	4) 数据表的属性数据填写率完全不符合地质资料拥有程度(极严重)、部分不符合地质资料拥有程度(严重、重、次重、轻、次轻)
	5) 数据表的属性数据填写格式符合规定程度	5) 数据表的属性数据填写格式完全不符合规定(极严重)、部分不符合规定(严重、重、次重、轻、次轻)
	6) 数据表约束条件符合规定程度	6) 数据表约束条件完全不符合规定(极严重)、部分不符合规定(严重、重、次重、轻、次轻)
	7) 其他相关规定符合程度	7) 其他相关规定符合程度(轻、次轻)
表格的属性数据	1) 表格的数据表的完整性程度	1) 表格缺少必要的数据表(极严重)、表格缺少次要的数据表(严重、重、次重、轻、次轻)
	2) 表格的数据表结构符合规定程度	2) 表格的数据表结构完全不符合规定(极严重)、部分不符合规定(严重、重、次重、轻、次轻)
	3) 表格的数据表的属性数据填写使用的资料数据有效性(包括来源、时效、比例尺、质量、精度等)符合要求程度	3) 表格的数据表属性数据填写使用的资料数据有效性完全不符合要求(极严重)、部分不符合要求(严重、重、次重、轻、次轻)
	4) 表格的数据表的属性数据填写内容符合现有地质情况研究程度	4) 表格的数据表的属性数据填写内容完全不符合现有地质情况研究程度(极严重)、部分不符合现有地质情况研究程度(严重、重、次重、轻、次轻)
	5) 表格的数据表的属性数据填写率符合地质资料拥有程度	5) 表格的数据表的属性数据填写率完全不符合地质资料拥有程度(极严重)、部分不符合地质资料拥有程度(严重、重、次重、轻、次轻)
	6) 表格的数据表的属性数据填写格式符合规定程度	6) 表格的数据表的属性数据填写格式完全不符合规定(极严重)、部分不符合规定(严重、重、次重、轻、次轻)
	7) 表格的数据表间约束关系符合规定程度	7) 表格的数据表间约束关系完全不符合规定(极严重)、部分不符合规定(严重、重、次重、轻、次轻)
	8) 其他相关规定符合程度	8) 其他相关规定符合程度(轻、次轻)
表格的元数据	1) 元数据内容填写符合规定程度	1) 元数据内容填写不符合规定(轻、次轻)
	2) 元数据格式(XML格式和自由文本格式)符合规定程度	2) 元数据格式(XML格式和自由文本格式)不符合规定(轻、次轻)
	3) 元数据填写内容与表格数据(表格结构组成和表格属性数据等)一致性符合程度	3) 元数据填写内容与表格数据(表格结构组成和表格属性数据等)不一致(轻、次轻)
	4) 其他相关规定符合程度	4) 其他相关规定符合程度(轻、次轻)

(2016 年至 2021 年)建设形成的全国矿集区调查及深部找矿预测大数据资产(总数据量约 6TB)。其

空间范围覆盖 60 个全国重要矿集区,找矿预测矿种或矿组包括铀、金、铁等 12 种以上,涉及专业包

表 7 表格类数据的质量检查项及缺陷与分级

Table 7 Quality check items, defects, and defect grading for table-type data entities		
表格类数据	质量检查项	可能存在的缺陷与分级
完整性	1) 相关资料附件齐全程度	1) 缺少必要的相关报告、附件、附表等 (极严重、严重)、缺少相关的统计表格、清单、卡片等 (重、次重、轻、次轻)
	2) 表格种类完整性程度	2) 缺少必要的表格类 (极严重、严重)、缺少相关的表格类 (重、次重、轻、次轻)
	3) 表格数据完整性程度	3) 缺少必要的表格 (极严重、严重)、缺少相关的表格 (重、次重、轻、次轻)、缺少相关的数据表 (重、次重、轻、次轻)
	4) 其他相关规定符合程度	4) 其他相关规定符合程度 (轻、次轻)
一致性	1) 与任务设计要求一致性程度	1) 提交内容与任务设计不一致 (极严重、严重)、部分内容与任务设计不一致 (重、次重、轻、次轻)
	2) 与规定技术要求一致性程度	2) 提交内容与规定技术要求不一致 (极严重、严重)、部分内容与与规定技术要求不一致 (重、次重、轻、次轻)
	3) 其他相关规定符合程度	3) 其他相关规定符合程度 (轻、次轻)
规范性	1) 内容规范程度	1) 提交内容与内容规范不一致 (极严重、严重)、部分提交内容与内容规范不一致 (重、次重、轻、次轻)
	2) 形式规范程度	2) 提交内容目录形式与内容目录规范不一致 (极严重、严重)、部分提交内容目录形式与内容目录规范不一致 (重、次重、轻、次轻)
	3) 管理规范程度	3) 提交内容未经过相应初审流程等 (极严重、严重)、初审流程不符合要求 (严重、重、次重、轻、次轻)
	4) 其他相关规定符合程度	4) 其他相关规定符合程度 (轻、次轻)

括中比例尺地质填图、大比例尺地质填图、物化遥地质测量、勘探工程、采样测试分析、矿产地调查、物化遥异常综合查证、地质背景研究、成矿规律研究、矿产预测研究等,数据内容主要包括 60 个矿集区找矿预测项目形成的原始、成果和实物资料数据。借助本研究建立的质量控制体系、质量控制软件以及经统一受训过的同等数量业内权威专家组支持,即:使用“同一把尺子”对该数据资产建设质量实施终审检查与评价,并分别按 62 个矿集区找矿预测项目形成正式质量终审意见、质量评述报告以及检查与评价记录明细附表附件等,实际所用时间为 5 个工作日(按传统方式测算需 6 个月以上时间方能完成)。较之传统方式,提高工作效率 25 倍以上。而且,据全国地质资料馆反馈:全国矿集区调查及深部找矿预测所汇交数据资料的质量等级为优秀(完整性、一致性、规范性、易用性),属于支撑国家新一轮找矿突破战略行动借阅利用率最高的数据资料。

(2)本研究为支撑全过程全要素质量检查与评价数智化体系运行制定了管理办法、操作规范、统一培训教材、验收技术要求以及一系列记录、意见和报告模版。

管理办法包括体系运行的组织架构、业务架构、技术架构以及体系运行规程与管理制度等;制定了质量控制体系、质量控制活动以及质量控制软

件等操作规范;编制了统一培训教材;制定了原始数据资料、成果数据资料以及实物资料数据验收技术要求;制定了适合于自检、互检、专检、野外工作验收(数据部分)、初审、终审、复核、确认接收以及项目期间组织的质量监督检查等质量控制活动的记录模版,野外工作验收数据部分意见模版及野外工作验收数据部分质量评述模版,初审意见模版及初审质量评述模版,终审意见模版及终审质量评述模版,督检报告模版、复核报告模版、确认接收报告模版等。

因此,本研究制定的管理办法、操作规范、统一培训教材、验收技术要求专门适用于质量管理者、数据建设者、质量检查与评价专家组;制定的一系列记录、意见和报告模版专门适用于质量控制软件。二者统一全过程全要素质量检查与评价数智化体系,支撑并实现质量控制目的。

(3)本研究提出的全过程全要素质量控制新范式的普适性基础。

全过程全要素质量控制新范式即以已重构并优化的数据生产及建设全过程(包括已重构并优化的领域研究工作标准规范和领域数据实体数据模型等)作为质量控制体系运行的“总流程”;采用满足“本体属性约束”的领域质量数据模型、领域数据实体数据模型以及领域质量控制模型基础框架(涵盖领域术语、知识、理论、模型、方法、技术、标准、规范、规程等)作为质量控制体系运行的“控制流”;

用大数据思维纳入与质量控制相关的全部要素,借助数智化技术实现手段(主要指质量控制软件)协同并驱动质量控制全部要素,开展数据生产及建设质量控制,包括质量自检、互检、专检、督检、野外工作验收、初审、终审、复核以及确认接收等全部质量控制活动,实现质量控制目标。

因此,本研究提出的全过程全要素质量控制新范式,源自地球科学最具代表性或典型性的领域,即矿产勘查预测领域。从质量控制而言,矿产勘查预测领域的代表性或典型性主要集中表现在研究对象的空间结构极度不确定性、关系结构异常复杂性。源自矿产勘查预测领域的全过程全要素质量控制新范式,很容易应用或扩展到空间结构和属性联系相对简单的地球科学其他领域或者地球科学以外的领域。

(4)本研究提出的数据模型符合“信息本体属性约束”的概念的具体涵义说明。

在计算机科学中,关于“信息本体”较为统一且公认的定义是“共享概念模型的明确的形式化规范说明”,即信息本体必需具备如下四个属性:①概念化(Conceptual),信息本体是通过抽象客观世界或一些现象而得的相关概念及概念模型,表示的含义独立于具体的环境状态。②明确的(Explicit),信息本体使用的概念、概念约束及概念间联系等均具明确的定义或规定。③共享的(Share),信息本体所反映的知识应该是共同认可的,是相关领域中公认的概念集合,它所针对的用户应该是团体而不是个体。④形式化(Formal),信息本体具有精确的数学描述形式,是计算机可读、可理解的。本研究提出数据模型应符合“信息本体属性约束”意即指定义的数据模型应满足信息本体规定的四个属性要求:概念化、明确的、共享的和形式化。领域质量数据模型、领域数据实体数据模型和领域质量控制模型本质上均属于数据模型范畴,本研究要求均应符合信息本体规定的概念化、明确的、共享的和形式化要求。

因此,本研究建立的图件类数据模型和表格类数据模型,均属于领域数据实体数据模型之一,二者的定义应符合“信息本体属性约束”。图件类数据模型,既适用于表述二维空间图件对象,也适用于表述高维空间图件对象;表格类数据模型,既适用于表述结构与关系相对简单的数据表,也适用于

表述结构与关系异常复杂的数据库。

(5)本研究研制的质量控制软件 GeoDQCBM 设计思想与实现理念独特。

GeoDQCBM 软件采用模型驱动及自动演进思想设计并研制,既有效且相对隔离了实体数据、数据模型、软件处理逻辑及其功能界面四者之间的复杂联系,又满足且控制了实体数据、数据模型、软件处理逻辑及其功能界面四者本身变化需求。四者本身变化需求具体表现分述为:1)实体数据的种类、内容及复杂度因领域不同而极度不同。2)数据模型因领域不同而完全迥异。3)软件处理逻辑解析数据模型、与数据模型语义一致。4)软件功能界面保持相对不变,即对所有领域数据资产建设质量控制总表现为:自检、互检、专检、野外工作验收、初审、终审、复核、确认接收以及项目期间组织的质量监督检查等质量控制活动。实体数据为领域实体数据,数据模型为领域数据实体数据模型、领域质量数据模型以及领域质量控制模型基础。

因此,质量控制领域的变化或扩展,只影响到需检查与评价的实体数据和驱动软件处理逻辑的数据模型两方面的变化或扩展,而不影响软件功能界面。采用模型驱动及自动演进思想设计并研制,确保质量控制软件 GeoDQCBM 可以从矿产勘查预测领域直接应用或扩展到地球科学领域或者地球科学以外领域。

(6)本研究提出质量控制活动特别是初审、终审均应采取“全检全评”方式的客观理由。

理由有两方面。一方面,矿产勘查预测领域特点之一就是其研究对象的空间结构极度不确定性和关系结构异常复杂性,特点之二是地质固有区域性特征。二者导致不同区域的地质内容不完全相同,每个地质数据或地质数据产品都是独一无二的。地质数据生产及地质数据产品,不能等同或视为具有可重复性的大规模工业生产及工业产品。因此,客观上要求对地质数据生产及地质数据产品的质量控制,采取全部检查、全部评价质量控制方式,即本研究主张的“全检全评”方式。另一方面,现代信息技术发展早已突破地质数据生产及地质数据产品的质量控制所需的存储容量和计算速度“瓶颈”。因此,技术上可实现“全检全评”方式的质量控制,而且,借助迅猛发展的大数据思维和数智

化技术,可以做得更好。推而广之,建议地球科学及以外领域的数字资产建设质量控制,均应采取“全检全评”方式。

(7)进一步研究内容和发展方向:本研究构建了全过程、全要素、全检全评方式的矿集区找矿预测及大数据实体建设质量检查与评价数智化体系,显著提高了质量控制效率并确保质量检查与评价结论的客观公正性。随着大数据思维和数智化变革向更高阶段发展演进,包括云计算、雾计算、边缘计算、物联网、无线网、区块链、大数据、人工智能、大模型、数字孪生等数智化信息技术,一方面应结合相关地质调查与地质科研项目,逐步丰富并完善质量控制模型库内容,使其适用于地球科学及以外领域数字资产建设质量控制应用需求,另一方面应密切关注数智化变革最新发展动态,持续迭代并完善全过程全要素质量控制数智化体系,以适应地球科学大数据爆发式增长、海量式集聚的发展模式,积极促进地球科学大数据蕴藏的巨大价值得以被发掘和利用。

5 结 论

(1)建立的质量检查与评价数智化体系具有原始创新性、实际应用效果和普适应用范围。其实现基础是系统而完备的数据划分、数据粒度、质量检查与评价模型理论与方法以及相关数智化技术巧用,展现出颠覆传统思维及做法的无可比拟的优越性;支持完成了全国矿产资源潜力评价及大数据实体建设、矿集区找矿预测及大数据实体建设的质量控制工作;可直接应用于新一轮找矿突破战略行动,支持新一轮找矿突破战略固体矿产资源大数据建设的质量控制工作。还可扩展应用到地球科学及以外领域的数字资产建设的质量控制工作。

(2)建立的质量检查与评价数智化体系具有高效性、客观性和全程实战性。高效性、客观性和全程实战性已在全国矿产资源潜力评价(2006—2013年)、全国矿集区调查及深部找矿预测(2016—2021年)中得到大规模实践验证,而且,较之传统方式,工作效率提高25倍以上。支持质量自检、互检、专检、野外工作验收、初审、终审、复核、确认接收以及项目期间组织的质量监督检查等质量控制活动,完全符合地质数据生产和质量管理实践

要求。

(3)制定的质量检查与评价数智化体系运行的管理办法、操作规范、统一培训教材、验收技术要求以及一系列记录、意见和报告模版,支撑质量检查与评价数智化体系实现质量控制目的,而且扩展到地球科学及以外广阔领域数字资产建设质量控制工作。

(4)提出的数据划分、数据粒度以及质量检查与评价模型理论与方法具有普适性,不仅适用矿产勘查预测领域数字资产建设质量控制工作,也适用于地球科学及以外领域数字资产建设质量控制工作。

(5)建立的全过程全要素质量控制新范式,已经全国范围、大规模、长周期项目实践验证,适用于地球科学及以外广阔领域数字资产建设质量控制工作,具有科学性、有效性和普适性。

(6)提出的数据模型定义应符合“信息本体约束”的概念具有普适性和实际意义,成为质量检查与评价数智化体系、全过程全要素质量控制新范式以及质量控制软件可扩展并应用到地球科学及以外广阔领域数字资产建设质量控制工作的基础。

(7)研制的质量控制软件显著提高了质量控制的规范性、客观性和工作效率,可直接应用到地球科学及以外广阔领域大数据资产建设质量控制。

(8)构建的质量控制全新范式及方法技术,为地球科学及以外广阔领域大数据资产建设质量控制管理体系构建和质量控制工作本身从传统模式加速向数智模式转型,提供了前瞻性实践探索案例。

致谢:衷心感谢中国地质调查局发展研究中心庞振山深部资源找矿预测研究团队提供的科研环境与全力支持、中国地质大学(武汉)汪新庆资源预测信息技术研究团队在实验测试与数据分析中的全力协作与探索实践;衷心感谢两位匿名审稿专家提出的深刻见解与细致修改意见,显著提升了本文的学术严谨性。同时,特别感谢《中国地质》编辑部在稿件评审与出版过程中的高效协调与专业支持。

References

- Ali A, A. N, A. S, R. A. 2019. An assessment of open data sets completeness[J]. International Journal of Advanced Computer Science and Applications, 10(6): 557–562.
- Bartky W. 1942. Statistical method from the viewpoint of quality

- control. by Walter A. Shewhart; W. Edwards Deming[J]. *American Mathematical Monthly*, 49(3): 188.
- Bianco S, Buzzelli M, Mazzini D, Schettini R. 2017. Deep learning for logo recognition[J]. *Neurocomputing*, 245: 23–30.
- Cai Lihua, Ni Daichuan. 2021. Review on evaluation of research data at home and abroad[J]. *Digital library Forum*, (11): 65–72 (in Chinese with English abstract).
- Carranza E J M, Laborde A G. 2015. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines)[J]. *Computers & Geosciences*, 74: 60–70.
- Chen C L, Zhang C. 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data[J]. *Information Sciences*, 275: 314–347.
- Chen Jianping, Li Jing, Xie Shuai, Liu Jing, Hu Bin. 2017. China geological big data research status[J]. *Journal of Geology*, 41(3): 353–366 (in Chinese with English abstract).
- Cheng Qiuming. 2021. What are mathematical geosciences and its frontiers?[J]. *Earth Science Frontiers*, 28(3): 6–25 (in Chinese with English abstract).
- Guo H, Wang L, Chen F, Liang D. 2014. Scientific big data and digital earth[J]. *Chinese Science Bulletin*, 59(35): 5066–5073.
- Guo H. 2018. Steps to the digital Silk Road[J]. *Nature*, 554(7690): 25–27.
- Guo Huadong. 2018. Scientific big data — A footstone of ational strategy for big data[J]. *Proceedings of the Chinese Academy of Sciences*, 33(8): 768–773 (in Chinese).
- Jonathan T O, Gerald A M, Bony S. 2011. Special online collection: dealing with data[J]. *Science*, 331(6018): 639–806.
- Kennedy M C, Hagan O. 2001. Bayesian calibration of computer models[J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3): 425–464.
- Li Guojie, Cheng Xueqi. 2012. Research status and scientific thinking of big data[J]. *Proceedings of the Chinese Academy of Sciences*, 27(6): 647–657 (in Chinese).
- Li Guojie. 2024. Big data and computing models[J]. *Big Data Research*, 10(1): 9–16 (in Chinese).
- Lü H, Ma H. 2019. Performance assessment and major trends in open government data research based on Web of Science data[J]. *Data Technologies and Applications*, 53(3): 286–303.
- Máková R, Hub M, Lnenicka M. 2018. Usability evaluation of open data portals[J]. *Aslib Journal of Information Management*, 70(3): 252–268.
- Marković D. 2015. Universalization and virtualization of the market economy, financial and real economy and the concept of global sustainable development[J]. *Zbornik radova Filozofskog fakulteta u Pristini*, 45(1): 123–142.
- Mayernik M S, Breseman K, Downs R R, Duerr R, Garretson A, Hou C S. 2020. Risk assessment for scientific data[J]. *Data Science Journal*, 19: 1–15.
- Sa Xu, Wang Jian, Fan Zhixuan, Liu Jianpin, Zhang Guilan, Xu Bo. 2020. Identification of scientific data quality evaluation indicators from the perspective of data journals peer review[J]. *Library and Information Service*, 64(17): 123–130.
- Schmidt C O, Struckmann S, Enzenbach C, Reineke A, Stausberg J, Damerow S, Huebner M, Schmidt B, Sauerbrei W, Richter A. 2021. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R[J]. *BMC Medical Research Methodology*, 21(1): 1–15.
- Song Jundian, Liu Fengyuan. 2018. A method and application study of data quality evaluation supporting[J]. *Computer Applications and Software*, 35(5): 328–333.
- Strong D M, Lee Y W, Wang R Y. 1997. Data quality in context[J]. *Communications of the ACM*, 40(5): 103–110.
- Tolle K M, Tansley D S W, Hey A J G. 2011. The fourth paradigm: Data-intensive scientific discovery [Point of View][J]. *Proceedings of the IEEE*, 99(8): 1334–1337.
- Tononi G, Sporns O. 2003. Measuring information integration[J]. *BMC Neuroence*, 4(1): 31.
- Wang Chao, Zhang Hui. 2020. Research and enlightenment of international open data evaluation index system in Europe and America[J]. *China Science & Technology Resources Review*, 52(5): 71–77.
- Wang Juanle, Chen Shenbin. 2006. Research on index and method of geosciences raster grid data quality evaluation[J]. *Science of Surveying and Mapping*, 31(5): 83–85, 82, 6.
- Wu Chonglong, Liu Gang. 2019. Big data and future development of geological science[J]. *Geological Bulletin of China*, 38(7): 1081–1088 (in Chinese with English abstract).
- Wu Chonglong, Liu Gang, Zhang Xialin, He Zhenwen, Zhang Zhiting. 2016. Discussion on geological science big data and its applications[J]. *Chinese Science Bulletin*, 61(16): 1797–1807 (in Chinese with English abstract).
- Zhai Mingguo, Yang Shufeng, Chen Ninghua, Chen Hanlin. 2018. Big data epoch: Challenges and opportunities for geology[J]. *Bulletin of Chinese Academy of Sciences*, 33(8): 825–831 (in Chinese with English abstract).
- Zhang H, Xiao J. 2020. Quality assessment framework for open government data[J]. *The Electronic Library*, 38(2): 209–222.
- Zhang Qi, Zhou Yongzhang. 2017. Big data will lead to a profound revolution in the field of geological science[J]. *Chinese Journal of Geology*, 52(3): 637–648 (in Chinese with English abstract).
- Zhang Z H, Zuo Q C. 2024. Definition and construction of data models of big data entity for prospecting prediction of ore concentration area [J/OL]. *Geology in China*. <https://link.cnki.net/urlid/11.1167.P.20240702.1551.011>.
- Zhang Zhihui, Zuo Qunchao. 2024. Development of application softwares for data models for prospecting prediction in mineralization focused areas based on model-driven and automatic evolution software theory[J]. *Mineral Exploration*, 15(8): 1478–1490 (in Chinese with English abstract).
- Zhang Zhihui, Zuo Qunchao. 2024. Framework of theories, methodologies and technologies system of ore-searching prognosis in mineralization focused areas[J]. *Geology in China*. <https://link.cnki.net/urlid/11.1167.P.20240523.0846.002>.
- Zhao Pengda. 2019. Characteristics and rational utilization of geological big data[J]. *Earth Science Frontiers*, 26(4): 1–5 (in

- Chinese with English abstract).
- Zhao Pangda, Chen Jianping, Zhang Shouting. 2003. The new development of "Three Components" quantitative mineral prediction [J]. *Earth Science Frontiers*, 10(2): 455–463 (in Chinese with English abstract).
- Zhao Pengda, Meng Xianguo. 1992. Quantification and geosciences[J]. *Earth Science*, (S1): 51–56 (in Chinese).
- Zhou Chenghu, Wang Hua, Wang Chengshan, Hou Zengqian, Zheng Zhiming, Shen Shuzhong, Cheng Qiuming, Feng Zhiqiang, Wang Xinbing, Lü Hairong. 2021. Geoscience knowledge graph in the big data era[J]. *Science China Earth Sciences*, 51(7): 1070–1079 (in Chinese).
- Zhou Yongzhang, Chen Chuan, Zhang Qi, Wang Gongwen, Xiao Fan, Shen Wenjie, Bian Jing, Wang Ya, Yang Wei, Jiao Shoutao, Liu Yanpeng, Han Feng. 2020. In troduction of tools for geological big data mining and their applications[J]. *Geotectonica et Metallogenia*, 44(2): 173–182 (in Chinese with English abstract).
- Zuiderwijk A, Pirannejad A, Susa I. 2021. Comparing open data benchmarks: Which metrics and methodologies determine countries' positions in the ranking lists?[J]. *Telematics and Informatics*, 62: 1–23.
- Zuo Qunchao, Yang Donglai, Song Yue, Ma Juan, Xiao Zhijian. 2013. The data quality control and technique of the mineral resources potential evaluation in China[J]. *Geology in China*, 40(4): 1314–1328 (in Chinese with English abstract).
- Zuo Qunchao, Pang Zhenshan, Xue Jianling, Chen Hui, Ye Tianzhu. 2022. An Application Guide to Theory, Methodology and Technical System of Ore-searching Prognosis in Mineralization Concentrating Area[M]. Beijing: Geological Publishing House (in Chinese).
- Zuo Qunchao. 2022. Methodology of Checking and Evaluation on Database Quality of Ore-searching Prognosis in Mineralization Concentrating Area[M]. Beijing: Geological Publishing House (in Chinese).
- Zuo Renguang, Peng Yong, Li Tong, Xiong Yihui. 2021. Challenges of geological prospecting big data mining and integration using deep learning algorithms[J]. *Earth Science*, 46(1): 350–358 (in Chinese with English abstract).
- 蔡丽华, 倪代川. 2021. 国内外科学数据评价研究综述[J]. *数字图书馆论坛*, (11): 65–72.
- 陈建平, 李靖, 谢帅, 刘静, 胡彬. 2017. 中国地质大数据研究现状[J]. *地质学刊*, 41(3): 353–366.
- 成秋明. 2021. 什么是数学地球科学及其前沿领域?[J]. *地学前缘*, 28(3): 6–25.
- 郭华东. 2018. 科学大数据—国家大数据战略的基石[J]. *中国科学院院刊*, 33(8): 768–773.
- 李国杰, 程学旗. 2012. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. *中国科学院院刊*, 27(6): 647–657.
- 李国杰. 2024. 大数据与计算模型[J]. *大数据*, 10(1): 9–16.
- 撒旭, 王健, 范智萱, 刘建平, 张贵兰, 徐波. 2020. 数据期刊同行评议视角下科学数据质量评价指标识别[J]. *图书情报工作*, 64(17): 123–130.
- 宋俊典, 刘丰源. 2018. 一种支持数据质量评价的方法与应用研究. *计算机应用与软件*, 35(5): 328–333.
- 王超, 张辉. 2020. 欧美开放数据评估指标体系调查研究及启示[J]. *中国科技资源导刊*, 52(5): 71–77.
- 王卷乐, 陈沈斌. 2006. 地质栅格数据质量评价指标与方法[J]. *测绘科学*, 31(5): 83–85, 82, 6.
- 吴冲龙, 刘刚, 张夏林, 何珍文, 张志庭. 2016. 地质科学大数据及其利用的若干问题探讨[J]. *科学通报*, 61(16): 1797–1807.
- 吴冲龙, 刘刚. 2019. 大数据与地质学的未来发展[J]. *地质通报*, 38(7): 1081–1088.
- 翟明国, 杨树锋, 陈宁华, 陈汉林. 2018. 大数据时代: 地质学的挑战与机遇[J]. *中国科学院院刊*, 33(8): 825–831.
- 张旗, 周永章. 2017. 大数据正在引发地球科学领域一场深刻的革命——《地质科学》2017 年大数据专题代序[J]. *地质科学*, 52(3): 637–648.
- 张志辉, 左群超. 2024a. 基于模型驱动和自动演进理论的矿集区找矿预测数据模型应用软件开发[J]. *矿产勘查*, 15(8): 1478–1490.
- 张志辉, 左群超. 2024b. 矿集区找矿预测大数据实体数据模型的界定与构建[J/OL]. *中国地质*. <https://link.cnki.net/urlid/11.1167.P.20240702.1551.011>.
- 张志辉, 左群超. 2024c. 矿集区找矿预测理论方法技术体系框架[J/OL]. *中国地质*. <https://link.cnki.net/urlid/11.1167.P.20240523.0846.002>.
- 赵鹏大, 陈建平, 张寿庭. 2003. “三联式”成矿预测新进展[J]. *地学前缘*, 10(2): 455–463.
- 赵鹏大, 孟宪国. 1992. 地质学的量化问题[J]. *地球科学*, (S1): 51–56.
- 赵鹏大. 2019. 地质大数据特点及其合理开发利用[J]. *地学前缘*, 26(4): 1–5.
- 周成虎, 王华, 王成善, 侯增谦, 郑志明, 沈树忠, 成秋明, 冯志强, 王新兵. 2021. 大数据时代的地质知识图谱研究[J]. *中国科学(地球科学)*, 51(7): 1070–1079.
- 周永章, 陈川, 张旗, 王功文, 肖凡, 沈文杰, 卞静, 王亚, 杨威, 焦守涛, 刘艳鹏, 韩枫. 2020. 地质大数据分析的若干工具与应用[J]. *大地构造与成矿学*, 44(2): 173–182.
- 左群超, 杨东来, 宋越, 马娟, 肖志坚. 2013. 中国矿产资源潜力评价成果数据质量控制及方法技术[J]. *中国地质*, 40(4): 1314–1328.
- 左群超, 庞振山, 薛建玲, 陈辉, 叶天竺. 2022. 矿集区找矿预测理论方法技术体系应用导论[M]. 北京: 地质出版社.
- 左群超. 2022. 矿集区找矿预测数据库质量检查与评价方法[M]. 北京: 地质出版社.
- 左仁广, 彭勇, 李童, 熊义辉. 2021. 基于深度学习的地质找矿大数据挖掘与集成的挑战[J]. *地球科学*, 46(1): 350–358.

附中文参考文献