

TC260-PG-2025NA

网络安全标准实践指南

——个人信息保护 个人信息去标识化指南

(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。

去标识化是个人信息保护的常用安全措施，可用于降低个人信息可识别性，经过去标识化处理的个人信息在利用过程中能够防止、减少因泄露或滥用对个人权益造成的威胁。本文件明确了脱敏与假名化两种主要的去标识化技术路径，给出了假名化的实现框架、流程步骤、安全保障措施，以及常见类型个人信息脱敏示例等，可为个人信息处理者实施去标识化提供指引。



目 录

1 范围	1
2 术语和定义	1
3 概述	3
3.1 去标识化概述	3
3.2 脱敏	4
3.3 假名化	4
4 假名化的实现	5
4.1 假名替换的实现方式	5
4.2 修改原始个人信息	6
4.3 防止未授权关联的技术和管理措施	7
4.4 链接假名数据	9
4.5 实施流程	12
5 去标识化适用	14
附录 A 标识符的识别规则和示例	16
附录 B 个人信息去标识化参考案例	20
附录 C 个人信息脱敏示例	25
参考文献	28





1 范围

本文件明确了脱敏与假名化两种主要的去标识化技术路径，给出了假名化的实现框架、流程步骤、安全保障措施，以及常见类型个人信息脱敏示例等。

本文件可为个人信息处理者实施去标识化提供指引。

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。

注：额外信息包括假名替换过程中产生的假名化密钥（例如密钥、查找表等）、原始数据等附加信息，以及特定环境下可用于与结果数据进行关联的其他数据集，外部或公开数据集，攻击者的背景知识等。

2.2 脱敏

一种针对原始数据中的敏感信息进行转换处理，从而降低数据可识别性的去标识化方法。

注1：本文件所称脱敏，是一种仅针对个人信息本身进行转换处理的去标识化。区别于假名化，其关注的是个人信息中的单个数据项，采用的处理方式不限于假名替换技术（还包括删除、遮蔽、泛化、随机化等），也不涉及其他配套的安全技术和管理措施。

注2：本文件所称敏感信息，主要指个人信息中的标识符，或其他任何有助于实现对个人唯一标识的数据属性（例如目标属性中的离群值）。

2.3 假名化

一种在对数据本身进行转换处理的基础上，保留了恢复识别能力、同时对恢复识别能力进行了管控的去标识化。

注：本文件所称假名化，是一种基于假名替换等技术，将原始数据转化为假名数据，单独保存附加信息（例如密钥、查找表等），并对附加信息实施管控措施，确保仅授



权方能通过逆向操作恢复识别特定个人信息主体的去标识化。

2.4 附加信息

能够用于将假名数据重新关联至已识别或可识别个人的、在处理者可控范围内的任何信息。

注：包括密钥、查找表等假名化密钥，原始数据，以及特定环境下可用于与结果数据进行关联的其他数据集，不包括外部或公开数据集，攻击者的背景知识等。

2.5 标识符

特定环境下单独或结合其他属性可以实现对个人唯一标识的数据属性。

注：标识符可用于在特定环境下识别特定自然人，分为直接标识符和准标识符。标识符的识别规则和示例见附录A。

2.6 直接标识符

特定环境下单独可以实现对个人唯一标识的数据属性。

注：例如姓名、公民身份号码、护照号、驾照号、详细住址、电子邮件地址、电话号码等。

2.7 准标识符

特定环境下结合其他属性可以实现对个人唯一标识的数据属性。

注：例如性别、出生日期或年龄、民族、职业、婚姻状况、国籍等。

2.8 假名

在假名替换过程中添加到数据中的标识符。

2.9 假名数据

个人信息经过假名替换处理后得到的结果数据。

注：其核心特征是在没有附加信息的情况下，无法恢复识别特定个人信息主体。

2.10 假名化域

由处理者划定的一个逻辑或物理环境，其目标是在该环境内有效阻止重识别行为的发生。



注：该“域”涵盖了环境内的人员、系统及可用的信息资源，但不包括被正式授权访问附加信息以执行逆向还原的人员。

2.11 假名化密钥

在假名替换过程中用于生成假名的数据。

注：通常指密钥或存储身份映射关系的查找表；它是附加信息的组成部分。

2.12 假名化处理者

采用假名替换技术对原始数据进行转换的个人信息处理者。

2.13 假名替换

一种使用假名来替换标识符的去标识化技术。

3 概述

3.1 去标识化概述

去标识化，是指个人信息经过处理，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。对个人信息进行去标识化处理，是个人信息处理者采取的风险控制措施，经过去标识化处理的个人信息在利用过程中能够防范因泄露或滥用对个人权益造成侵害的风险。由于个人信息进行去标识化处理后，仍可以借助额外信息重新识别特定自然人，因此去标识化后的信息仍然属于个人信息。个人信息去标识化参考案例见附录B。

根据处理方式的不同，去标识化主要分为基于脱敏技术的简单去标识化（简称“脱敏”），以及基于假名替换技术的复杂去标识化（简称“假名化”）两种。



3.2 脱敏

脱敏是一种仅针对个人信息本身进行转换处理的去标识化。

脱敏的目的是通过对个人信息进行处理，降低个人信息的“敏感程度”。敏感程度通常可能体现在标识个人身份的程度，也可能体现在所包含信息的私密程度（本文件指前者）。脱敏技术有时被认为是一系列此类处理技术的总称，包括删除、遮蔽、泛化、随机化、假名替换等。在个人信息保护领域，脱敏技术可用于实现去标识化，其中最典型的脱敏技术是遮蔽技术，例如将姓名、移动电话号码、公民身份号码中部分或全部的位数使用星号替代。常见类型个人信息脱敏示例见附录C。

3.3 假名化

假名化基于假名替换等技术，是一种在对个人信息进行转换处理的基础上，保留了恢复识别能力、同时对恢复识别能力进行了管控的去标识化。

在处理过程中通常会伴随附加信息的产生和使用，该附加信息可被用于识别特定自然人。

假名化通常包括三大步骤：

- a) 使用假名替换等技术修改原始个人信息；
- b) 单独保存附加信息；
- c) 采取技术和管理措施，防止附加信息的未经授权访问和使用。

处理者通常基于风险分析，定义在何种环境下使用假名替换技



术，以防止信息被关联到个人，处理者还会对附加信息采取技术和管理措施，以确保在该环境下的人员无法将假名数据关联回个人信息主体。本指南将该环境（包括人员、物理、组织、可用的IT资产等）称为“假名化域”。处理者需要采取一系列技术和管理措施，确保假名数据无法离开假名化域，以及能够恢复识别个人的假名化密钥等附加信息无法进入假名化域。

根据假名化的目标及相关的重识别风险，可以将假名化域定义为仅处理者单个组织、单个外部接收者、所有取得授权的合法接收者，或可能试图在未经授权的情况下访问数据的一系列或所有外部实体等。如果假名数据仅在处理者内部处理，则假名化域并非涵盖处理者整个组织，而是仅涵盖具有处理假名数据授权的人员、这些人员掌握的信息，以及他们使用的系统和服务。

4 假名化的实现

4.1 假名替换的实现方式

4.1.1 基于密码算法的假名替换

用于假名替换的密码算法主要包括密码单向函数（例如基于哈希的消息认证码，HMAC）以及加密算法。

通常优先考虑密码单向函数，因为即使已知相关参数，其逆向也较为困难。然而，根据具体场景的需求，特别是在授权场景中需要逆向假名化，可能需要使用加密算法。如果使用哈希函数作为假名替换中使用的密码单向函数的构建块，则建议使用专为安全密码认证而设



计的哈希函数。

4.1.2 基于查找表的假名替换

用于假名替换的查找表包含标识符与假名的对应关系。每当遇到一个新的原始标识符时，会生成一个随机且唯一的假名，并将“原始标识符—假名”这一对应关系存入一个表中。为防止攻击者通过少量假名推测生成规则，处理者应使用更安全的假名生产方式，例如硬件随机数生成器或密码学安全伪随机数生成器。

由于查找表能够帮助识别个人信息主体，因此查找表属于个人信息。

4.2 修改原始个人信息

4.2.1 针对直接标识符的替换或删除

选择替换还是删除直接标识符，取决于处理的目的和假名化的目标。

如果不需要进行记录链接，则需要删除“长期”的直接标识符，用假名替换“短期”的直接标识符¹。

如果需要进行记录链接，例如进行纵向分析，则需要使用假名替换长期标识符，同时删除其他直接标识符。

4.2.2 针对准标识符的删除、泛化、随机化

将数据重新关联到个人的一种方法是借助数据中包含的若干属性，这些属性往往揭示了有关个人身体、生理、遗传、心理、经济、

¹ 长期标识符指，长期稳定不变的、可以持续追踪个人的 ID，比如身份证号、IMEI、姓名、用户 ID；短期标识符是短期有效、关联链接能力有限的 ID，比如订单 ID、工单号等。



文化、社会身份等方面的信息，如果这些属性的组合足以将至少部分假名数据关联到特定个人，则它们被称为准标识符。典型准标识符是人口统计学属性，常见准标识符包括年龄、性别、地区等。如果处理假名数据的人员掌握了假名数据中某些个人此类属性的值，那么可能在无需使用假名化密钥的情况下重识别个人。

对准标识符进行处理的方式有三种：

- a) 删除；
- b) 泛化、随机化处理；
- c) 通过对假名化域中可用的信息进行最小化处理，从而减少需要被视为准标识符属性的数量。

第三种方法适用于假名数据在处理者内部处理的情况。当假名化的目标是保护数据免受处理者之外的未经授权的第三方带来的安全风险时，此方法不适用。

4.3 防止未经授权关联的技术和管理措施

为了防止假名数据被未经授权地恢复识别，应从三个方面采取措施：

- a) 选择合适的假名化设计。确保假名化密钥具备适当的安全级别，防止假名替换被逆转；
- b) 妥善处理准标识符；
- c) 应确保其关于假名化域范围、假名数据的使用，以及其中相关信息源可访问性的假设得到满足。



4.3.1 防止假名替换被逆转

为了使假名化有效，应确保假名数据无法通过合理的努力被逆转：

a) 保障转换处理本身的强度。使用查找表进行假名替换时，需选择随机生成的假名；若采用密码算法，需采用足够强大的加密算法，例如带密钥、具备抗原像性的密码学单向函数（如HMAC）。如果使用查找表或可逆加密算法，还需要对假名化密钥进行保密（为了增强安全性，也可以通过秘密共享²等方式对其进行拆分，并由不同的实体存储）。

b) 对执行假名替换的系统及其接口进行严格的访问控制，以确保处理系统和服务本身的完整性和保密性。可用的技术措施包括：网络分段、在硬件安全模块中存储密钥、应用程序编程接口（API）访问的安全认证，以及对假名替换，特别是对逆向转换的执行进行速率限制和日志记录。

c) 采取适当的管理措施。包括聘用经过审查、获得专门授权的人员，负责操作执行假名替换和存储假名化密钥的相关系统，并确保这些人员以及所有负责与个人信息主体互动和访问假名数据的人员都经过适当的培训。

4.3.2 保护假名化域

保护假名化域的措施包括：

² 秘密共享是一种实现多方安全计算的技术方案，通过将参与计算的数值分割成多个随机碎片发送给不同方来隐藏该数值，随后通过各方利用其碎片来重构该数值。



a) 将假名化域与其他信息隔离，采取适当措施，确保其他信息无法进入假名化域，同时防止假名数据流出假名化域，从而将假名数据限制在原始处理者或一组明确定义的接收者范围内。

b) 控制假名数据的流动，明确假名数据的披露对象以及披露范围。建立访问控制系统，并保护API免受未经授权使用。只有在获得授权后，才将假名数据传输给其他实体。

c) 通过合同协议等形式明确相关方的责任，其中应体现将假名数据保留在假名化域中的必要性、限制可能将假名数据重新关联到个人信息主体的信息的流入或访问、需要调整假名化域时应遵循的流程即审批环节等，并确保合同协议的有效执行。

4.4 链接假名数据

4.4.1 控制假名数据链接的范围

为了允许将指向同一个人信息主体的多条假名数据与相同的假名关联起来，基于假名化的目标，处理者需要定义哪些数据集将被一致地假名化。例如可能需要对同一天收集的所有数据进行一致地假名化，以允许将属于同一个人信息主体并在同一天收集的两条数据记录关联起来，但阻止将不同日期收集的属于同一个人信息主体的数据记录关联起来。

根据假名的可链接范围，可以分为个人假名、关系假名、交互假名三种。

a) 个人假名是指，一个个人信息主体在个人信息处理者处理其



所有相关个人信息时，始终使用同一个、长期不变的假名。无论个人信息是在何时收集、用于何种目的，只要是关于同一个人的，都会被标记上这个唯一的假名。该假名的生命周期是长期的，与该个人信息主体的整个生命周期或与个人信息处理者的整个关系周期绑定。当且仅当可能需要将与同一个人相关的不同假名数据关联起来，并且在这种情况下是合法的，才允许使用此类假名化。

b) 关系假名是指，一个个人信息主体根据其与个人信息处理者建立的不同“关系”或“角色”，被分配不同的假名。该假名的生命周期与关系存续期绑定。当一个关系结束时（如员工离职），该假名及其对应的密钥/条目也应随之失效或销毁。只有在需要将与同一个人有关的不同假名数据与处理者关联起来并且在这种情况下合法的情况下，才允许使用这种假名化。

c) 交互假名是指，对于个人信息主体的每一次独立的交互，都生成一个全新的、一次性的假名。该假名的生命周期极短，仅在一次交互中有效，交互结束后即失效。这种形式的假名化非常适合于缓解与非法或未经授权披露假名数据相关的风险。

在与处理的性质、范围、上下文和目的相符的情况下，处理者应优先选择交互假名，而非其他类型的假名。

4.4.2 链接不同处理者假名化的数据

在某些情况下，两个或多个处理者可以合法地关联其持有的不同假名数据，实现方式包括：



- a) 在处理者之间共享假名化密钥;
- b) 共同委托受信任的服务提供者执行假名化;
- c) 将两者结合, 对部分密钥进行拆分, 由处理者与服务提供者分别持有, 且服务提供者不会获知个人信息主体身份。此外, 还可通过密码学方法在不暴露直接标识符或长期假名的情况下, 计算共同假名(如私有集合求交)。

所有上述做法应满足两个前提: 一是各处理者采用相同的假名替换方式; 二是每个处理者生成的假名均基于各个数据集中共同存在的原始标识符。

第一种方式是最为简单, 但也存在诸多缺陷: 一是假名化密钥存储在多个位置, 增加了未经授权访问和使用的风险; 二是所有处理者不仅可以将自己的假名数据记录归因于特定个人, 还可以将其他处理者的假名数据记录归因于特定个人。三是增加了更新假名化密钥的复杂性。因此, 通常不建议采用这种方法。

第二种方式要求处理者之间共同签订合同, 并单独与受信任的服务提供商签订合同。该可信第三方仅需知道用于计算假名的个人信息主体的标识符, 而无需任何其他数据。因此, 处理者应只传输这些标识符, 并附带分配给相应记录的临时编号。该服务提供商对这些标识符应用一个对所有处理者都统一的假名替换, 从而得到假名。然后将这些假名连同各自的记录编号返回。随后即可用记录编号将假名写回数据记录, 再删除编号。这种方法的优势在于, 即便每个处理者获得



完整的关联数据集，也只能重新识别自己贡献的记录。

第三种方式中，各处理者商定一个共同的假名化密钥，使用这个共同的秘密信息计算出第一级假名，之后将第一级假名传输给可信服务提供商，后者再使用其自己的假名化密钥计算出第二级假名。这种方式的优点在于，可信服务提供商不会（也无法）获知原始标识符。此外，两个假名化密钥分散存储在不同实体处，使得未经授权的逆向还原更加困难。

4.5 实施流程

4.5.1 规划阶段

首先确定希望通过假名化实现的目标，定义假名化域，并决定哪些数据集需被一致地处理。

在确定处理方式时，需要对数据进行分析：

- a) 待假名化的个人信息中，哪些属性可以单独或组合使用以直接识别个人信息主体（即标识符，包括直接标识符、准标识符）；
- b) 根据链接范围（使用个人、关系还是交互假名），决定用哪个或哪些属性作为生成假名的输入；
- c) 确定使用的假名化方法（例如基于加密算法或查找表）和参数（例如加密算法的密钥长度、分组大小）；
- d) 哪些信息需要作为附加信息（例如假名化密钥）被保留，以便用于将假名数据重新关联到特定个人信息主体；
- e) 在假名化域内，考虑可从域内通过合理努力获取的信息，分



析个人信息中是否包含以及包含哪些属性，可以被单独或组合使用，以直接或间接地将部分数据重新归因于假名域内的个人信息主体；

f) 应使用何种方法修改或删除这些属性，以确保在不使用附加信息的情况下，个人信息不会被关联到已识别或可识别的自然人，同时保留对生成的假名数据进行一般分析的能力。可用的方法包括但不限于：删除、泛化和随机化；

g) 识别相关方和角色分工，包括处理者、受托的可信第三方等，由谁来执行假名替换操作（例如单独或共同执行）；

h) 确定假名化密钥或其他附加信息的存储位置，以及相应采取的安全保护技术和管理措施，确保它们无法在假名化域内访问和使用，且仅能在获得授权的情况下用于将假名数据归因于数据主体。

在定义了假名替换之后，处理者还需要评估在假名化域内发生重识别的风险，并确认该风险是微不足道的。

4.5.2 执行阶段

在应用假名替换时，处理者：

a) （可选）确定哪些数据记录属于同一个人信息主体，并为这些数据记录分配相应个人信息主体的唯一标识符；

b) 通过应用规划阶段确定的处理方式，将用于标识个人信息主体的所选属性和之前添加的唯一标识符（如果已插入）替换为假名，删除所有其他标识符，并将在此过程中生成的假名化密钥与假名数据分开存储；



c) 通过应用规划阶段确定的处理方式，修改或删除准标识符。

4.5.3 维护阶段

完成假名替换处理后，所有参与的处理者都应对用于重新识别的附加信息（特别是假名化密钥）采取相应技术和管理措施，严格限制对假名化密钥的访问与使用。

所有接收方都应采取适当的技术和组织措施确保假名数据无法离开假名化域，并同时确保任何已知能实现重识别的信息无法进入该假名化域。

最后，处理者应将对假名数据的处理限制在必要的范围内，以降低任何残余的逆转还原风险。

5 去标识化适用

去标识化作为一种降低个人信息可识别性的安全技术措施，有助于在确保个人信息可用的同时，降低处理活动对个人权益的影响，保障所处理个人信息的安全，防范未经授权的访问以及个人信息的泄露、篡改、丢失等风险：

a) 降低个人信息处理活动对个人权益的影响。去标识化有助于贯彻最小化处理个人信息原则，可以防止后续相关方直接获取个人的直接标识符，降低变更个人信息处理目的等风险。

b) 降低个人信息未经授权访问的风险。对个人信息进行去标识化处理，在遭受未经授权访问或因系统权限设置不当而被意外访问时，仅可访问、查看到去标识化的个人信息，可以降低个人相关信息



泄露的风险。

c) 降低数据泄露导致的个人信息泄露风险。当发生数据泄露，因个人信息采取了严格的去标识化措施，不借助额外信息的情况下无法识别特定自然人，可以降低相应数据泄露导致个人信息泄露的风险。

对个人信息采取适当的去标识化措施，可以作为最小化处理个人信息、采取必要措施保障所处理个人信息安全等履行个人信息保护义务的证明之一，同时在开展个人信息保护影响评估时可重点考量是否采取了适当的去标识化措施。当发生或者可能发生个人信息泄露、篡改、丢失的，相关个人信息去标识化措施能够确保个人信息的可识别性，以及被关联、复原、重识别的风险极低的，可以作为个人信息处理者采取措施避免个人信息泄露、篡改、丢失造成危害的证明举措之一。





附录 A 标识符的识别规则和示例

A.1 识别因素

A.1.1 标识符的基本特征

标识符包括以下基本特征。

a) 唯一性：具有确定的全局唯一性，或具有个性，但在特定环境下唯一或重复率极低。例如公民身份号码、唯一设备识别码具有确定的唯一性，姓名在特定范围内唯一或重复率极低。

b) 稳定性：具有不可变、不易变的性质，或通常不会发生改变。例如不可变唯一设备识别码不可更改，可变唯一设备识别码需要特定方式才能更改，姓名通常不会改变等。

c) 辨识性：易于辨别或易于检索。例如全脸照片易于辨别，移动电话号码、用户账号ID易于检索等。

d) 普遍性：在特定范围或特定行业领域内具有通用性和普遍性。例如公民身份号码在全国范围内通用、学号在全校范围内通用；姓名、移动电话号码、性别、年龄、地区等的收集在各种行业内普遍。

e) 与个人关联：与个人相关或与个人形成关联。包括来自个人固有的区分特征（如个人生物识别信息），或专门构建用于标识个人（如公民身份号码），以及本身是物品设备等的唯一标识，但与个人形成数据集层面的关联，通常用于标识个人（例如唯一设备识别码）。

A.1.2 直接标识符的识别因素

识别直接标识符时，还需考虑的因素包括：



- a) 单独实现唯一性；
- b) 属于或源于个人的区分特征。例如样貌特征属于个人区分特征，个人生物识别信息源于个人区分特征；
- c) 为了标识（特定身份下的）个人而构造、编码、生成。例如公民身份号码、学号、账号等；
- d) 属于设备、物品、账号、网络等的唯一标识，但通过与个人形成关联，可用于标识个人。例如唯一设备识别码、设备指纹、银行卡号、MAC地址等；
- e) 能够直接与个人取得联系。例如移动电话号码、即时通信账号等只对应一人。

A.1.3 准标识符的识别因素

识别准标识符时，还需考虑的因素包括：

- a) 结合其他属性可实现唯一性，且自身取值不唯一；
- b) 属于个人所属特定群体、所在区域的区分特征，通过与其他此类属性叠加，能够不断缩小个人身份的可能范围，直至实现唯一性。例如性别、年龄、地区、昵称（可重复）、头像，以及表征个人兴趣爱好、用户画像等的各类标签信息等；
- c) 属于设备、物品、账号、网络等的基本参数，但通过与个人形成关联，并基于特定算法进行加工处理，能够生成与个人关联的、具有唯一性的属性。例如已与个人形成关联的屏幕分辨率等设备参数信息。



A.2 识别原则

识别个人的前提是该属性与个人已形成关联，因此识别标识符遵循先关联再识别的原则，即首先对可能形成关联的数据集所具备的属性进行分析，通常分为以下情形：

a) 个人信息处理者内部共享：内部使用方所掌握各数据集中，是否存在同样具有该属性且与个人身份相关联的数据。如果存在，视为形成关联；

b) 向其他个人信息处理者提供：外部使用方所掌握各数据集中，是否存在同样具有该属性且与个人身份相关联的数据。如果存在，视为形成关联；

c) 公开个人信息：视为形成关联。

A.3 识别规则

在形成关联的基础上，标识符识别遵循以下规则。

a) 满足A.1.1中所有基本特征的个人信息可能属于标识符，其中任一基本特征不满足则不属于标识符。

b) 在满足A.1.1中所有基本特征的基础上：

1) 具备A.1.2中任一因素的，识别为直接标识符；

2) 具备A.1.3中任一因素的，识别为准标识符。

A.4 常见直接标识符

常见直接标识符（非全集）示例见表A.1。



表A.1 常见直接标识符（非全集）示例

类别	示例
个人基本资料	姓名、照片（全脸图片图像和其它任何可比对的图像）等
网络身份标识	用户账号、网络昵称（不可重复）、IP 地址、MAC 地址等
个人联系方式	移动电话号码、即时通信账号、电子邮件地址、住址等
个人身份信息	身份证、军官证、护照、驾驶证、工作证、出入证、社保卡、居住证、港澳台通行证等证件号码、证件照片或影印件
个人生物识别信息	生物识别原始信息（如样本、图像等）和比对信息（如特征值、模板等），如人脸、指纹、步态、声纹、基因、虹膜、笔迹、掌纹、耳廓、眼纹等
金融账户信息	银行卡号、支付账号、证券账户、基金账户、保险账户、公积金账户、公积金联名账号等
唯一设备识别码	Android ID、IDFA、IDFV、OAID 等可变更的唯一设备识别码，IMEI、IMSI、MEID、设备 MAC 地址、硬件序列号、ICCID 等不可变更的唯一设备识别码，车辆识别代号（仅在车厂间）等

A.5 常见准标识符

常见准标识符（非全集）示例见表A.2。

表A.2 常见准标识符（非全集）示例

类别	示例
个人基本资料	生日、年龄、性别、民族、国籍、语言、地区、籍贯、婚姻状况等
网络身份标识	网络昵称（可重复）、头像等
健康状况信息	体重、身高、体温、肺活量、血压、血型等
个人教育信息	学校、学历、学位、所学专业等
个人工作信息	职业、职位、职称、工作单位、所在部门、工作地点等
个人标签信息	基于个人上网记录等各类个人信息加工产生的用于对个人用户分类分析的描述信息，如 App 偏好、关系标签、终端偏好、内容偏好等标签信息



附录 B 个人信息去标识化参考案例

B.1 案例一：隐私面单

B.1.1 处理背景和目的

电商平台和快递公司为了保护收寄件人的个人信息（如姓名、电话、地址）在包裹流转过程中不被无关人员轻易获取，普遍采用“隐私面单”技术。其目的是在确保包裹能被快递员准确派送的前提下，最大限度地减少个人信息的直接暴露，降低信息泄露和被滥用的风险。

B.1.2 直接标识符和准标识符

收寄件人姓名、移动电话号码、详细收货地址（精确到门牌号）等直接标识符。

B.1.3 去标识化处理

收寄件人姓名遮蔽1个汉字以上（例如仅保留姓氏）；联系电话遮蔽6位以上；地址遮蔽单元户室号。

B.1.4 重识别风险

对于掌握了“额外信息”的快递员或内部员工而言，可以通过专用设备、App等扫描包裹条码将包裹与特定收件人重新关联。此外，即便仅从面单信息来看，在特定的小区或办公楼内，结合姓氏和部分地址信息，对于具备相关背景知识的人（例如邻居、同事）仍有较大概率能定位到具体个人。



B.2 案例二：行政执法信息公示

B.2.1 处理背景和目的

对行政处罚的相关执法对象基本信息、执法内容、法律依据等进行公示。

B.2.2 直接标识符和准标识符

被处罚人姓名、公民身份号码、家庭住址等直接标识符；单位名称等准标识符。

B.2.3 去标识化处理

对被处罚人姓名进行遮蔽仅保留姓氏，对公民身份号码保留前六位其余位遮蔽，对家庭住址保留至路段其余信息遮蔽；单位名称不进行处理。

B.2.4 重识别攻击风险

一般公众无法单凭页面信息确定具体自然人，但结合工商信息、媒体报道，以及家庭住址路段附近相关知情人士背景知识等，可能确定相关人员身份。

B.3 案例三：企业内部人力资源分析

B.3.1 处理背景和目的

一家大型企业的人力资源部门在符合处理个人信息的合法性基础前提下，希望对过去三年的员工离职情况进行统计分析，以找出离职率与员工年龄、司龄、部门、职级等因素的关联性，从而优化人才保留策略。分析工作由内部数据分析团队执行，为保护员工隐私，



HR部门不希望分析团队接触到能识别具体员工的信息。

B. 3.2 直接标识符和准标识符

员工姓名、员工编号、公民身份号码等直接标识符；所属具体部门、年龄、入职日期等准标识符。

B. 3.3 去标识化处理

删除员工姓名、公民身份号码等直接标识符。使用加盐哈希技术将“员工编号”转换为一个不可逆的假名；将精确年龄（如28岁）替换为年龄段（如26-30岁）；将具体的末端部门（如“市场推广二组”）泛化为上级业务单元（如“市场部”）；将精确入职日期替换为入职年份。

B. 3.4 附加信息隔离

将“员工编号”与假名的查找表存储在由HR信息系统管理员严格控制的独立数据库中，数据分析团队无任何访问权限，从而实现对附加信息的隔离。

B. 3.5 重识别风险

数据分析师拿到的数据集中，直接标识符已被替换，准标识符也被泛化。即使攻击者拥有一定的背景知识（例如，知道市场部去年有名28岁的员工离职），由于年龄已被泛化为“26-30岁”，无法唯一确定是哪位员工。除非攻击者能同时获取分析数据集和被隔离的查找表，否则难以重识别。



B.4 案例四：医疗数据开放科研

B.4.1 处理背景和目的

一个由多所大学组成的科研联盟建立了一个“数据中心”，旨在开展一项大型纵向健康研究。在符合处理个人信息的合法性基础前提下，该研究需要安全地汇集并链接来自两个完全独立来源的个人信息：一是各大学附属医院提供的患者医疗数据，二是从政府劳工机构获取的同一批人的职业暴露风险数据。核心目的是在保护参与者隐私的前提下，实现跨领域数据的安全融合与分析，同时保留在发现重大个人健康风险等极端情况下，能够联系到参与者的能力。

B.4.2 直接标识符和准标识符

参与者的公民身份信息、医院内部ID（MedID）、劳工机构内部ID（HR-ID）等直接标识符；年龄、性别、职业、所在地区等准标识符；具体的医疗诊断记录、治疗方案、职业暴露类型和时长等目标属性。

B.4.3 去标识化处理

联盟委托一个独立的“信任中心”作为可信第三方来执行相关的假名化操作。医院和劳工机构各自将其内部的用户ID（MedID和HR-ID）安全地发送给信任中心，由“信任中心”为项目中的每一位个人信息主体生成一个统一的、与真实身份无关联的研究对象ID（SubjID）。



B. 4. 4 附加信息隔离

包含“MedID/HR-ID \Leftrightarrow SubjID”完整映射关系的查找表仅由信任中心加密保管。

B. 4. 5 重识别攻击风险

由于能够复原身份的附加信息(查找表)由完全独立的实体保管,除非攻击者能同时攻破数据中心和信任中心,否则无法完成重识别。





附录 C 个人信息脱敏示例

表 C.1 个人信息脱敏示例

信息项	脱敏目标	推荐方法	具体要求与示例	适用场景/备注
姓名	降低个人可识别性	遮蔽	保留姓氏，隐藏名字。例如：张三 -> 张* 或 张**；诸葛亮 -> 诸葛**	用于前端界面展示、日志记录等，用户本人能识别即可。
		泛化	用称谓替换。例如：张三 -> 张先生	用于客服记录、统计报表等。
		假名替换	生成一个虚假姓名。例如：张三 -> 赵伟	用于开发测试环境的数据生成。
公民身份号码	消除唯一性，隐藏生日和地区等信息	删除/抑制	删除或抑制屏蔽。例如：110101199001011234 -> *****	无需展示具体信息的场景（如仅提示“已实名”）。
		遮蔽	保留前6位、前4位或前2位，其余遮蔽。例如：110101199001011234 -> 110101*****	仅需保留大致地区信息的场景。
		截断	仅保留后4位。例如：110101199001011234 -> 1234	内部核验等非公开场景。
移动电话号码	消除唯一性，防止骚扰	遮蔽	保留前3位和后4位。例如：13812345678 -> 138****5678	用于用户识别和信息展示。
		哈希/加密	使用加盐哈希（HMAC-SHA256）生成唯一标识符。例如：13812345678 -> a9b8c7...	用于用户行为分析、跨表数据关联。
电子邮箱	消除唯一性，防止垃圾邮件	遮蔽	保留首尾字符和域名。例如：username@example.com -> u*****e@example.com	用于前端展示。
		部分遮蔽	隐藏用户名部分字符。例如：username@example.com -> user****@example.com	用户识别度更高。
家庭住址	降低地理位置精度	泛化	模糊到街道、区或市级。例如：北京市海淀区中关村大街1号 -> 北京市海淀区中关村大街 或 北京市海淀区	用于地区统计、物流区域分析。



信息项	脱敏目标	推荐方法	具体要求与示例	适用场景/备注
		随机偏移	对经纬度坐标进行随机偏移。例如：(39.98, 116.31) -> (39.97, 116.32)	用于基于位置的服务（LBS）数据分析，保护精确落点。
银行卡号 / 支付账号	防止金融欺诈	遮蔽	保留前6位和后4位。例如：6222020100123456789 -> 622202*****6789	用于支付信息展示等。
		截断	仅保留后4位。例如：6222020100123456789 -> 尾号 6789	安全要求较高的场景。
车牌号	降低车辆可识别性	遮蔽	保留前两位（省份和地区）和后两位。例如：京 A88888 -> 京 A****88	用于违章记录、停车系统展示。
设备ID (IMEI, MAC地址)	降低设备可识别性	哈希 / 加密	使用加盐哈希生成唯一标识符。例如：A1-B2-C3-D4-E5-F6 -> f5e4d3...	用于设备激活统计、用户行为分析。
		抑制	直接删除或置空。	如果不需要设备粒度的分析，可直接删除。
出生日期	降低时间精度，保留年龄特征	泛化（年龄）	计算为年龄。例如：1990-05-15 -> 34岁	最常用的用户画像分析方法。
		泛化（年龄段）	计算为年龄段。例如：1990-05-15 -> 30-35岁	进一步降低精度，用于统计分析。
		泛化（年份）	保留出生年份。例如：1990-05-15 -> 1990年	用于按年代分析用户群体。
性别 / 国籍 / 民族	消除个体精确属性，保护少数群体隐私	抑制 / 删除	在不需要此维度进行分析时，直接删除该字段或置为NULL。例如：男 -> NULL	当该字段非业务分析核心，或为最大化保护隐私时采用。
		添加噪声 / 随机扰动	按一定比例（如5%）随机反转部分记录的属性值。例如：数据集中随机将5%的男改为女，5%的女改为男。	用于统计分析场景，可在保护个体精确值的同时，基本维持整体数据的统计分布特性。
邮政编码	降低地理位置精度	泛化 / 截断	保留前2位或3位。例如：100084 -> 100xxx 或 100	用于较大范围的地理分布分析。



信息项	脱敏目标	推荐方法	具体要求与示例	适用场景/备注
工作单位/职业	降低身份关联性	泛化	替换为行业类别或职业大类。例如：XX 科技有限公司 软件工程师 -> 信息技术/软件开发	用于职业分布、行业分析。
密码	防止凭证泄露	哈希	使用加盐的强哈希算法（如 Argon2, bcrypt, scrypt）。例如：password123 -> \$2b\$12\$. . .	严禁明文存储或可逆加密存储密码。
收入/财产信息	保护个人隐私	泛化	替换为收入范围。例如：月收入 12500 元 -> 10000-15000 元	用于市场分析、信用评估模型训练。
		抑制	直接删除。	如果与业务目的无关，应直接删除。
病历/健康信息	保护个人健康隐私	抑制	删除所有可识别信息（姓名、ID 等）。	用于医学研究的基础要求。
		哈希、加密或令牌化	为每个患者生成唯一的、无内在含义的标识符（假名），以代替真实身份信息。	用于需要长期跟踪研究的场景。
		泛化	疾病名称使用上层分类。例如：2 型糖尿病 -> 内分泌系统疾病	用于宏观疾病统计。
通话/短信记录	保护通信内容和关系隐私	抑制	删除通话内容、短信正文。	用于通信行为分析时，应删除内容。
		哈希	对对方号码进行哈希处理。	用于社交网络分析，研究关系强度。
网页浏览历史	保护个人兴趣和行为隐私	泛化	将具体 URL 替换为网站类别或标签。例如：https://.../product/123.html -> 电商/手机	用于用户兴趣画像、推荐系统。
GPS 轨迹	保护行踪隐私	降采样	降低轨迹点的密度。	减少数据点，保留大致路径。
		随机偏移	对所有坐标点进行统一或随机的偏移。	用于基于位置的服务（LBS）数据分析，保护精确落点。
		聚合	将轨迹数据聚合为区域停留信息（如“在 A 区域停留 2 小时”）。	用于分析城市热点、通勤模式。



参考文献

- [1] 中华人民共和国个人信息保护法
- [2] GB/T 37964—2019 信息安全技术 个人信息去标识化指南
- [3] GB/T 41833—2022 快递电子运单
- [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 25237:2017. Health informatics – Pseudonymization
- [6] ISO/IEC 20889:2018. Privacy enhancing data de-identification terminology and classification of techniques
- [7] ISO/IEC 27559:2022. Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework
- [8] ISO/IEC 29100:2024. Information technology – Security techniques – Privacy framework
- [9] European Data Protection Board (EDPB). Guidelines 01/2025 on Pseudonymisation. 2025
- [10] Information Commissioner's Office (ICO). Anonymisation and Pseudonymisation guidance. 2025