

审计全覆盖背景下大数据审计的特征及发展

夏江华¹

(嘉兴南湖学院 商学系, 浙江 嘉兴 314001)

【摘要】: 从审计全覆盖对审计工作效率的要求入手, 总结了大数据审计的三个特征, 围绕这三个特征的结合点, 即内外部多维数据相关分析, 进一步讨论大数据审计开展过程中存在的数据采集和人才短缺问题, 提出了依靠审计人员、技术人员有效协作, 分阶段提升、推进大数据审计的思路。

【关键词】: 审计全覆盖 大数据审计 特征 发展

【中图分类号】: F239. 1 **【文献标志码】:** A **【文章编号】:** 1671—3079(2021)02—0124—07

大数据时代的来临必然会带来审计环境和模式深刻的变化, 这已成为审计领域的共识。在审计全覆盖的要求下, 大数据审计在国家审计领域发展迅速,^[1]相关研究方兴未艾。从现有文献看, 一是着眼于实践经验总结, 聚焦工作流程、模型构建、数据处理、技术方法等研究,^[2-3]二是着眼于未来发展趋势的分析, 聚焦于非结构化数据、机器学习、智能化审计等前沿概念的探讨。^[4]前者缺乏对大数据审计特征的理论探讨, 不利于整体推进大数据审计的发展; 后者忽略了大数据审计发展的路径, 致使审计实践与发展目标脱节。因此, 本文基于审计全覆盖背景, 探究大数据审计的特征, 进而讨论制约大数据审计发展的关键问题, 为形成大数据审计的发展路径提供建议。

一、审计全覆盖背景下大数据审计的特征

在大数据审计发展过程中, 存在着不少需要辨析的疑问。现从审计全覆盖对工作效率的要求使审计工作中心环节发生转移、工作中心的转移要求向基于多维数据相关分析的“总体审计”转变、多维数据高效采集和处理必然要求向“联网审计”发展 3 个方面探究大数据审计的特征。

(一) 审计全覆盖对工作效率的要求使审计工作中心环节发生转移

2014 年, 李克强同志明确提出, “凡使用财政资金的单位和项目, 都要接受审计监督”。¹审计全覆盖的提出, 对审计工作效率提出了极高的要求, 这是大数据审计在国家审计领域迅速发展的重要原因。传统审计模式下, 现场实施阶段是审计工作的中心环节, 在挖掘审计线索时, 数据分析方法的地位和重要性与检查、观察、询问、外部调查等其他审计方法没有明显差异, 并没有成为审计工作的主导环节。而受检查、观察、询问、外部调查这些非信息化方法工作效率的制约, 国家审计机关以往只能将一部分被审计对象纳入年度审计范围, 通过几年一轮的方式来实现审计全覆盖。

审计信息化是提高工作效率的有效途径,^[5]如不改变以现场实施为中心环节、数据分析方法与非信息化方法并行的工作模式, 就不能适应审计覆盖面提升所带来的数据量急剧增长的现状。大数据审计模式下, 工作效率的提高不仅在于信息化带来的数据处理速度的提升, 更重要的是在于工作模式的变化。非现场数据采集、整理和分析成为大数据审计工作的中心环节, 数据分析

¹作者简介: 夏江华(1974-), 男, 浙江海宁人, 嘉兴南湖学院商学系讲师, 研究方向为审计理论与实务。

基金项目: 浙江省审计厅重点课题(201901012)

成为发现审计疑点、线索的主导方式，而其他非信息化审计方法则主要起到验证、延伸、反馈等作用，如此才能大大提升工作的效率并扩大审计覆盖面。因此，大数据审计与传统审计的数据分析方法最大的区别，在于数据分析环节是否主导了审计工作。如果只是在审计工作过程中引入一些大数据分析技术或创新一些信息展示方式，是不能称之为大数据审计的。

(二)工作中心的转移要求向基于多维数据相关分析的“总体审计”转变

由于工作效率的原因，传统审计模式下的检查、询问、外部调查等方法只能采取抽样的方法开展。有研究指出，大数据审计将从“抽样审计”向“总体审计”“全量审计”转变。^[6-7]但如何理解“总体审计”或“全量审计”，仅仅是指对被审计单位财务、业务系统的全部数据进行检索和分析吗？事实上，借助计算机审计工具，传统审计模式下就能实现各系统全部数据的检索以及系统间全部数据的比对。大数据审计模式下的“总体审计”或“全量审计”与传统审计模式下对全部数据分析的区别，在于观察视野和取证思路上的不同。

传统审计模式下，不论是检查、询问还是数据分析方法，审计观察的视野主要集中于单位内部财务、业务系统。外部调查方法受到工作效率的影响，能纳入审计的外部数据极其有限。不论是检查、询问还是数据分析方法，其工作思路往往都是先从单一的财务系统入手开展，进而依据于业务流程因果关系，前后延伸到其他业务系统以获取审计线索和证据，如图 1 所示。同时，传统审计模式下，依靠综合运用检查、询问等多种方法能弥补不同方法各自的不足，从而形成比较完善的审计取证链，但缺点在于不能适应审计全覆盖对工作效率的要求。

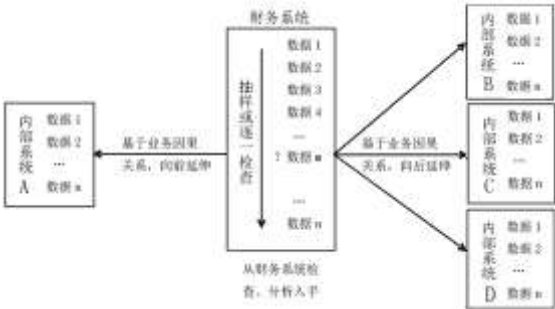


图 1 局限于内部系统数据，基于业务因果关系的传统审计取证思路

为提高工作效率，数据分析将成为发现审计疑点、线索的主导方式。但如果数据分析的视野仍局限于单位内部财务、业务系统，仅仅依靠数据分析而缺乏其他方法对物理证据、行为证据进行采集，是难以形成完善的审计证据链条的。因此，审计人员必须将审计视野从单位内部延伸到单位外部，通过内外部相关数据的比对，形成新的审计证据链。因此，大数据“总体审计”“全量审计”的取证思路，是从基于单位内部财务、业务系统间的“因果关系”分析向跨部门、跨层级、跨单位的多维数据“相关关系”分析转变，^[8]其工作思路如图 2 所示。

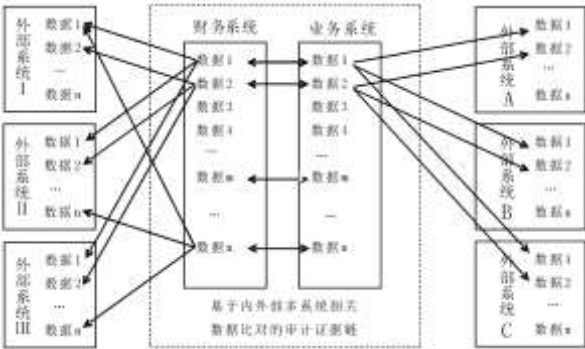


图 2 基于内外部多系统相关数据比对的“总体审计”取证思路

要注意的是，现阶段“总体审计”和“全量审计”作用的发挥受到数据技术和数据供给的制约，过多着眼于其发展远景不利于推进大数据审计的发展。比如，有研究指出，大数据审计应该纳入文本、图像和视频等非结构化数据。^[4]但从现有的数据技术看，以图片、影像等方式存储的非结构化数据缺乏有效的技术手段进行转化处理，要大量纳入大数据分析还需要长期探索；而以文档方式存储的非结构化数据的处理，目前主要采用的是关键字检索、标签云等工具，只能提供简单的信息而不能深入。因此，当前大数据审计工作的重点仍应该放在多维结构化数据相关关系的挖掘上。又如，有研究认为，基于内外部“全量数据”能形成无遗漏的审计证据链条，因而大数据审计将全面取代传统审计，计算机将全面替代审计人员。^[4]该种观点未免过于乐观，因隐私、保密、安全等原因要取得“全量数据”几乎是难以实现的，即使在国家审计工作中要想获取相对独立、垂直管理的如银行、税务等系统数据，也必须有明确理由并获得批准才能查询获取部分而非全部数据。因此，在可预见的较长时间内，大数据审计仍将依赖审计人员的职业能力以补充、延伸查证工作。

（三）多维数据高效采集和处理必然要求向“联网审计”发展

从发展的角度看，以多维数据相关分析为基础的大数据审计要求建立高效的数据采集、分析机制，这与“联网审计”侧重强调通过网络自动采集、存储数据，并依赖审计模型自动分析异常情况以提示审计线索相一致。两者的共同点在于：一是数据采集和分析成为主导其他审计活动的前置中心环节；二是两者都需要借助一定的审计模型自动开展数据分析以提升工作效率，也为“持续审计”的开展奠定基础。当前，大数据审计实践受到数据网络建设、数据可得性等方面的制约，数据批量导入方式仍普遍存在，但“联网审计”必然是未来的发展趋势。

需要注意的是，不论是“联网审计”还是“大数据审计”，目前来看，受到模型数据视野、分析逻辑和审计人员经验的限制，还难以应对被审计单位出现的新问题、新风险。因此，审计全覆盖要求下的大数据审计绝不是简单的“大数据处理技术+分析模型+联网审计”，更重要的是将尽可能多的相关维度数据纳入到审计视野，通过不断地发现新关系、新特征，为审计工作提供新思路、新线索。

综上所述，审计全覆盖背景下的大数据审计应以数据采集和分析为中心环节，主导其他审计活动；通过内、外多维数据“相关关系”的分析，构建起不同于传统审计模式的审计线索、证据链条，并将“联网审计”作为发展的主要目标。

二、大数据审计发展需解决的问题

大数据审计模式下，多维数据的纳入、获取和甄别是推进大数据审计发展的关键。同时，大数据审计必然对审计人员的知识和技能提出新的要求。大量的研究表明，复合型审计人才的短缺已成为制约大数据审计发展的一个瓶颈。^[9]因此，如何解决复合型审计人才短缺问题，也是推进大数据审计发展必须思考的。

（一）规范化、标准化采集多维数据

1. 多维数据采集存在的主要问题

大数据审计的一个重要特征就是开展多维数据相关分析，因此，需要采集的数据范围大大超过传统审计，不仅包括被审计单位内部数据、财务数据和结构化数据，还涉及外部数据、非财务数据和部分非结构化数据。因此，能否按工作需要规范、完整、高效地采集多维数据是工作质量的重要保证。目前，国家审计大数据分析的数据来源除被审计单位的财务与业务数据外，已能涵盖财政、社保、工商、税务等多个部门。但从实际操作看，仍存在数据的可得性、数据的可靠性和数据的相关性问题。

数据的可得性。从实践看，一方面有些数据由于存在保密性、安全性、隐私性和法律法规要求等方面的限制而难以取得，如银行数据等；另一方面，有些数据特别是非结构化的数据(如视频、图像、文档等)受到技术条件的制约，难以被纳入审计分析系统。

数据的可靠性。大数据审计虽然不要求数据有很高的精确性，但还是要保证数据在一定程度上的可靠性。在实践中，由于各被审计单位财务、业务等领域的系统尚未统一，往往出现数据重复(如重复在不同系统中采集某些数据点)、数据不一致(如相同数据点在不同系统中的值不一致)、数据丢失(如数据字段为空)、数据不明确(如在不同系统对相同数据的定义不同，从而不能与其他系统中的数据建立联系)等情况。如此，在开展大数据分析时可能会导致产生大量的“例外”数据，反而掩盖了可能存在的“错误”数据，需要审计人员投入较多的时间加以辨别难免影响审计的效率和质量。

数据的相关性。从理论上来说，被审计单位内外部不同领域、不同维度的所有数据都是与审计工作相关的，利用大数据分析技术对被审计单位全量业务数据进行探索式的全维度数据分析，是有可能发现有意义的审计新线索的。但由于审计时间和成本限制，这种“大海捞针”式的探索分析实际上并不可行。必须坚持问题、风险导向，结合业务逻辑、数据可得性等因素来规划好应获取哪些维度的数据。如何在时间和成本约束下，使多维数据的相关性水平保持在一个较优水平，避免因数据不充分、不全面、代表性不足从而影响审计质量，这些工作非常依赖审计人员的职业能力。

2. 信息标准化是解决问题的唯一途径

目前，大数据审计需采集不同部门、不同单位、不同业务系统的数据，但由于数据库技术、数据表结构、数据格式以及业务描述的不同，极大地影响到数据采集的效率和质量。

因此，只有从源头上，即在系统开发、数据库建设、数据输入等环节加强信息标准化建设，才能最终实现高效快速、规范标准的数据采集。当然，信息标准化仅仅只是依靠审计部门是难以推动的。在政府各部门信息化过程中，如果一开始缺乏顶层设计，只需满足各自部门对数据的需求，采用的技术方案也各不相同，就容易形成一个个“信息孤岛”。在这种情况下，如果没有政府或相关强力部门牵头协调，引入专业部门并投入相当的资源，完全依靠审计部门推进是不现实的。以嘉兴市为例，在财政部门推动下，构建了基于应用支撑平台的、业务标准统一的财政管理信息系统，从而为开展大数据审计、联网审计奠定了良好的基础。

(二)如何解决复合型审计人才短缺问题

开展大数据审计，需要审计人员具备一定的大数据分析能力。基于此，很多学者均提出要培养兼具审计和大数据分析能力的复合型审计人才的建议。^[9]这个观点无疑正确，但并不能满足当下大数据审计发展的迫切需要。相关研究对于需要怎样的复合型审计人才尚无明确、深入的认识，普遍停留在“审计技术+大数据技术”层面上。事实上，因为大数据技术的急速发展，要培养既具有较高审计专业能力又能跟上大数据技术发展前沿的复合型人才是极其困难的，人才供给必然是稀缺的。受到人员编制、数据处理能力等的限制，大量的数据采集、整理等工作可能仍然需要外包完成。因此，思考如何发挥现有审计人员的作用，研究培养怎样的复合型审计人才是一个重要课题。

1. 短期内，建立良好的协作机制是解决人才短缺问题的有效途径

复合型人才的培养目前仍处在探索阶段，短期内难以取得显著效果。当下，面对大数据审计发展的需求，往往将数据采集、整理等工作外包给专业技术公司来完成。但基于大数据的多维数据相关性分析，涉及被审计单位的内外部财务、业务系统。一方面，数据技术人员由于对相关业务接触较少，难以构建数据关联逻辑，必须依靠审计人员对业务的熟悉和审计逻辑，才能有效整合相关数据资源，关联和拼接各种数据。另一方面，审计人员缺乏大数据的处理能力，难以将数据需求与单位财务、业务

系统内的实际数据存储相对应，同时对各系统的后台运算逻辑不够了解，难以有效分析实际数据和预计结果存在的偏差，必须依赖数据技术人员的技术支持。因此，良好的审计、数据技术人员协作机制是大数据审计顺利开展的基础。审计人员应根据审计逻辑建立多维数据需求清单，由数据技术人员根据需求建立明确的业务系统数据提取路径指向，才能保证数据提取是完整、可靠、准确、快捷的。

2. 长远看，明确复合型审计人员的能力构成要素是解决人才短缺问题的根本

从长远看，培养复合型审计人才是解决人才短缺问题的根本途径。复合型人才应当在审计工作能力之外具备一定的大数据分析能力，但具有怎样的知识程度值得讨论。鉴于大数据技术发展极其迅猛，如同专业人员一样全面掌握、跟踪数据分析技术是极其困难也无必要的。复合型审计人才的知识储备，关键在于能否形成大数据审计所需的职业判断能力和沟通协调能力。

职业判断能力。在时间和成本、效率和质量的约束下，基于全体相关数据进行大数据审计分析是难以承受的。因此，工作中往往需要做出选择，比如，是要开发新的模型，还是增加新的数据来改进现有数据分析的效率和效果？是从采集难度高的系统中获取直接证据，还是从采集难度低的系统中获取间接证据？这些问题，只有具备复合型知识才能做出合理的职业判断。简而言之，在审计准备阶段，要从技术角度权衡审计数据的可得性、完整性，从审计角度确定所采集的数据的可靠性和相关性；在审计实施阶段，要从技术角度结合审计环境特点及审计目标需求选择或组合运用与之匹配的数据分析方法，从审计角度对数据分析发现的“例外”线索进行判断是“误报”还是“风险”；在审计报告阶段，要从技术角度选择最合适的展示方式(如可视化技术)以达到最好的信息传递效果，从审计角度判断“风险”的性质、产生的背景和原因。因此，复合型审计人才的知识储备首先必须满足职业判断的需要。

沟通协调能力。审计工作的顺利开展，不仅需要大数据分析模型，还需要审计人员与大数据专业技术人员间的沟通协调。但从现有的审计实践来看，由于审计人员和大数据专业技术人员之间专业界限清晰，都有各自的专业语言和思维方式，存在一定的沟通障碍。因此，复合型审计人才，需要既有阐释审计逻辑和数据需求的能力，也要有引介数据分析和解析数据分析结果的能力，从而协调双方的工作。

三、审计全覆盖背景下大数据审计发展的建议

审计全覆盖背景下，以多维数据相关性分析为特征的大数据审计，受数据资源、人才资源、数据技术发展等的限制，其发展高度依赖审计人员、技术人员的有效协作。因此，大数据审计的发展，必须依托审计业务能力和数据技术能力两条主线分阶段逐步提升、推进，其思路如图 3 所示。

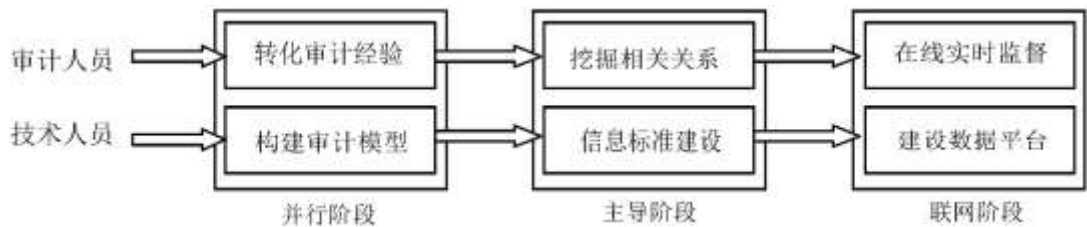


图 3 审计全覆盖背景下大数据审计发展建议

(一)并行阶段：经验转化，重构审计证据链条

大数据模式下，审计工作的中心环节发生转移，建立在多种方法组合运用基础上的传统审计证据链条将被打破，能否通过

引入内外部多维数据重构审计证据链条是推进大数据审计的关键所在。受到数据供给、技术能力等的制约，贸然地全面推进大数据审计，难免会出现原有审计模式难以转变、数据分析效率低下、风险难以全面覆盖等问题。因此，在大数据审计发展初期，应当有一个与传统审计模式并行的阶段，以便于审计人员验证做法、积累经验，为实现审计模式的转变奠定基础。具体形式，可以在常规审计项目中纳入大数据分析环节，或者采用大数据审计试点项目的方式。

并行阶段的首要任务，是基于以往审计发现的问题，依托审计人员熟悉不同部门数据供给的职业优势，在可获取内外部数据的前提下，构建多维数据相关分析，逐步实现对传统审计取证逻辑的替代。在此阶段，审计人员要思考大数据相关分析的逻辑是否可行，基础数据来源是否满足大数据分析的需求，数据采集采用怎样的方式开展，提取的数据是否可靠、完整，获取的审计证据证明能力是否充分，如何更好地与传统审计方法整合等等问题。

并行阶段的第二个任务，是要借助数据技术人员开发、固化审计模型，不断提高采集、整理、分析数据的效率。技术外包是当期审计部门解决数据提取、整理问题的主要方式，但每次审计都需要依托技术人员进行数据处理，不但影响工作效率，也不利于大数据审计的推广。面对关系错综复杂的海量数据，唯有将数据采集整理、对比分析逻辑、风险判定规则等固化成审计模型，才能满足不同知识背景的审计人员反复使用的需要。

（二）主导阶段：维度拓展，实现工作重心转移

通过并行阶段的验证和积累，审计项目开展过程中的大数据分析环节比重日渐提升、试点项目范围日益扩大，审计工作开始进入主导阶段。在此阶段，大数据分析逐渐成为主导审计工作的中心环节，但其他非信息化审计方法仍将长期存在并起到补充作用，对数据分析发现的线索、疑点进行验证、反馈。

并行阶段多维数据相关性分析，主要依赖于审计人员的经验建立数据分析逻辑和风险判定规则，是针对已知业务关系、已知风险特征的数据扫描。而经验化模型往往容易受到原有观察视野、分析逻辑的局限制约，很可能难以应对改革中出现的新现象、新趋势、新问题，也难以应对日益隐蔽的舞弊手法。因此，在主导阶段部分采用传统审计模式开展一定的审计项目，有助于审计人员加深对被审计单位的业务认知，不断挖掘新的数据关系、数据特征，使大数据分析能够更全面地揭示风险。

因此，主导阶段的首要任务，是要充分发挥审计人员的创新意识和能动性，突破原有经验思维限制，思考如何不断地将跨部门、跨层级、跨领域的的数据纳入审计视野，深度挖掘离散存储于不同系统平台中海量数据间的相关联系，进一步充实数据分析维度，不断创新数据分析模型，提高审计效率和精准发现问题能力，力求实现“风险全覆盖”。在基于业务流程检查、问题因果逻辑的数据挖掘思路之外，审计人员可尝试将聚类分析、社会网络分析、可视化技术等引入审计工作以获取事物之间的内在联系，^[10]揭示非常规、隐蔽性强的风险。

主导阶段的第二个任务，是在政府的领导和支持下，在数据技术人员的协助下，开展数据标准化建设。大数据审计能纳入的数据维度越多、审计取证的链条越完整，才能更好地发现线索和问题。从现有数据环境看，不同部门、不同系统间差异巨大，不仅是采用的数据库技术、数据结构和格式等存在不同，甚至同一信息的数据描述都不一致、不完整，数据整理和清洗的工作量较大，直接影响大数据分析的效果。因此，一方面须寻求政府主管部门的支持，从整体层面开展数据标准化建设；另一方面，需要依靠数据技术人员，基于当前审计工作面对的数据环境，通过数据关联构建常用的、标准的关联数据表，便于审计人员日常使用。如浙江省审计厅，分基础资料、预算编制、国库支付、总预算会计、预算单位会计、非税收入管理、政府采购、总决算、工资系统、固定资产、财政供养人员等 11 类制定财政审计数据规划，并按统一数据规划归集全省财政数据，实现了全省财政数据标准化收集。

（三）联网阶段：在线监控，常态化全覆盖监督

通过搭建大数据审计平台,实时运行审计数据分析模型实施在线分析和监控,^[11]是大数据审计发展的高级阶段。搭建大数据审计平台,建立基于审计数据分析模型的在线实时风险评估体系,充分发挥审计前瞻性整体数据分析功效,是大数据审计第三阶段要完成的任务。根据功能的不同,平台设计至少应包括两个方面:一是要构建具备多类型数据处理能力的综合数据平台;二是要构建基于审计数据分析模型的智能分析平台,实现审计由事后查处向事中控制、事前预防转变。

四、结语

大数据审计是一个新兴事物,审计部门必须对其特征有明确的认识,对其存在的问题有解决的思路,并通过制定合理的发展步骤有效推进。基于审计全覆盖对审计工作效率的要求,大数据审计应以数据采集和分析为中心环节,并通过内外部的多维数据“相关关系”分析构建新的审计取证链条,并最终向“联网审计”发展。同时,受到数据采集和人才培养这两个因素的制约,大数据审计的发展必须依靠审计人员和技术人员的有效协作,分并行、主导和联网三个阶段推进,才能最终满足审计全覆盖的需要。

参考文献:

- [1]王彪华.大数据审计理论与实践研讨会综述[J].审计研究,2020(2):52-56.
- [2]程平,白沂,胡雪莹.重庆市审计局基于大数据技术的预算执行审计方法[J].财务与会计,2018(6):77-79.
- [3]张亚丽.财务共享服务模式下的审计框架探究[J].会计之友,2019(21):21-24.
- [4]张敏.大数据审计:五大趋势与五大挑战[J].会计之友,2020(8):2-11.
- [5]王玉珊.审计全覆盖背景下提升审计质量的思考[J].重庆科技学院学报(社会科学版),2017(6):55-57.
- [6]秦荣生.大数据、云计算技术对审计的影响研究[J].审计研究,2014(6):23-28.
- [7]程平,张雅頔.大数据时代基于云会计的审计实施框架构建[J].会计之友,2015(24):134-136.
- [8]杨柔坚.深入推进大数据审计应用的实现途径[N].中国审计报,2020-09-14(008).
- [9]阳杰,应里孟.审计大数据分析人才需求及其培养——基于CDIO理念的模式建构[J].财会月刊,2019(4):108-119.
- [10]山东省审计学会课题组,栾心勇,孟祥宇.山东大数据审计探索与实践[J].审计研究,2020(3):29-35.
- [11]杨丽.从微观审计走向宏观服务——大数据背景下南京市部门预算执行审计转型经验谈[J].审计月刊,2016(6):30-31.

注释:

1 李克强在听取审计署工作汇报时强调:加强依法审计,实行全覆盖监督,切实看住管好公共资金,详见:<http://politics.people.com.cn/n/2014/0110/c1024-24075903.html>。