

《公共数据 数据治理规范 第1部分：数据归集》地方标准编制说明

一、工作简况

（一）任务来源

2020年7月30日，山东省市场监督管理局印发《2020年度地方标准制（修）订计划项目》的通知（鲁市监标字〔2020〕249号），本标准列为推荐性地方标准，计划编号：2020年度地方标准制（修）订计划项目—159。本文件由山东省大数据局提出并归口。

（二）起草单位、起草人及任务分工

本标准的起草单位为山东省大数据局、山东省大数据中心、烟台市大数据中心、山东新一代标准化研究院有限公司、山东省计算中心（国家超级计算济南中心）、浪潮数字（山东）建设运营有限公司、山东优数网络科技有限公司、山东省新型智慧城市大数据工程技术研究院、中国标准化研究院、山东省标准化研究院。

本标准的主要起草人为王伟、林庆、朱薪安、姜青岳、王铁、相树志、王茜、郑慧、关新雨、柯林森、郑奇、何敬明、石伟、刘福才、史丛丛、王溪、李杨、王庆明、许德民、李童、钱恒、郑祎、张媛、邹丰义、刘佳佳、许潇文、相吉利、王亚楠。

起草人任务分工：王伟、林庆、朱薪安、姜青岳负责标准总体设计、技术把关工作，王铁、相树志、王茜、郑慧、关新雨负责标准资料调研、汇总等工作，柯林森、郑奇、何敬明、石伟、刘福才负责技术资料分析、标准框架搭建等工作，史丛丛、王溪、李杨、王庆明、许德民负责标准内容起草、征求意见及修改、专家研讨及意见修改、送审材料编写等工作，李童、钱恒、郑祎、张媛、邹丰义、刘佳佳、许潇文、相吉利、王亚楠等负责参与标准各阶段的技术讨论、标准的试验论证等工作。

（三）起草过程

1. 调查研究阶段（2020年12月-2021年3月）

标准编写组启动标准调研工作，对数据归集相关的标准和政策文件进行收集、研究，重点搜集数据归集相关的国家标准、行业标准、地方标准等资料。召开多次内部讨论会，充分借鉴国家和各省市公共数据归集提供的经验做法。结合我省公共数据归集架构和数据归集实际情况，研究分析山东省公共管理和服务机构归集数据至省一体化大数据平台的总体要求、数据归集流程和归集方式等，确定了标准范围，并对标准框架和内容整理编制思路。

2. 标准起草阶段（2021年4月-2021年5月）

根据前期文件调研以及本系列标准制定情况，为进一步发挥本标准在山东省数据治理工作规范化开展中的保障作

用，参考已发布系列国家标准GB/T 34960《信息技术服务 治理》，经讨论论证，该系列标准名称由《政务信息资源数据》改为《政务信息资源 数据治理规范》，同时将标准名称“政务信息资源数据 第5部分：数据归集”调整为“政务信息资源 数据治理规范 第1部分：数据归集”。标准编写组根据收集的资料和分析结论，充分借鉴、吸收国家和其他各省在数据治理工作中数据归集的经验做法，深入调研我省数据归集的现状，搭建标准框架，确定标准框架和主要技术内容，编写形成标准草案。针对数据归集的标准名称、数据归集总体要求、数据归集架构、数据归集流程、数据归集方式、数据更新和安全等问题，标准编写组先后组织召开了多次多方共同参与的内部研讨会，对标准适用性、标准技术内容等开展研讨，根据研讨意见对标准草案多轮次迭代优化，形成征求意见稿初稿。

3. 征求意见阶段（2021年6月-2021年8月）

2021年6月以来，标准编写组多次召开内部研讨会议，讨论标准征求意见稿初稿，对数据归集的标准名称、数据归集总体要求、数据归集架构、数据归集流程、数据归集方式、数据更新和安全等相关内容进行研讨，主要包括接口协议是否需要细化、数据更新频率、数据更新方式等内容，标准编写组按照实际要求对标准草案进行了修改完善，形成征求意见稿。面向25个省直部门，16个地市广泛征求意见，收到

15 个省直部门，13 个地市的征求意见回函，回函并有建议或意见的单位数 10 个。本标准一共收取 18 条意见，采纳 16 条，部分采纳 2 条，其中对部分采纳意见进行了详细的解释说明。主要意见和意见处理如下：

（1）目前数据归集架构图，对数据源、原始库、主题库、服务库等的定义未进行详细说明。标准编写组与相关技术团队讨论并明确数据归集架构中数据源位于公共管理和服务机构端，原始库、标准库、服务库等位于省一体化大数据平台端，修改表述并在标准中对各术语进行修改；

（2）数据归集方式中库表、服务接口、文件的技术要求单薄，且未说明是对哪一方的要求。标准编写组与相关技术团队对标准技术内容进行充分研讨，增加库表、服务接口、文件归集方式的对账机制和对账表，补充服务接口内容示例和 xml 格式数据示例等，补充数据传输要求中压缩传输和加密传输等内容；

（3）需精简表述，标准不体现管理要求。标准编写组内部讨论后对标准整体表述进行精简，删除“三定职责”等内容，不体现管理相关要求。

标准编写组根据征求意见情况作了认真修改，形成标准研讨初稿。

4. 专家研讨阶段（2021 年 8 月）

2021 年 8 月 17 日，标准编写组邀请省民政厅、省工信厅、教育部门、省市场监管局、省住建厅等相关专家，在济

南召开标准研讨会，会议中对标准的名称、框架和技术内容进一步研讨确认。会议提出的主要问题为标准范围、标准使用相关方、标准使用场景不明确。标准编写组充分调研山东省一体化大数据平台、全省公共管理和服务机构数据归集现状，多次沟通标准需求，明确标准范围、相关方和应用场景，将标准范围界定为数据源到贴源层，即数据从公共管理和服务机构到省一体化大数据平台的过程。

5. 标准验证阶段（2021年9月-2022年11月）

标准编写组整理归纳征求意见单位、专家研讨会各专家的反馈意见，组织召开多次内部讨论会，逐条对意见反馈情况进行处理，并同步修改完善标准。同时在全省农业农村厅、省教育厅向省级一体化大数据平台数据归集工作中，济南市、烟台市、淄博市3个试点市的市级一体化大数据平台数据归集工作中开展标准验证，参照验证结果修改标准技术内容。补充数据高速通道、文件传输、对账机制相关的技术要求，对标准进一步修改完善，形成标准送审稿。

6. 专家审查阶段（2022年12月）

2022年12月8日，山东省大数据局通过线上会议组织召开了山东省地方标准专家审查会议，省市场监管局对审查会议进行监督指导。来自省建设培训与执业资格注册中心、济南市大数据局、省社会救助服务指导中心、省商务厅、省财政研究和教育中心、山东青年政治学院、省市场监管监测

中心、省工业和信息化厅工业行业离退休干部服务中心、山东健康医疗大数据管理中心等单位共 9 名专家组成了审查委员会。审查委员会听取了标准编制情况汇报，对标准文本进行了逐章、逐条审查，对标准编制说明等进行了审查。会议提出的主要问题和处理结果如下：

（1）缺少大文件拆分相关技术要求，建议给出大文件拆分记录表。标准编写组与技术支撑团队充分沟通后，增加文件拆分信息表，记录每个批次非结构化文件的拆分信息。文件拆分信息表包括记录 ID、文件信息表记录 ID、文件名、文件路径、文件大小、文件完整性校验、更新时间等字段；

（2）文件信息表中建议去除“文件 MD5 校验码”，考虑国产密码的应用。标准编写组与技术支撑团队讨论后确认采纳该条意见，将“文件 MD5 校验码”替换为“文件完整性校”；

（3）建议标准名称中前缀由“政务信息资源”修改为“公共数据”。基于国家、山东省相关政策文件要求，标准编写组与审查专家团队充分讨论后，确认采纳该条意见，并修改标准名称。

标准编写组认真修改完善标准，形成标准报批稿，按照相关程序对标准进行报批。

二、地方标准制定目的和意义

标准制定目的：本标准制定的目的主要有三点，一是明

确公共管理和服务机构向省一体化大数据平台归集数据的方式和相关流程，二是规范数据归集的相关技术要求，包括对账机制等，三是规范了数据归集中的安全要求，为数据归集的安全保驾护航。

标准制定意义：本标准发布后，可对公共管理和服务机构向省一体化大数据平台归集数据的流程和要求进行统一规范，指导公共管理和服务机构向省一体化大数据平台归集数据，按照“按需归集、应归尽归”原则归集数据，推动数据治理能力提升，支撑数字强省的建设。

三、地方标准编制原则、主要技术内容和确定依据

（一）标准编制原则

本标准依据 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》给出的规则起草，该标准编写的主要原则如下：

1. 科学性原则：本标准的制定充分吸取相关国家标准、行业标准，借鉴学习其他省市在数据归集系统建设过程、归集过程中的经验做法，结合全省数据归集实际的建设情况，确定标准的结构框架，确保编制的科学性。

2. 需求主导原则：本标准的制定充分考虑山东省公共管理和服务机构向省一体化大数据平台归集数据的实际情况，调研分析各级一体化大数据平台、公共管理和服务机构数据归集的需求，确定数据归集标准范围和内容，确保标准与实

实际需求相符。

3. 紧密结合工程建设原则：标准研制紧密结合我省一体化数据资源体系实际建设情况、山东省一体化大数据平台建设情况，充分调研省市县三级一体化大数据平台中数据归集工作的实际开展情况，综合考虑省市县三级一体化大数据平台在数据归集工作的共性要求，并在标准编制过程中及时根据建设中反馈的问题完善标准。

（二）本标准主要技术内容

本文件规定了公共数据的数据归集流程、数据归集要求、数据更新要求、数据归集安全等。本文件适用于指导各级公共管理和服务机构归集数据至省一体化大数据平台。本文件主要内容包括：

（1）术语与定义：给出公共数据资源目录、省一体化大数据平台、公共管理和服务机构的定义。如公共数据资源目录是按照一定的分类方法进行排序和编码的一组信息，用于描述各个公共数据特征及组织方式，便于公共数据的组织、检索、定位、发现与获取。

（2）缩略语：给出CSV、FTP、GZIP、HTTP等缩略语的解释，如CSV：逗号分隔文件格式，英文名称为Comma-Separated Values。

（3）数据归集流程：给出数据归集的流程和要求，主要包括明确归集范围、明确归集方式、开展数据归集、归集

结果确认。

首先公共管理和服务机构应根据业务职责，编制并动态维护公共数据资源目录，按照“按需归集、应归尽归”的原则，明确归集范围为本机构所有可归集公共数据。其次结合归集数据的范围、数据传输要求等，确定数据的归集方式。公共管理和服务机构应将数据库表结构发送至大数据工作主管部门，并根据数据归集方式和数据实际情况，做好数据归集前准备工作并开展数据归集。最后大数据工作主管部门应通过对账数据等对已归集数据进行核对，确保公共管理和服务机构提供的数据与已归集的数据、公共数据资源目录一致。

（4）数据归集要求：给出库表方式、文件方式、服务接口方式三种数据归集方式的技术要求。如库表归集方式中，前置库应支持国产数据库，字符编码应为 UTF-8；业务数据表除业务字段外应包含记录 ID、批次号、业务操作标识、更新时间字段等扩展信息。文件归集方式中，文件系统可采用 FTP、SFTP，并对文件的命名约束、文件大小、文件格式、文件内容等提出要求。服务接口归集方式中，规定服务接口调用形式为：<Protocol>://<Hostname>:<Port>/<URI>，服务接口资源应采用 Schema 架构说明的标准 XML 文件方式进行描述，并明确了服务接口主要包括增删改类数据服务接口和只增类数据服务接口。

(5) 数据更新要求：给数据更新需要遵守的要求，如应按照公共数据资源目录的更新频率进行数据更新。对于初始化同步、公共数据资源目录变更、定期全量备份校验或数据特性不适合增量更新等情况，可采用全量更新方式；对于每次仅归集近一段时间的数据时，可采用增量更新方式。

(6) 数据归集安全：给出数据归集中遵守的安全要求。数据归集安全应符合 GB/T 22239—2019 中等级保护三级的要求，个人信息安全应符合 GB/T 35273—2020 要求，其他安全要求应符合 GB/T 37973—2019。

(7) 附录 A：给出库表归集相关的业务数据表、对账数据表的约束及示例。

(8) 附录 B：给出文件归集相关的文件信息表、文件拆分信息表、对账数据表的约束及示例。

(9) 附录 C：给出服务接口数据示例及内容描述。

(三) 本标准主要技术内容的确定依据

1. 相关标准和技术文献

— 《山东省“十四五”数字强省建设规划的通知》（鲁政字〔2021〕128号）

— 《深化数据赋能建设“无证明之省”实施方案的通知》（鲁政办字〔2022〕50号）

— 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

- GB/T 30850.4—2017 电子政务标准化指南 部分：信息共享
- GB/T 22239—2019 信息安全技术 网络安全等级保护基本要求
- GB/T 34960.5—2018 信息技术服务 治理 第5部分：数据治理规范
- GB/T 39477—2020 信息安全技术 政务信息共享数据安全技术要求
- GB/T 38664.2—2020 信息技术 大数据 政务数据开放共享 第2部分：基本要求
- GB/T 36073—2018 数据管理能力成熟度评估模型
- GB/T 36344—2018 信息技术 数据质量评价指标
- GB/T 35295—2017 信息技术 大数据 术语
- GB/T 35589—2017 信息技术 大数据 技术参考模型
- GB/T 34952—2017 多媒体数据语义描述要求
- GB/T 33850—2017 信息技术服务质量评价指标体系
- DB37/T 3521.3—2019 政务信息资源目录 第3部分：编制指南
- DB37/T 3522—2019 政务信息资源 数据交换规范

— DB37/T 4225—2020 政务信息资源 数据服务接口规范

2. 标准编制背景及依据

2021年7月，山东省人民政府印发《山东省“十四五”数字强省建设规划的通知》（鲁政字〔2021〕128号），提出营造富有活力的数字生态，构建完善全省一体化大数据资源体系，提升数据治理能力。实施“数治工程”，推动源头治理，按照统一标准要求实现数据规范化、标准化生产，推动数据汇聚、存储、应用等各环节责任主体开展联合治理，保证数据的及时性、完整性和准确性。2022年5月，山东省人民政府办公厅印发《深化数据赋能建设“无证明之省”实施方案的通知》（鲁政办字〔2022〕50号），重点提出深化数据整合共享，依托省一体化大数据平台开展数据源头治理，分级分类推进各领域数据资源统一汇聚和服务，提出完善一体化大数据平台，提升全省一体化大数据平台数据治理和开发服务能力，按需建设省市节点“数据服务中台”，支撑各级各部门开展数据源头治理和创新应用。

3. 标准主要内容的确定

（1）标准框架的确定

一方面标准编写组调研了国家和其他省市正在起草和已发布的数据归集、数据采集、数据采集系统相关规范等标准框架，总结出数据归集通常从数据归集流程、数据归集要

求、数据更新要求、数据归集安全等几个方面进行规范描述。另一方面对我省的省市县三级一体化大数据平台中公共管理和服务机构数据归集实际情况展开调研。

结合以上调研情况，标准编写组搭建了标准框架，主要包括数据归集流程、数据归集要求、数据更新要求、数据归集安全等内容。一是通过规范数据归集的流程，主要包括明确归集范围、明确归集方式、开展数据归集、归集结果确认；二是明确数据归集三种方式的技术要求，即库表归集、文件归集、服务接口归集的技术要求，全省公共管理和服务机构能更加清晰的了解公共数据的归集方法和技术要求；三是通过规范数据更新要求，可以为全省公共管理和服务机构数据更新提供依据；四是通过规范数据归集安全要求，为公共数据归集的安全提供保障。

（2）数据归集方式和相关技术要求的确定

本标准数据归集方式和相关技术要求的确定，从理论调研、实践应用和技术发展三个维度综合考量。一是详细调研数据归集相关国家标准、行业标准、地方标准，参考国家、行业、其他省相关标准对公共数据归集方式的规定和要求；二是充分调研省市县三级一体化大数据平台归集公共管理和服务机构数据时实际使用的数据归集方式和要求；三是从技术发展的角度，增加文件归集方式、库表和文件归集方式中的对账机制，避免因对账机制、归集方式等技术要求的缺

失限制省市县三级一体化大数据平台数据归集技术的发展，以及相关工作的开展。最后，经标准编写组内部多次讨论，并与数据归集技术人员多次沟通，最终明确公共管理和服务机构向省一体化大数据平台归集数据的方式应包括库表、文件、服务接口三种方式，以及对账机制等相关技术要求。

（3）数据归集流程的确定

本标准数据归集流程的确定，是在数据归集国家标准、行业标准、地方标准充分调研的基础上，结合公共管理和服务机构向省一体化大数据平台数据归集工作的实际流程和实际工作情况，确定数据归集流程主要包括明确归集范围、明确归集方式、开展数据归集、归集结果确认。公共管理和服务机构首先需要确定归集范围，根据业务职责，编制并动态维护公共数据资源目录，按照“按需归集、应归尽归”的原则，将本机构所有可归集公共数据持续归集至省一体化大数据平台；然后结合归集数据的范围、数据传输要求等，明确数据的归集方式，主要包括库表归集、文件归集、服务接口归集三种；按照数据归集方式和数据实际情况，做好数据归集前准备工作并开展数据归集；最后是通过数据对账来确认归集结果。

（4）数据更新要求的确定

本标准数据更新要求的确定，一是在详细调研数据更新相关国家标准、行业标准、地方标准的基础上充分了解有关

数据更新的工作机制和技术要求；二是采取线上线下相结合的调研方式，对省市县三级一体化大数据平台数据更新工作情况和要求进行调研，在不限数据更新技术发展的前提下，研讨确认数据更新方式包括全量更新和增量更新，同时给出数据更新的通用要求，即公共管理和服务机构应按照公共数据资源目录中约定的数据更新频率将更新数据归集至省一体化大数据平台。

（四）本标准的实验论证情况

标准编写组基于标准研讨稿，一是在省农业农村厅、省教育厅向省级一体化大数据平台数据归集工作中，对本标准的数据归集方式和相关技术要求进行全面论证。通过标准论证，对前置机中前置库和数据表的命名要求进行调整，其命名应由英文、数字、下划线组成，不应使用中文、特殊符号。

二是在济南市、烟台市、淄博市 3 个试点市的市级一体化大数据平台数据归集中开展标准验证，主要验证了归集流程和更新要求的合理性和适用性。通过标准验证，补充增量更新的技术要求，包括时间戳字段设置、触发器设置、日志解析等。

四、与现行相关法律、行政法规和其他标准的关系

本标准遵循法律、法规，符合国家有关现行法律、法规和强制性国家标准的规定，与相关法律、法规和国家标准、行业标准相协调，无冲突。

五、重大分歧意见的处理过程、处理意见及其依据

无重大分歧意见。

六、对地方标准自发布日期至实施日期之间的过渡期（以下简称过渡期）的建议

本标准为您推荐性地方标准，建议过渡期为一个月。建议过渡期间进行本标准的宣贯培训工作，根据本标准的适用范围，将主要面向省直部门和 16 地市进行标准的培训与宣贯，采用专家讲座、系列课程、交流答疑、发放宣贯材料等方式，积极推进标准实施后的应用。标准宣贯的目的在于使相关人员能更好的理解、执行本标准，推进标准的贯彻和实施。

七、其他需要说明的内容

《山东省“十四五”数字强省建设规划的通知》（鲁政字〔2021〕128号），提出构建完善全省一体化大数据资源体系，提升数据治理能力。《深化数据赋能建设“无证明之省”实施方案的通知》（鲁政办字〔2022〕50号）重点提出深化数据整合共享，依托省一体化大数据平台开展数据源头治理，分级分类推进各领域数据资源统一汇聚和服务。《山东省大数据发展促进条例》《山东省公共数据共享工作细则（试行）》《山东省公共数据开放办法》（省政府令第344号）等最新政策文件重点给出公共数据的相关管理要求。同时，参考已发布系列国家标准GB/T 34960《信息技术服务 治理》，经讨论论证，该系列标准名称由《政务信息资源数据》

改为《公共数据 数据治理规范》。根据前期调研，以及本系列标准制定情况，标准编写组与主管部门和标准负责人深入沟通后确定数据归集是数据治理中重要且必要的一步，将标准名称“政务信息资源数据 第5部分：数据归集”调整为“公共数据 数据治理规范 第1部分：数据归集”，标准内容无修改。

