国 际 电 信 联 盟

ITU-T X.1148

(09/2020)

国际电信联盟
电信标准化部门

X系列：数据网、开放系统通信和安全性

安全应用和服务（1）– 网页安全

电信服务提供商的去身份识别进程框架

ITU-T X.1148建议书

国际电信联盟

ITU-T X 系列建议书

数据网、开放系统通信和安全性

公用数据网	X.1–X.199
开放系统互连	X.200–X.299
网间互通	X.300–X.399
消息处理系统	X.400–X.499
号码簿	X.500–X.599
OSI组网和系统概貌	X.600–X.699
OSI管理	X.700–X.799
安全	X.800–X.849
OSI应用	X.850–X.899
开放分布式处理	X.900–X.999
信息和网络安全	
一般安全问题	X.1000–X.1029
网络安全	X.1030–X.1049
安全管理	X.1050–X.1069
生物测定	X.1080–X.1099
安全应用和服务（1）	
组播安全	X.1100–X.1109
家庭网络安全	X.1110–X.1119
移动安全	X.1120–X.1139
网页安全	**X.1140–X.1149**
安全协议（1）	X.1150–X.1159
对等网络安全	X.1160–X.1169
网络身份安全	X.1170–X.1179
IPTV安全	X.1180–X.1199
网络空间安全	
网络安全	X.1200–X.1229
反垃圾信息	X.1230–X.1249
身份管理	X.1250–X.1279
安全应用和服务（2）	
应急通信	X.1300–X.1309
泛在传感器网络安全	X.1310–X.1319
智能电网安全	X.1330–X.1339
验证邮件	X.1340–X.1349
物联网（IoT）安全	X.1360–X.1369
智能交通系统（ITS）安全	X.1370–X.1389
分布式账簿技术安全	X.1400–X.1429
分布式账簿技术安全	X.1430–X.1449
安全协议（2）	X.1450–X.1459
网络安全信息交换	
网络安全概述	X.1500–X.1519
漏洞/状态信息交换	X.1520–X.1539
事件/事故/启发式信息交换	X.1540–X.1549
政策的交换	X.1550–X.1559
启发式和信息请求	X.1560–X.1569
标识和发现	X.1570–X.1579
确保交换	X.1580–X.1589
云计算安全	
云计算安全概述	X.1600–X.1601
云计算安全设计	X.1602–X.1639
云计算安全最佳做法和指导原则	X.1640–X.1659
云计算安全实施方案	X.1660–X.1679
其他云计算安全	X.1680–X.1699
量子通信	
术语	X.1700–X.1701
量子随机数发生器	X.1702–X.1709
QKDN安全框架	X.1710–X.1711
QKDN安全设计	X.1712–X.1719
QKDN安全技术	X.1720–X.1729
数据安全	
大数据安全	X.1750–X.1759
5G 安全	X.1800–X.1819

欲了解更详细信息，请查阅ITU-T建议书目录。

ITU-T X.1148建议书

电信服务提供商的去身份识别进程框架

摘要

电信组织收集、管理、使用和共享包括个人身份信息在内的个人数据。因此，他们可利用数据去身份识别技术为个人数据提供保护。ITU-T X.1148建议书阐述了一个配有操作步骤的去身份识别进程框架，依据数据生命周期模型和利益攸关方的角色，规定了电信服务提供商去身份识别进程的数据发布模式和各个数据阶段。

历史沿革

版本	建议书	批准日期	研究组	唯一识别码*
1.0	ITU-T X.1148	2020-09-03	17	11.1002/1000/14249

关键词

数据主体、去识别、去识别过程、k-匿名、l-多样性、PII-保护、释放模型、t-接近度。

* 欲查阅建议书，请在您的网络浏览器地址域键入URL http://handle.itu.int/，随后输入建议书的唯一ID，例如，http://handle.itu.int/11.1002/1000/11830-en。

前言

国际电信联盟（ITU）是从事电信、信息和通信技术（ICT）领域工作的联合国专门机构。国际电信联盟电信标准化部门（ITU-T）是国际电信联盟的常设机构，负责研究技术、操作和资费问题，并且为在世界范围内实现电信标准化，发表有关上述研究项目的建议书。

每四年一届的世界电信标准化全会（WTSA）确定ITU-T各研究组的研究课题，再由各研究组制定有关这些课题的建议书。

WTSA第1号决议规定了批准建议书须遵循的程序。

属ITU-T研究范围的某些信息技术领域的必要标准，是与国际标准化组织（ISO）和国际电工技术委员会（IEC）合作制定的。

注

本建议书为简明扼要起见而使用的“主管部门”一词，既指电信主管部门，又指经认可的运营机构。

遵守本建议书的规定是以自愿为基础的，但建议书可能包含某些强制性条款（以确保例如互操作性或适用性等），只有满足所有强制性条款的规定，才能达到遵守建议书的目的。“应该”或“必须”等其它一些强制性用语及其否定形式被用于表达特定要求。使用此类用语不表示要求任何一方遵守本建议书。

知识产权

国际电联提请注意：本建议书的应用或实施可能涉及使用已申报的知识产权。国际电联对无论是其成员还是建议书制定程序之外的其它机构提出的有关已申报的知识产权的证据、有效性或适用性不表示意见。

至本建议书批准之日止，国际电联已收到实施本建议书可能需要的受专利保护的知识产权的通知。但需要提醒实施者注意的是，这可能并非最新信息，因此大力提倡他们通过下列网址查询电信标准化局（TSB）的专利数据库： http://www.itu.int/ITU-T/ipr/。

目录

引言

随着基于互联网的信息通信技术及服务的快速发展，不但产生了海量数据且数据的传输和存储亦出现了爆炸式增长。生成数据的来源多种多样：不仅有传感器、照相机或网络设备，还包括网页、电子邮件系统或社交网络以及诸多其他来源。数据集变得如此庞大复杂且增长速度如此之快，以至于传统数据处理方法和工具力不从心。在可容忍的延迟范围内开展高效数据分析变得极具挑战性。为解决上述问题，人们正在开发一种大数据分析范式。

电信组织收集、管理、使用和共享包括个人身份信息在内的个人数据。因此，他们可利用数据去身份识别技术为个人数据提供保护。参与数据交换而产生数据流的各方之间的关系，将对究竟是在收集数据之前、收集数据之后但在数据保存之前，还是仅在数据交换过程中与下一参与方共享数据之前，执行数据去识别操作产生影响。因此，电信服务提供商需要以及时、高效和安全的方式，向数据客户提供数据去识别服务。

ITU-T X.1148建议书

电信服务提供商的去身份识别进程框架

1 范围

本建议书依据数据生命周期模型，概要阐述了一个配有操作步骤的去身份识别进程框架以及各利益攸关方在去身份识别进程中扮演的角色。此外，本文还深入讨论了去身份识别进程的数据发布模式和各个数据阶段，在附件和附录中展示了各种去身份识别的方法。

本建议书不涉及有关监管的问题。

2 参考文献

无。

3 术语和定义

3.1 他处定义的术语

本建议书使用了下列他处定义的术语：

3.1.1 **聚合数据**[b-ISO/IEC 20889]：表示一组数据主体的数据，例如该组数据统计属性的集合。

3.1.2 **匿名化**[b-ISO/IEC 29100]：一种个人身份信息（PII）出现不可逆转变化的过程。在这种过程中，再也无法单独通过 PII 控制程序或与任何其它方协作，直接或间接识别 PII 主体。

3.1.3 **属性**[b-ISO/IEC 20889]：固有特性。

3.1.4 **数据集**[b-ISO/IEC 20889]：一组数据。

3.1.5 **去身份识别**[b-ISO 25237]：通用术语，用于描述旨在移除一系列识别数据与数据主体之间关联的所有进程（参见第 3.2.4 节）。

3.1.6 **去身份识别进程（de-identification process）**[b-ISO/IEC 20889]：旨在移除一系列识别属性与数据主体之间关联的进程。

3.1.7 **去身份识别技术** [b-ISO/IEC 20889]：用于转换数据集的方法，其目的是减少信息与单个数据主体关联的能力。

3.1.8 **去身份识别的数据集** [b-ISO/IEC 20889]：应用去身份识别进程而形成的数据集。

3.1.9 **去身份识别信息** [b-NISTIR 8053]：此类信息已经移除或掩藏了足够多 PII 的记录，经过这些操作后，剩余信息无法识别个人且没有合理的依据相信该信息可以用于个人识别。

3.1.10 **差分隐私** [b-ISO/IEC 20889]：一种正式的隐私衡量模型，可确保统计分析输出的概率分布最多相差一个指定值，而无论输入数据集是否表达了任何特定的数据主体。

注 – 具体而言，差分隐私提供：

a) 一种有关隐私的数学定义，该定义假设应认为任何统计分析的结果都提供隐私保护，人们无法区分原始数据集的分析结果与将任何数据主体添加到数据集或从数据集中删除后获得的结果；且

b) 一种隐私保护措施，能够监控累积的隐私信息损失并设置损失上限（或“预算”）。这种措施的正式定义如下。设ε为正实数，M为以数据集作为输入内容的随机化算法。如果单个元素（即一个数据主体的数据）以及M范围内的所有子集S和mml_m1（用概率取代了算法使用的随机性）中的所有数据集D1和D2并不相同，则认为算法M实现了差分隐私。

3.1.11 标识符[b-ISO/IEC 20889]：数据集内的一组属性，用于以唯一的形式标识特定操作环境下的数据主体。

注 – 有关该定义与其他标准所给出定义关系的讨论，请参见附件B。

3.1.12 识别属性[b-ISO/IEC 20889]：数据集内的属性，有助于以唯一的形式标识特定操作环境下的数据主体。

3.1.13 隐私利益攸关方[b-ISO/IEC 29100]：自然人或法人、公共管理部门、机构或任何其他机构，他们可能会影响与个人身份信息（PII）处理相关的决定或活动，受到此类决定或活动的影响或认为自己受到这方面的影响。

3.1.14 假名化[b-ISO/IEC 20889]：一种去识别技术，使用假名取代数据主体标识符以隐藏数据主体的身份。

3.1.15 准标识符[b-ISO/IEC 20889]：数据集中的属性，在与数据集中的其他属性结合考虑时，可标识一个数据主体。

3.1.16 记录[b-ISO/IEC 20889]：关于单个数据主体的一组属性。

3.1.17 重新识别[b-ISO/IEC 20889]：将去识别数据集中的数据与原始数据主体中的数据建立关联的过程。

注 – 本定义中包含确定特定数据主体存在于某数据集的过程。

3.1.18 遴选[b-ISO/IEC 20889]：通过观察一组可唯一识别此数据主体的已知特征，将属于数据集内某数据主体的记录分离出来。

3.1.19 第三方[b-ISO/IEC 29100]：除个人身份信息（PII）主体、PII控制程序和PII处理程序，以及在PII控制程序或PII处理程序直接授权下获得数据处理权的自然人之外的隐私利益攸关方。

3.1.20 受信任的第三方[b-ISO/IEC 18014-1]：在与安全相关的活动方面得到其他实体信任的安全机构或其代理。

3.1.21 k-匿名算法[b-ISO/IEC 20889]：正式的隐私衡量模型，可确保数据集中的每个标识符都有至少包含 K 条记录的一个对等类别。

3.1.22 l-多样性[b-ISO/IEC 20889]：正式的隐私衡量模型，可确保对于选定的属性，每个等价类别至少有L个展现良好的值。

注 – L-多样性是数据集的一个属性，给出了有关所选属性等价类别共享值多样性的保证下限L。

3.1.23 t-保密算法 [b-ISO/IEC 20889]：正式隐私衡量模型，可确保等价类别中所选属性的分布与该属性在整个表中的分布间距不超过门限值 T。

注 – 如果包含选定属性的所有等价类别都有T-保密算法，则认为表中提供关于该属性的T-保密算法。

3.2 本建议书中定义的术语

本建议书定义了下列术语：

3.2.1 **数据控制方**：确定处理数据的目的和方法的利益攸关方（或隐私利益攸关方），而非出于个人目的使用数据的自然人。

3.2.2 **数据处理方**：代表数据控制方并根据其指令处理数据的利益攸关方。

3.2.3 **数据保护员**：由个人身份信息（PII）控制方任命的人员，以独立的方式确保隐私法律或法规的要求得到遵守。

注 – “PII控制方”是“数据控制方”的同义词。

3.2.4 **数据主体**：与数据相关的实体。

注 – “数据主体（data subject）”与“PII主体”和“数据主体（data principal）”是同义词。

3.2.5 **过程**：就信息或数据而言，过程是指获取、记录或保存信息或数据，或对信息或数据执行任何操作或一组操作，包括：

– 信息或数据的组织、修改或变更；
– 信息或数据的检索、查询或使用；
– 通过传输、传播的方式披露信息或数据，或以其他方式提供信息或数据，或
– 信息或数据的一致性保持、组合、封锁、擦除或销毁。

4 缩写词和首字母缩略语

本建议书使用了下列缩写词和首字母缩略语：

DP 差异隐私
DPO 数据保护员
PII 个人身份信息
TTP 受信任的第三方

5 惯例

无。

6 去身份识别进程概述

去身份识别进程的宗旨是保护主体数据的机密性。鉴于这些数据可能包含个人身份信息（PII），所以在进行数据分析前后，为提取有意义的信息，数据分析师必须将安全因素考虑在内。

本节定义了数据分析环境、数据生命周期模型、各实体在去身份识别进程和其他去身份识别考虑中的作用。

6.1 数据生命周期模型和去身份识别阶段

通常，某个组织会为实现 V 隐私和安全的目的而设定去身份识别目标。本节定义了数据生命周期，并描述了根据此数据生命周期模型应在何时考虑去身份识别进程。

数据生命周期概念用于根据对重新识别可能性的分析来选择施加适当的控制。本建议书按第 6.1.1 至 6.1.5 节所述定义数据生命周期。

图 1 概述了数据生命周期模型中的去身份识别进程。

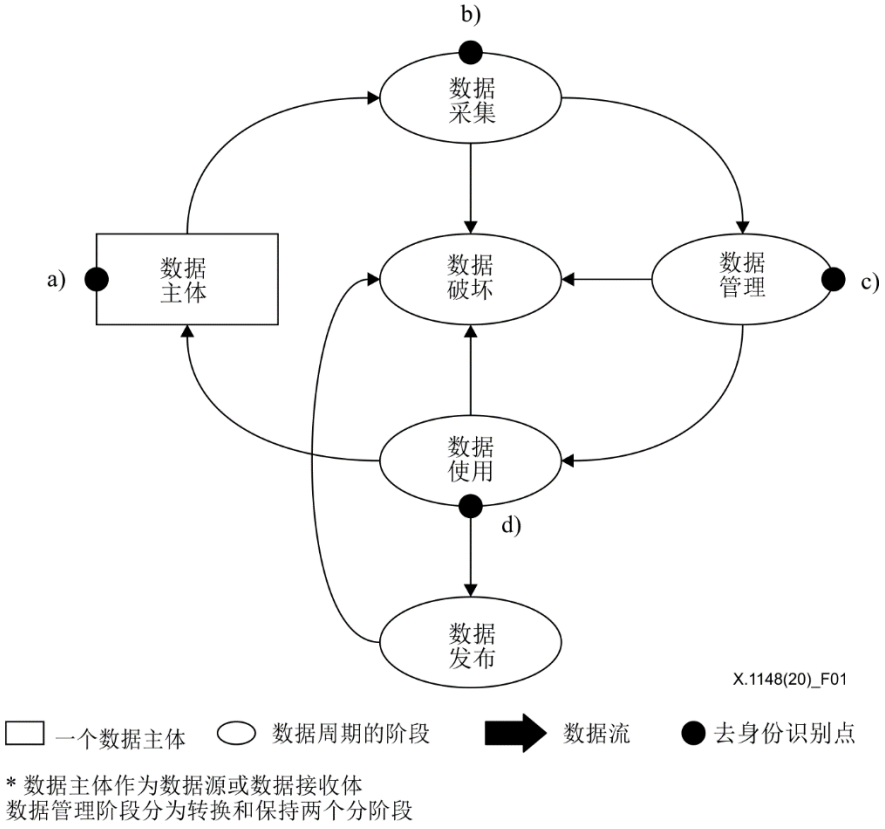


图 1 – 数据生命周期模型中的去身份识别进程

6.1.1 数据采集阶段

数据是从数据主体，即数据所提及的人员那里收集而来。此类数据收集生成的数据集可能包括 PII。去身份识别会创建一个新的数据集，从中删除所有 PII。建议相关组织尽可能在内部使用去身份识别的数据集，而非原始数据集。

使用此模型，去身份识别可能在出现以下两种情况之一时发生：

– 在数据采集过程中，即图1中的（b）；或

– 虽然采集了数据但实际上并不需要标识符，即图1中的（a）。

不应采集数据管理（数据转换和数据保存）并不需要的标识符。

6.1.2 数据管理阶段

为了避免对标识符进行存档，应在数据转换之后和保存数据之前实施去身份识别操作，参见图 1 中的(c)。建议相关组织考虑重新识别的可能性，并设置明确的访问控制权限、最大保存限制和数据删除策略，以最大限度地降低去身份识别数据之间存在关联的可能性。亦建议相关组织在设定用途允许的情况下，考虑使用数据聚合等匿名化技术。

6.1.3 数据使用阶段

如果某组织内部需要使用 PII 进行数据管理，则建议在将数据作为数据共享数据集发布之前，对其实施去身份识别操作，见图 1 中的(d)。

6.1.4 数据发布阶段

可与受“数据共享”协议等附加管理控制约束的第三方共享数据。此外亦可能发布去身份识别的数据集。去身份识别数据的发布分为三种模式：公开、半公开或非公开。根据选择发布模式的不同，所需去身份识别的数量可能会有所不同。

6.1.5 数据销毁阶段

数据销毁可以在任何阶段进行，即数据采集、数据管理、数据使用和数据发布阶段。应使用经验证的措施销毁数据，以避免数据恢复。特别是在检测到存在重新识别的可能性时，应考虑销毁数据。

6.2 去身份识别方面的考虑

在整个数据生命周期中应用去身份识别可以提高数据生命周期的有效性。然而，数据流各参与方之间关系的性质，将对究竟是在收集数据之前（即图 1 中的(a)）、在收集数据之后（即图 1 中的(b)）但在数据保存之前（即图 1 中的(c)），还是仅在数据交换过程中与下一参与方共享数据之前（即图 1 中的(d)），执行数据去识别操作产生影响。这一决定反过来又会影响安全性和采取其他组织措施的可行性，这些其他措施旨在提高特定去身份标识技术在各用例中的有效性。虽然在使用目的不支持采用匿名化技术的情况下，去身份标识可能是一种保护主体数据机密性的有用技术，但其本身不足以保护主体数据，因而必须将其视作全面数据保护框架的一部分。本节描述了各阶段的特点和考虑。

6.2.1 数据采集

本地去身份标识（或在源头去身份标识）是一种最为重要的方法，允许个人（或为个人处理数据的控制方）在发布用于分析的数据之前移除所有 PII。

与数据采集阶段直接相关的一个去身份标识问题是数据最小化。负责采集主体数据的各数据控制方均需精确定义哪些数据对预期使用目的而言是绝对必要的，并将数据采集工作仅限于收集那些定义的参数。

应出台具体流程，将不必要的 PII 排除在数据采集或传输之外，以减少数据字段的数量。

去身份识别的另一个问题是数据聚合。数据控制方需在各种情况下考虑数据聚合，其中使用目的并不严格要求遴选出单个数据主体。

6.2.2 数据管理

6.2.2.1 数据转换

数据转换阶段可能包括应用聚合、统计披露限制、加密等去身份识别技术。数据转换可以在一个或多个阶段应用，例如在采集数据之后和长期保存之前直接应用，在相当长的保存期之后和访问之前应用，或与访问一起应用。

数据编辑或聚合的常见转换可以在数据采集后至发布前的任何时间使用。如果在采集后立即应用，编辑或汇总数据或可减少数据泄露给数据主体造成的潜在危害；然而，这样做也削弱了编辑后进行关联、合并或更新数据的可能性。

数据转换方法的选择，应在审慎考虑暴露于数据主体可能带来的潜在危害后做出。转换决策还应虑及必须得到今后数据使用目的支持的分析，因为旨在降低披露风险的技术会给此后可能的使用和分析造成影响。

6.2.2.2 数据保存

数据保存是指由数据控制方或在其指导下行动的执行方，将数据（包括 PII）存储到任何形式非易失性存储器的过程。鉴于信息安全和隐私控制集中于保存阶段，因此本节总结了控制措施，但没有提供详细的考虑因素[b-ISO/IEC 27001]。有些信息安全和隐私控制方式在保存阶段很常见，如访问控制、维护、安全评估、认证程序、事件监控和响应以及审计。

相关组织特别应该遵循有关最长数据保存期和删除的政策，以确保数据保存时间绝不超过实现使用目的所需的时间，并且会在最长保存期过后数据将数据完全销毁。例如，数据共享协议通常规定接收方必须在特定时间，例如在接收后一年内，销毁数据。此外法律也可能要求制定这样的合同条款。

6.2.3 数据的使用

可以为一系列目的和应用采集、存储或共享已去身份识别的数据，这些数据均依赖于去身份识别后保留的某些数据属性。发布去身份识别数据集的主要原因之一是，为其他人提供研究原始数据价值和属性的机会[b-ISO/IEC 20889]。因此，在保护个人隐私的同时，去身份识别还应该尽可能保留信息的实用性。去身份识别的这种双重目的使得其成为一种值得考虑的重要方法，供人们在数据发布模式等诸多环境下使用。

在发布已去身份识别的数据时，相关组织必须就考虑与发布有关的数据主体所产生的潜在影响做出决定，且决策机构通常是一个涵盖广泛利益攸关方的专家委员会。风险评估和核对清单一般用于指导评估，并确定有利于降低重新识别风险的适当发布机制。

去身份识别技术的选择，取决于相关技术针对特定用例的适用性或“效用”。

7 去身份识别进程框架

本节描述了一个通过四步提供去身份识别 PII 的去身份识别进程框架，如图 2 所示[b-KOREA]。

步骤1 – 初步审查

步骤 1 涉及验证目标数据是否为 PII。如果数据中的确包含 PII，请继续步骤 2。有必要实施去身份识别操作。

步骤2 – 去身份识别

步骤 2 涉及对数据进行去身份识别操作，以防止有人从目标数据集中推断特定的个体信息。此步骤调用相应的方法，旨在全部、部分移除或转换 PII 元素。PII 元素包括标识符、准标识符和敏感属性。

步骤3 – 充分性评估

步骤 3 涉及评估已去身份识别的数据集的充分性，其中就涵盖 PII 元素。相关考虑包括目标数据集是否仍包含 PII、直接重新识别的可能性、能够导致重新识别的潜在关联。

步骤4 – 后续管理

步骤 4 涉及为防止重新识别而进行的管理和技术安全性衡量。

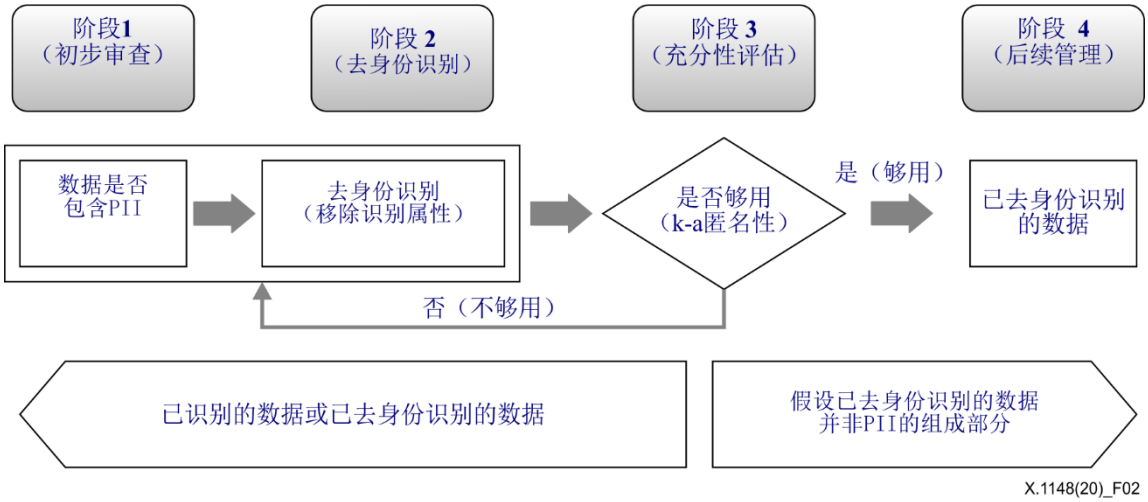


图 2 – 去身份识别进程

第 7.1 至 7.4 节进一步阐述了这些步骤。

7.1 步骤1 – 初步审查

拟使用或为各种目提供数据的组织，应首先确定其政策和标准。推荐采用的政策和标准包括以下内容：

– 已去身份识别的信息的目的和预期用途是什么？
– 已去身份识别的数据包含哪些数据属性？
– 去身份识别使用什么技术？
– 重新识别的风险水平和不利影响是什么？
– 如果某个特定个人被重新识别，有何应对解决方案？
– 如何评估重新识别的水平？
– 如何确定去身份识别所需人力和成本？

初步审查的具体考虑因素可能因数据类型和预期用途而异。然而，我们建议应建立一套相应的标准。

拟为多种目的处理数据的组织应参考适当的标准，以验证特定数据是否为 PII 数据。即使数据未被确定为 PII，该组织亦需考虑可用数据之间任何潜在的关联性，并采取适当措施将尽量降低此风险。如果判定结果是肯定的，则有必要执行去身份识别步骤。

PII 的判断标准包括：

– 数据的类型、形式、特征和格式没有特别限制；
– 如果数据控制方能够使用某种数据识别个人，则认为这种数据是PII；

– 数据必须关乎个人。由多个个体构成的组的统计值并非PII；
– 可通过与附加信息的结合来对个人进行识别的数据属于PII。附加信息通常指公开或容易获得的信息。

7.2 步骤2 – 应用去身份识别操作

7.2.1 标识符的去身份识别

“标识符”是一种数据，例如分配给个人的唯一值或名称或与个人相关的事物。总体而言，应尽量少收集“标识符”，并应删除数据集中包含的任何标识符。

然而，对于预期目的而言，绝对必要的标识符可能包括如下数据：

– 唯一标识符（居民登记号、社会保险号（SSN）、护照号、外国人身份证号、驾照号等）；
– 姓名（中文字、英文名称等）；
– 详细地址（门牌号、街道地址等）；
– 日期（生日、周年纪念日（婚礼等）、证书日期等）；
– 电话号码（手机、家庭、办公室、传真等）；
– 病历号、国家医疗保险号、福利领取人号等；
– 银行账号、信用卡号等；
– 照片（静态图片、视频、闭路电视视频等）；
– 生物数据（指纹、声音、虹膜等）；
– 电子邮件地址、IP地址、媒体访问控制地址（MAC）、主页统一资源定位符等；
– 身份识别码（员工号、客户号等）；
– 其他唯一识别号码（兵役号码、商业登记号码等）。

7.2.2 准标识符和高可识别属性的去身份识别

一般而言，如果数据集中包含的准标识符与数据的用途无关，则应将其移除。如果与数据使用相关的准标识符拥有可识别的要素，则应使用诸如假名化和聚合等去身份识别技术。

行为信息等可能携带高可识别性信息的数据，必须尽可能使用匿名技术实施去身份识别。

7.2.3 去身份识别技术

包括假名化、聚合、数据抑制和数据屏蔽在内的一系列技术，既可以单独使用亦可组合使用。仅应用假名技术可能不足以实现去身份识别。

各种类型的技术均可供使用。我们应根据数据使用的目的以及每种特定技术的优缺点来选择使用最合适的技术。去身份识别完成后，继续下一步。

7.3 步骤3 – 去身份识别进程的充分性评估

当去身份识别不充分时，可以通过结合其他数据或使用各种推断技术来识别个体。

为了降低重新识别的风险，在使用之前，有必要对已去身份识别的数据进行充分性评估。此评估的内容包括以下问题：

– 此去身份识别请求的目的是什么？
– 去身份识别涉及哪种类型的数据属性（包括标识符还是不包括标识符）？
– 去身份识别的适当级别？

这种充分性评估可以由数据保护员（DPO）即受信任的第三方（TTP）执行或由外部评估组执行。

在评估充分性时，K-匿名算法模型与其他隐私保护模型一起使用。K-匿名算法模型是一种基本评估手段。可在必要时应用其他评估模型（l-多样性、t-保密算法、差异隐私等）。

关于充分性评估的更多详细信息，请参考附件 A。

7.4 步骤4 – 后续管理

7.4.1 为已去身份识别的数据提供的保护措施

实施保护措施是为了防止在已去身份识别的数据遭泄露和/或与其他数据相结合的情况下，出现重新识别这些数据的可能性。这些措施包括：

– 管理保护措施：指定一名人员负责已去身份识别的数据文件，确定是否要共享已去身份识别的数据，并在达到使用目的后销毁这些数据；
– 技术保护措施：限制对已去身份识别的数据文件的访问，管理访问记录，安装和运行安全程序；

此外，安全措施还包括在已去身份识别的数据泄露的情况下采取的保护措施。这些措施包括：

– 分析泄漏原因，实施管理和技术安全措施，防止进一步泄漏；
– 撤回并销毁泄露的已去身份识别数据。

7.4.2 监控重新识别的可能性

若拟使用已去身份识别的数据或将其提供给第三方的数据控制方，应定期监控重新识别的可能性。

在检测到重新识别的可能性时，必须向已为其提供了去身份识别数据的数据控制方申请暂停数据处理、撤回和销毁操作。

7.4.3 与第三方签约的要求

在向第三方提供或委托第三方使用已去身份识别的数据时，合同中须包括重新识别风险管理。重新识别风险的管理包括：

– 向第三方披露数据时通知数据主体；
– 尽可能向第三方提供匿名数据；
– 禁止重新识别：规定已获得或受托处理去身份识别数据的数据控制方，不得通过与其他数据结合的方式重新识别数据；
– 对重新提供或重新委托的限制：在提供或委托处理去身份识别数据时，应在合同中规定允许重新提供或重新委托的范围；
– 有关重新识别风险的通知：规定有义务终止数据处理，并在重新识别数据或重新识别数据的可能性变高时，将重新识别数据问题通知委托人和受托人。

7.4.4 重新识别的应对措施

如果已去身份识别的数据被重新识别，则应停止数据处理，并采取必要措施防止PII泄漏。

须立即销毁重新识别的数据。

8 已去身份识别的数据的使用

8.1 去身份识别数据的各个阶段

本节定义了可以用数据类型表示的数据去身份识别阶段，借以描述通过数据直接识别个人的程度以及个人如何与数据特征（属性）建立关联。在数据使用或数据处理的背景下，数据规范的内容不仅应包括数据类型，还应阐述数据识别个人的能力或将个人与数据中的一组特征联系起来的程度。

图 3 展示了在去身份识别过程中，从实施身份数据识别到去身份数据识别的各个阶段。每个阶段都存在不同的重新识别风险的可能性。

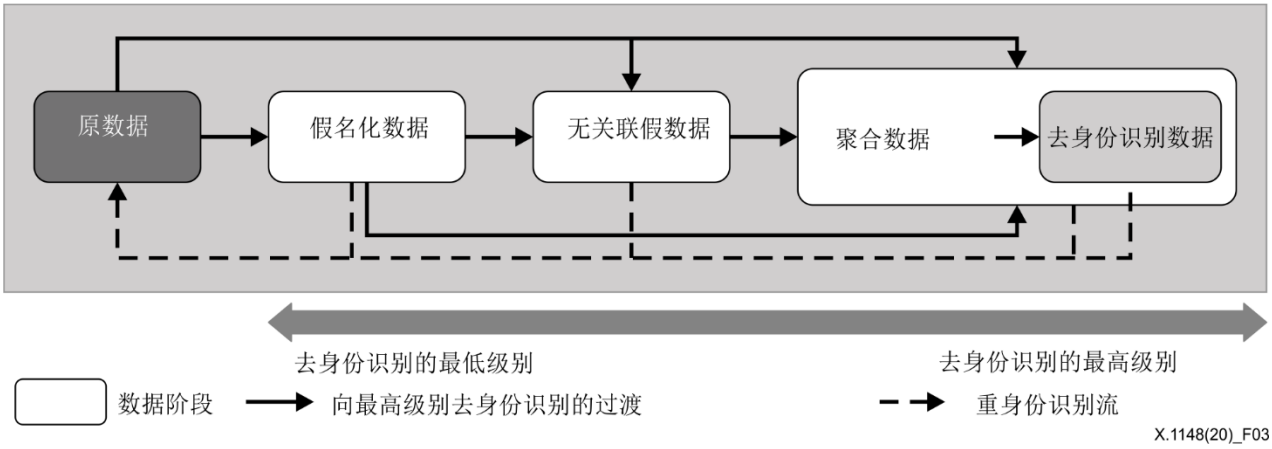


图 3 – 已去身份识别的数据的各个阶段

如图 3 所示，去身份识别阶段存在各种数据。右侧（最高级别的去身份识别）是与个人无关的去身份识别数据（例如，历史天气记录），因此不存在隐私风险。左端（最低级别的去身份识别）是与特定个人直接关联的识别数据。在这两个数据阶段之间，存在可以通过努力建立关联的数据。这些数据只能与人群建立关联，且虽然基于个人，但却不能与个人进行反向关联。总体而言，去身份识别过程旨在将数据推向右侧，同时保留一些期望的效用，并降低向更广泛人群或公众分发已去身份识别的数据的风险。

8.1.1 原始数据阶段

在识别数据的初始阶段，由于能在信息中观察到个人信息，因此这些数据可以明确地与特定个人建立关联。有关哪些信息可视作标识符的指南请参见 [b-ISO/IEC 29100]的第 4.4.1 节。

8.1.2 假名化数据阶段

在假名化数据阶段，鉴于所有标识符均由别名替代，因此除别名分配的执行方外，其他任何人的合理努力都无法逆转数据。然而，假名化的数据仍然可以通过与其他数据建立关联重新识别。

这与第 3.1.14 节定义的“假名化”数据相对应。

8.1.3 无关联的假名化数据阶段

在无关联的假名化数据阶段，所有标识符均被删除或由别名所替代，且其分配功能亦被删除或不可逆。这样，任何人（包括执行这些操作的一方）都无法通过合理的努力都重新建立关联。然而，无关联的假名化数据仍然可以通过与其他数据建立关联重新识别。

8.1.4 聚合数据阶段

在聚合数据阶段，数据会生成足够多的关于不同个人的信息，以至于无法将个体属性推断为不包含个体条目的统计数据，因此要对其加以组合。使用聚合技术时，如果某些变量组合的给定交叉单元大小能够导致特定个人被识别，则相应门限值以下的所有聚合数据均不会达到可识别的程度。

这与第 3.1.1 节定义有“聚合数据”数据相对应。

8.1.5 去身份识别数据阶段

在去身份识别数据阶段，通过数据关联的解除和属性以及其它改变（例如，属性值已随机化或泛化），使得去身份识别存在合理的置信度，即已无法通过单独的数据或与其他数据的组合，直接或间接地对个人加以识别。

8.2 数据发布模式

根据数据分析上下文的情况，将去识别的数据发布模式分为三类模型[b-UKAN]。

有三种发布模式可用于交付已去身份识别的数据：公开、半公开或非公开数据。

每种发布模式可提供不同级别的信息可用性和信息保护。根据数据发布的目的和/或法律要求，每种模型的适用性可能会有所不同。发布模式在去身份识别进程中发挥着重要作用，因为根据所选发布模式的不同，需要的去身份识别数量可能会存在差异。

第 8.2.1 至 8.2.3 节对三种发布模式进行了逐一考察。

8.2.1 公开数据发布模式

在传统的公开数据发布中，任何人均可在无需注册的情况下或无条件访问这些数据。此类发布的示例包括，由相关组织提供的公开可用数据以及在开放访问数据库（如门户网站）发布的数据。相关组织主动发布数据集，允许所有人自由使用并重新发布这些数据。

公开发布数据时，通常的做法是尽可能少地对信息设限，包括谁可以访问以及如何访问。因此，在无法识别下载数据集的个人时，应将这些披露信息视作公开数据发布处理。

尽管第 8.2.2 节所述访问信息请求，应作为公开数据发布来处理，但前提是不要求请求信息者同意有关信息处理、隐私或安全的条款或条件。

8.2.2 半公开数据发布模式

半公开数据共享模式比公开数据发布模式限制性更强，且其使用是在相关方提出获得数据访问权的正式请求和批准过程中发生。在这种情况下，数据接收方可能会同意某些使用条款或签署“点选”（click-through）合同。点选合同即在线使用条款，可能会对如何处理数据以及如何处置数据施加限制。无论如何，任何人都可以下载这些数据。

在响应访问数据集信息的请求时，去身份识别或许能够发挥作用。通过使用去身份标识，相关组织能够以保护隐私的方式对请求做出响应，同时保留信息的效用。在通过信息系统共享数据时，组织可以对某些限制施加访问控制，例如：

– 要求所有用户在访问数据之前注册并提供联系信息；
– 使用认证协议验证个人身份；
– 使用分层访问系统，根据个人的从属关系或证书等，授予各方不同级别的访问权限。

利用此类信息系统或可将交互式查询系统提供给研究人员，而原始数据可能会提供给少数经仔细筛选过程批准的分析师。

此外，当分析师要求数据控制方代表他们进行分析时，会出现不需要任何数据共享的数据访问情况。因此，这种情况可能不涉及组织共享数据。

8.2.3 非公开数据发布模式

只有在国家监管指导允许披露的情况下，包含 PII 的数据集才能在组织内部和组织之间共享数据。如果披露未获允许，但机构仍然希望共享数据集，那么必须删除所有 PII。非公开数据发布的数据可用性最低，但提供的保护程度较高，所需的去身份识别操作较少。

在组织间共享信息时，由于共享仅限于组织有权访问数据集，因此可以通过数据共享协议设置并实施有关信息隐私和安全的要求。对作为非公开数据处理的数据发布，双方必须达成数据共享协议。数据共享协议是这些发布所用风险缓解策略的重要组成部分，这其中包括一些常见术语，例如：

– 允许访问的规范（接收方控制）；
– 数据安全要求（基础设施控制）；
– 使用限制，特别是禁止与其他文件建立关联和有意进行重新识别（其他数据和治理控制）；
– 使用完成后销毁数据的要求（治理控制）。

数据共享协议有三重目的：

– 清楚地区分数据控制方信任和不信任的个人或组织；
– 作为规范访问条件的框架；
– 规范个人/组织违反这些访问条件时的制裁或处罚。

8.2.4 数据发布模式对比

在数据流环境下，限制重新识别的方法之一是控制获取和使用数据的方式。这些控件可以根据不同的数据发布模式加以分类，其优势和风险各有千秋。相关组织也可以选择应用分层访问的方法，这种方法结合了相关模型中的几种为各不同用例和隐私威胁提供的解决方案。此外，发布模式应考虑是否可以多次或定期发布数据。现有若干已命名的模型，有的没有限制而有的限制严格。表 2 提供了数据发布模式的对比情况。

表2 – 数据发布模式的对比

	公开发布模式	半公开发布模式	非公开发布模式
访问权限	• 每个人都可以自由访问发布的数据	• 受限制的个人或组织访问发布的数据（或子集）	• 个人或组织的下属可以访问发布的数据
用例	• 通过门户网站不受限制地访问数据。即任何人都可以免费获得	• 现场安全设置 • 交付访问 • 远程虚拟访问 • 通过分析服务器访问	• 组织内部和组织之间的共享
权限	• 不受限制的重复使用和重新分发数据权	• 可供获得授权的个人或组织使用	• 禁止重复使用、重新发布或分发数据
重新识别攻击	• 针对公开性的示威攻击	• 蓄意的内部攻击 • 熟人无意中发现了数据集中的朋友 • 数据泄漏	

8.3 数据发布模式与数据阶段的关系

8.3.1 非公开数据发布模式

当我们将数据从数据源共享到非公开发布模式时，需要对数据进行去身份识别操作。在正常情况下，尽管采用了非公开发布模式，但仍会使用未关联的假名化数据和更高级别的去身份识别数据。在这种情况下，可以使用诸如假名、密码、合成、抑制等去身份识别工具。

但是，如果双方之间存在特殊的合同或法律，那么此阶段可以使用假名化的数据来分析和存储数据。

8.3.2 半公开数据发布模式

当我们将数据从数据源共享到半公开发布模式时，需要比非公开发布模式更高级别的去身份识别。在此情况下，将通过统计处理禁止重新识别。此后，聚合数据和更高级别的去身份识别
数据可以用半公开发布模式发布。更具体地说，可以使用统计、随机化等去身份识别工具。

如表 2 所示，与公开发布模式相比，这种模式允许使用相对较低级别的去身份识别，因为只有有限的个人或组织可以访问相关数据。

8.3.3 公开数据发布模式

当我们将数据从数据源共享到公开发布模式时，需要比半公开发布模式更高级别的去身份识别。这种模式所执行的过程旨在获取去身份识别数据。在此过程结束之后，结果可以用于公开发布模型，如表 2 所示。

附件A

充分性评估程序

（此附件为本建议书不可分割的组成部分。）

A.1 本附件提供了一个充分性评估程序模型[b-KOREA]。参见图 A.1

以下为充分性评估程序中的步骤：

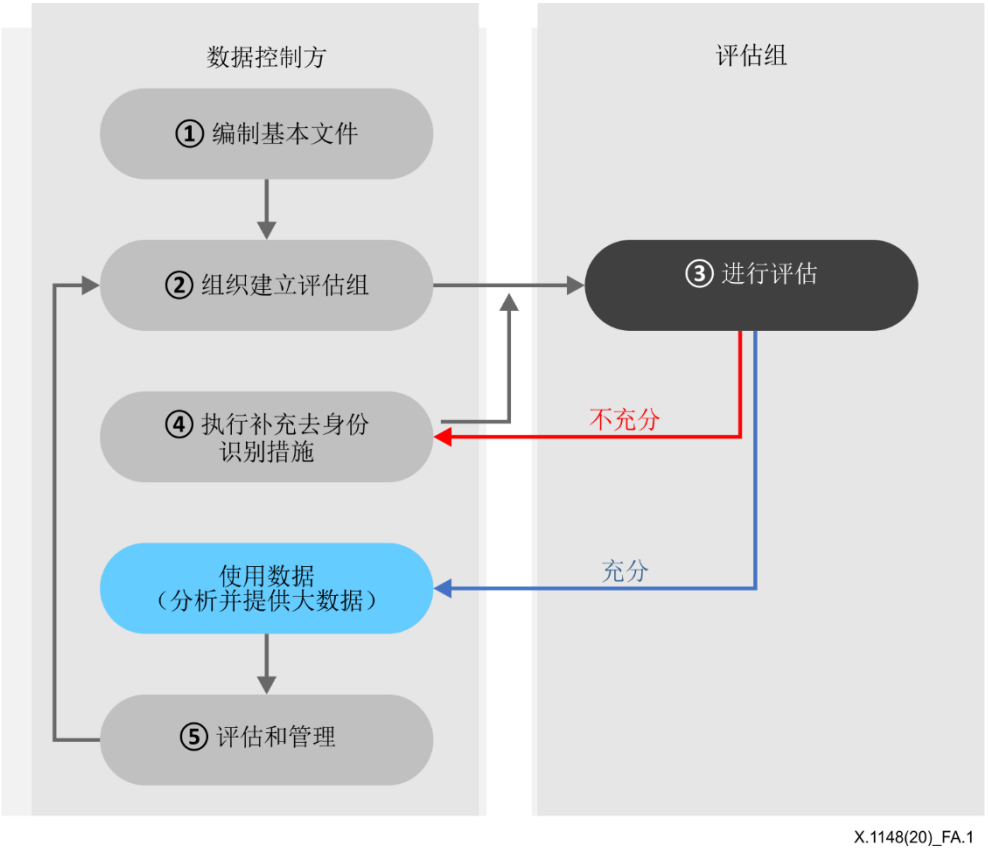
– 编制基本文件。数据控制方须准备充分性评估所需的基本文件，例如有关数据声明、去身份识别状态和用户组织管理水平的文件。“用户组织”是指拟在去身份识别后利用已去身份识别的数据的组织。

– 组织建立 一个评估组。隐私保护员既可成立评估组，亦可请DPO或TTP进行评估。

– 进行评估。评估组应利用PII管理员编写的基本文件，评估去身份识别级别的充分性。

– 执行补充去身份识别措施。如果评估结果不充分，数据控制方须以体现评估参与者意见的方式，实施额外的去身份识别操作。

– 使用数据。如果去身份识别的评估是充分的，则数据可以用于大数据分析之类的目的。



图A.1 – 去身份识别程序的充分性评估

A.1 编制基本文件

数据控制方应编制充分性评估所需的基本文件，例如有关评估主体的数据声明、识别状态和用户组织管理水平的文件。

A.2 组织建立评估组

隐私保护员可以成立评估组。在委托外部专业人员进行评估时，应从各领域专门机构的专家库中指定一名以上的法律和去身份识别专家。

评估组由与数据使用目的无直接利益关系的成员组成。

A.3 进行评估

评估组根据基本文件并使用 k-匿名算法模型对去身份识别的充分性开展评估。

– 初步审查。审查数据控制方编写完毕的基本文件并审核面谈记录，以检查数据集中是否有个人身份识别要素，以及使用目的和去身份识别技术是否适当。

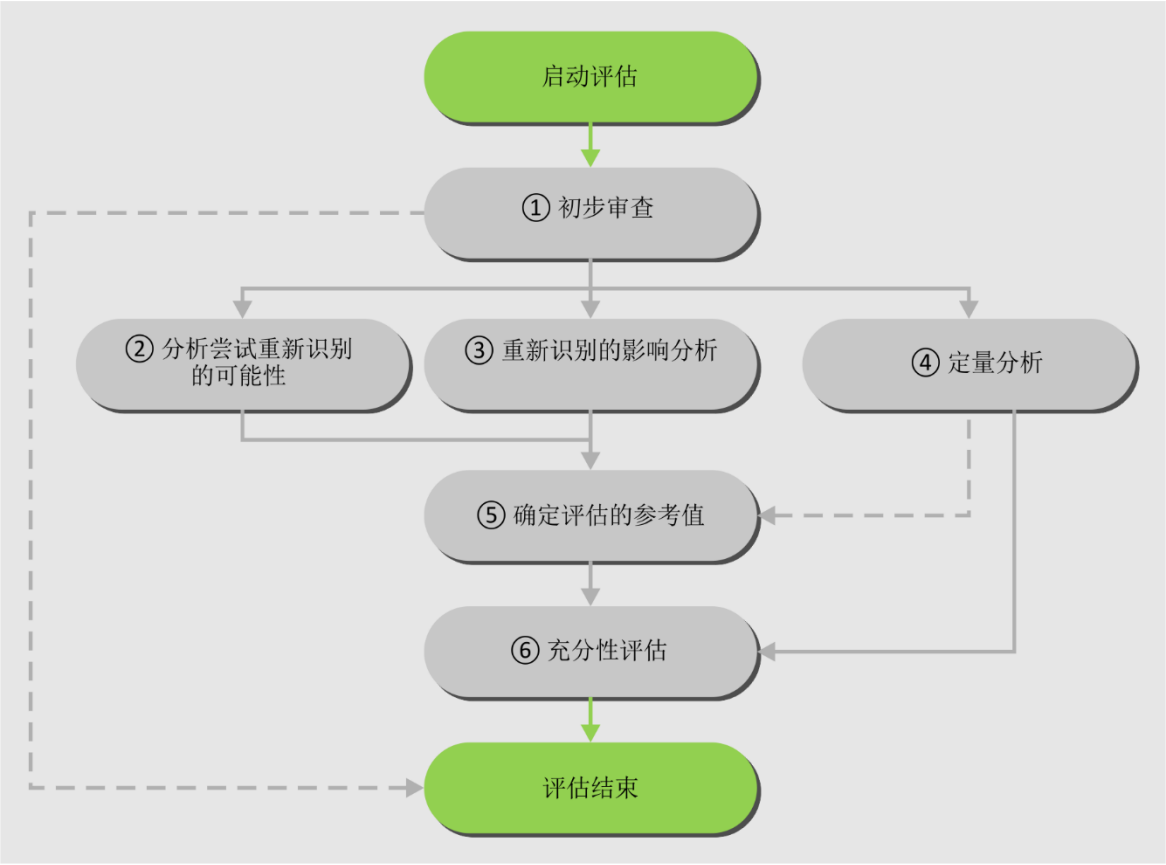
– 分析尝试重新识别的可能性。分析进行重新识别尝试的可能性，包括意图、PII保护级别以及负责使用或接收数据的数据控制方的能力。

– 重新识别的影响分析。当有意或无意地重新识别数据时，评估其可能给数据主题造成的影响。

– 定量分析。验证数据控制方所提供K值的准确性。

– 确定评估的参考值。评估组通过考虑重新识别的可能性、重新识别的影响、定量分析的结果和数据使用的目的，全面确定评估的参考值。

– 充分性评估。通过比较由平均参考值和定量分析得出的计算值，确定去身份识别的充分性。



图A.2 – 充分性评估程序

A.4 执行补充去身份识别措施

– 如果评估结果不充分，数据控制方须根据评估组的反馈执行补充去身份识别措施。

– 数据控制方执行完补充去身份识别措施后，评估组须开始重新评估。

A.5 使用数据

– 在大数据分析中使用已去身份识别的数据，或者如果评估（重新评估）认为去身份识别充分，则允许将其提供给第三方。

– 原则上，如果由于重新识别的风险很高，因此数据发布模式没有适当的数据风险缓解策略，则应禁止向公众或非签约数据用户提供或披露数据。

– 一旦使用数据的目的已经达成或已不再需要相关数据，则应销毁数据。

– 在以去身份识别数据的形式有效使用数据的过程中，应遵守后续管理步骤。

附件B

非结构化的去身份识别方法

（此附件是本建议书不可分割的组成部分。）

与结构化数据的去身份识别不同，非结构化数据的去身份识别机制应用于原始数据而非结构化的数据字段。对于下面的照片，去身份识别意味着移除人脸或使用其他信息加以替换，如图 B.1 所示。

图B.11 – 人脸去身份识别示例

非结构化数据有四种类型：

1） 非结构化文本数据：网络数据、报告文档、博客、新闻等。

2） 非结构化视频数据：视频数据均为非结构化，且一些标签信息提供了规范化的数据；

3） 非结构化音频数据：音频数据均为非结构化，且一些标签信息或识别的音频被转换为文本数据；

4） 非结构化日志数据：机器生成的日志数据是非结构化数据，但通常具有某种格式，可转换成结构化的形式。

为了表示文本、话音、图像和视频等非结构化数据的句法信息，去身份识别处理系统应包括以下三个单元：

1） 多媒体信息检测单元，用于从输入多媒体数据中的检测文本元信息；其中

- 包括负责将语音输入转换成文本的语音检测程序，用于跟踪语音输入中包含的对象或活动；
- 包括负责从图像输入中提取字符的光学字符识别检测程序；
- 包括视觉检测程序，该程序负责提取图像输入中包含的对象或活动，或者负责从图像输入或运动图片输入中移除图像；
- 包括从图像或运动图片输入中提取文本句子的视觉语句检测程序。

2） 基于知识的赋形单元将文本化的元信息和上下文信息分成表示外部配置的句法和表示内部信息的语义：

– 句法信息包括生成多媒体数据的源信息、由源生成的多媒体数据信息、以及从含义区提取的对象检测信息；

– 语义信息包括负责配置多媒体数据和上下文信息的含义区内的事件信息。

3） 去身份识别单元将可识别的PII从知识库和文本化的元信息中移除。

非结构化数据，去身份识别机制应定义如下相关要求和安全强度：

– 去身份识别的目标：针对应用或在线服务，应保护哪个目标对象？

– 如何实施去身份识别：确定应采用哪种机制去身份识别。去身份识别有哪些级别（例如黑箱、像素处理、模糊化）？

– 去身份识别与重新识别：应确定恢复身份识别或重新识别的必要性。当某项政策要求在犯罪调查中使用原始照片时，是否可以恢复已去身份识别的照片？

附录I

典型去身份识别技术示例

（此附录非本建议书不可分割的组成部分。
本附录提供了一些典型去识别技术的示例和说明）

I.1 去身份识别技术的统计工具

– 抽样：发布整个数据集的样本而非释放整个数据集的过程。如果发布的是子样本，则重新识别的概率会降低。

– 聚合：生成整个数据集表示值的一组统计函数。

I.2 去身份识别技术的加密工具

– 确定性加密[b-ISO/IEC 11770]：一种加密方案，对于给定的明文和密钥， 在单独执行加密算 法时，总是产生相同的密文。

– 保序加密[b-AGRAWAL]：一种加密方案，其中明文的数字排序得以保留。

– 同态加密[b-ISO/IEC 18033-6]：一种加密方案，允许对密文进行计算，从而生成加密结果，该结果与解密时对明文执行的操作结果相匹配。

– 保格式加密[b-NIST 800-38G]：一种加密方案，其中密文与明文格式相同。

– 同态秘密共享[b-ISO/IEC 18033-6]：一种秘密共享算法，其秘密使用同态加密方式加密。

I.3 抑制技术

– 掩码：用值替换或删除字段的过程。抑制技术的例子包括用星号或随机生成的假名替换电话号码。

– 局部抑制：从选定记录中抑制或移除特定属性值的过程。移除数据会增强隐私保护，但可能会降低数据集的效用。

– 记录抑制：从数据集中删除一个或多个完整记录的过程。

I.4 假名化技术

一种既移除与数据主题的关联，又在与数据主题相关的一组特定特征和一个或多个假名之间添加关联的过程。通常，这种技术以用假名（例如随机生成的值）替换直接标识符的方式实现假名化。直接标识符的例子包括姓名、电子邮件地址和政府发布的号码。所有直接标识符和潜在附加属性或所有其它识别属性均替换为假名。

I.5 泛化技术

– 舍入：用另一更短、更简单或更明确的近似值替换某数值的过程。

– 顶部和底部编码：一个编码过程，将高于上限（或低于下限）的属性设置为最大（或最小）门限值。

I.6 随机化技术

– 噪声添加：将无法预测的随机值添加到数据集选定属性的过程。

– 置换：在数据集的记录之间交换选定属性值而不进行修改的过程。

– 微聚合：用某种算法计算出的平均值取代所有连续属性值的过程。

I.7 合成数据

合成数据是一种通过人工生成微观数据来表示预定义统计数据模型的方法。根据定义，合成数据集不包含从现有数据主体收集的数据，但看起来确与预期目的相符。

附录II

去身份识别进程的方法

（此附录非本建议书不可分割的组成部分。）

本附录提供了去身份识别进程方法的一些示例和细节。

II.1 以数据为中心的去身份识别方法

鉴于去身份识别技术会修改原始数据以防止 PII 的泄露，效用与隐私之间显然会出现紧张关系。我们面临的挑战是如何以最小的准确性损失保护隐私：理想情况下，数据用户应该对已去身份识别的数据进行分析，而在对原始数据进行分析时，不会对这些分析的结果造成准确性损失。

实践中很难在不损害数据集效用的情况下，完美的实现去身份识别。对于大数据，这个问题会因数据的数量和变化而增加。一方面，低水平的去身份识别（例如，去身份识别仅是抑制直接标识符）通常无法确保不可识别性。另一方面，太强的去身份识别可能会阻止将取自不同来源的同一个人（或相似个人）的数据联系在一起，因此无法获得许多大数据潜在的好处。

本节描述了两种以数据为中心的去身份识别方法，目的是处理效用和隐私之间的紧张关系。我们可以利用数据专用和一般效用措施，解决如何衡量已去身份识别且已发布的数据集的效用。

II.1.1 效用优先的去身份识别方法

大数据中采集的个人信息通常是来自几个独立来源。因此，与属于相同（或相同类型/相似）个人的记录建立关联的能力，是创建大数据的核心。

在效用优先的去身份识别方法中，首先在微数据集上运行采用“启发式参数选择”和“适当效用保持特性”的去身份识别技术，之后再衡量披露的风险。因此，效用优先的去身份识别方法生效慢，且缺乏正式的隐私保证。例如，重新识别的风险可通过尝试记录原始数据集与去身份识别的数据集之间的联系，做出经验性估算。如果认为现存风险太高，则必须使用更加严格的隐私参数重新运行去身份识别技术，在可能会牺牲更多效用的同时反复改变参数，直至经验性披露风险低至官方统计中通常显示的数值。

当然，尽管从效用角度看能够建立关联是可取的，但这亦对隐私保护产生了威胁：在已去身份识别的数据集中，建立关联的准确性应该比在原始数据集中低得多。与去身份识别技术或去身份识别隐私模型兼容的关联性的数量，决定了分析师是否以及如何关联针对同一个人的独立去身份识别数据（在该技术/模型下）。

II.1.2 隐私优先的去身份识别方法

隐私模型的实施使用一个参数，该参数可同时保证重新识别披露风险以及属性披露风险的上限。模型实施是通过使用针对特定模型的去身份识别技术实现的，该技术中存在从模型参数导出的参数。众所周知的 k-匿名算法及其扩展以及 ε 差分隐私等隐私模型，经常导致较差的数据效用/关联性。

在隐私优先去身份识别方法中，如果得到的去身份识别数据效用太低，则应使用对效用损害较小的替代性去身份识别技术运用隐私模型，或应选择不太严格的隐私参数，抑或甚至使用不同的去身份识别隐私模型。

II.2 以角色为中心的去身份识别方法

本节分别描述了三种类型的方法，它们在去身份识别进程各司其职。以角色为中心的方法可以概括为对“谁”、“什么”和“在哪里以及如何”这些问题做出回答：

– 谁有权访问数据？

– 可以开展哪些分析或不可以进行哪些分析？

– 在哪里访问/分析数据，如何获得访问权限？

II.2.1 集中式去身份识别

统计披露控制过程侧重于集中进行去身份识别，由能够访问整个原始数据集的数据控制方执行。这种集中式方法的优缺点如表 II.1 所示。

表II.1 – 集中式去身份识别的特性

	细节
优点	• 个人无需为其提供的数据记录进行去身份识别。数据控制方拥有更多的计算资源，并且可能在去身份识别方面有更多的专业知识，因此有理由期望数据员能够充分实现整个数据集的去身份识别。 • 数据控制方拥有原始数据集的全局视图，因此最适合对数据效用和现行披露风险进行折衷。
缺点	• 提供原始数据的各方必须信任数据控制方（因为控制方有权访问所有原始数据）。虽然这在官方统计中不是问题（在官方统计中，数据控制方是一个国家统计机构），但在典型的大数据场景中，这可能是一个主要障碍。例如，当组装若干数据源的数据控制方仅是一家私人公司（例如，数据经纪人）时。 • 尤其是在大数据时代，对单个控制方而言，去身份识别带来的计算负担或许太重。 • 许多控制方都会参与单独的大数据处理场景，因而造成集中式的方法难于管理。

本地去身份识别方法和通过协作进行去身份识别操作，为上述优缺点提供了补充。

II.2.2 本地去身份识别

本地去身份识别是一种替代性的披露限制方法，适用于个人（数据主体）不信任（或仅部分信任）负责组装数据的数据控制方的情况（包括大数据）。每个主体在将自己的数据交给数据控制方之前，均要对这些数据进行去身份识别操作。

考虑到隐私保护问题，应在提供由给定源采集的数据之前，在源头进行去身份识别操作。然而，由于相关主体是在看不到其他主体数据的情况下对其数据进行去身份识别操作，因此每个来源施行的去身份识别操作，会比集中式去身份识别操作损失更多的信息。换言之，相关主体不掌握数据集的全局视图，这使得他们难以在已实现的披露风险限制与信息损失之间达成良好的折衷。

II.2.3 通过协作去身份识别

通过协作去身份识别的进程集两种优势于一身，即集中式去身份识别的低效用损失和本地去身份识别的高主体隐私性。集中式去身份识别的一个问题是，如果数据主体不信任数据控制方能够正确使用和/或去身份识别他/她的数据，则他/她有可能会决定提供错误数据（因此导致响应偏差）或者根本不提供数据（从而导致非响应偏差）。因此，相关主体可以通过协作确定与其数据相关的披露风险，然后以分布式协作的方式在本地应用适当的保护级别，而这需要利用两个主要属性：

– 在隐私级别相同的情况下，不会导致丢失比集中式方法数据集的更多信息。通过协作去身份识别的方式优于本地身份识别的方法，因为这种方法产生的信息损失较少。

– 数据主体和数据控制方对任何其他特定数据主体机密属性的了解，都不会超过已去身份识别的数据集存储的知识。与数据采集程序相比，这种方式还提供了隐私保护，因此优于集中式方法。

此外，通过协作方法可以使相关协议在没有外部强制机制的情况下顺利工作。在微观数据去身份识别操作中，某主体获得的隐私保护会影响其他主体获得的隐私保护。为提高协作方法的共同效用，需对电子协议进行安全的多方转换，使得两方或多方能够以一种特殊方式执行涉及两个数据集的转换，在这种方式中没有任何一方需要明确地将数据集交给其他方。鉴于安全的多方转换允许在无需集中所有数据存储的情况下转换查询，因此其降低了数据泄露的危害，并允许不完全互信的各方进行交叉计算。某些情况下，多方计算可以提供更好的隐私和实用性。

参考资料

[b-ISO/IEC 11770] ISO/IEC 11770（所有部分）信息技术 – 安全技术 – 密钥管理

[b-ISO/IEC 18033-6] ISO/IEC 18033-6，信息技术安全技术 – 加密算法 –第6部分：同态加密

[b-ISO/IEC 20889] ISO/IEC 20889 (2018)，隐私增强型数据去身份识别术语和技术分类

[b-ISO/IEC 27001] ISO/IEC，27001 (2018)，信息技术 – 安全技术 – 信息安全管理系统

[b-ISO/IEC 29100 ISO/IEC 29100 (2011)，信息技术 – 安全技术 – 隐私框架。

[b-NIST 800-38G] NIST特别出版物800-38G（2016年），分组密码操作模式推荐：保持格式的加密方法

[b-NISTIR 8053] NISTIR 8053（2015），个人信息的去识别

[b-AGRAWAL] Agrawal, R., Kiernan, J., Srikant, R.和 Xu, Y.（2004 年），数字数据的保序加密，2004 年 6 月在法国巴黎举行的 ACM SIGMOD 国际数据管理大会的 SIGMOD '04 会议记录，第 563-574 页

[b-KOREA] 韩国内政部，《去身份识别措施指南》，2016 年 6 月。
<http://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000821178&fileSn=2&nttId=7187&toolVer=&toolCntKey_>
（韩语，2019 年 7 月 26 日最后一次访问）
<https://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000827161&fileSn=0>
（英语，2020 年 12 月 12 日最后一次访问）

[b-UKAN] 英国匿名化网络，匿名化决策框架，2016 年
<https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf>

ITU-T系列建议书

系列A	ITU-T工作的组织
系列D	资费及结算原则和国际电信/ICT的经济和政策问题
系列E	综合网络运行、电话业务、业务运行和人为因素
系列F	非话电信业务
系列G	传输系统和媒介、数字系统和网络
系列H	视听及多媒体系统
系列I	综合业务数字网
系列J	有线网络和电视、声音节目及其他多媒体信号的传输
系列K	干扰的防护
系列L	环境与ICT、气候变化、电子废物、节能；线缆和外部设备的其他组件的建设、安装和保护
系列M	电信管理，包括TMN和网络维护
系列N	维护：国际声音节目和电视传输电路
系列O	测量设备的技术规范
系列P	电话传输质量、电话设施及本地线路网络
系列Q	交换和信令，以及相关的测量和测试
系列R	电报传输
系列S	电报业务终端设备
系列T	远程信息处理业务的终端设备
系列U	电报交换
系列V	电话网上的数据通信
系列X	**数据网、开放系统通信和安全性**
系列Y	全球信息基础设施、互联网协议问题、下一代网络、物联网和智慧城市
系列Z	用于电信系统的语言和一般软件问题

瑞士印刷
2021年，日内瓦