

审计数据归集规范

Specification for audit data collection

2023 - 04 - 18 发布

2023 - 05 - 18 实施

前 言

本标准按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本标准的某些内容可能涉及专利。本标准的发布机构不承担识别专利的责任。

本标准由浙江省审计厅提出、归口并组织实施。

本标准起草单位：浙江省标准化研究院、浙江省大数据局发展中心、杭州市审计局、宁波市审计局、温州市审计局、台州市审计局、丽水市审计局。

本标准主要起草人：王家华、吴肖浦、邱文强、孙雅和、郑培、孟一丁、金加和、赵程遥、俞巍滔、杜辉、甄理、洪飞云、王维佳、戴鹏军、刘薇、刘卉、牛李金梁、韩伟、郑腾腾、周长明、何阿健、吴希凑、李前义。

本标准为首次发布。

审计数据归集规范

1 范围

本标准规定了审计数据归集的基本原则、归集体系、数据申请与审批、数据交换与归集、数据治理和安全控制等内容。

本标准适用于收集、产生公共数据的信息系统和公共数据平台通过相关流程向国家审计大数据中心浙江分中心平台归集审计数据。

2 规范性引用文件

下列标准中的内容通过文中的规范性引用而构成本标准必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本标准；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本标准。

GB/T 22239—2019 信息安全技术网络安全等级保护基本要求

GB/T 31916.1 信息技术 云数据存储和管理 第1部分：总则

GB/T 36326 信息技术云计算云服务运营通用要求

GB/T 39477 信息安全技术 政务信息共享数据安全技术要求

DB33/T 2349 数字化改革公共数据目录编制规范

DB33/T 2350—2022 数字化改革术语定义

DB33/T 2351 数字化改革 公共数据分类分级指南

DB33/T 2359—2021 公共数据交换技术规范

DB33/T 2487—2022 公共数据安全体系建设指南

DB33/T 2488 公共数据安全体系评估规范

3 术语和定义

GB/T 36326、DB33/T 2349、DB33/T 2350、DB33/T 2351界定的以及下列术语和定义适用于本标准。

3.1

公共数据 public data

国家机关、法律法规规章授权的具有管理公共事务职能的组织以及供水、供电、供气、公共交通等公共服务运营单位（公共管理和服务机构），在依法履行职责或者提供公共服务过程中收集、产生的数据。

[来源：DB33/T 2487—2022，定义3.1，有修改]

3.2

一体化智能化公共数据平台 integrated and intelligent public data platform

以云计算、大数据、人工智能、互联网等技术为支撑，是省域治理全过程数据感知、数据共享、数据计算的基础平台，是支撑数字化改革的集成运行平台。

[来源：DB33/T 2350—2022，定义3.2.2.1]

3.3

国家审计大数据中心浙江分中心平台 Zhejiang audit data center

审计机关建设的为全省审计人员提供数据共享、数据服务、数字创新的软硬件平台和信息系统。

3.4

审计专题库（省级） audit special database (provincial)

省一体化智能化公共数据平台中审计机关专用的政务云数据仓，专门用于存放和处理已归集的审计数据，并配套建立审计数据资源目录。

3.5

审计归集库（市县级） audit collection database (city/county level)

市、县级一体化智能化公共数据平台中审计机关专用的政务云数据仓，专门用于存放和处理已归集的审计数据。

3.6

数源机构 data source organization

根据法律、法规、规章确定的某一类公共数据的法定收集部门，包括国家机关、法律法规规章授权的具有管理公共事务职能的组织以及供水、供电、供气、公共交通等公共服务运营单位等。

3.7

数据归集 data collection

面向各信息系统、公共数据平台及其他特定场景的数据，进行收集、传输、清洗加工和整合的活动。

3.8

数据治理 data governance

基于数据生命周期，进行数据全面质量管理、资产管理、风险管理等统筹与协调管控的过程。

注1：多学科视角下，数据治理需要解决数据权属关系问题，明确数据利益相关方的角色。权利和权益及其责任关系和工作任务，避免数据风险，提高数据质量，确保数据资产能长期有序地、可持续地得到管理和利用。

注2：信息技术视角下，数据治理指对数据进行管控，处置、格式化和规范化的过程。数据治理是数据和数据系统管理的基本要素。数据治理涉及数据全生命周期管理，无论数据是处于静态，动态，未完成状态还是交易状态。

[来源：GB/T 37043—2018，定义2.3.1]

4 基本要求

4.1 唯一性

一项数据有且只有一个法定数源机构，审计机关根据各业务领域审计工作需要与内部规范，明确归集的数据内容和标准。可以通过公共数据平台共享获取数据的，审计机关不应重复收集；共享数据无法满足履行职责需求的，审计机关可经数源机构授权后，向数源机构获取，或向公共数据主管部门提交数据需求清单，由公共数据主管部门与相关数源机构协商解决。

4.2 完整性

被审计单位或公共数据平台应提供满足审计需求的完整审计数据，包括历史数据、过程数据和增量数据等。

注：数据归集的范围依据《中华人民共和国审计法》《中华人民共和国审计法实施条例》《浙江省审计条例》等法律法规规定，包括但不限于被审计单位在履职过程中产生的财务、会计资料以及与财政收支、财务收支有关的业务、管理等资料的电子数据和有关文档。

4.3 按需归集

审计数据的归集对象为所有收集、生产公共数据的数源机构，各级审计机关应按需求归集审计数据。

4.4 准确性

被审计单位或公共数据平台提供的审计数据应与其关联的财政收支、财务收支有关的业务、管理实际相一致。对审计数据的数据治理应贯穿整个归集过程，确保数据真实有效。

4.5 时效性

审计机关应根据各业务领域审计工作需要，向被审计单位或公共数据平台确定数据归集的时间跨度范围及其有效性。

4.6 可追溯性

数源机构应配合公共数据主管部门和审计机关建立健全数据全流程质量管控体系，保证数据准确完整，及时更新已变更、失效数据，实现数据及其变化过程全程可追溯。

5 归集体系

5.1 基本架构

应按照图 1 所示流程归集审计数据，包括数据申请、审核确认、数据交换与归集、数据治理等：

- a) 审计数据申请：审计机关应在审计专题库（省级）中按照 5.2.2 的要求确定审计数据资源目录中各项审计数据的归集范围与属性，并根据唯一性原则，按照 6.2 的要求向数源机构发起数据归集申请；
- b) 审批确认：数源机构的审计业务责任部门收到审计机关发起的申请后，在 3 个工作日内确认是否同意；

注：审计业务责任部门为数源机构中对接审计工作的相关责任部门。

- c) 数据交换与归集：数源机构确认数据目录后，公共数据主管部门按照第 7 章的要求将所需数据交换至审计专题库（省级），并与各级审计机关共同完成数据归集，其中：
 - 1) 省级数源机构的数据由省级公共数据主管部门交换至审计专题库（省级），并完成数据归集；
 - 2) 市、县级数源机构的数据由同级公共数据主管部门交换至相应审计归集库（市县级）后，由各级公共数据主管部门逐级上报，最后由省级公共数据主管部门统一交换至审计专题库（省级），完成数据归集。
- d) 数据治理：审计机关将汇聚到审计归集库（市县级）、审计专题库（省级）的相关数据按第 8 章的要求进行数据治理后，最终将审计数据归集到国家审计大数据中心浙江分中心平台。
- e) 安全可靠：审计数据归集全链条中各相关方应按照第 9 章的要求，确保审计数据申请审批、交换归集和治理过程中各自的网络与数据安全。

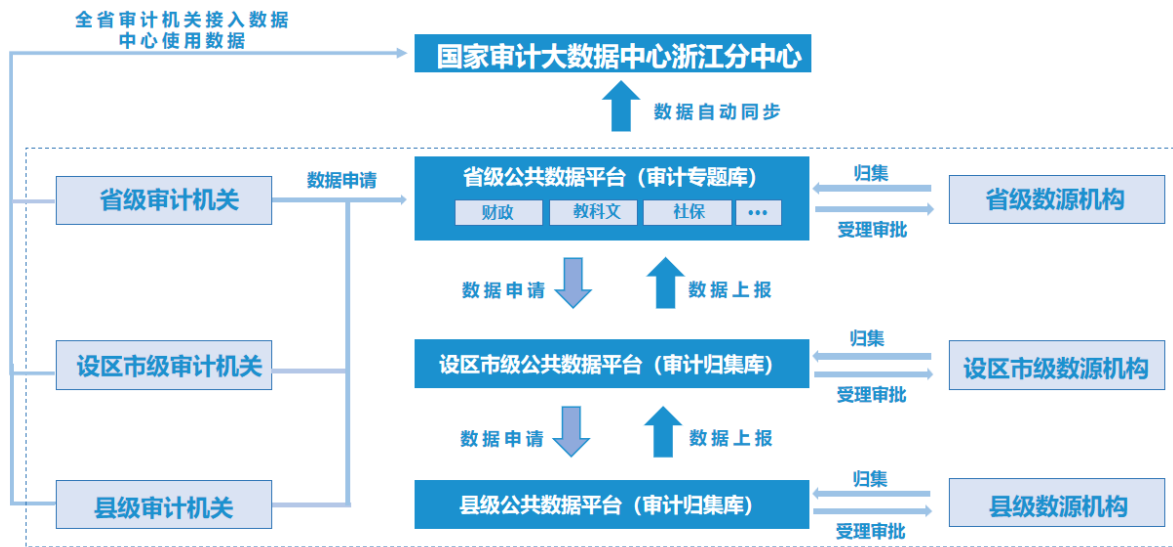


图1 审计数据归集架构图

5.2 归集数据内容

5.2.1 数据资源目录编目

资源目录应包括省本级数源机构审计数据、设区市级数源机构审计数据和县级数源机构审计数据，以及这些信息的来源、分类领域、更新频率、上次更新时间、申请入口等，相关表信息可参考附录A。

5.2.2 审计业务分类

全量审计数据可按来源分为12个领域，具体如下：

- a) 财政领域：主要包括总预算决算审计、财政转移支付资金审计等；
- b) 教科文领域：主要包括高校主要领导干部经济责任审计、宣传审计、文化和旅游审计、教育审计、科技审计、广播电视审计和体育审计等；
- c) 农业农村领域：主要包括农业农村审计、扶贫开发审计等；
- d) 固定资产投资领域：主要包括投资建设项目审计、重大公共工程项目审计、援建资金和项目审计和重大灾后重建跟踪审计等；
- e) 社保领域：主要包括基本养老审计、基本医疗审计、失业保险审计、工伤保险审计、社会救助审计、社会福利审计、慈善捐赠审计、养老服务体系审计、公共卫生服务审计、社会优抚审计、社会事务审计、应急管理体系审计和残疾人权益审计等；
- f) 自然资源和生态环境领域：主要包括领导干部自然资源资产离任（任中）审计、自然资源管理审计、污染防治审计和生态保护与修复情况等；
- g) 金融和外贸领域：主要包括金融机构审计、援助贷款项目审计等；
- h) 企业领域：主要包括国有企业审计、驻外非经营性机构的财务收支审计等；
- i) 市县党政主要领导干部领域：主要包括市县党政主要领导干部经济责任审计等；
- j) 党的机关和政法部门领域：主要包括法院审计、检察院审计、公安局审计等；
- k) 行政事业领域：主要包括部门预算执行审计、部门主要领导经济责任审计等；
- l) 信息系统领域：主要包括信息系统审计、信息化建设审计、网络安全审计等。

6 数据申请与审批

6.1 向省级数源机构申请

按图2所示流程，审计机关申请省级数源机构的审计数据应通过数源机构和省级公共数据主管部门审批同意，随后由省级公共数据主管部门按照5.1 c)和d)的要求归集数据至审计专题库（省级）。

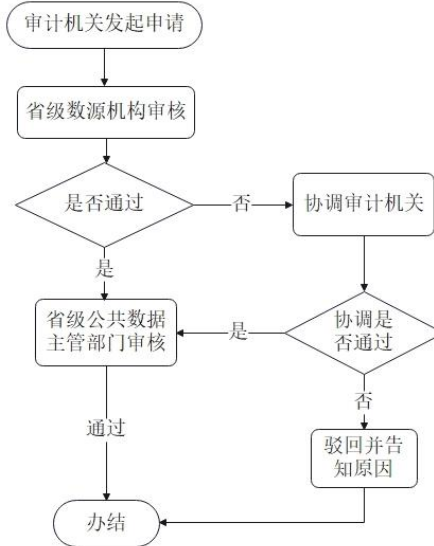


图2 审计机关申请省级数源机构数据流程图

6.2 向设区市级数源机构申请

按图3所示流程，审计机关申请设区市级数源机构的审计数据应通过数源机构和设区市级公共数据主管部门审批同意，并向省级公共数据主管部门报备，随后由市级公共数据主管部门按照5.1 c)和d)的要求归集数据至审计专题库（省级）。

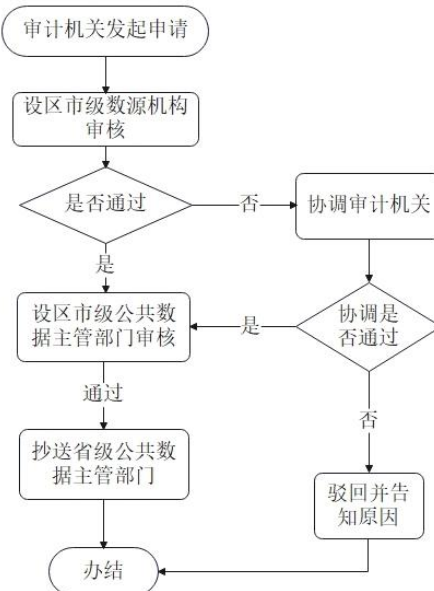


图3 审计机关申请设区市级数源机构数据流程图

6.3 向县级数源机构申请

按图4所示流程，审计机关申请县级数源机构的审计数据应通过数源机构审批和县级公共数据主管部门审批同意，并逐级向设区市级公共数据主管部门和省级公共数据主管部门报备，随后由县级公共数据主管部门按照5.1 c)和d)的要求归集数据至审计专题库（省级）。

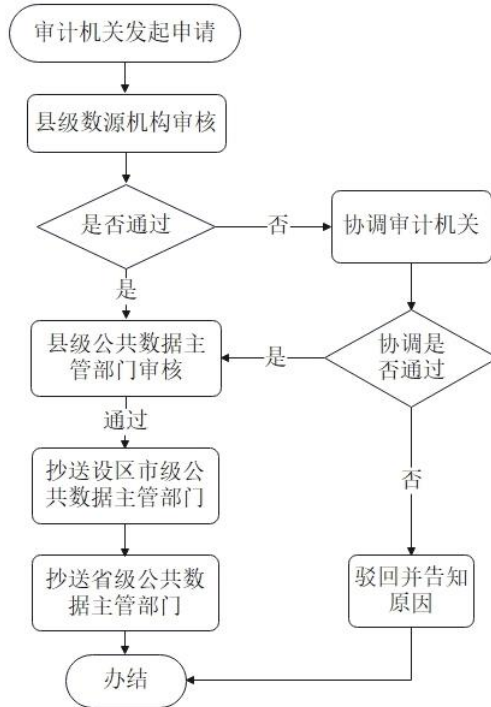


图4 审计机关申请县级数源机构数据流程图

7 数据交换与归集

7.1 交换准备

数源机构应在首次审批完成后与公共数据主管部门确认需要归集审计数据的数据库、数据类型和通讯方式，主要包括：

- a) 数据类型：可支持电子文件、电子表格、数据库、矢量图斑/图形图像、音视频/流媒体、传感数据等不同类型的数据库，其中数据库可支持结构化数据和非结构化数据属性；
- b) 通讯方式：可支持传输控制协议（TCP）、用户数据报协议（UDP）、文件传输协议（FTP）、超文本传输协议（HTTP/HTTPS）等通信协议；可支持 JavaScript 对象表示、Java 数据库连接、开放数据库互联、消息队列和遥测传输等连接方式。

注：Java、JavaScript分别是不同的面向对象的计算机编程语言。

7.2 交换能力

应具有数据目录编目、全量自动扫描归集、增量自动扫描归集、依据分类批量申请上报等功能，保障在复杂网络环境下、不同异构数据源之间高速、稳定、弹性的数据移动和同步能力：

- a) 支持历史数据与全量数据采集：应提供数据传输服务、高并发的离线数据上传下载服务，支持 TB/ PB 级别的批量数据导入及导出；
- b) 支持实时或定时增量数据采集：可提供实时同步、定时采集、数据订阅等服务；
- c) 支持条件过滤：按照指定条件进行指定过滤采集，例如字段内容；

- d) 支持采集作业管理和调度：采集作业支持条件触发、并发调度、周期循环调度等模式；支持对作业启动、停止、暂停、恢复等操作。

7.3 归集方式

按照DB33/T 2359—2021中第5章至第8章的相关要求或通过前置数据传输、数据接口、数据高铁等方式，通过一体化智能化公共数据平台自动归集与汇聚数源机构的审计数据。

注：数据高铁是一种以分布式算法、流式计算技术为核心，集成数据日志的实时归集、解析和数据的清洗、比对等功能，支撑海量数据高速传输的数据实时交换系统，可实现公共数据跨层级、跨地域、跨部门、跨系统、跨业务的实时同步共享。

7.4 归集周期

审计机关应按5.2.2中的审计业务分类，分别确定数据归集的周期，并在数据申请阶段告知数源机构和公共数据主管部门，数据归集周期主要包括：

- a) 按年度归集；
- b) 按半年度归集；
- c) 按季度归集；
- d) 按月度归集；
- e) 实时归集；
- f) 临时归集等。

8 数据治理

8.1 数据质检与处理

数源机构向审计归集库（市县级）、审计专题库（省级）归集数据，审计专题库（省级）的数据归集至国家审计大数据中心浙江分中心平台后，应分别对归集的审计数据进行质量检测与处理。包括但不限于：

- a) 建立内部数据质量审核管理制度,包括数据质量查验、保障人员、工作内容、工作周期和时间节点等；
- b) 归集的数据应进行校检处理，并提出相应的解决方案。校验内容：空值的检测和处理、非法值的处理、不一致数据的检测和相似重复记录的检测；
- c) 应对校验后的数据编码转化，统一数据的代码、格式和类型等；
- d) 应对转换后的数据通过数据源验证、数据聚合、数据归类、数据关联等方式进行分析与预处理，形成上下文完整有效的数据，不合格的依据数据质量检测规则退回整改后重新提交。

8.2 数据存储

审计归集库（市县级）、审计专题库（省级）、国家审计大数据中心浙江分中心平台应满足GB/T 31916.1的相关要求。

8.3 特殊情况处理

各数源机构如发生如下情况，应通知并重新组织全量数据归集：

- a) 软件系统升级，主要业务数据发生变化；
- b) 归集数据内容不符合审计相关“唯一性”定义；

- c) 增量归集机制出现故障，无法继续识别增量数据；
- d) 因数据出现重大变化导致无法按照正常流程进行数据归集的其他情况等。

9 安全可靠

9.1 审计数据归集全链条全系统，及各相关方系统稳定性应不低于 99.9%（即全年故障时间不超过 8 小时），并制定备份、应急处置预案和定期维护计划，保存维护记录，记录保存期限不少于 5 年。

9.2 数据归集过程中应符合 GB/T 22239—2019 中的第三级安全要求，GB/T 39477 中第 5 章、第 6 章，以及 DB33/T 2487、DB33/T 2488 对网络系统与数据应用安全的相关要求，按照 DB33/T 2488 的相关要求建立审计数据分类分级制度，实现审计数据的授权访问、可定位溯源、数据加密、保密传输、安全审计监测，全方位防御，避免病毒、攻击与非授权的访问，同时应保障访问记录等的审查和监督。

附录 A

(资料性)

审计数据资源目录表信息

审计数据资源目标表信息示例和信息资源要素见表A.1和表A.2。

表A.1 审计数据资源目标表信息示例

字段名称	中文名称	是否必须	描述
type	操作类型	是	insert:新增 update:修改
data	目录对象	是	—
id	目录唯一id	是	—
backFlow	数据回流标识	否	0 未回流; 1 省市回流; 2 省市县回流; 3 市县回流(默认0)
catalogs	关联数源目录id	否	—
ignoreCatalogs	免审核目录	否	—
catalogOrgs	数源机构	是	—
approveOrgs	被审计单位	是	—
shareType	共享属性	是	1受限共享; 2无条件共享; 3非共享
shareCondition	共享条件	是	—
openType	开放属性	是	1无条件开放; 2受限开放; 3禁止开放
openCondition	开放条件	为开放时必须	—
tableName	英文表名	是	—
themeDBInfoId	所属审计库基础信息id	是	—
kind1	专题库一级分类ID (专题库名称)	是	—
kind2	专题库二级分类ID	是	—
kind3	专题库三级分类ID	是	—
kind4	专题库四级分类ID	是	—
resCode	资源编码	是	—
appCatalogId	关联应用建档id	是	暂无为: -1
catalogName	信息资源中文名称	是	—
summary	信息资源摘要	是	要求20个字~300个字
catalogFormat	信息资源格式	是	详细要素见表A.2(01电子文件; 02电子表格; 03数据库; 04矢量图斑/图形图像; 05音视频/流媒体; 06传感数据及其他自描述格式)
updateFrequency	数据更新频率	是	每月; 每年; 实时; 临时; 不更新
columns	数据项列表	是	—
orgId	数据项数源机构	是	—

表A.1 审计数据资源目标表信息示例（续）

字段名称	中文名称	是否必须	描述
columnCN	数据项中文名	是	—
columnEN	数据项英文名	是	—
columnType	字段类型	是	—
description	字段描述	否	—
length	字段长度	否	—
defaultValue	默认值	否	—
shareType	共享属性	是	1受限共享；2无条件共享；3非共享
shareCondition	共享条件	是	—
openType	开放属性	是	1无条件开放；2受限开放；3禁止开放
openCondition	开放条件	为开放时必须	—
isDic	字典项	否	0否；1是
isNull	空值	是	0否；1是
isKey	主键	是	0否；1是

表A.2 信息资源要素

序号	字典码	信息资源要素分类	数据资源要素格式名称
1	0100	电子文件	其他
2	0101		doc
3	0102		docx
4	0103		html
5	0104		OFD
6	0105		pdf
7	0106		ppt
8	0107		txt
9	0108		wps
10	0109		xml
11	0110		dpf
12	0111		bmp
13	0200	电子表格	其他
14	0201		et
15	0202		xls
16	0203		xlsx
17	0300	数据库	其他
18	0301		access
19	0302		db2
20	0303		dbase
21	0304		dbf

表 A.2 信息资源要素（续）

序号	字典码	信息资源要素分类	数据资源要素格式名称
22	0305	数据库	Dm
23	0306		KingbaseES
24	0307		oracle
25	0308		sqlServer
26	0309		sybase
27	0310		mysql
28	0311		GBase
29	0312		Hbase
30	0313		hdfs
31	0314		EsgynDB
32	0315		redis
33	0316		mongodb
34	0400		矢量图斑/图形图像
35	0401	bmp	
36	0402	gif	
37	0403	jpg	
38	0500	音视频/流媒体	其他
39	0501		mpg
40	0502		rm
41	0503		swf
42	0600	传感数据及其他自描述格式	其他
43	0601		表格驱动码