

# 重要地质钻孔数据库质量检查方法研究

梁银平, 李杰, 王斌, 岳鹏

(国土资源实物地质资料中心, 河北三河 065201)

**摘要:** 重要地质钻孔数据库数据量大、信息丰富,数据质量的优劣直接影响数据库建设的成效以及后续服务利用水平。本文从现代数据质量概念入手,提出重要地质钻孔数据库数据质量控制要求,阐述了数据质量检查内容并提出一些能提高效率和准确度的实用技术和方法,以期通过交流,共同提高钻孔数据库质量检查的技术水平。

**关键词:** 地质钻孔数据库; 数据质量; 质量检查方法

中图分类号: P628 文献标识码: A 文章编号: 1004-4051(2015)S1-0072-03

## Methods to checking data-quality of important geological drilling database

LIANG Yin-ping, LI Jie, WANG Bin, YUE Peng

(Cores and Samples Centre of Land and Resources, China Geological Survey, Sanhe 065201, China)

**Abstract:** The important geological drilling database involves abundant information, and its data-quality may directly determine construction result and decision-making. Here, the data-quality requirements were put forward from modern data quality conception, the contents of the data-quality check were described and some data quality check methods were suggested to improve the data-quality check.

**Key words:** geological drilling database; data quality; data-quality check method

全国重要地质钻孔数据库涵盖了新中国成立以来我国开展的大规模地质勘查和地质科学研究工作所获取的丰富地质钻孔资料,蕴藏着巨大的潜在价值,可为基础地质调查、矿产资源评价、国土资源合理开发利用、国民经济建设、制定区域规划、保护人类赖以生存的地质环境及其他相关学科研究提供有效的基础地质资料,对于保障我国矿产资源持续供应和可持续发展也具有重要的数据支撑作用。在当前钻孔数据库建设中,数据质量的优劣直接影响数据库建设的成效,也会成为决定基于该数据库的决策水平的重要因素。明确数据库数据质量要求,建立完善的数据质量控制体系,并对数据质量控制中存在的问题提出一套正确合理的排查办法非常必要。

## 1 数据质量及要求

### 1.1 数据质量

数据质量即数据的可靠性。全国重要地质钻孔数据库建库人员众多、涉及资料量多且时间跨度大。在建库过程中,由于钻孔数据采集软件功能限制、数据录入人员操作失误或专业知识欠缺、原始数据质量限制、资料重复利用及多地保管等多方面的因素,会导致各种数据质量问题。从入库数据内容来看,地质钻孔数据库数据质量问题主要分为属性项结构化数据和图表非结构化数据质量检查。属性项结构化数据质量检查主要包括重要属性项填写错误或填写不规范、钻孔重复录入、相关备注填写不完善等。图表非结构化数据质量问题主要包括入库图件错误、扫描图件不清晰、图件整饰不规范、入库图表不完整等。

### 1.2 数据质量要求

现代数据质量概念注重从用户的角度衡量数据质量,强调用户对数据的满意程度,一般以完整性、规范性、准确性、一致性、及时性等指标来描述数据质量<sup>[1]</sup>,同时根据具体的数据评估需求对数据质量评估指标进行取舍。援引现代数据质量概念对数据

收稿日期: 2015-05-04

基金项目: 中国地质调查局项目全国重要地质钻孔数据库平台建设  
与试点应用项目资助(编号: 1212011220350)

作者简介: 梁银平(1982-),女,工程师,主要从事实物地质资料综合研究。E-mail: yinpingliang@163.com。

质量的评价指标,我们对全国重要地质钻孔数据库提出完整性、规范性、准确性、一致性、唯一性、客观性等数据质量要求。

完整性关注的重点是数据库中是否存在多余数据、重复记录或缺失记录、字段的情况,不关注数据的规范性和准确度。

规范性要求数据库中数据项的格式及填写符合相关规范要求,图件的扫描质量及图面整饰符合《图文地质资料扫描数字化规范(试行)》要求。

准确性指数据的准确度和可信赖度,考量的是数据的客观性,要求入库数据项与其对应的原始地质资料实际情况相符。

一致性要求某一数据项在数据库中出现多次的,涉及该字段的数据必须完全一致。例如,钻孔信息中的终孔深度应当与柱状图中的终孔深度及图内的累积深度一致。

唯一性要求数据库中不存在重复保管单位、项目及钻孔记录。

## 2 数据质量检查的内容

质量检查是数据质量控制的重要环节。在质量检查过程中,首先要对数据质量在整体上进行把握,主要从代表保管单位、项目和钻孔的文件夹数分别与实际入库记录条数的对应性以及入库图件总数与图件设定数总和的对应性两方面对数据库的完整性进行检查。其次是对钻孔资料保管单位信息、钻孔所属项目信息和钻孔信息从完整性、规范性、准确性、一致性、唯一性等几方面作具体检查。

完整性主要检查保管单位信息中的行政区划代码、组织机构代码、保管单位名称、最高地勘资质等级、所属行业部门、通讯地址、邮政编码、联系人姓名、联系电话等,项目信息中的组织机构代码/保管单位名称、项目名称、资料名称、成果资料档号、项目结束时间、密级等和钻孔信息中的原始资料档号、钻孔编号、钻孔类型、坐标系、孔口高程、终孔深度、终孔日期、施工单位、测井报告、原始地质记录表、钻孔岩心等必填数据项是否为空;工程布置图、勘探线剖面图、钻孔柱状图(以下简称三图)和样品分析结果表(以下简称一表)入库文件数与设定数是否一致,图表文件有无缺失;分幅扫描的三图是否完整及能否正确衔接,样品分析结果表有无缺页、备注是否齐全。其中,备注完整性的检查主要涉及以下内容:①缺失真实坐标的钻孔,建库人员通过矿区平面地形地质图获取钻孔坐标或者将公开网络查找的矿区中心点坐标赋予此项目下的所有钻孔坐标的;②老旧资料,经查找确无原始资料档号和省馆资料档号,用

成果资料档号代替的;③历史老旧资料,按要求扫描后达不到质量检查的清晰度要求的;④对于部分历史资料,经查实省馆及地勘单位确实没有“三图一表”相关图件资料的,是否在备注中对具体情况进行相关说明。

规范性检查主要包括保管单位组织机构代码填写是否规范;项目的工作程度、比例尺、主要矿种等的填写是否规范,工程布置图、勘探线剖面图和样品分析结果表的文件格式、扫描分辨率是否符合《全国重要地质钻孔数据库建设工作技术要求》,资料密级的确定是否符合定密规范、栅格文件能否达到倾斜度小于1%或者图件与原图的对角线误差和边长误差小于±0.1%的图像整饰要求等。

准确性主要检查行政区划代码填写是否正确,组织机构代码填写是否正确,入库图表是否正确,钻孔坐标是否正确,坐标系与孔口坐标XY是否匹配等。

一致性检查主要包括保管单位组织结构代码与保管单位名称是否一致,孔口坐标和高程、孔深与钻孔柱状图中的信息是否一致,高程、坐标等非来源于柱状图或与柱状图不符的是否对数据来源进行备注。

唯一性主要检查数据库中是否存在重复保管单位、项目及钻孔记录。

值得提出的是,在数据库建设过程中,由于建库人员知识背景限制,存在将工程布置图外其他图件录入工程布置图的情况。因此,图表正确性的检查不仅要检查入库的图表是否属于本项目/钻孔的资料,还要检查实际入库的图表是否为本数据项要求的图表。

## 3 数据质量检查方法

根据检查内容的不同,全国重要地质钻孔数据库数据质量检查方式主要包括计算机软件自动检查、人工判断检查和人机交互的计算机辅助检查三类。

### 3.1 计算机软件自动检查

全国重要地质钻孔数据库数据主要借助地质钻孔数据采集系统(以下简称采集系统)采集入库。采集系统对数据项是否必填、数据条目的格式及长度等设置了相应的校验规则,在数据采集入库时可以实时对必填项的完整性以及组织机构代码、工作程度、比例尺、主要矿种等规范性进行检查。如填写内容与该校验规则不符,系统会出现报错提示。根据报错提示,建库人员对出错项及时修改完善,从而在建库最初环节对数据质量进行把关。同时,采集系统还对保管单位和项目分别设置了数据滤重处理逻辑,在数据汇总合并时进行滤重检查,避免了因多个

建库人员录入数据造成的数据重复。

### 3.2 人工判断检查

人工检查是数据检查人员利用自身的专业知识、工作经验或参阅相关文件、规范要求,对入库各类数据对照数据采集所用的原始资料进行核对,从而对数据质量进行全面检查。

数据检查人员通过检查从代表保管单位、项目和钻孔的文件夹数分别与实际入库记录条数的对应性,可以对数据的完整性在整体上进行把握。通过查阅《中华人民共和国行政区划代码》《中央党政机关人民团体及其他机构名称代码》可以判断行政区划代码填写是否正确,组织结构代码与保管单位名称是否对应。对照《全国重要地质钻孔数据库建设工作技术要求》,判断“三图一表”格式、图像整饰是否符合规范。

数据检查人员通过仔细阅读图件信息,可以检查钻孔坐标、终孔深度、高程与柱状图上的信息是否一致;对照入库图表名称与项目名称致,柱状图上钻孔编号与入库钻孔编号是否一致性,可以判断入库图件是否正确;对分幅扫描的图件相邻图件边缘的线条、色彩、地质要素能否完全衔接以及柱状图上的层号、钻孔回次是否连续,判断入库图件有无缺失。

我国的经度范围西起 $73^{\circ}$ 东至 $135^{\circ}$ ,可分成6度带11个(13~23号带),3度带22个(24~45号带),检查人员通过检查坐标系与坐标中的带号是否对应,可以判断坐标系或地理坐标填写是否正确。

### 3.3 人机交互检查

人机交互检查是质量检查人员在借助采集系统、Excel和Access软件、Mapgis软件,以及计算机系统某些自带功能等对数据进行一定处理后,再通过人工检查对数据质量做出判断的过程。主要检查图件的扫描质量,钻孔记录的唯一性,钻孔坐标的正确性等。

通过操作系统搜索筛选功能,查找出备查数据中的所有图件,并对其按“水平分辨率”升序排列。对于分辨率低于300dpi的图件采用人工检查方式,判断原图和扫描图在1:1情况下,各类线条是否连续,图面中的小数点是否清晰。

钻孔编号、孔口坐标、高程和终孔深度对于钻孔来说是非常重要的信息。对重复钻孔的排查首先是将Access数据库的钻孔信息导出为.xlsx格式,然后通过Excel软件“条件格式”功能,对“孔口坐标X”列设置“仅对唯一值或重复值设置格式”规则后,对所有记录进行重新排序,找出坐标相同的钻孔,结合原始资料档号、孔口高程和终孔深度等重要字段以及钻孔柱状图等信息,判断坐标相同的钻孔是否为重复记录。

利用上述方法对河北省2013年重要地质钻孔数据进行检查,发现存在钻孔编号、孔口坐标和高程相同的两个钻孔,但其终孔深度不同,在进一步对照柱状图后,确定两个钻孔不是重复钻孔,且其中一个钻孔编号填写错误。

钻孔坐标正确性检查可以通过采集系统筛选出某一项目的钻孔,检查该项目所有钻孔的坐标否在相近范围,对钻孔坐标的正确性进行初步判断。但最有效的方法是按钻孔坐标分属的不同投影带分别投影成钻孔点位图,将点位图与对应坐标系的地理底图套合,详细检查钻孔是否沿道路、沿河、岸带等地貌地物分布(沿道路、沿河、岸带施工的钻孔),是否按勘探线或勘探网度规则排列(矿产勘查钻孔)以及对照纸质工程布置图,判断有无离群钻孔点等来判断钻孔位置是否合理、正确。

## 4 结语

全国重要地质钻孔数据库建设是一项十分复杂的工作,具有信息量大、建库人员众多、投入高等特点。数据的质量控制是数据库建设中极其关键的环节,明确数据质量要求和钻孔数据质量检查内容,制定详细的质量检查操作细则是保证入库数据客观真实、完整可靠的有力措施。只有建立高质量的数据库,才能确保全国重要地质钻孔数据库数据的正确性和规范性,并最终更好地为钻孔资料社会化服务利用提供数据支撑保障。

### 参考文献

- [1] 张华义,刘新,罗涛.数字油田建设中的数据质量控制方法研究[J].天然气勘探与开发,2015,38(1):88-93.