

国家人工智能应用中试基地（医疗）·浙江

关于印发健康医疗数据流通管理办法（试行） 等4个数据流通制度的通知

各市卫生健康委、疾控局，省级医疗卫生单位：

根据《浙江省医疗卫生机构对外数据合作管理办法（试行）》（浙卫发〔2026〕15号）精神，为加快推动健康医疗数据资源化价值化，经省卫生健康委同意，现将《健康医疗数据流通管理办法（试行）》《健康医疗数据资产价值评估工作指引（试行）》《健康医疗数据去标识化与匿名化工作指引（试行）》《健康医疗数据流通收益分配工作指引（试行）》印发给你们，请结合实际参照执行。

- 附件：1. 健康医疗数据流通管理办法（试行）
2. 健康医疗数据资产价值评估工作指引（试行）
3. 健康医疗数据去标识化与匿名化工作指引（试行）
4. 健康医疗数据流通收益分配工作指引（试行）

国家人工智能应用中试基地（医疗）·浙江

2026年10月10日



健康医疗数据流通管理办法 (试行)

为充分发挥国家人工智能应用中试基地（医疗）（以下简称“中试基地”）的支撑能力，推进健康医疗数据资源化价值化，根据《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》、《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》、《医疗卫生机构数据安全和个人信息保护管理办法（试行）》等法律政策，制定本办法。

第一条 适用范围

中试基地是健康医疗数据要素市场化配置的支撑平台，承担全省健康医疗数据“统一归集入口、统一标准化治理、统一产品孵化、统一交易出口”的枢纽职能。本办法适用于依托中试基地开展的各类健康医疗数据流通及相关活动。

第二条 术语定义

本指引所称健康医疗数据，是指单独或者与其他信息结合后能够识别特定自然人或者反映特定自然人生理或者心理健康的相关电子数据，以及由上述数据加工处理之后得到的健康医疗相关电子数据。

本办法所称行业可信数据空间，是指由中试基地建设运营，支撑健康医疗数据资源安全流通利用的基础设施。

本办法所称数据产品，是指自然人、法人或者非法人组织对其合法获取的健康医疗数据资源，经过实质性加工和创新性劳动后形成的，可满足特定需求的数据加工品和数据服务。

本办法所称数据出域，是指经合规处置的健康医疗数据及数据产品，脱离行业可信数据空间安全边界，向外提供落地应用的行为。

本办法所称数据出境，是指向境外机构、组织、个人提供在中华人民共和国境内收集、产生的健康医疗数据及数据产品的行为。

第三条 基本原则

（一）价值释放。以需求为导向，推动各类主体协同联动，建立合规的数据开发利用、价值评估、交易流转与收益分配机制，规范数据资源化到价值化全流程，提升数据供给质量与应用质效。

（二）高效流通。整合各类资源，加强要素保障，优化服务流程，精简审批环节，在严守安全底线前提下提升数据要素流转效率，及时响应各方合规用数需求。

（三）公平公正。保障各类主体拥有依法平等获取与使用数据的权利，杜绝歧视性定价、排他性协议及其他不正当竞争行为，构建权责对等、竞争有序、诚信规范、共治共享的数据流通生态。

（四）安全可控。坚持发展与安全并重，严格落实分类分级、最小授权、按需开放管理要求，严控数据使用边界、

流转范围与共享权限，制定差异化安全策略，防范数据泄露、篡改与滥用风险。

第四条 参与主体与职责

中试基地建设主体：统筹负责中试基地基础设施和技术平台建设，制定流通、安全、治理规范，对健康医疗数据流通开展业务指导和监管。

中试基地运营机构：指经授权承担中试基地日常运营与统筹管理工作的主体，对中试基地整体运行、流通管控、权限审计追溯负主体责任。主要负责健康医疗数据流通供需对接与全流程闭环管理，统筹提供数据流通所需的运行环境、传输通道、存储资源、算力支撑、安全防护、去标识化等基础保障；协同第三方专业服务机构开展数据治理加工、权属核定、产权登记、产品交易、全链条合规核验等业务工作。

数据供给机构：指各级各类医疗卫生机构、医疗科研机构及其他合规持有健康医疗数据的主体，对源头数据的合法性、真实性、完整性、合规性负主体责任。主要负责数据治理、伦理审查、安全审核、备案、授权管理、数据提供等工作。

数据加工机构：指具备健康医疗数据治理、加工、标注、产品开发专业能力的技术服务机构，对加工成果的质量、合规性、安全性负主体责任。主要负责在中试基地环境内，依规开展数据治理、匿名化处理、数据产品开发等工作。

数据使用机构：指申请使用中试基地健康医疗数据产品的医疗卫生机构、高校、科研机构、企业等用数主体，对数

据产品使用合规性、场景适配性、使用边界管控负主体责任。严格按照协议约定的场景、范围、权限合规使用数据产品，全面落实数据安全、保密、防泄露要求，不得擅自复制、扩散、传播、篡改数据，不得反向推导、还原自然人原始个人信息。

第三方专业服务机构：指依法合规提供数据交易撮合、合规评估、权属登记、匿名化认证、数据质量核验、价值评估、法律咨询、公证鉴证等服务的专业机构，包含数据交易机构、律师事务所、第三方检测认证机构、评估机构、登记服务机构等，对其出具的专业报告、认证结果、评估结论的真实性、合规性、有效性承担相应法律责任。

第五条 供需对接

中试基地运营机构、数据供给机构、数据使用机构等按照各自职责，开展数据资源发布、用数需求申报、供需匹配工作，全流程落实数据分类分级管控要求。

（一）数据资源发布。数据供给机构经合规审查后，通过中试基地门户与数据交易机构双渠道发布资源信息，同步提交数据资源清单。中试基地运营机构对供给资源开展合规性初核，符合要求的纳入中试基地健康医疗数据资源统一管理，用于供需撮合对接。

（二）用数需求申报。数据使用机构向中试基地提交用数申请，中试基地运营机构在 5 个工作日内完成资质与场景合规性前置审核，审核通过后开放样本数据探查权限，由数据使用机构核验数据质量与场景匹配度，确认最终需求方

案。

（三）供需匹配。中试基地建设主体指导中试基地运营机构，依托基地数据资源目录开展数据供需撮合，组织供需双方制定实施方案，必要时组织专家论证。中试基地建设主体依规向省卫生健康委报备数据合作情况。

第六条 数据汇聚

中试基地运营机构、数据供给机构各司其职、协同联动，规范开展健康医疗数据前置预处理、安全传输归集、入库核验及密态存储工作，落实隔离存储、分级防护等安全管控要求，保障入库数据真实有效、安全可控。

（一）数据预处理。数据供给机构按照实施方案，完成数据的质量筛查、初始去标识化处理，通过专线专网、可信设备，将数据集定向归集至中试基地可信数据空间。

（二）入库核验。中试基地运营机构对归集入库的数据开展核验，核验通过的数据集纳入中试基地，实行专库管理。

第七条 数据清洗加工

中试基地运营机构统筹治理加工任务，组织数据加工机构在可信数据空间内开展数据去重、异常处理、结构化转换、多源数据关联整合、匿名化等标准化治理工作，形成标准化基础数据资源，纳入中试基地统一管理。

第八条 数据开发利用

数据加工机构依托中试基地基础设施，对文本、图片、影像等多模态数据开展精细化标注等工作，构建细分领域专科专病数据集，开发形成高质量、可复用的场景化衍生数据

产品。中试基地运营机构组织第三方专业服务机构开展匿名化合规检测并出具正式报告，按规定纳入数据产品目录。

第九条 数据产品登记

中试基地运营机构协同各参与主体开展数据权属梳理、资料归集等工作，清晰界定数据权属，同时组织第三方专业服务机构开展合规评审和数据产权登记工作，核验全流程合规性，并出具医疗健康数据产品流通登记证明、数据产权登记证书等。

第十条 数据产品交易

中试基地运营机构组织第三方专业服务机构，依据中试基地数据流通价值评估相关指引开展数据价值评估，作为交易定价参考依据。供需双方签署标准化数字合约或书面交易合同，明确交付标准、使用范围、安全责任、交易金额与收益分配、违约责任、争议处置等核心条款，全程留痕归档。

第十一条 数据产品交付

数据产品以域内交付模式为主，需出域的实行“一事一议”审批。

（一）域内交付。中试基地运营机构对数据使用机构资质、使用场景进行复核后开通对应权限，监控域内数据应用行为，对超范围、超权限、异常访问行为实时预警、即时阻断，所有操作日志留存期限不少于 3 年。

（二）出域审批。数据产品确需出域的，由数据使用机构向中试基地运营机构提交书面出域申请，详细说明出域必要性、具体使用场景、使用期限、接收方安全防护能力及保

密管控措施。中试基地运营机构组织安全评估，评估通过后，报中试基地建设主体核准，通过专属安全通道定向交付。涉及数据出境的，严格执行国家及浙江省数据出境相关管理规定，依法完成数据出境安全评估、备案等法定程序。

第十二条 收益分配

各参与主体严格遵循“谁投入、谁贡献、谁受益”原则，按照中试基地健康医疗数据流通收益分配相关工作指引，依规开展数据流通交易收益核算、比例拆分、对账结算、台账归档等工作，建立权责清晰、分配规范、公平合理、全程可溯的收益分配机制，保障数据流通收益分配合规透明、账目完整、有据可查。

第十三条 安全保障与风险管控

各参与主体应建立健全数据安全管理制度，配备对应安全管理人员，采取加密、身份认证、权限管控、操作审计等必要安全技术措施，建立风险监测预警与应急处置机制。中试基地运营机构应建立覆盖全流程的审计体系，同时建立常态化报送机制，按月向中试基地建设主体报送数据流通运营报告，发生重要数据外泄、大规模个人信息泄露等重大安全事件及其他重大事项，应第一时间上报省卫生健康委。中试基地建设主体在省卫生健康委指导下，建立跨部门协同监管机制，开展动态合规核查、受理违规线索，督促主体限期整改，规范各类数据流通行为，保障健康医疗数据合规有序流转。

附：数据流通流程图



健康医疗数据资产价值评估工作指引 (试行)

一、总则

(一) 编制目的与依据

为规范国家人工智能应用中试基地(医疗)(以下简称“中试基地”)健康医疗数据资产价值评估工作,建立科学合规的健康医疗数据价值评估体系,根据《国家数据局等部门关于促进企业数据资源开发利用的意见》、《信息技术 大数据 数据资产价值评估》(GB/T 46353—2025)、《高质量数据集质量评测规范》(TC609-5-2025-04) 等文件,制定本指引。

(二) 适用范围

本指引适用于中试基地范围内开展健康医疗数据合规流通所涉及的数据供给机构、数据使用机构、第三方专业服务机构等主体,用以指导健康医疗数据资产价值评估工作。

(三) 核心术语定义

1.数据资产价值评估: 相关专业机构及其专业人员遵守法律、行政法规、标准和行业准则,根据委托对评估基准日特定目的下的数据资产价值进行评定和估算,并出具资产评估报告的专业服务行为。

2.收益期限：数据资产在经济寿命年限内持续发挥作用并产生经济利益流的期限。结合健康医疗数据迭代特性明确参考标准：普通慢病标准化数据集收益期限为 5-6 年；肿瘤、罕见病长期随访队列、专病科研数据集收益期限为 8-10 年。

3.重置成本：评估基准日当前市场条件、合规监管及行业标准要求下，重新完成合法采集、合规处理、标准化加工、安全管控，形成与被评估数据资产功能等效、质量相当、合规状态一致的健康医疗数据产品所需的全部合理成本。涵盖前期调研、数据合规采集、去标识化与匿名化处理、清洗标注、存储运维、安全防护、合规审查、软硬件及算力分摊、专业人工成本、第三方技术服务费、相关税费及必要合规成本等全周期投入。

二、数据价值影响要素评价

从数据质量、数据生成、数据应用三大核心维度构建量化评价体系，评价结果为数据资产价值评估所涉及的核心系数确定提供依据。

1.数据质量要素评价：从准确性、一致性、完整性、规范性、时效性、可访问性等六个维度评价数据质量，通过加权计算得出数据的质量系数，为数据资产价值评估提供核心基础依据。

2.数据生成要素评价：围绕数据全生命周期资源投入、管理治理、安全管控等开展综合评价分析，体现数据资产的生成成本与合规程度，作为重置成本核算和合规风险判定的主要依据。

3.数据应用要素评价：结合数据使用范围、应用场景以及供求关系、产品化程度等市场化属性，评估数据资产的应用潜力与变现价值，作为流通系数、稀缺系数等市场类参数的主要测算依据。

三、价值评估方法与测算模型

结合数据应用场景，选用成本法、市场法、收益法等估值方法。

（一）核心测算模型（成本法）

鼓励采用成本法开展健康医疗数据资产价值评估，评估结果可在数据流通中作为价值参考。公式如下：

$$\text{数据资产估值} = \text{重置总成本} \times \text{质量系数} \times \text{流通系数} \times \text{稀缺系数} \times (1 - \text{合规风险减值系数}) \times \text{容量系数} \times \text{期日调整系数} \times \text{剩余经济寿命系数}$$

注：公式内全部系数取值均为无量纲系数；若某系数实际条件不适用，按基准值 1 参与计算。

（二）核心参数解释与取值规范

1.重置总成本：归集数据集全周期投入成本，含规划咨询、采集建设、软硬件部署、算力支撑、日常运维、人工治理、数据脱敏、第三方技术服务等，按数据集实际占用资源比例分摊。成本归集口径按照财政部《企业数据资源相关会计处理暂行规定》要求。

2.质量系数：取值区间为 0-1，基于质量系数由准确性、完整性、一致性、时效性、规范性、可访问性等要素评价的核心指标加权测算得出，质量越高系数越趋近于 1。质量系

数小于 0.6 的数据集，质量不达标，不予开展价值评估及价值转化工作。

3.流通系数：仅数据供给机构内部使用、不对外共享流通的数据集取值 0.3；可对外开展合规科研共享、有限商用的数据集取值 0.7-0.9；可公开流通交易的数据集取值 0.9-1.0。

4.稀缺系数：肿瘤、罕见病、特殊专科等稀缺性数据集取值 1.3-1.8；常见检验、影像数据集取值 0.9-1.2；普通常见病、慢病标准化数据集取值 0.5-0.8。

5.合规风险减值系数：高敏感健康医疗数据集取值 0.35；受限科研数据集取值 0.2；公开脱敏数据集取值 0.05。

6.容量系数：常规诊疗数据以万条基础数据为基准，基准规模系数为 1，根据数据集服务器资源占用率、数据体量逐级递增；罕见病、专病队列数据以区域同类数据集平均规模为基准，基准系数为 1，结合实际数据体量、样本质量上下浮动。

7.期日调整系数：如果重置成本取自往期历史造价或往年数据采购合同，需要补充日期修正。当年内评估、重置成本为当期发生，期日系数直接赋值 1；1-2 年取值 1.04-1.08；2-3 年取值 1.08-1.12；3 年以上取值 1.12-1.2。若行业下行、数据要素贬值，则下修系数，最低不得低于 0.9。

8.剩余经济寿命系数：剩余经济寿命与总使用年限比值。基因组类数据集核心信息长期稳定，价值衰减缓慢，剩余经济寿命系数取值较高；医学影像等时效性较强的数据集价值衰退较快，剩余经济寿命系数取值偏低。

（三）交叉验证要求

满足以下任一条件的数据集，鼓励采用两种及以上估值模型开展交叉验证：

- 1.成本法预估价值超过 500 万的高价值数据集；
- 2.拟开展场内交易、作价入股、资产质押的合规数据产品；
- 3.稀缺性高、科研附加值高、市场争议度大的专病队列数据集。

多模型测算结果差异超过 20%的，必须全面复核参数取值、成本归集、指标评分、合规判定依据，补充核查佐证材料，重新测算修正，直至结果合理可控，确保估值公允精准。

四、数据价值评估实施流程

遵循标准化、合规化、可审计评估原则，对委托、审查、确权、核验、建模、测算、出证实施全流程闭环管控，切实做到权属清晰、流程合规、指标科学、结果真实有效。

（一）评估委托

数据供给机构作为委托方，向具备合法资质的第三方专业服务机构提交评估申请，明确评估目的、评估基准日、评估范围、应用场景，同步提交数据权属证明、伦理批复、知情同意书、合规备案文件、质量溯源记录等全套基础材料，对材料真实性、完整性、合法性负责。

（二）合规性审查

对拟评估数据集开展合规性审查，未通过审查的数据集，终止评估流程。

（三）资产边界梳理

全面梳理待评估数据集的权属、业务用途、数据来源、覆盖范围及字段构成，划定数据应用边界，明确评估对象与评估范围，排查权属争议与合规风险隐患。

（四）数据质量全维度核验

由数据供给机构参照《高质量数据集 质量评测规范》（TC609-5-2025-04）实施标准化核验与打分，各项指标参考权重设置如下：结构完整性 0.15、内容真实性 0.15、内容一致性 0.14、内容干净性 0.14、格式规范性 0.12、标注规范性 0.12、安全规范性 0.09、类型一致性 0.09。同步开展数据清洗、匹配、纠错与优化工作，编制形成《数据质量核验报告》。

（五）评价模型搭建与指标校验

结合数据类型、应用场景、流通属性，择优选取适配评价指标，搭建分层量化评分模型，开展指标权重校验、数据匹配核验与模型试运行，确保评价体系契合评估工作需求。

（六）价值测算与结果核验

依托核心测算模型完成价值初步核算，对高价值数据开展多模型交叉验证，结合合规风险、市场稀缺性、应用潜力完成结果修正，形成正式《数据资产估值报告》，经数据供给机构、第三方专业服务机构共同确认作为流通交易、资产入表、科研转化的合规依据。

五、价值评估安全保障

严格落实健康医疗数据全生命周期安全管理要求，强化作业人员、工具系统、作业环境的安全管控，防范隐私泄露、

数据滥用及各类安全违规风险。

1.人员安全管控：所有参与数据评估、数据处理、成果编制的工作人员，应签订保密协议，严格遵守医疗数据隐私保护与安全管理规定，严禁泄漏患者个人信息、涉密数据及数据资产评估核心信息。

2.工具系统安全：针对评估工作所使用的系统平台、测算工具、存储设备，应开展事前安全漏洞检测、事中风险监测、事后安全审计，防范系统漏洞、工具后门引发的数据安全风险。

3.档案全程留存：针对评估全过程操作记录、原始数据底稿、测算过程资料、评价报告、审批文件等全部材料实行电子化、纸质双存档，实现价值评估全程可追溯、可核查。

六、评估结果应用与管理

完成合规估值的健康医疗数据资产，可依法依规开展市场化运作、内部管理应用、科研成果转化等合规业务，充分释放数据要素价值。

（一）流通交易定价参考

评估结果作为健康医疗数据合规流通的公允定价依据，为交易双方定价磋商、合同签订、价款结算提供标准化参考，规范健康医疗数据交易市场秩序。

（二）科研成果转化应用

以正式评估结果为基准，签订数据授权使用、成果转化合作协议，明确数据授权费用、收益分配比例、成果权属划分，规范医疗科研数据、AI数据产品的价值化转化流程。

（三）内部数字化考核管理

将数据供给机构各专项数据集的资产估值结果、质量评价等级、价值转化成效，纳入数据供给机构数字化建设、数据资产管理、科研创新工作考核指标，助力单位内数据资产精细化管理。

（四）价值动态重估管理

建立数据资产价值动态迭代管理机制，原则上每两年开展一次全面价值重估。出现数据集规模大幅扩容、数据质量显著提升、应用场景重大拓展、合规管控体系优化、市场价值大幅波动等情形时，应及时启动专项重新评估，确保数据资产价值与实际状态、市场水平、合规要求精准匹配。

（五）示范推广与标准输出

依托中试基地实践经验，持续优化健康医疗数据资产价值评估指标体系、测算模型、合规流程，总结形成可复制、可推广的健康医疗数据资产价值评估案例，为全国健康医疗数据要素市场化配置、行业标准完善、地方政策迭代提供实践支撑与示范参考。

附：健康医疗数据资产价值评估案例

案例一 影像数据集评估

依托 2024 年购买的影像设备采集形成 10000 条影像学资料，采用成本法开展数据资产价值测算。

1.重置总成本：将软件、硬件设备、人力、网络、电力、算力、数据脱敏治理等成本及税费、合理利润统一归集，按资源耗用占比分摊至本数据集。

硬件设备分摊：260 万元；

软件投入（数据库、脱敏软件、数据管理平台等）分摊：140 万元；

算力、备份处理、网络、电力能耗成本分摊：100 万元；

脱敏治理专项支出分摊：40 万元；

相关税费及合理利润：30 万元；

合计 570 万元。

2.质量系数：1。

3.流通系数：对外科研共享，取值 0.7。

4.稀缺系数：属常见检验、影像数据，取值 1。

5.合规风险减值系数：受限科研数据，取值 0.2。

6.容量系数：常规万条级数据集，系数为 1。

7.期日调整系数：重置成本发生于 2024 年，期日系数赋值 1.12。

8.剩余经济寿命系数：剩余经济寿命系数=剩余经济寿命

÷总使用年限。医学影像数据受影像设备迭代、临床诊断标准更新影响，数据效用衰减相对较快，总使用年限为 8 年，剩余经济寿命 5 年，剩余经济寿命系数 0.625。

计算公式：

$$\begin{aligned}\text{数据资产估值} &= 5700000 \times 1.12 \times 0.7 \times 1.0 \times (1 - 0.2) \times 1 \times 0.625 \\ &= 223.44 \text{ 万元}\end{aligned}$$

案例二 病历文本数据集评估

基于医院 HIS 系统生成 10000 条病历文本数据集，采用成本法开展数据资产价值测算。

1.重置总成本：将软件、硬件设备、人力、网络、电力、算力、数据脱敏治理等成本及税费、合理利润统一归集，按资源耗用占比分摊至本数据集。

硬件设备分摊：38 万元；

软件投入（数据库、脱敏软件、数据管理平台等）分摊：24 万元；

算力、备份处理、网络、电力能耗成本分摊：22 万元；

脱敏治理专项支出分摊：12 万元；

相关税费及合理利润：4 万元；

合计 100 万元。

2.质量系数：1。

- 3.流通系数：对外科研共享，取值 0.7。
- 4.稀缺系数：属常见检验数据，取值 1。
- 5.合规风险减值系数：受限科研数据，取值 0.2。
- 6.容量系数：常规万条级数据集，系数为 1。
- 7.期日调整系数：重置成本于本年度发生，期日系数赋值 1。

8.剩余经济寿命系数：剩余经济寿命系数=剩余经济寿命÷总使用年限，总使用年限为 8 年，剩余经济寿命 7 年，剩余经济寿命系数 0.875。

计算公式：

$$\begin{aligned}\text{数据资产估值} &= 1000000 \times 1 \times 0.7 \times 1 \times (1 - 0.2) \times 1 \times 1 \times 0.875 \\ &= 49 \text{ 万元}\end{aligned}$$

健康医疗数据去标识化与匿名化工作指引 (试行)

一、总则

(一) 编制目的与依据

为规范国家人工智能应用中试基地（医疗）（以下简称“中试基地”）范围内健康医疗数据去标识化、匿名化全流程操作，明确标准化技术流程、风险管控与效果评估要求，根据《中华人民共和国个人信息保护法》《信息安全技术 个人信息去标识化指南》（GB/T 37964—2019）、《信息安全技术 个人信息去标识化效果评估指南》（GB/T 42460—2023）、《高质量数据集质量评测规范》（TC609-5-2025-04）等法律政策，制定本指引。

(二) 适用范围

本指引适用于中试基地范围内健康医疗数据的去标识化和匿名化处理及评估，为健康医疗数据合规流通与利用筑牢基础保障。

二、去标识化与匿名化路径选型

开展健康医疗数据去标识化或匿名化处理，应结合处理目的、数据可识别程度、使用环境和合法性基础进行需求分析，确定处理路径。不具备处理必要性的，应调整处理需求

或终止流程；具备处理必要性的，需进一步结合业务目的进行个人信息处理。路径选型流程如图 1 所示。

确需识别个人信息的，应在处理前依法履行告知义务，采用取得同意等合规方式，制定并组织实施去标识化处理策略。经去标识化效果评估合格的，即完成个人信息去标识化处理；评估不合格的，应调整处理策略后重新开展处理。

无需识别个人信息的，应采用匿名化处理，制定并组织实施匿名化处理策略。经匿名化效果评估合格的，即完成个人信息匿名化处理；评估不合格的，应调整处理策略后重新开展处理。

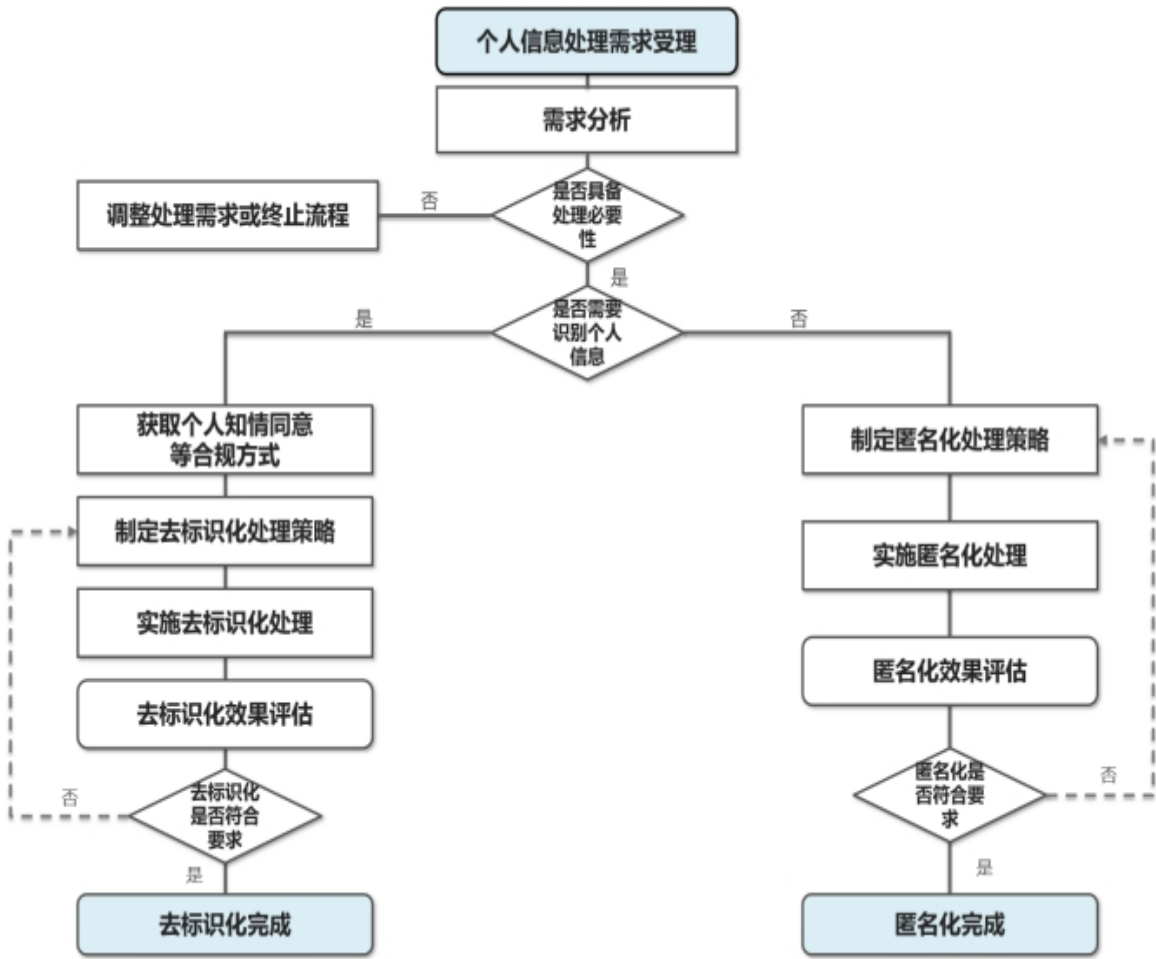


图 1 数据去标识化与匿名化路径选型流程

三、数据去标识化处理

参照GB/T 37964—2019 实施数据去标识化，核心目标是消除直接标识符，对准标识符进行处理，将重识别风险控制在可接受阈值内。具体处理流程如图 2 所示。

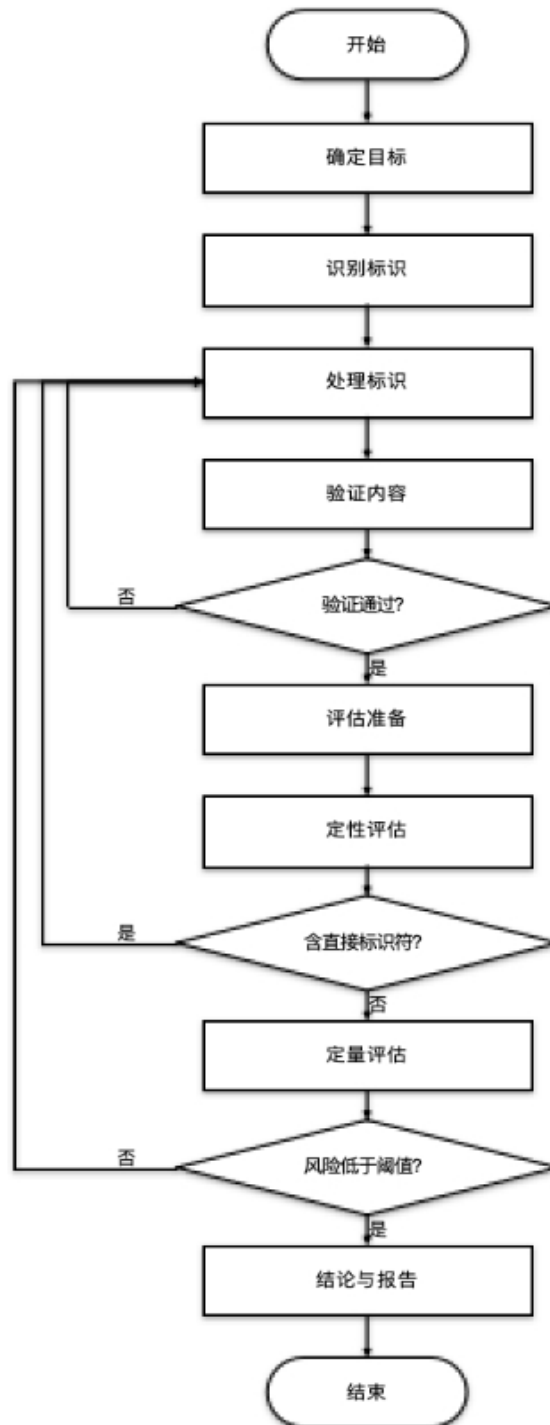


图 2 数据去标识化处理流程

（一）前置工作

1.需求分析：明确使用环境、提供方式、数据范围、标识符，及处理后数据应满足的使用需求。

2.合法性审查：按国家及浙江省相关法律法规和标准，开展数据分类分级、个人信息保护影响评估、统一授权审查。以个人同意作为处理合法性基础的，应核验告知与同意获取材料的完备性；涉及敏感个人信息的，应取得单独同意。

3.策略制定：确定去标识化目标，参照附录A确定个人信息标识度目标等级；确定去标识化处理技术，参照附录B选择适宜的技术与模型；确定风险阈值，结合流通情形参照附录C明确模型参数，并计算风险阈值。去标识化处理策略应兼顾数据安全与数据可用性。

（二）处理实施

去标识化处理应按照处理标识、验证内容、效果评估的流程执行。

1.处理标识

（1）预处理：规范化数据以保证一致性，必要时抽样减少数据量。

（2）实施去标识化：按照策略制定阶段确定的去标识化技术和参数开展去标识化处理。存在多项标识符需处理的，应结合数据特征和业务特点设定处理顺序，有序实施并生成结果数据集。

2.验证内容

验证内容包括验证个人信息安全与数据可用性。

（1）验证个人信息安全：检查去标识化结果是否达到预定目标，分析残余风险；计算重识别风险并与风险阈值比对。

（2）验证数据可用性：对比处理前后的样本量、分布、缺失率等指标，确认数据仍满足预定用途。

具体验证方法参照GB/T 37964—2019 执行。验证未通过的，返回调整参数并重新开展处理。

（三）效果评估

参照GB/T 42460—2023，去标识化效果评估按评估准备、定性评估、定量评估、形成评估结论的流程执行。

1.评估准备

确定待评估的数据集及其使用环境；组建评估团队，宜包括个人信息保护合规专家、去标识化技术专家、卫生健康业务专家；制定评估方案，包括评估依据、重识别风险计算方法及判定阈值。判定阈值应取处理实施前确定的风险阈值，不得在现场重新确定。

2.定性评估

参照GB/T 37964—2019 识别标识符，结合字段语义逐字段人工研判，确认数据集是否含有标识符。经研判仍含直接标识符的，评定为不合格，退回重新处理；不含直接标识符，仅含准标识符的，进入定量评估。

3.定量评估

参照GB/T 42460—2023 计算重识别风险：

（1）确定每行记录的重标识风险。按准标识符属性值对数据行分组，每行记录的重标识风险为其所在等价类记录数的倒数。

（2）计算数据集重标识风险。取各记录重标识风险的最大值与平均值，分别为最大值风险与平均值风险。小规模数据集及受控流通从严场景采用最大值风险作为判定口径，大规模发布场景可采用平均值风险，选定口径应在评估方案中载明。

（3）计算环境重标识风险概率。按流通情形，对内部故意攻击、数据集包含熟人、数据泄露三类概率取最大值，具体取值由评估方结合本地外部数据环境确定。

（4）计算重标识总体风险。总体风险为数据集重识别风险（按选定口径）与环境重识别风险概率之积。总体风险与风险阈值比较：低于阈值的，评估通过；高于或等于阈值的，重新处理后复评。

4.形成评估结论

综合定性评估与定量评估结果，形成评估报告。报告可注明参照GB/T 42460—2023 评定的标识度等级，作为风险水平的描述性信息。

四、数据匿名化处理

匿名化处理目标是实现处理后数据无法识别特定自然人且不能复原，不再构成个人信息。具体处理流程如图3所示。

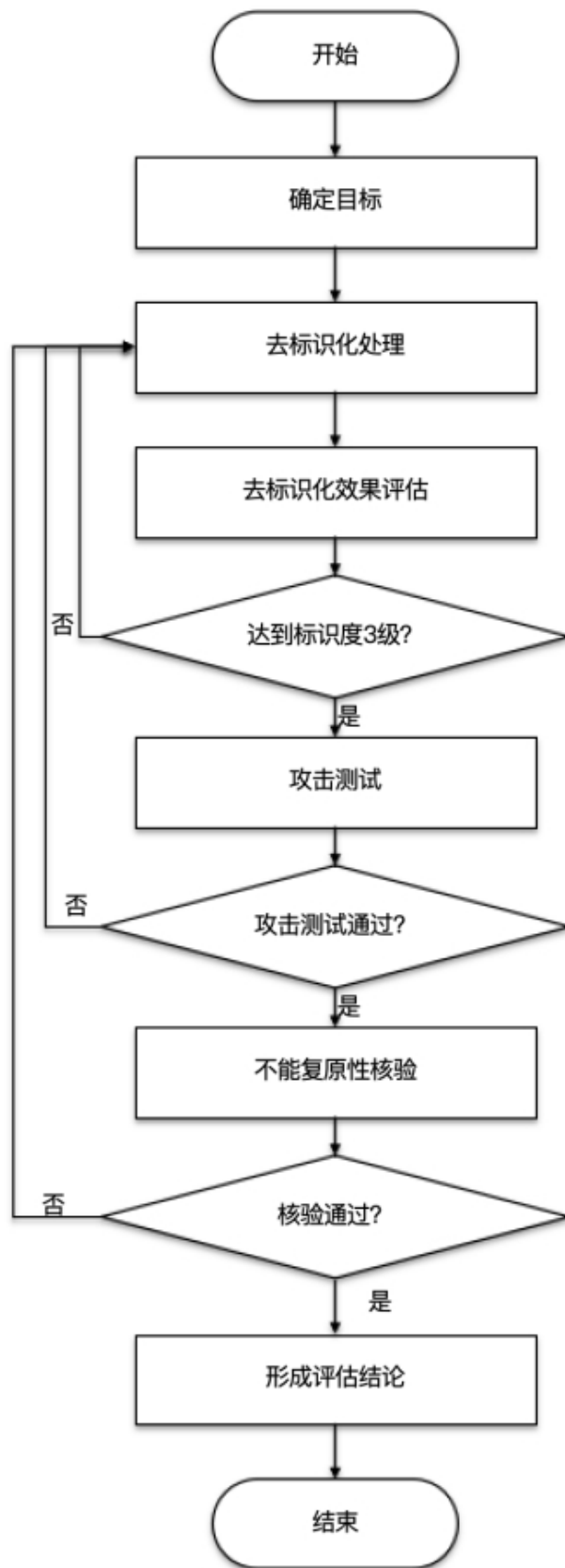


图 3 数据匿名化处理流程

（一）前置工作

1.需求分析：明确数据使用目的、公开范围与输出形式，论证匿名化处理后数据仍能满足预设用途。经论证无法满足的，应调整数据范围或流通方式。

2.合法性审查：匿名化处理本身属于个人信息处理活动，处理前应完成个人信息保护影响评估。

3.策略制定：根据应用场景确定匿名化处理技术，参照附录B选择适配的技术方法，不同技术方法可组合使用；确定数据可用性，处理后数据仍应满足预期用途。

（二）处理实施

实施去标识化处理，结果应达到个人信息标识度第3级。涉及全生命周期多模态数据关联需求的，应在匿名化技术处理前完成关联。

（三）效果评估

匿名化效果评估以去标识化处理结果达到个人信息标识度第3级为前提，依次开展攻击测试与不能复原性核验。两项核验均通过的，方可视为完成匿名化处理。

1.攻击测试

模拟攻击者利用合法可得的外部辅助数据，对结果数据尝试重标识（链接攻击、成员推断、属性推断）。测试按以下步骤实施：

（1）定义模拟攻击者，明确数据来源、动机、技能与资源；

（2）识别可辅助重识别的关键变量；

(3) 收集合法可得的外部数据;

(4) 实施攻击尝试;

(5) 判定结果。

出现下列情形之一的，攻击测试不通过:

(1) 重识别风险超过预设阈值;

(2) 存在可稳定区分、可链接的持久标记，可支持对同一人实施差异化影响;

(3) 敏感属性推断或成员推断风险增量超过阈值。

攻击测试不通过的，返回重新处理。

3.不能复原性核验

从以下三个方面核验，三者同时满足的，核验通过:

(1) 技术不可逆: 处理方法层面不存在可逆映射，或已执行不可逆销毁并留存处置证据;

(2) 关键辅助材料不可得: 还原所需的密钥、对照表等辅助材料已销毁或隔离至无法获取的状态，且控制措施强度不低于国家密码管理相关要求;

(3) 环境不可越权: 原始数据区、处理区、结果区有效隔离，访问控制与日志审计机制有效运行，结果环境无法直接访问原始数据或关键辅助信息。

任一方面不满足的，核验不通过，返回重新处理。

4.形成评估结论

攻击测试与不能复原性核验均通过的，由具备专业能力的第三方评估机构出具匿名化评估报告。报告出具前实行复核，出具人员与复核人员相互独立。

五、安全要求

数据供给机构应监督数据加工机构的处理过程，查验相关处理记录，持续跟踪外部数据环境、数据共享范围或使用目的发生的变化，对处理形成的结果数据开展常态化风险监测，并定期开展重新评估，以应对重标识技术和攻击手段的迭代演变。当风险水平显著升高时，应及时采取重新加工、暂停数据供给等处置措施。

数据加工机构应在隔离专用环境内开展数据处理工作，实现分区管控。接触原始数据的人员应保持相对固定，完成信息登记并签署保密承诺。除约定期限允许留存的情形外，不得私自保存原始数据及中间产物，可逆处理要素按与数据供给机构的约定进行处置。数据处理全流程应完整留痕，形成可追溯文档。

中试基地运营机构应提供满足隔离管控与审计要求的处理环境，配套相应安全防护能力。如发现违反本指引的数据处理行为，应采取暂停、阻断等操作，并同步告知数据供给机构。

附录A：个人信息标识度分级

分级标准：依据《信息安全技术 个人信息去标识化效果评估指南》（GB/T 42460—2023），个人信息标识度分为4级：

表A-1 个人信息标识度分级

等级	判定条件
1 级	含直接标识符，在特定条件下可直接识别个人信息主体
2 级	不含直接标识符但含准标识符，重识别风险高于或等于可接受风险阈值
3 级	不含直接标识符但含准标识符，重识别风险低于可接受风险阈值
4 级	不含任何标识符

附录 B：通用技术方法

列出去标识化和匿名化处理中常用的技术与模型，具体操作细节可参照GB/T 37964—2019。

表B-1 通用技术

技术名称	技术说明	适用场景	注意事项
属性删除	删除无需保留的标识符字段（如姓名、身份证号、病历号等）	数据集中包含姓名、身份证号、病历号、联系方式等直接标识符，且这些字段对当前分析目标无贡献；适合对外共享前的第一道清理。统计分析场景，统计区域Top10 发病率，与患者姓名身份证、病历号无关，可以直接删除无关属性。	确保删除后不影响分析目标
可逆假名化	使用国密算法（如SM4）将标识符加密转换为假名，在获取算法和密钥后可将假名还原为原标识符	需要在不同机构、不同时间点之间关联同一患者记录，且后续经授权可还原身份。如医联体数据互通、多中心随访、需要回溯核对的受控数据交换。	密钥须独立安全管理；处理后仍属个人信息
不可逆假名化	使用密码杂凑算法（如SM3）将标识符单向转换为无法还原的假名	仅需在同一数据集内保持个体记录可关联，但无需还原真实身份。如科研数据集发布、回顾性研	取值种类少的字段（如性别）须配合泛化或加

		究、模型训练集构建。	盐，防彩虹表攻击
泛化	降低数据精度，将精确值替换为范围值或更高层级的概念值	需要保留数据统计特征但降低个体可识别性。如将年龄转为年龄段、精确日期转为年份或月份、详细地址转为地市/区县。	平衡泛化粒度与数据可用性，避免过度泛化导致数据失效
扰动	对数值施加微小随机变化	数值型字段需保留整体分布但削弱精确对应关系。如入院日期随机偏移、检验数值添加微小噪声、年龄微调。	控制扰动幅度，过小则保护不足，过大则失真
聚合	将个体记录汇总为统计值（计数、均值、中位数等）	仅需输出群体统计结果、不涉及个体记录。如疾病发病率统计、区域健康指标汇总、药品使用量分析。	分组样本量小于5时，可能被反向推断出个体值，应审慎使用
数据合成	基于原始数据统计特征人工生成模拟数据，不体现任何特定记录	需要保留原始数据的统计分布和变量关系，但不使用任何真实患者记录。如应用开发测试、算法验证、教学演示。	—
随机化	随机修改数据属性值（数据扰动/	需要保留数据整体统计分布特征，同时使个体记	—

	噪声添加、数据置换），使处理后值区别于真实值	录值区别于真实值以降低重识别风险。如数据扰动、记录置换后的受控发布。	
--	------------------------	------------------------------------	--

表B-2 通用模型

模型名称	技术说明	适用场景	注意事项
K-匿名	确保每条记录与至少K-1条其他记录在准标识符组合上不可区分	适用于以表格形式逐条对外提供、且准标识符组合可能识别特定个体的个体级数据集（如患者明细表）。	推荐 $K \geq 5$
L-多样性	在K-匿名基础上，要求每个等价类中敏感属性至少包含L个不同值	在已满足K-匿名的数据集中，敏感属性（如疾病诊断、用药记录）取值单一、易被推测。如防止通过等价类推断出某患者所患疾病。	与T-接近性择一或并行使用
T-接近性	在K-匿名基础上，要求每个等价类中敏感属性的分布与整体数据集的敏感属性分布之间的差异不超过阈值T	在已满足K-匿名的数据集中，敏感属性分布与整体分布差异过大、可能被分布攻击。如需要发布更接近真实总体分布的医疗统计数据。	与L-多样性择一或并行使用
差分	在分析结果中添加	统计发布和机器学习的	参照 GB/T

隐私	加经校准的随机噪声，防止推断个体记录是否在数据集中	高保障方案	37964—2019 附录B.2
----	---------------------------	-------	------------------

表B-3 常见标识符处理方法

标识符	可用方法	适用说明
姓名	属性删除、泛化（如“张某某”）、随机化（随机替代）、不可逆假名化、数据合成	需保留姓氏特征时用泛化；需保持个体关联时用不可逆假名化；开发测试场景可用数据合成生成模拟姓名
身份证号	属性删除、泛化（保留地址码或出生年代段）	按地域或年代统计分析时，可泛化保留地址码或出生日期段
银行卡号	属性删除、泛化（保留前6位或后4位）、可逆假名化	需校验发卡行或账户匹配时部分保留；否则属性删除
地址	泛化（省、市、区县）、属性删除	按分析粒度选择泛化层级；地址无分析价值时属性删除
电话号码	属性删除、泛化（保留前3位或后4位）、扰动	需联系校验场景保留部分位段；无联系需求时属性删除
数值型标识符	泛化（取整、顶层/底层编码）、扰动（噪声添加）、随机化（置换）、属性删除、聚合	需保留分布特征供建模时用扰动；仅需区间分布时泛化；仅需总体统计结果时聚合为统计值
日期	泛化（年、年月）、扰动（按固定规则偏移）、聚	需保留时序分析价值（病程、季节分布）时用扰动偏移；仅

	合	需时段统计时泛化或聚合为计数
地理位置	泛化（省、市、区县）、 属性删除	与静态地址处理逻辑相近；轨迹类数据按分析精度逐级泛化

附录C：不同场景的模型参数建议

去标识化处理的量化水平取决于流通情形，参数基准见下表；效果评估判定参照GB/T 42460—2023 执行。

表C-1 模型参数建议

流通场景	K-匿名 (K)	L-多样性 (L)	T-接近性 (T)	差分隐私 ϵ
完全公开 共享	大于等于 20	大于等于 5	小于等于 0.05	小于等于 1
受控公开 共享	大于等于 5	大于等于 3	小于等于 0.10	小于等于 2
领地公开 共享	大于等于 3	大于等于 2	小于等于 0.20	小于等于 8

a) 完全公开共享：数据一旦发布,很难召回,一般通过互联网直接公开发布。

b) 受控公开共享：通过数据使用协议对数据的使用进行约束。

c) 领地公开共享：在物理或者虚拟的所辖范围内共享,数据不能流出到领地范围外。

d) 参数选择应结合数据敏感度、外部数据可得性、影响面与攻击模型确定，并在报告中说明理由。

附录 D：标识符示例

参照GB/T 42460—2023，标识符示例如下。

表D-1 直接标识符示例

类别	示例
姓名	本人姓名、曾用名、别名、家属姓名
证件号码	身份证号、护照号、驾照号、社保卡号、工作证号
医疗编号	病历号、健康卡号、住院号、检查号
联系方式	电话号码、电子邮箱、社交媒体账号
详细住址	精确至门牌号的居住地址
生物识别信息	指纹、声纹、掌纹、虹膜、面部特征、DNA序列

表D-2 准标识符示例

类别	示例
基础指标	性别、年龄、身高、体重、血型、婚育状况
事件时间	门诊/住院/检查/手术/出院等各类事件的发生时间
事件地点	就诊/住院/检查等事件发生的地理位置
单位信息	工作单位、学习单位、所在部门
社会经济特征	民族、受教育水平、职业、收入水平

健康医疗数据流通收益分配工作指引

（试行）

一、总则

（一）编制目的与依据

为规范国家人工智能应用中试基地(医疗)(以下简称“中试基地”)健康医疗数据流通的全链条收益分配管理,提升收益分配规范性、透明度与可落地性,根据《中华人民共和国会计法》《会计基础工作规范》《医疗卫生机构数据类科技成果转化管理专家共识》《关于加强高校院所科技成果转化的实施意见》等法律政策,制定本指引。

（二）适用范围

本指引适用于中试基地范围内,依法开展健康医疗数据产品供给、数据服务输出、授权运营、流通交易活动的全部参与主体,包括中试基地建设主体、数据供给机构、数据使用机构、中试基地运营机构、数据加工机构、第三方专业服务机构等。覆盖数据流通项目收益核算、差异化收益分配、激励资金使用、扶持优惠落地、资金台账管理、全流程监督稽核等相关工作。

排除情形:未完成脱敏、伦理审查、权属认定、合规备案,存在隐私泄露、国家安全、公共卫生风险,法律法规禁

止市场化流通的数据，不得适用本指引开展收益核算与分配。

二、收益分配规则

（一）分主体差异化基准分配比例

根据参与主体的实际贡献程度，以项目实际交易收入为核算基数，对数据供给机构、数据加工机构、中试基地运营机构、第三方专业服务机构等主体实行差异化比例分配，基本按照“5-3-1-1”的比例确定分配区间。各方实际占比总和不高于 100%。具体参考基准如下：

序号	数据流通参与主体	基准分配比例 (交易收入为基数)	浮动分配比例 (交易收入为基数)
1	数据供给机构	50%	30%-60%
2	数据加工机构	30%	15%-40%
3	中试基地运营机构	10%	5%-20%
4	第三方专业服务机构及其他	10%	1%-15%或固定费用

（二）动态调节修正系数

结合数据资产稀缺程度、质量水平、市场应用价值设置动态调节修正系数，对分配收益进行动态修正，实现高价值数据、高投入贡献主体收益倾斜。一般情况下，修正后的分配比例在浮动分配比例区间内。计算公式：

主体最终分配收益=项目交易收入×基准分配比例×动态调节修正系数

（三）特殊情形

部分特殊数据流通情形允许参与主体突破浮动分配比例区间，应由中试基地运营机构报中试基地建设主体组织论证审议确定分配比例。

三、收益使用管理规范

公立医疗卫生机构参照国家卫生健康委医药卫生科技发展研究中心牵头编制的《医疗卫生机构数据类科技成果转化管理专家共识》执行。医疗卫生机构根据合同金额测算成本与净收入，其中不低于 50% 的净收入作为报酬或奖励给予完成人以及转化实施人员；剩余资金全额用于院内数据集标准化治理、数据质量提升、智慧医疗与临床科研平台建设，不得挪作其他经营性开支。

公共数据参照公共数据授权运营相关管理办法执行。其他企事业单位的健康医疗数据流通收益分配可以参照上述收益分配规则执行。

四、监督管理与风险防控

（一）建立常态化联合稽核对账机制

由中试基地建设主体牵头组织中试基地运营机构、第三方专业服务机构建立年度联合稽核制度，每年核查数据流通项目服务内容、收益核算依据、分配比例执行、优惠政策落地、应收实收资金对账匹配情况，确保账证、账实、账协议完全一致，稽核报告、对账凭证按照相关规定留存。

（二）落实标准化信息公开公示要求

中试基地运营机构依规常态化公开数据产品收费目录、

基准定价标准、收益分配通用规则、扶持政策、投诉监督渠道，公示信息同步报送卫生健康、数据、财政等部门，主动接受行政监管与社会公众监督。

（三）全流程常态化行政监督检查

省卫生健康委联合省财政厅、省数据局等部门对基地内数据流通收费执行、收益分配落地、公益反哺资金使用、优惠政策兑现开展定期检查与随机抽查。对发现的违规行为，责令责任主体限期全面整改。