

数 字 山 东 工 程 标 准

SZSD01 0015—2024

人才数据建设和应用指南

Guidelines for talent data construction and application

2024 - 04 - 15 发布

2024 - 06 - 01 实施

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 总则 1

5 数据建设 1

 5.1 参考模型 1

 5.2 数据采集 2

 5.3 数据存储 2

 5.4 数据治理 3

 5.5 数据资源 3

 5.6 数据应用 4

 5.7 管理制度 4

 5.8 标准规范 4

6 数据安全 5

 6.1 数据传输 5

 6.2 数据存储 5

 6.3 数据管理 5

附录 A（规范性） 人才数据分类和标准化示例 1

 A.1 人才基础数据 1

 A.2 人才数据 3

 A.3 单位、高校数据 5

附录 B（资料性） 人才数据分析示例 8

参考文献 9

前 言

本文件按照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。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由济南市人才服务中心提出。

本文件由济南市大数据局归口。

本文件起草单位：济南市人才服务中心。

本文件主要起草人：高硕、纪鹏飞、宋晓倩、赵廷春、钟玉资、孙婷。

人才数据建设和应用指南

1 范围

本文件提供了人才数据建设和应用的总则、数据建设、数据应用、数据安全等方面的指导和建议。本文件适用于指导各级政务部门人才数据的建设和应用。

2 规范性引用文件

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

GB/T 22239—2019 信息安全技术 网络安全等级保护基本要求
GB/T 35273—2020 信息安全技术 个人信息安全规范
GB/T 39477—2020 信息安全技术 政务信息共享 数据安全技术要求

3 术语和定义

本文件没有需要界定的术语和定义。

4 总则

按照合法合规、安全可控、经济实用的原则，开展人才数据的建设和应用，包括数据建设、数据应用和数据安全。

5 数据建设

5.1 参考模型

人才数据建设包括数据采集层、数据存储层、数据治理层、数据资源层、数据应用层5个层次，以及标准规范、管理制度、安全保障3个支撑体系。人才数据建设参考模型见图1。

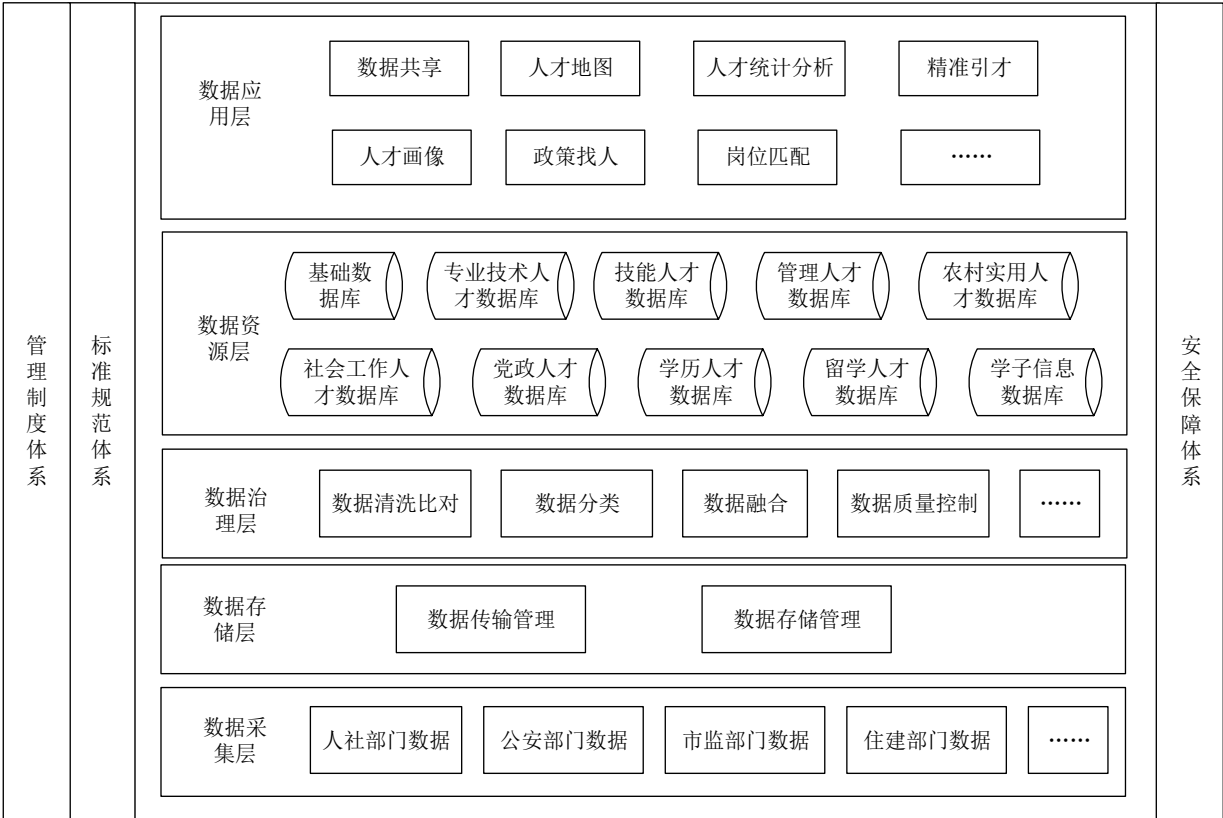


图 1 人才数据建设参考模型

5.2 数据采集

5.2.1 采集范围

人才数据涉及各部门、各领域的相关数据，包括结构化数据和非结构化数据，宜按照最小必要审慎原则采集人才数据，采集范围包括但不限于：

- a) 人社部门数据：包括社保缴费数据、考试数据、流动人员档案数据、专业技术人员继续教育数据等；
- b) 公安部门数据：包括人口基本数据、户籍数据、学历数据等；
- c) 市监部门数据：包括人才单位数据、专利数据等；
- d) 教育部门数据：包括人才高校数据等；
- e) 住建部门数据：包括人才房产数据、人才补贴数据等；
- f) 民政部门数据：包括人才社工数据等。

5.2.2 采集方式

人才数据采集方式包括但不限于：

- a) 库表方式：主动推送数据至前置节点或通过数据视图等方式主动抓取数据；
- b) 接口方式：将数据封装为接口，通过调用接口获取数据；
- c) 文件方式：通过即时通讯、FTP 服务、邮件、光盘、移动硬盘、其他介质等途径获取数据；
- d) 数据直报方式：通过人才填写表单并提交获取数据。

5.3 数据存储

人才数据存储包括数据传输和数据存储管理，建设内容见6.1、6.2。

5.4 数据治理

5.4.1 数据清洗比对

宜针对人才身份信息、学历信息、人才认定等建立数据清洗比对规则，对人才数据进行清洗和比对，人才数据清洗比对的规则包括但不限于：

- a) 人才数据清洗比对工作宜以身份证件号码作为唯一标识，宜保持18位统一身份证号编码格式，身份证件号码为15位的宜转变为18位编码格式；
- b) 国内学历人才基本信息主要来源社保数据、人口数据、考试数据和流动人员档案数据；国外学历人才基本信息主要来源人才项目申报和人工采集。
- c) 技能人才认定宜调用人社部技能人员职业资格证书查询接口获取人才技能考试情况，并统一验证人才数据，对于存在考试记录的认定为技能人才。宜逐条保留技能考试记录，当前技能人才状态以最高等级技能职业资格为准。

5.4.2 数据分类

5.4.2.1 宜按照人才分类、人才政策、人才认定模型或应用需求将数据进行分类和标准化，数据分类和标准化示例见附录A。

5.4.2.2 宜按照人才分类、单位、产业类型、人才政策类型、所在区域、职称、服务类型、个性化服务等维度建立数据标签。

5.4.3 数据融合

宜对不同来源的人才数据进行融合处理，通过业务规则使用关联、统计、挖掘等方法，对人才数据、单位数据、高校数据进行融合交互，支撑跨层级、跨部门、跨行业的人才数据应用。

5.4.4 数据质量控制

宜建立数据质量管理闭环机制，并根据数据质量特性分别制定人才数据质量控制方法和规则：

- a) 准确性：描述数据是否与其对应客观实体的特征一致，宜通过数据校验、数据清洗比对等方式保证人才数据准确性；
- b) 完整性：描述数据是否存在缺失记录或缺失字段，宜通过日志跟踪机制、标志位轮询方式等方法保证人才数据的完整性；
- c) 一致性：描述同一实体同一属性的值在不同的系统中是否一致，宜通过数据标准化方式保证人才数据的一致性；
- d) 有效性：描述数据是否满足用户定义的条件或在一定的取值范围内，宜通过有效值规则制定、数据校验等方式保证人才数据有效性；
- e) 唯一性：描述数据是否存在重复记录，宜通过设定唯一标识，只保留一条重复记录的方式保证人才数据的唯一性；
- f) 及时性：描述数据的产生和供应是否及时，宜通过时间戳字段设置、触发器设置、日志读取等数据更新方式保证人才数据更新的及时性。

5.5 数据资源

经数据治理后，宜根据数据资源现状和应用实际进行扩展建设或选择建设人才数据资源库，人才数据资源库包括但不限于：

- a) 基础数据库；

- b) 专业技术人才数据库；
- c) 技能人才数据库；
- d) 学历人才数据库；
- e) 管理人才数据库；
- f) 农村实用人才数据库；
- g) 社会工作人才数据库；
- h) 党政人才数据库；
- i) 留学人才数据库；
- j) 学子信息数据库。

5.6 数据应用

5.6.1 概述

宜建立人才信息数据应用规则，依据数据应用场景，提供数据共享、数据分析等应用，支撑跨部门、跨层级、跨行业、跨领域间的协调联动和信息共享。

5.6.2 数据应用场景

人才数据宜根据具体应用需求、应用场景进行挖掘分析，人才数据应用场景包括但不限于：

- a) 人才画像：通过人才的基础信息、学历信息、工作信息、项目申报信息、享受人才服务信息、信用信息等绘制准确率高、时效性强的人才画像；
- b) 人才统计分析：对全量人才或各类型人才进行多维度统计分析，包括人才增减员变化统计、人才项目申报数据统计、人才求职供需统计等，构建区域人才价值指数、流动指数、薪酬指数，精准掌握人才发展趋势；
- c) 人才地图：基于人才户籍地址、人才单位地址数据、人才高校地址等数据，将人才所在位置转换为经纬度，并在实体地图上进行标注并展示人才总体分布、产业人才分布以及人才流入流出情况；
- d) 政策推送：根据人才类别或人才属性，构建人才分类服务模型，向人才精准推送相关政策和项目申报通知动态等；
- e) 政策找人：通过构建政策和人才标签体系构建自动匹配政策的人才数据库，筛选符合政策但未申报的人才，经线上消息提醒、线下对接指导等方式实现政策找人；
- f) 政策支持：对人才供给、需求、引进、流动数据指标等进行监测和预警，帮助政策制定部门发现问题、预测趋势、把握规律，支撑人才政策规划建设、调整和迭代优化；
- g) 岗位匹配：通过对不同维度人才数据进行关联分析，根据人才求职意向和岗位匹配度分析结果向人才精准推送适配岗位和就业创业指导等服务；
- h) 精准引才：通过对全国高校中本地学籍生源的数量、专业领域、学历层次、毕业年限等数据进行分析，结合引才单位需求制定精准引才计划，实现从毕业院校到引才单位的精准引才。

人才数据分析示例见附录 B。

5.7 管理制度

宜制定人才数据审核和备案管理制度、运维管理制度、应急响应机制、安全管理制度等。

5.8 标准规范

宜建设人才基础数据标准、人才指标数据标准、人才元数据标准、人才主数据标准、人才数据质量标准、人才数据管理标准和人才数据应用标准等。

6 数据安全

6.1 数据传输

- 6.1.1 宜根据传输的数据类型、级别和应用场景，制定人才数据安全策略并采取保护措施。
- 6.1.2 宜采取密码技术、校验技术、安全传输通道等措施，保障数据传输过程安全、可控、及时、高效。

6.2 数据存储

- 6.2.1 宜单独划分区域存储或单独建库存储人才数据，并进行存储备份。
- 6.2.2 人才数据存储采取脱敏、加密、访问控制等安全措施。
- 6.2.3 宜存储在政务外网中，并设置最小访问权限。

6.3 数据管理

- 6.3.1 制定内部管理制度和操作规程，合理确定人才信息处理的操作权限，定期对相关人员进行安全教育和培训，制定并组织实施安全事件应急预案等，防止未经授权的访问以及人才数据泄露、篡改、丢失。
- 6.3.2 应对人才数据进行分类分级，将人才身份证件号码、护照号码等信息定义为高风险等级信息，建立基于分类分级的人才数据流通安全策略，确保在安全可控的环境下，促进人才数据流通。
- 6.3.3 开展数据脱敏脱密、数据水印、数据防泄漏、数据加密、数据去标识化、权限控制等安全技术和安全策略，确保人才数据的安全。
- 6.3.4 按照 GB/T 22239—2019、GB/T 35273—2020 的要求开展信息安全等级保护建设和个人信息安全保护建设，定期组织安全检查和监测评估。
- 6.3.5 建立数据全量和增量的安全备份管理机制，定期进行数据备份。
- 6.3.6 人才数据应用完成后，宜对数据进行封存，封存期满后对数据进行分类分级评估，采取继续封存或销毁等安全处理措施，并关闭相关数据应用。

附 录 A
(规范性)
人才数据分类和标准化示例

A.1 人才基础数据

人才基础数据见表A.1。

表A.1 人才基础数据

类别	数据项名称	数据项格式
人才基础信息	姓名	字符型
	性别代码	数值型
	身份证件类型代码	字符型
	身份证件号码	字符型
	出生日期	日期型
	民族代码	数值型
	国籍代码	数值型
	户籍(洲)代码	数值型
	户籍(国家)代码	数值型
	户籍(省)代码	数值型
	户籍(市)代码	数值型
	户籍	字符型
	政治面貌代码	数值型
	婚姻状况代码	数值型
	生存状况代码	数值型
	居住地址	字符型
	手机	字符型
	联系电话	字符型
	电子邮箱	字符型
	邮编	字符型
	最高学历	数值型
	毕业院校	字符型
	毕业时间	日期型
	学历代码	数值型
	学位代码	数值型
	专业(一级学科)代码	字符型
	专业(二级学科)代码	字符型
	专业(三级学科)代码	字符型
	职称等级代码	数值型
	技能等级代码	数值型
	当前工作单位	字符型
	工作单位统一社会信用代码	字符型
人才学习经历信	毕业院校	字符型

类别	数据项名称	数据项格式
息	入学时间	日期型
	毕业时间	日期型
	学历代码	数值型
	学位代码	数值型
	专业(一级学科)代码	字符型
	专业(二级学科)代码	字符型
	专业(三级学科)代码	字符型
	研究方向	字符型
	是否双一流高校	数值型
	是否双一流专业	数值型
	是否全球 TOP200 高校	数值型
人才工作经历信息	工作开始时间	日期型
	工作结束时间	日期型
	工作单位	字符型
	单位统一社会信用代码	字符型
	单位地点	字符型
	工作部门	字符型
	工作职责	字符型
	当前社保缴费状态	字符型
	社保机构	字符型
	社会保障号码	字符型
	首次参保日期	日期型
	是否领取养老金	数值型
	离退休日期	日期型
	实际缴费月数	数值型
人才承担科研项目情况信息	项目名称	字符型
	项目来源	字符型
	项目下达单位	字符型
	项目级别	数值型
	项目经费	数值型
	项目开始时间	日期型
	项目结束时间	日期型
	本人排名	数值型
人才知识产权产出情况信息	知识产权类别	数值型
	论文（著作）名称	字符型
	论文（著作）类型	字符型
	出版社刊物	字符型
	发表时间	日期型
	位次	数值型
	收录情况	字符型
	影响因子	字符型

类别	数据项名称	数据项格式
	他引情况	字符型
	软件著作权证书号	字符型
	软件名称	字符型
	授权国家	数值型
	著作权人	字符型
	开发完成日期	日期型
	首次发表日期	日期型
	权力范围	字符型
	登记号	字符型
	专利证书号	字符型
	专利类型	数值型
	专利名称	字符型
	发明（设计）人	字符型
	专利号	字符型
	授权国家	数值型
	专利申请日	日期型
	专利权人	字符型
	授权时间	日期型
	本人排名	数值型
	批准机构	字符型
人才获奖情况和 荣誉称号信息	奖励名称	字符型
	奖励等级	字符型
	颁奖单位	字符型
	获奖时间	日期型
	本人排名	数值型
	荣誉称号	字符型
	授予单位	字符型
	授予时间	日期型
	荣誉层级	字符型

A. 2 人才数据

人才数据见表A. 2。

表A. 2 人才数据

类别	类别定义	数据项名称	数据项格式
专业技术人才信息	专业技术人才主要包括在省内各类企业、事业单位及民办非企业单位中从事专业技术工作的人员，持有专业技术人员职业资格证书的人员。	专业技术职务系列	字符型
		职称级别	数值型
		现专业技术职称资格	字符型
		考试名称	字符型
		考试专业	字符型
		证书级别	字符型

类别	类别定义	数据项名称	数据项格式
		证书管理号	字符型
		通过年月	字符型
技能人才信息	技能人才主要包括省内各类企业、事业单位、机关及民办非企业中在工勤技能岗位工作、具有技能水平的人员，以及持有职业技能等级证书的人员。	职业	字符型
		证书等级	数值型
		证书编号	字符型
		证书核发日期	日期型
		职业方向	字符型
		理论知识考试成绩	数值型
		技能考核成绩	数值型
		评定成绩	字符型
		文化程度代码	字符型
		职业技能鉴定机构	字符型
		发证机关	字符型
		数据责任单位	字符型
学历人才信息	学历人才主要包括中专、大专、本科、硕士、博士学历人才。	入学时间	日期型
		毕业时间	日期型
		毕业学校	字符型
		毕业院系	字符型
		学历代码	数值型
		学位代码	数值型
		专业(一级学科)代码	字符型
		专业(二级学科)代码	字符型
		专业(三级学科)代码	字符型
		是否双一流高校	数值型
		是否双一流专业	数值型
		是否国外院校	数值型
		研究方向	字符型
		数据来源	字符型
管理人才信息	管理人才主要包括省内各类企业和民办非企业单位中从事经营管理工作的人员。	工作单位	字符型
		工作单位统一社会信用代码	字符型
		单位地点	字符型
		工作部门	字符型
		职务	字符型
		职称等级代码	数值型
		管理工作开始时间	日期型
		管理工作结束时间	日期型
农村实用性人才信息	农村实用人才主要包括具有一定知识或技能，为农村经济和科技、教育、卫生、文化等各项社会事业发展提供服务、做出贡献，起到示范或带动作用的农村劳动者。	所在村居	字符型
		人才类型	字符型

类别	类别定义	数据项名称	数据项格式
党政人才信息	党政人才主要包括公务员、参照公务员法管理的群团机关工作人员和试用期人员、事业单位人才等。	工作单位	字符型
		工作单位统一社会信用代码	字符型
		单位地点	字符型
		工作部门	字符型
		职务	字符型
		职称等级代码	数值型
		工作开始时间	日期型
		工作结束时间	日期型
		是否缴纳机关保险	数值型
		是否国外院校	数值型
		研究方向	字符型
		数据来源	字符型
社会工作人才信息	社会工作专业人才主要包括具有一定社会工作专业知识和技能，在相关领域直接提供社会服务的专业人员，以及持有社会工作者国家职业资格证书的人才。	职称等级代码	字符型
		工作单位	字符型
		工作单位统一社会信用代码	字符型
		证书级别	字符型
		证书管理号	字符型
留学人才信息	留学人才指通过教育部留学服务中心国外学历认证的人才。	通过年月	字符型
		毕业院校	字符型
		毕业时间	日期型
		留学国家	数值型
		学历代码	数值型
		学位代码	数值型
		专业(一级学科)代码	字符型
		专业(二级学科)代码	字符型
		专业(三级学科)代码	字符型
		职称等级代码	数值型
		工作单位	字符型
学子信息	学子人才为本地户籍，年龄 18-25 岁，通过学信网学籍查询接口验证返回数据的人才。	工作单位统一社会信用代码	字符型
		入学日期	日期型
		学制	字符型
		学籍状态	字符型

A.3 单位、高校数据

单位、高校数据见表A.3。

表A.3 单位、高校数据

类别	数据项名称	数据项格式
单位基础信息	单位名称	字符型
	统一社会信用代码	字符型
	企业（机构）类型	字符型

类别	数据项名称	数据项格式
	注册号	字符型
	注册资本	数值型
	成立日期	日期型
	登记机关	字符型
	法定代表人	字符型
	经营（业务）范围	字符型
	经营（驻在）期限自	日期型
	经营（驻在）期限至	日期型
	行业门类	字符型
	行业名称	字符型
	企业状态	字符型
	单位所在地址	字符型
	工商股东	字符型
	变更记录	字符型
	分支机构	字符型
	联系电话	字符型
	社保参保人数	数值型
	是否失信被执行企业	数值型
	企业所得税	数值型
	企业营业收入	数值型
	知识产权类别	数值型
	论文（著作）名称	字符型
	论文（著作）类型	字符型
	出版社刊物	字符型
	发表时间	日期型
	位次	数值型
	收录情况	字符型
	影响因子	字符型
	他引情况	字符型
	软件著作权证书号	字符型
	软件名称	字符型
	授权国家	数值型
	著作权人	字符型
	开发完成日期	日期型
	首次发表日期	日期型
	权力范围	字符型
	登记号	字符型
	专利证书号	字符型
	专利类型	数值型
	专利名称	字符型
	发明（设计）人	字符型

类别	数据项名称	数据项格式
	专利号	字符型
	授权国家	数值型
	专利申请日	日期型
	专利权人	字符型
	授权时间	日期型
	批准机构	字符型
	是否大型企业	数值型
	是否中型企业	数值型
	是否小型企业	数值型
	是否微型企业	数值型
	是否国有控股企业	数值型
	是否股份制企业	数值型
	是否外商及港澳台商投资企业	数值型
	是否私营企业	数值型
	是否新型研发机构	数值型
	是否高新技术企业	数值型
	是否金融机构	数值型
	是否民政服务机构	数值型
	是否医疗卫生机构	数值型
	是否境外企业（机构）	数值型
	是否规模以上工业企业	数值型
	是否有资质的建筑业企业	数值型
	是否限额以上批发和零售业	数值型
	是否限额以上住宿和餐饮业企业	数值型
	是否房地产开发经营业企业	数值型
	是否规模以上服务业企业	数值型
高校基础信息	高校名称	字符型
	高校外文名称	字符型
	地址(洲)代码	数值型
	地址(国家)代码	数值型
	地址(省)代码	数值型
	地址(市)代码	数值型
	地址	字符型
	是否双一流高校	数值型
	是否科研院所	数值型
	是否全球 TOP200 高校	数值型

附录 B
(资料性)
人才数据分析示例

人才数据分析可以从数量统计、年龄分布、性别比例、行业分布、专业分布、学历分布、流动性、区县分布、人才类型分布等维度进行分析。

以济南市人才数据行业分布为例，通过数据统计分析得出，人才集聚的前十行业分别为公共管理社会保障和社会组织、租赁和商务服务业、批发和零售业、制造业、居民服务业、信息传输软件和信息技术服务业、建筑业、科学研究和技术服务业、教育、交通运输仓储和邮政业，流入人才主要分布于社会保障和社会组织、租赁和商务服务业、批发和零售业、信息传输软件和信息技术服务业，流出人才主要分布于租赁和商务服务业、建筑业、居民服务业和批发零售业，见图B.1。

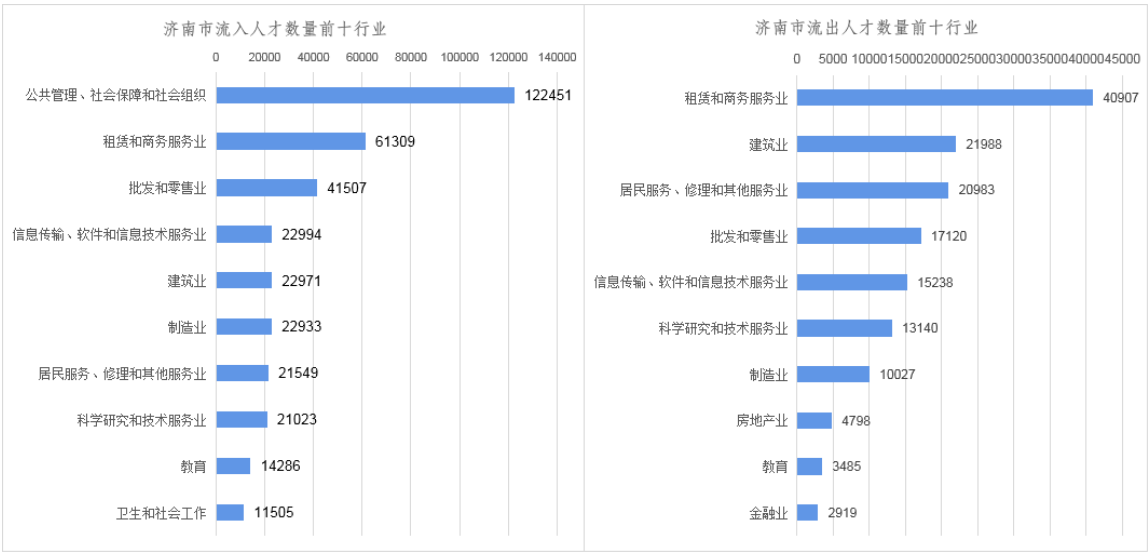


图 B.1 济南市流入人才和流出人才数量前十行业（单位：人）

参 考 文 献

- [1] 中华人民共和国. 中华人民共和国网络安全法. 中华人民共和国主席令第五十三号 (2016-11-7) [2017-6-1]. https://www.gov.cn/xinwen/2016-11/07/content_5129723.htm
- [2] 中华人民共和国. 中华人民共和国数据安全法. 中华人民共和国主席令第八十四号. (2021-6-10) [2021-9-1]. http://www.gov.cn/xinwen/2021-06/11/content_5616919.htm
- [3] 中华人民共和国. 中华人民共和国个人信息保护法. 中华人民共和国主席令第九十一号. (2021-08-20) [2021-11-1]. http://www.gov.cn/xinwen/2021-08/20/content_5632486.htm
-