

第三章 大数据审计分析数据基础概述

目录
CONTENTS

- 第一节** 大数据审计分析数据基础概述
- 第二节** 审计数据采集
- 第三节** 审计数据预处理

第一节

大数据审计分析数据基础概述

3.1 大数据审计分析数据基础概述

数据作为一种重要的信息资源，是大数据审计分析的“原料”，接下来将从大数据审计分析的**数据特征、数据来源、数据类型、数据处理流程**四个方面来阐述大数据审计分析的数据基础知识。

3.1.1 大数据审计分析的数据特征

(1) 来源广泛

大数据审计分析的数据来源既包括互联网中的媒体咨询、搜索引擎、社交网络数据等**交互性数据**，也包括被审计单位信息系统的财务业务数据等**货币性数据**以及经营战略、管理经验、组织结构、社会关系等**非货币性数据**，还包括来源于其他渠道的一些数据，比如物联网、传感器数据等。

(2) 数据量大

传统审计数据以**MB**和**GB**的单位进行储存，而**审计大数据**储存单位可以直接达到**TB**甚至**PB**级。

(3) 关联性强

大数据背景下，急剧增加的数据量将推动被审计单位内部数据与被审计单位业务往来单位数据（如政府数据、企业数据、行业标准数据等）无限关联，从而提升审计大数据的关联性，增加挖掘潜在审计价值的可能性。

(4) 结构多元

审计大数据既包括**结构化数据**，又包括文档、音频、视频、图像等半结构化和**非结构化数据**。

3.1.2 大数据审计分析的数据来源

(1) 被审计单位

财务数据。包括但不限于来源于被审计单位财务系统、税务系统、ERP系统的财务账套数据、原始凭证等。

业务数据。包括但不限于来源于被审计单位ERP系统、CRM系统、WMS系统、HR系统、OA系统、生产管理系统、物联网系统、内部控制系统、预算系统、业务运行系统的业务数据表、业务单据等。

IT数据。包括但不限于被审计单位信息系统操作用户信息、信息系统用户操作日志等。

文本数据。包括但不限于被审计单位的内控文档、制度文件、生产记录、会议记录、内刊咨询、过往审计报告等。

视频与图像数据。包括但不限于被审计单位的访谈照片、访谈视频、生产经营场所的视频信息、存货存储场所视频信息、生物资产活动场所的视频、图片信息等。

音频数据。包括但不限于被审计单位的会议录音、访谈录音、客服录音等。

其他信息。包括但不限于被审计单位经营场所的客户、员工、运输工具、生物资产等的地理信息系统（GIS）和全球定位系统（GPS）数据。

3.1.2 大数据审计分析的数据来源

(2) 社交媒体

被审计单位与被审计单位相关的广义利益相关者会通过社交媒体主动或被动的产生与被审计单位有关的数据，包括但不限于被审计单位的企业微信，广义利益相关者的QQ、微信、微博、朋友圈、邮件，论坛讨论贴等。

(3) 外部网络

被审计单位所处外部环境的其他社会关系参与者也会产生与被审计单位有关的数据，包括但不限于主流媒体新闻报道，宏观层面的政策文件，中观行业信息，监管部门问询函，被审计单位及其关联公司、上下游公司的财务报告等。

(4) 专有数据库

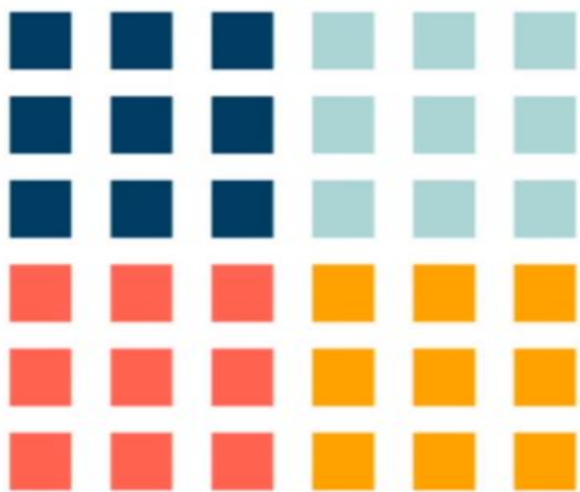
一些专有数据库也会保存与被审计单位相关的数据，包括但不限于国泰安CSMAR数据库，Wind数据库，上交所XBRL数据，深交所数据，国家统计局数据，不动产登记数据，工商注册数据，车辆管理数据，税务数据，银行数据。

3.1.3 大数据审计分析的数据类型

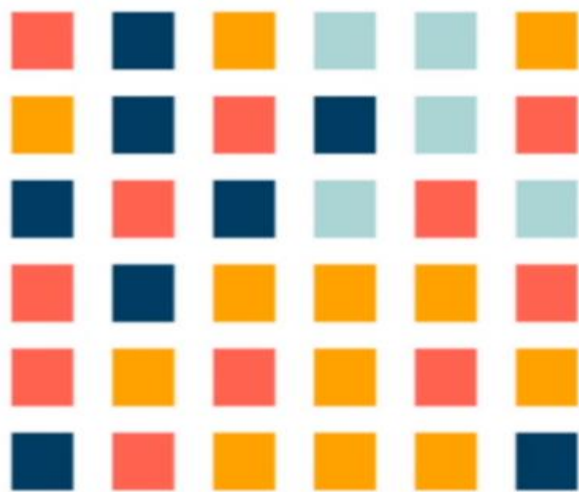
结构化数据：是指能够用统一的设计结构予以存储的数据，如数值符号等。任何以固定格式储存、访问和处理的数据都称为结构化数据。也称为行数据，一般存储在数据库中，可以用二维表来逻辑表达实现。

半结构化数据：介于结构化数据和非结构化数据之间的数据。半结构化数据具有一定的结构性，如HTML文档、XML文档。

非结构化数据：是指字段长度可变，每个字段的记录又可以由可重复或不可重复的子字段构成的数据，如文本、图像、声音、网页、视频等。非结构化数据不方便用数据库二维逻辑表来实现。



结构化数据



半结构化数据



非结构化数据

3.1.4 大数据审计分析的数据处理流程

大数据审计分析的数据处理流程包括审计数据采集、审计数据储存、审计数据清洗、审计数据分析、审计数据可视化。

- (1) 审计数据采集:**审计人员为了完成既定审计目标,在进行大数据审计分析时,按照审计需求从被审计单位数据库或其他来源获取审计数据的过程。
- (2) 审计数据储存:**审计人员将收集、加工后的审计数据储存在特定数据库,以供作为审计分析对象或审计证据的过程。
- (3) 审计数据清洗:**审计人员对采集的审计数据进行清理、转换、验证、整合,并将其转化为可被理解利用的审计分析信息源的过程。
- (4) 审计数据分析:**审计人员利用数据分析公式或逻辑表达式,建立审计事项内在的数量关系,从而对被审计对象相关活动的真实性、合法性、效益性做出科学判断的过程。
- (5) 审计数据可视化:**审计人员使用计算机图形学、图像处理技术、人机交互技术等,将经过处理后的审计信息转化为易于被人们接受和理解的图形、图像、视频、动画等视觉表达形式,从而增强审计信息使用者对审计事项认知的过程。

第二节

审计数据采集

3.2 审计数据采集

审计数据采集是大审计数据分析的第一个步骤，采集的数据质量直接决定了后续工作是否有效。

3.2.1 数据质量特点

(1) 全面性

全面性是指数据的数量较多，而且能够代表整体数据的各个部分，数据量足够支撑分析需求。比如要对员工的差旅费用是否合规进行分析，就需要采集差旅费发生的时间、地点、金额、公司费用标准等数据。

(2) 多维性

多维性是指审计人员能灵活、快速自定义数据的多种属性和不同类型，从而满足不同的分析目标。比如分析某项产品收入的合理性，可以与生产该产品的成本进行关联，从而判断收入是否合理。

(3) 真实性

数据必须真实准确地反映客观的实体存在或真实的业务，真实可靠的审计数据是审计工作的根基，是对审计结论作出合理保证必不可少的第一手资料。

(4) 准确性

准确性是用于分析和识别哪些是不准确的或无效的数据，不准确的数据可能会造成严重的问题，会导致有缺陷的方法和错误的决策。

(5) 时效性

时效性是指对审计分析具有效用的、比较新的数据。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/115032340234011040>