

TC260-PG-2025NA

网络安全标准实践指南

——个人信息保护 个人信息匿名化指南

(征求意见稿 v1.0-202511)

全国网络安全标准化技术委员会秘书处

2025 年 11 月

本文档可从以下网址获得：

www.tc260.org.cn/



全国网络安全标准化技术委员会
National Technical Committee 260 on Cybersecurity of SAC



前 言

《网络安全标准实践指南》（以下简称《实践指南》）是全国网络安全标准化技术委员会（以下简称“网安标委”）秘书处组织制定和发布的标准相关技术文件，旨在围绕网络安全法律法规政策、标准、网络安全热点和事件等主题，宣传网络安全相关标准及知识，提供标准化实践指引。

本文件起草单位：中国电子技术标准化研究院、北京理工大学等。

本文件主要起草人：



全国网络安全标准化技术委员会
National Technical Committee 260 on Cybersecurity of SAC

声 明

本《实践指南》版权属于网安标委秘书处，未经秘书处书面授权，不得以任何方式抄袭、翻译《实践指南》的任何部分。凡转载或引用本《实践指南》的观点、数据，请注明“来源：全国网络安全标准化技术委员会秘书处”。



全国网络安全标准化技术委员会
National Technical Committee 260 on Cybersecurity of SAC



摘 要

为帮助理解和实施个人信息保护政策法规，针对个人信息保护关键内容和难点堵点，依据《中华人民共和国个人信息保护法》《网络数据安全条例》等法律法规，参照个人信息保护相关国家标准，制定个人信息保护系列实践指南，为个人信息处理者提供具体、可操作的实施细则，保护个人信息权益，规范个人信息处理活动，促进个人信息合理利用。

相对于去标识化，匿名化去除个人身份特征更为彻底，或者所采用的技术方法更为复杂，使匿名化后的信息在不付出高额成本的情况下就难以复原，具有较高的安全性。《个人信息保护法》规定匿名化处理后的信息不属于个人信息。本文件给出了匿名化的判断规则、实现方式、适用场景，以及个人信息匿名化参考案例、匿名化处理涉及的技术等，可为个人信息处理者进行有效的匿名化处理提供指引。



目 录

1 范围	1
2 术语和定义	1
3 匿名化判断规则	1
4 匿名化实现方式	3
4.1 匿名化的可行性判别	3
4.2 匿名化的相关考虑要素	4
4.3 匿名化的实现流程	5
4.4 匿名化相关技术风险分析	7
5 匿名化适用场景	7
附录 A 个人信息匿名化参考案例	9
附录 B 匿名化处理涉及的技术	15
参考文献	19





1 范围

本文件给出了匿名化的判断规则、实现方式、适用场景，以及个人信息匿名化参考案例、匿名化处理涉及的技术等。

本文件可为个人信息处理者实施匿名化提供指引。

2 术语和定义

2.1 去标识化

个人信息经过处理，使其在不借助额外信息的情况下无法识别特定自然人的过程。

2.2 匿名化

个人信息经过处理无法识别特定自然人且不能复原的过程。

3 匿名化判断规则

匿名化，是指个人信息经过处理无法识别特定自然人且不能复原的过程。个人信息经过匿名化处理后不再属于个人信息，但仍需采取技术措施和其他必要措施保障数据安全，随着技术条件等发展变化，相关信息不再符合匿名化条件时，属于个人信息，应当按照个人信息进行保护。

有效的个人信息匿名化处理，应当同时满足以下三个条件。

a) 不可单独挑出。无法从数据集中挑选出属于某个人的部分或者全部记录。“单独挑出”风险不要求直接辨认出个人的真实身份，其关注的是数据中记录的可区分性，因此具体而言“不可单独挑出”



要求无法将数据集中的任何一个个体通过其独特的属性值或属性值组合与其他个体区分出来。例如，某公司发布了一份员工健康数据集，其中隐去了姓名，但保留了部门、年龄和性别字段，一名攻击者知道同事张三在“财务部”工作、今年“47岁”、性别为“男”，他利用此背景知识查询该数据集发现只有一条记录与之匹配，则张三的记录被单独挑出。

b) 不可链接。无法将关于同一个体或同一组个体的不同记录进行链接，无论这些记录来自同一数据集或不同数据集。如果攻击者可以判断出两条或多条记录属于同一个体或同一组个体，则存在“可链接”风险；如果攻击者可以确定两条记录属于同一组个体，但无法从该组中单独挑出个体，则可以抵御“单独挑出”风险但不能抵御“可链接”风险。链接攻击通常依靠的是相对持久不变的唯一标识，或多个数据集所共同拥有的属性（通常为准标识符）。例如，两个存在相同个体记录的“匿名”数据集中都含有用户精确的出生时间和出生地邮政编码，攻击者就可能通过这两个共同属性将两个数据集中关于同一个体的记录进行关联。

c) 不可推断。无法以相当大的概率推断出数据集中某个个体的相关信息。例如，假设一个匿名化数据集记录了某个区域的所有居民的“健康状况”都是“患有某种疾病”，那么只要知道某人是该区域居民，就能推断出其患有该种疾病。

在做具体判断时，应当充分考虑个人信息处理者及任意第三方，



在运用所有可能合理采取的手段后，仍然无法单独挑出特定个体、不可链接、不可推断个体相关信息，即可认为个人信息经过了匿名化处理。应当采取动态评估的方式对所有可能合理采取的手段进行评估，当评估发现不再同时满足判断匿名化的三个条件时，相关信息转为个人信息，应当按照个人信息进行保护。

可能合理采用的手段，是指在当时技术条件下，合理可采用的识别手段¹，包括技术能力、经济成本、时间投入和获取信息的渠道等，理论上存在但实际不太可行的极端手段不纳入考虑。如果经过综合评估，表明实现单独挑出、关联或者推断的可能性可以忽略不计，则认为达到了匿名化。

4 匿名化实现方式

4.1 匿名化的可行性判别

在某些情况下，匿名化可能无法实现。

a) 情形一：使用个人信息的目的建立在识别、关联、推断个人之上，与匿名化提供的保护目的相悖。例如，提供个性化广告、个性化信息推送等个性化服务，必须与用户关联；金融风控场景下，必须需要知道是谁在贷款，信用评价如何，以便管理风险。此类情形下可能会保留长期稳定的标识信息（例如设备标识、各类ID）、利用多维度信息生成“指纹”，或保留复原回可识别状态的能力，用于需要唯一标识、关联和追踪个人的使用目的。

¹ 既包括合法的手段，也应当防范非法手段。



b) 情形二：个人信息本身具有较强唯一性或独特性。例如，生物识别信息，其本质就是为了唯一识别个人，对这类数据进行匿名化，可能破坏其实用价值；高维度、稀疏且需要保持个体或较细粒度的个人信息，例如对个人的多维度（如成百上千维度）画像标签，即使删除了姓名、联系电话等直接标识符，标签中包含的性别、年龄、地区、购物记录、搜索历史等信息的组合，具有极高的独特性，在此种情况下，为了满足 k -匿名等模型，需要进行极高强度的泛化、抑制处理，从而导致个人信息丧失实际使用价值。

c) 情形三：处理个人信息的上下文环境具有较大固有风险。例如个人信息处理链条长、接收方数量众多，需要保留个体细粒度个人信息（非统计/聚合情形），且无法完全删除所有可能起到标识作用的属性的场景（例如对直接标识符未做删除、抑制等处理，或保留了部分准标识符属性列，特别是涉及保留各类行业普遍收集的移动电话号码、性别、年龄、地区等属性的），由于面临潜在众多攻击者和无法预计的背景知识信息等情况，在满足数据可用的前提下，重识别风险难以低到可以忽略不计。

4.2 匿名化的相关考虑要素

在匿名化可行的前提下，匿名化的实现也不仅仅与个人信息本身的转换处理有关，还与场景风险、技术和管理措施等相关。匿名化的有效性，往往取决于个人信息处理、技术和管理措施等一系列措施的综合保障强度相对于场景风险是否充分、是否相适应，以使得单独挑



出、可链接、推断风险均被消除或低到可以忽略不计。

a) 场景风险：场景风险反映初始状态下，个人信息在场景下遭到重识别的可能性的**高低**，以及一旦遭到重识别对个人权益的影响，需要考虑的因素包括个人信息情况、个人信息使用目的、个人信息处理方式和范围、场景相关方等。

b) 个人信息处理程度：个人信息处理程度反映个人信息经过转换处理后的可识别性强弱，例如是否最小化处理、是否包含直接标识符、准标识符的泛化程度等。

c) 安全技术措施：安全技术措施反映保障个人信息处理安全的措施水平，例如是否采取有效的访问控制、多方安全计算等安全技术措施。

d) 匿名化管理措施：匿名化管理措施反映对相关方匿名化措施的体系性、完备性的持续保障，以及对相关合规和风险的管理和控制。例如是否具备相关管理制度、制度流程、应急处置机制、管理措施是否具备有效约束力等。

4.3 匿名化的实现流程

在匿名化的具体实现方面，主要遵循以下流程：

a) 初步判断匿名化的可行性：若匿名化不可行，则应采取相应的去标识化技术措施，并通过取得个人同意等方式获得处理个人信息的合法性基础。

b) 开展场景分析：分析个人信息中的直接标识符、准标识符以



及数据集中的其他信息，明确场景涉及的处理目的、方式和范围、相关方等，识别可能面临的个人信息处理活动风险和数据安全风险，确定场景所需的匿名化处理措施。

c) 开展匿名化准备：在场景分析基础上，制定匿名化处理策略，选择合适的技术（例如泛化、随机化）和模型（例如 k -匿名、差分隐私），明确数据安全技术措施，防范来自场景内外的安全风险，并在实施前对策略进行审核评估。

d) 实施匿名化处理：按照匿名化处理策略，选择相应技术和模型开展处理转换，实施相应数据安全技术措施等。

e) 结果验证和效果评估²：验证实际开展的转换处理、采取数据安全技术措施，是否与匿名化处理策略一致，并通过标识符识别、隐私模型参数验证、重识别风险度量指标计算等方式进行匿名化效果评估。

f) 匿名化风险管理：常见的管理措施包括管理制度建设、匿名化策略管理、规则公开、记录留存、接收方评估、相关方约束、应急机制、监督审计等方面。例如对匿名化过程和效果进行记录和证明，通过具有法律约束力的文件等明确相关方的数据安全义务、数据的使用目的范围限制、禁止开展重识别、是否允许接收方将数据进行二次提供及其条件、相关应急处置机制等。

规范落实上述流程和相关要求，经评估确定其综合措施能够确

² 匿名化效果评估方法包括筛查结果数据中是否仍包含直接标识符，核验结果数据是否满足既定的模型，通过计算检察官风险等重识别风险度量指标判断是否达到足够低的风险，基于攻击测试验证是否能够成功重识别或复原等。



保，在运用所有可能合理采取的手段后，仍然无法单独挑出、不可链接、不可推断特定个体相关信息，则可以认定达到了匿名化效果。

匿名化并非一劳永逸，随着后续的使用以及技术的发展，原先匿名化的数据如果被证明可以重新识别到特定自然人，那么这些数据仍将被视为个人信息，因此处理者还应当持续评估匿名化处理后的个人信息重识别风险。

个人信息匿名化参考案例见附录A。

4.4 匿名化相关技术风险分析

匿名化处理需要对各项技术进行综合应用。主要包括噪声添加、置换、聚合或 k -匿名、 l -多样性、差分隐私、哈希/标记化等。常见的匿名化相关技术在实现匿名化过程中都存在一定的可识别风险，表1对常见匿名化相关技术与单独挑出、可链接、推断风险的对应关系作了梳理。匿名化处理涉及的常见技术见附录B。

表1 常见匿名化相关技术与单独挑出、可链接、推断风险的对应关系

常见技术或模型	单独挑出风险?	可链接风险?	推断风险?
噪声添加	是	不一定	不一定
置换	是	是	不一定
聚合或 k -匿名	否	是	是
l -多样性	否	是	不一定
差分隐私	不一定	不一定	不一定
哈希/标记化	是	是	不一定

5 匿名化适用场景

常见典型的匿名化适用场景包括：

a) 科研与医疗领域：科研机构 and 医疗机构对个人信息进行匿名



化处理后再加以分析利用。例如，医学研究中对患者个人信息进行匿名化处理，以研究疾病发展趋势和群体健康特征，在保护患者隐私的同时获得有意义的研究结论。

b) 统计分析与公共政策：对人口、经济等数据进行匿名化处理，用于统计分析、公共政策制定等。例如，人口普查数据在匿名化处理后发布，用于分析社会经济状况和制定宏观政策。

c) 大数据与人工智能训练：在互联网、大数据和人工智能领域，企业对收集的个人信息进行匿名化处理后用于模型训练和算法开发。通过使用匿名化后的数据集开展训练，既能够提高模型对群体行为的预测能力，又能避免处理可识别的个人信息，从而降低个人信息权益相关风险。





附录 A 个人信息匿名化参考案例

A.1 案例一：智慧交通出行

A.1.1 处理背景和目的

通过产学研多方协作，共同研究城市智慧交通领域的共性问题，如道路拥堵预测等，推动智慧出行服务平台的技术开发与应用。

A.1.2 匿名化可行性判断

本场景旨在分析群体性、规律性的交通现象（如特定路段在特定时段的车流量、平均车速等），不涉及识别、关联或追踪特定个人（例如出行轨迹）。业务目标聚焦于宏观统计和趋势预测，不依赖于可识别的个人信息，处理后的聚合数据仍能满足科研需求，不影响业务价值，因此具备匿名化的可行性。

A.1.3 场景分析

场景相关方：数据提供方为出行服务企业；数据使用方为科研院所、高校等合作机构。

数据类型：涉及用户出行行程数据、路网拓扑信息及车辆相关数据等。包括用户名、用户ID等直接标识符；行程ID、用户设备ID等间接标识符，精确的上下车时间戳、精准地理位置信息（例如包含精确经纬度坐标形成的轨迹点）等准标识符，以及POI（兴趣点）信息等目标属性/敏感属性。

风险分析：原始数据包含精确的时空轨迹，属于敏感个人信息，直接共享存在较高的重识别风险。



A.1.4 匿名化准备

策略制定：针对不同类型数据采取差异化处理。对直接标识符进行删除；对行程ID等间接唯一标识进行假名替换；对时空数据进行聚合与泛化。

处理技术选择：

- a) 删除/抑制：用于处理直接标识符。
- b) 假名替换：对行程ID等间接标识符进行不可逆的假名替换。
- c) 泛化/聚合：将精确时间泛化为时段（如15分钟），将经纬度坐标聚合至地理网格或路段级别等。

安全环境：研究人员在数据提供方提供的隔离安全环境中使用数据开展研究，确保数据无法离开受控环境。

A.1.5 实施匿名化处理

删除直接标识符。删除数据集中的用户名、用户ID等字段。

假名替换间接标识符。对行程ID、设备ID等进行不可逆的假名替换，例如生成随机ID替换原始行程ID，同时不保留查找表，确保未经授权无法还原到原始行程。

泛化聚合时空数据：时间戳泛化为时段（例如将精确到秒泛化为“15分钟”），将连续的经纬度轨迹点聚合为承载在特定路段上的流量信息，或映射到预设的地理网格中。不保留个体轨迹细节。

POI信息最小化处理。去除POI（兴趣点）具体信息，仅保留POI ID，然后对POI ID进行不可逆的假名替换，仅保留转换后的ID（例如



生成随机ID替换POI ID，同时不保留查找表）。

A.1.6 结果验证和效果评估

输出验证：验证研究方可访问的结果数据集中不包含任何直接标识符和原始准标识符，数据以聚合统计形式呈现，不包含个体级记录。

效果评估：通过模拟攻击进行评估。假设攻击者掌握“某用户在某日某时段从A地到B地”的背景知识，检验其是否能在结果数据集中单独挑出该用户的记录。由于时空精度已大幅降低且个体轨迹被聚合，攻击者无法进行有效匹配，确认达到了“不可单独挑出”的效果。

A.1.7 匿名化风险管理

协议约束：与合作机构签订数据使用与保密协议，明确数据来源、使用范围及保护责任等。

安全环境：研究人员在数据提供方提供的隔离安全环境中操作，数据无法离开该环境，且所有操作均有日志记录，便于审计。

审核评估：数据开放全流程经过多方评估与审核，确保安全可控。

记录留存：对匿名化处理的关键步骤、参数选择、评估过程和最终测试结果进行文档化记录，以履行合规证明责任。

A.1.8 小结

本案例综合运用删除、假名替换、泛化与聚合等技术，结合安全环境控制，将包含敏感个人信息的原始出行数据转化为服务于公共研究的宏观交通分析数据，在保障数据价值的同时有效控制重识别风险。



A.2 案例二：公共卫生统计数据公开

A.2.1 处理背景和目的

某地疾病预防控制中心需要向社会公开发布关于某类传染病的统计数据，以帮助公众了解疫情态势、指导科研机构研究，并为政府决策提供依据。数据需要包含不同区域、不同人群的发病情况，但不能泄露患者的个人信息。

A.2.2 匿名化可行性判断

处理目的在于公共知情和科学研究，不需要识别到任何个体患者。

A.2.3 场景分析

场景相关方：疾病预防控制中心。

数据类型：原始病例数据库包含患者姓名、身份证号、手机号、详细家庭住址等直接标识符，以及年龄、性别、职业、所属区县、确诊日期等准标识符，患病类别等目标属性/敏感属性。

风险分析：原始数据包含精确的个人患病情况，属于敏感个人信息，若直接公开，重识别风险较高。

A.2.4 匿名化准备

匿名化策略制定：明确对不同类型的数据采取不同的处理方式。直接标识符完全删除；准标识符进行不可逆转换或泛化处理；时空数据进行聚合与泛化。

技术与模型选择：



a) 删除/抑制：患者姓名、身份证号、手机号、详细家庭住址等直接标识符进行删除/抑制。

b) 泛化：对年龄、性别、职业、所属区县、确诊日期等准标识符进行泛化。例如将“年龄”从具体岁数（如37岁）泛化为年龄段（如30-39岁）；将“确诊日期”从具体日期（如3月5日）泛化为时间段（如3月上旬）。

c) 隐私模型：公开数据集满足 k -匿名模型以及 l -多样性模型，其中 k 取20， l 针对患病类别，取5。

A.2.5 实施匿名化处理

删除直接标识符。删除数据集中的患者姓名、身份证号、手机号、详细家庭住址等字段。

设定隐私模型和参数，应用匿名化相关算法，迭代地对准标识符进行泛化、抑制，直至结果数据满足 k -匿名模型以及 l -多样性模型。

A.2.6 结果验证和效果评估

输出形态验证：公开发布之前，确认发布数据已不包含任何直接标识符和原始准标识符，等价类大小至少为20，所有等价类中患病类别这一敏感属性至少有5个不同的取值。

效果评估：

a) 不可单独挑出： k -匿名模型保证了任何记录都至少与其他 $k-1$ 条记录无法区分，攻击者无法单独挑出某个特定个体。

b) 不可链接：由于直接标识符已删除，且准标识符被泛化，不



易将此数据集中的记录与其他数据集中的个体信息进行有效链接。

c) 不可推断：攻击者即使知道某一个体在数据集中，也只能推断出他属于一个至少有20个人的群体，且该群体中至少有5个不同的患病类别取值，因此无法获得关于该个体的确定性的信息。

A. 2.7 匿名化风险管理

需定期评估 k 值、 l 值的适当性。随着更多相关数据集的公开，可能会出现新的链接风险，届时需要重新评估匿名化策略或提高 k 值、 l 值，确保持续符合匿名化要求。

A. 2.8 小结

本案例通过应用 k -匿名、 l -多样性等隐私模型，对数据集进行泛化、抑制等处理，形成了重识别风险较低的结果数据，并公开发布。





附录 B 匿名化处理涉及的技术

B.1 处理技术

B.1.1 泛化

泛化技术是指一种降低数据集中所选属性粒度的去标识化技术，对数据进行更概括、抽象的描述。泛化技术实现简单，能保护记录级数据的真实性。使用泛化技术的目标是减少属性唯一值（更概括地说，是指多个属性值的组合集的唯一值）的数量，使得被泛化后的值（或多个值的集合）被数据集中多个记录所共享，从而增加某特定个人信息主体被推测出的难度。因此，通常选择对标识符属性进行泛化，但是根据具体情况也可考虑对任何属性（特别是敏感属性）进行泛化。

B.1.2 噪声添加

噪声添加是一种随机化技术，通过添加随机值、“随机噪声”到所选的连续属性值中来修改数据集，同时尽可能保持该属性在数据集的原始统计特性。该类统计特性包括属性的分布、平均值、方差、标准偏差、协方差以及相关性。

B.1.3 置换

置换是在不修改属性值的情况下对数据集记录中所选属性的值进行重新排序的一种技术。因此，置换保持了整个数据集中所选属性的准确统计分布。

置换技术适用于数字与非数字值。因为观察到的不一致性可能有助于对置换算法实施逆向工程，需要考虑如何来确保生成的数据集是



一致的。

不同置换技术的区别在于方法与复杂性的差别。在保持所选属性之间原有关联性的情况下，置换算法可用于单个或多个属性。

通常情况下，采用逆向工程可以将数据恢复到原始状态，从而加大受控重标识的可能性，因此把随机化算法引入到置换中会增强对抗重标识攻击的能力。

B.1.4 数据聚合

数据聚合作为一系列统计技术（如求和、计数、平均、最大值与最小值）的集合，应用于微数据中的属性时，产生的结果能够代表原始数据集中的所有记录。

对数据抽样技术选择和使用应注意以下几个方面：

a) 数据聚合可能会降低数据的有用性；因为得到的是统计值，无法反映独立数据记录的特征。

b) 数据聚合对重标识攻击非常有效；数据聚合的输出是“统计值”，该值有利于对数据进行整体报告或分析，而不会披露任何个体记录。

例如：2012年我国18岁及以上成年男性平均身高1.67m。如果数据集以平均身高来标识数据集中每个人的身高值，则记录（男，本科，北京，1.67m，1980年9月1日）中，身高属性值对攻击者识别身份主体没有什么作用。



B.2 隐私模型

隐私模型是对数据在可识别性方面或反映其重识别风险的一些量化标准，典型代表有 k -匿名、 l -多样性、 t -接近性和差分隐私：

B.2.1 k -匿名

k -匿名是一种隐私模型，该利用了等价类的概念，基本思想是“将个体隐藏在一个大小至少为 k 的等价中，使得该个体与其他至少 $k-1$ 个个体不可区分”。

k -匿名模型通过保证数据集中每个等价类至少包含 k 条记录，防止了攻击者将某条记录与单一个人相对应，从而降低“单独挑出”个体的风险。

在 k -匿名基础上，还可引入 l -多样性、 t -接近性等衍生模型来进一步降低“属性推断”风险。

B.2.2 l -多样性

l -多样性是针对属性值差异性不大的数据集提出的一种增强概念。为防止确定性推导， l -多样性要求在 k -匿名的基础上，实现每一等价类在每一敏感属性上存在至少 l 个不同值。在数据分布很不均衡时，防止推导性攻击的能力受到限制。

B.2.3 t -接近性

t -接近性是 l -多样性的增强概念，适用于发布数据集的敏感属性分布要尽可能接近整个数据集的敏感属性分布。针对属性值分布不规则、属性值范围很小或已被分类的数据集，为防止概率性推导，要求



任何等价类中敏感属性的分布与整个数据集中相应属性的分布之间的距离小于阈值 t 。

B.2.4 差分隐私

差分隐私是一种对隐私性的定义，它限定了满足该定义的随机算法对于仅在一条记录上有差别的两个数据输入而言，得到任何输出的概率都是相近的，从而无法辨别随机算法的输出具体是来自哪个输入。

差分隐私通过在查询统计结果中注入随机噪声的方式，严格限制任何单条记录对输出结果的影响，从而保证攻击者无论是否掌握某个个体的数据，观察统计结果都难以推断出该个体的存在与否。





参考文献

- [1] 中华人民共和国个人信息保护法
- [2] 网络数据安全管理办法
- [3] GB/T 37964—2019 信息安全技术 个人信息去标识化指南
- [4] GB/T 42460—2023 信息安全技术 个人信息去标识化效果评估指南
- [5] ISO/IEC 20889:2018. Privacy enhancing data de-identification terminology and classification of techniques
- [6] ISO/IEC 27559:2022. Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework
- [7] ISO/IEC 29100:2024. Information technology – Security techniques – Privacy framework
- [8] Article 29 Data Protection Working Party (WP29). Opinion 05/2014 on Anonymisation Techniques. 2014
- [9] Information Commissioner's Office (ICO). Anonymisation and Pseudonymisation guidance. 2025