

浅析国有企业大数据审计的风险识别与控制问题

徐永阳

响水县国有企业和数据审计中心，江苏 盐城 224600

DOI:10.61369/ER.2025050012

摘要：当前大数据的广泛应用为国有企业内审带来了一定的优势，他使得内部审计成本更低、速度更快、数据更准确，同时也使得内审与企业管理的联系更加紧密。但是大数据在为建筑类企业内审带来便利的同时也带来了更多挑战，在此过程中由于数据异构性、算法黑箱以及业务逻辑和技术规则的脱节，造成审计风险具有隐蔽性和传导性的特征。因为多源数据采集失真、分析模型构建逻辑错误、疑点核查不闭环等缘故，须要搭建起风险识别体系。加强源头数据标准化校验、完善分析模型参数阈值设置、打通跨系统全链条数据比对路径，可防止出现技术性误判，提高审计结论的客观性、准确性，保障国有资产监督效能持续提升。

关键词：大数据审计；风险识别；控制机制；数据校验

Analysis of Risk Identification and Control Issues in Big Data Audit of State-owned Enterprises

Xu Yongyang

Xiangshui County State owned Enterprise and Data Audit Center, Yancheng, Jiangsu 224600

Abstract : The widespread application of big data has brought certain advantages to the internal audit of state-owned enterprises. It makes the cost of internal audit lower, faster, and the data more accurate. At the same time, it also makes the connection between internal audit and enterprise management closer. However, while big data brings convenience to internal audits of construction enterprises, it also brings more challenges. In this process, due to data heterogeneity, algorithmic black boxes, and disconnection between business logic and technical rules, audit risks have the characteristics of concealment and transmission. Due to distortions in multi-source data collection, logical errors in model construction, and incomplete verification of suspicious points, it is necessary to establish a risk identification system. Strengthening the standardization and verification of source data, improving the threshold setting of analysis model parameters, and establishing a cross system and full chain data comparison path can prevent technical misjudgments, improve the objectivity and accuracy of audit conclusions, and ensure the continuous improvement of the efficiency of state-owned asset supervision.

Keywords : big data audit; risk identification; control mechanism; data validation

随着数字化时代的到来，大数据被应用于各行各业的生产和办公中，各公司的办公系统也越来越完善，庞大的数据库为公司各部门的办公提供了便捷，大大缩短了各部门间的数据沟通时间^[1]。在扩大审计范围和加深审计程度的时候，也重新塑造了审计风险的产生机理和表现形式。面对体量庞大、来源多样、结构各异的电子数据，传统的依靠人工核查的经验很难直接复用，数据本身的真实性、算法逻辑的匹配性、技术路径的合法性成了新的风险爆发点。不依赖主观经验，依靠技术手段精准地发现数据链条上的异常波动并控制审计全流程，已经成了目前深化国有企业监督职能必须解决的紧迫问题。

一、大数据审计环境下的业务数据特征

国有企业业务规模庞大、涉及的行业领域多，造成审计所面对的数据环境十分复杂。数据不再只来源于财务系统，而是分散到采购、生产、销售、库存等各个业务子系统中，形成多源异构的存储状态。这些数据的体量通常达到TB级甚至PB级，更新

速度非常快，有些关键业务数据达到毫秒级刷新。数据类型包括结构化的数据库记录、半结构化的日志文件、非结构化的合同文本和影像资料。在如此高维度的数据空间里，仍然存在着单一的数据孤岛现象，各个系统的接口标准不统一，字段定义存在争议，造成数据在跨系统流转过程中很容易出现丢失或者变形的情况^[2]。高并发、多维度的业务数据特征客观上加大了审计取证的

作者简介：徐永阳（1990.01—），男，汉族，江苏灌南人，硕士研究生，职称：中级会计师、审计师，研究方向：审计、会计。

难度，且企业面临着更加复杂和多变的风险和挑战，企业要进一步加强风险防控机制的建设，确保企业在复杂多变的市场环境中稳健运营。

二、数据采集与预处理阶段的风险识别

(一) 源头数据真实性校验的缺失

由于被审计单位信息系统架构各不相同，直接从数据库底层获取的原始数据没有经过严格的整体性检查。部分业务系统在设计之初就缺少了严密的逻辑约束，可以录入空值或者非法字符，造成关键字段信息缺失。如果不进行针对性的校验，在采集时不加区分地把残缺数据直接输入分析库。供应商的统一社会信用代码位数不统一，金额字段中混入文本符号等都会使后续的关联分析失效。源头上的数据失真具有很强的隐蔽性，一般只有在审计模型运行出错或者得出荒谬结论的时候才被发现，造成审计资源的极大浪费。

(二) 非结构化数据转换的偏差

国有企业存在着大量的非结构化数据，工程图纸、会议纪要、合同扫描件等都属于此类。将这些非结构化信息转换成计算机可以处理的结构化数据时，要依靠OCR识别技术或者人工补录。转换的时候，因为图像清晰度，排版格式多种类等缘故，很容易出现识别失误的情况。例如把合同金额中的小数点识别成逗号，把重要日期的年份识别错误。如果没有经过二次人工复核或者逻辑规则校验，那么就会造成转换偏差，从而使得生成的审计中间表包含错误信息，进而影响到审计证据的证明力。

(三) 数据传输接口的对接壁垒

审计系统同被审计单位业务系统间的数据传输，一般会采用标准接口或者中间库的方式。但是由于系统开发商的不同，数据字典的定义标准存在差别。当通过接口传输数据的时候，如果没有统一的字段映射规则，很容易造成数据截断、乱码或者字段错位等问题。尤其是对于跨地域、跨板块的集团型企业数据，不同的分子编码规则不统一，在采集的时候没有进行统一的转码清洗，会导致数据在物理上是完整的，但是在逻辑上不能进行有效的横向对比，审计分析的范围会非预期地缩小。

(四) 历史数据清洗逻辑的失真

为了满足审计分析的要求，需要对收集到的原始数据进行清洗，剔除无效或者多余的信息。清洗规则设置过严，会把一些可疑数据当作有效数据删除。在去除重复记录的时候，没有考虑到高频交易本身就是一种业务异常，因此将高频交易当作系统冗余数据去除。反之，清洗规则太宽松，保留了很多测试数据、回滚数据，就会干扰正常的分析结果。清洗逻辑黑箱化造成审计人员很难察觉数据在预处理环节是否已经被“整形”，进而产生数据可用性风险。

(五) 采集范围与业务脱节的风险

大数据审计要求全覆盖，但是由于权限或者技术上的限制，采集的范围一般只限于核心业务系统，而忽略了外围辅助系统或者线下台账。只采集了ERP系统中的采购订单，却没有采集招投

标系统的过程文件，不能对采购流程的合规性进行全面的核查。这种采集范围的片面性会造成审计视野的盲区，使得一些通过体外循环完成的违规操作无法在数据分析中发现，造成审计结论的片面性，不能真实反映企业的全貌。

三、审计模型构建与分析环节的控制

(一) 算法逻辑与业务规则的冲突

在大数据环境下，建立精准的风险模型是实现企业动态监控的重要手段。审计分析模型是大数据审计的核心，审计分析模型的基础是把业务规则转化为计算机能执行的代码。但是技术人员对于复杂的业务场景理解不深入，造成算法逻辑同实际业务规则存在偏差。核查资金支付合规性的算法设置只有一个金额阈值，而没有考虑特定时期、特定合同条款下的支付例外规则。会造成大量的伪疑点跑出，增加核查的工作量，并且会遮蔽真实的违规线索。必须创建起业务专家和技术人员深度合作的机制，保证每一行代码都完全映射出业务逻辑。

(二) 模型参数设定的主观性偏差

在建立预警模型的时候，阈值的确定常常依靠审计人员的经验来判断，带有较大的主观性。阈值过高会漏掉很多违规线索，阈值过低会造成大量误报，因为参数设置不合理而造成审计资源的浪费。不同板块、不同规模的业务单元如果用同样的参数标准来衡量，那么分析的结果就会失真。下表1为在某次物资采购审计中，各个业务板块模型阈值设定与疑点检出率对比表，可以直观的看出参数偏移的影响。

表1 业务板块模型阈值设定与疑点检出率对比表

业务板块类别	设定异常阈值 (%)	样本总量 (万条)	检出疑点数量	伪疑点占比 (%)	实际违规确认数
工程建设板块	>15	45.2	128	82.5	22
物资贸易板块	>5	89.6	4560	96.2	173
装备制造板块	>10	32.1	85	45.8	46
能源化工板块	>8	56.4	310	68.4	98

(三) 多维数据关联分析的伪相关

大数据分析的一个重要手段就是多维关联，通过挖掘不同数据表之间的相关性来发现疑点。但是数据之间的统计相关性并不等同于因果关系。在跨部门、跨系统的关联分析中很容易出现伪相关。将员工考勤记录与报销记录关联起来，仅因为时间上巧合就判定为虚假报销，而忽略了远程办公或者特殊出差的情况。这种缺少业务背景支持的机械化联系，会产生很多没有实质证据支撑的疑点，大大降低审计分析的准确性，需要更多的角度进行交叉检验^[3]。

四、全流程审计数据链条的闭环控制

(一) 电子数据全生命周期留痕

为了保证审计结论具有抗辩能力，对电子数据从采集、清

洗、转换到分析的全过程进行留痕控制。用哈希算法这些技术手段，给每个处理环节的数据状态做数字签名，产生不能篡改的操作记录。防止数据被恶意篡改，也为了在审计复核阶段还原数据处理的每一个步骤。经由创建涵盖全部生命阶段的留痕体系，任一分析成果都可以追溯至最初的记载数据，就技术层面而言，这就保证了审计证据链条的完备和可信度，从而规避由于数据处理环节缺乏透明度所造成的合规性问题。

（二）跨系统数据比对的勾稽关系

单一系统的数据容易被粉饰，构建跨系统数据勾稽关系是发现虚假信息的一种有效方法。应重点创建财务系统同业务系统、内部数据同外部数据之间的比对路径。将财务账面记录的运费支出同物流系统里的车辆轨迹、过磅记录加以比对，把采购入库记录同生产领用记录、废料处理记录实施总量平衡分析。经过这样多维度交叉比对之后，可以找出单个系统里不容易察觉的逻辑矛盾之处，依靠数据间的制衡关系去定位异常情况，从而明显增强识别风险的穿透能力。

（三）数字化审计底稿的溯源机制

传统的审计底稿主要是纸质的或者静态的电子文档，不能承载大数据分析的复杂逻辑。创建数字化审计底稿，把分析代码，中间过程表以及疑点核查记录打包到同一个数字化容器里。底稿上既写明了“发现了什么”，又写明了“怎样发现的”。建立点击即穿透的溯源机制，复核人员可以查看某一个疑点背后算法逻辑以及原始数据支持。这样会大大提高审计工作的透明度，使审计

人员在分析时必须使用严谨的逻辑思维，减少由于随意操作而产生的风险。

（四）总体分析与个体验证的结合

大数据审计虽提倡全样本审计，但也不能脱离个体的验证。控制风险最好的办法就是坚持“总体分析、发现疑点、分散核查、系统研究”的路子。在发现总体异常趋势之后，必须选取典型样本做穿行测试，验证数据反映的问题在物理世界中是否真实存在。借助个体验证来调节总体分析的偏差，把个体验证的结论反过来修正总体分析的逻辑。宏观和微观相结合的控制手段可以防止陷入纯粹的数字游戏，保证审计发现的问题有扎实的业务事实依据^[4]。

五、结语

国有企业大数据审计是项系统工程，风险存在于数据流转的每个环节。识别并控制风险不是去追求技术的极端复杂，而是建立一个逻辑严密、流程规范的校验体系。对准采集源头、动态调整分析模型、闭环追踪数据链条，就能较好地处理好技术与业务脱节引发的次生风险。随着数字化程度的加深，审计工作必须不断加强对数据规律的深刻认识，用标准化的控制来驾驭大量的数据，在日益复杂的经济活动中，准确地抓住违规线索，客观地评价经营效能，切实履行好国有资产看门人职责。

参考文献

- [1] 王雅丽.大数据环境下财务风险管理及审计策略分析 [J].企业改革与管理,2024,(13):152-154.
- [2] 毛亚东.大数据背景下企业审计的风险与应对措施 [J].财经界,2024,(36):141-143.
- [3] 刘欢,王丽云.大数据在企业审计风险识别及应对中的有效应用 [J].市场瞭望,2024,(19):109-111.
- [4] 张力夫.新形势下审计风险识别策略研究 [J].河北企业,2024,(02):122-124.