数据脱敏1 | “数据脱敏”是一个法律概念或技术概念吗？

合规科技系列文章 **Compliance-Tech Series**

高速发展的时代背景下，一方面行业分工在层层细化，一方面跨学科的交叉研究又越来越不可或缺。科技与法律表面上是两个相去甚远的专业领域，但就数据治理与隐私保护而言，只有跨界互通才可能找到最佳的解决方案。

“合规科技专题文章”旨在兼顾科技与法律的双重视角，深度解读数据技术的逻辑原理与数据合规的法律要求，从而促进技术人与法律人的双向理解，探讨数据利用与个人权益协调发展的可行方案。

“大数据”已然从热词变成日常，而数据在释放无限潜力的同时，也引发了隐私泄露的巨大隐患。从若干年前科技公司野蛮生长，到近年来数据立法接踵而至，信息社会正在两极之间寻求平衡。数据脱敏提供了这样一种可能性——通过降低数据与主体之间的关联，可以同时保留较高的隐私保护程度和较大的数据利用价值。

“数据脱敏”专题文章将梳理匿名化、去标识化、假名化等一系列相关概念，分析中国、欧盟、美国等法域对不同概念的法律评价，介绍数据脱敏的技术方案与隐私模型，探讨各个业务场景下的行业实践案例与法律落地方案，以推动数据利用和隐私保护的平衡发展。

“数据脱敏”专题往期文章链接

- 数据脱敏1｜“数据脱敏”是一个法律概念或技术概念吗？

本文是“数据脱敏”专题文章的第一篇，首先需要回答最基础的概念问题，也是实践中容易混淆的问题——数据脱敏是一个法律概念或技术概念吗？数据脱敏虽然是业界热词，但它并不是一个法律概念，也不是一个技术概念，甚至不是一个具体、特定的概念。实际上，数据脱敏一词的辐射范围非常广泛，它可以涵盖一系列多层次的法律概念和技术概念。

一. 数据脱敏不是一个法律概念或技术概念

数据脱敏通常是指对敏感数据进行技术处理，去除或降低其敏感度。数据脱敏是行业中的常用话术，也出现在一些效力层级较低的法律文件中。

关于加快构建全国一体化大数据中心协同创新体系的指导意见

[国家发展和改革委员会,中共中央网络安全和信息化委员会办公室,工业和信息化部,国家能源局] [发改高技〔2020〕1922号] [2020.12.23 发布] [2020.12.23 实施]

[命中频次]: 数据脱敏 2

摘要: 通、交易等环节的制度法规和机制化运营流程。建立完善数据资源质量评估与价格形成机制。完善覆盖原始数据、脱敏处理数据、模型化数据和人工智能化数据等不同数据开发层级的新型大数据综合交易机制。探索有利于超大规...

关于加强全民健康信息标准化体系建设的意见

[国家卫生健康委员会,国家中医药管理局] [国卫办规划发〔2020〕14号] [2020.09.27 发布] [2020.09.27 实施]

[命中频次]: 数据脱敏 1

摘要: 据联通共享的安全需求，从个人信息安全、重要数据安全、跨境数据安全三个方面，研究编制数据分类分级、数据脱敏、去标识化、数据跨境、风险评估等标准。 3.推进行业应用安全标准研制。为指导行业应用安全规...

工业和信息化部办公厅关于开展2020年网络安全技术应用试点示范工作的通知

[工业和信息化部] [工信厅网安函〔2020〕190号] [2020.07.30 发布] [2020.07.30 实施]

[命中频次]: 数据脱敏 1

摘要: 方案，以及结合海量网络数据汇聚存储、流动共享等安全需求，在数据资产识别、分类分级防护、数据加密、数据脱敏、泄露追溯等方面的解决方案。 6.物联网安全。结合智慧家庭、智能抄表、零售服务、智能...

工业和信息化部关于工业大数据发展的指导意见

[工业和信息化部] [工信部信发〔2020〕67号] [2020.04.28 发布] [2020.04.28 实施]

【官方解读】

[命中频次]: 数据脱敏 1

摘要: 环管理，全面保障数据安全。 （十五）加强工业数据安全产品研发。开展加密传输、访问控制、数据脱敏等安全技术攻关，提升防篡改、防窃取、防泄漏能力。加快培育安全骨干企业，增强数据安全服务，培育良好安全产业生态...

交通运输部关于印发《推进综合交通运输大数据发展行动纲要（2020—2025年）》的通知

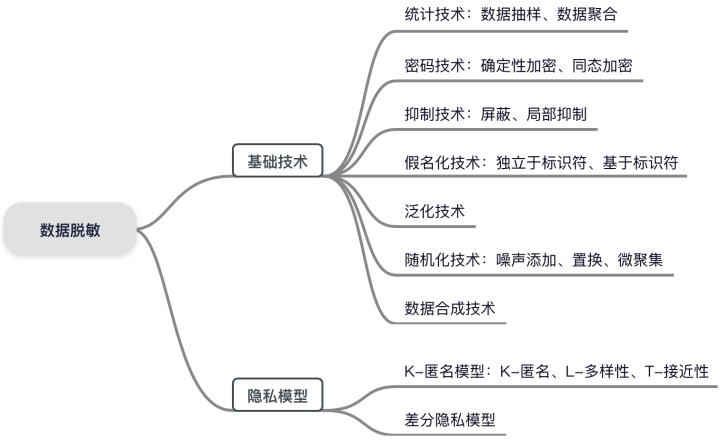
[交通运输部] [交科技发〔2019〕161号] [2019.12.09 发布] [2019.12.09 实施]

[命中频次]: 数据脱敏 1

摘要: 完善数据安全保障措施。推进交通运输领域数据分类分级管理，加强重要数据和个人信息安全保护，制定数据分级安全管理、数据脱敏等制度规范。推进重要信息系统密码技术应用和重要软硬件设备自主可控。（部办公厅、科技司、各省级交通...

但严格来说，数据脱敏并不是一个法律概念。在个人信息保护的法律体系中，与数据脱敏相关的法律概念主要包括匿名化（**anonymization**）、去标识化（**de-identification**）、假名化（**pseudonymization**）等。数据脱敏的起点是个人信息（**personal information**），即以电子或者其他方式记录的、与已识别或者可识别的自然人有关的各种信息。中国、欧盟、美国等法域对个人信息的定义基本一致，均强调了可识别性（**identifiable**），即该信息能否单独或与其他信息相结合而识别特定的自然人。脱敏处理后的数据能否具有特定的法律地位，主要取决于其具体实现的程度和效果，于个人信息而言，即脱敏后在多大程度上仍可识别特定个人。

数据脱敏也不是一个严格的技术概念。从技术上而言，为了降低或去除数据与个人之间的联系，可以采用统计、密码、抑制、假名化、泛化、随机化、数据合成等基础技术，并通过K-匿名、差分隐私等模型进行隐私度量。近年来，联邦学习、多方安全计算等技术也开始应用于实践。



二. 数据脱敏的两个维度

在个人信息保护的语境下，数据脱敏一般是指对个人信息进行技术处理，去除或降低数据与个人之间的关联，导致个人在一定程度上不可识别。数据脱敏并不是一个严格的法律概念或技术概念，但是，对数据脱敏的理解离不开技术（过程）和法律（效果）这两个维度。

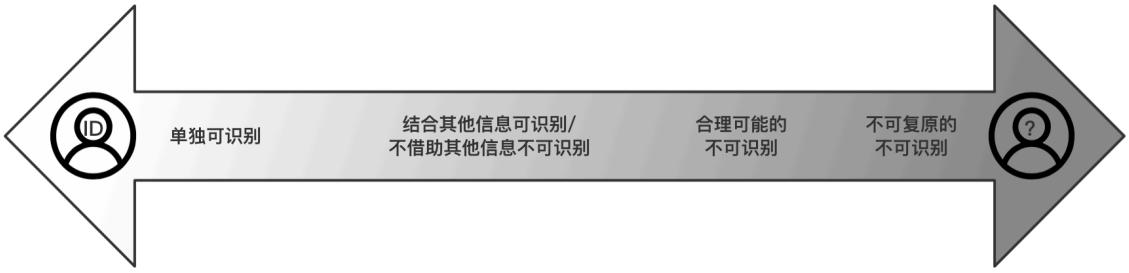
其一，数据脱敏是一种技术处理的过程，包括选择哪种脱敏技术或其组合，以及每种技术的实际实施强度。数据脱敏只是一个大类概念，实践中需要基于具体的业务场景和需求，综合考虑数据主体的授权与要求、数据的性质与类型、数据处理的方式与目的、重识别的风险与后果、当前可用的技术水平、合理的成本投入等因素，选择合适的技术实现方案。

其二，数据脱敏是一种对处理效果的法律评价，即经过技术处理后的数据，具体实现了哪种程度的不可识别。数据脱敏是一个渐进的光谱，根据程度的递增，例如从假名化、去标识化到匿名化，法律将作出差异化评价。

实践中常见的误区是：将数据脱敏静态地视为特定的技术处理，并概括认为个人信息脱敏后即与个人脱离关联。技术的视角有利于具象化地理解数据脱敏，但不可识别的程度才是法律上定义和区分数据脱敏相关概念的本质特征。

三. 从本质特征理解数据脱敏

可识别性是个人信息的本质特征，不可识别性是数据脱敏的本质特征。在个人信息与非个人信息之间，各国法上存在假名化、去标识化、匿名化等概念。从本质上说，相关概念的不可识别程度是渐进的，包括单独可识别、结合其他信息可识别/不借助其他信息不可识别、合理可能的不可识别、不可复原的不可识别。



数据脱敏的一端是单独可识别的个人信息，典型例子是直接标识符，即在特定环境下可以单独识别个人的属性，例如姓名、身份证号、电话号码、地址、邮箱、银行卡号、学生证号、车牌号、设备标识符、生物识别码、IP地址等。

数据脱敏的另一端是不可识别个人的非个人信息，如经匿名化处理的信息。应特别注意的是，各国对匿名化的要求并不相同，例如：欧盟《通用数据保护条例》（GDPR）的匿名化是基于“合理可能”（reasonably likely）标准而言的——综合考虑技术、成本、时间等因素，数据控制者或其他人采用了所有合理可能的方法，仍无法直接或间接识别数据主体；而我国现行法下不仅要求个人信息主体无法被识别，还要求匿名化处理后的信息不能被复原。

数据脱敏的中间状态，是结合其他信息可识别、而不借助其他信息不可识别。这在我国被称为“去标识化”，它的不可识别是可复原的，而它的可识别依赖于额外信息。例如，1997年，美国马萨诸塞州公布的健康数据删除了姓名、地址、社会保险号等直接标识符，因此无法直接识别个人，但哈佛大学教授Sweeney将该健康数据与当地的选民名册进行对比，轻松找出了Weld州长——因为与他具有相同的出生日期、性别、邮政编码的人只有他自己。

总体而言，数据脱敏的相关概念主要以不可识别的程度为标准，随着程度的量变而实现概念的质变，但各个概念之间并非泾渭分明，而有赖于具体场景下的综合判断。

本期小结与下期预告：数据脱敏并非一个特定的法律概念或技术概念，而是一个包罗万象的框架性概念。可以从两个维度理解数据脱敏，它既包括技术处理的过程，也包括对处理效果的法律评价。脱敏处理之后所实现的不可识别程度，是区分匿名化、去标识化、假名化等法律概念的本质特征。然而，不同法域对同一法律概念的定义和尺度并不相同，极易造成混淆。下期文章将为您解读中国、欧盟、美国等法域下对匿名化、去标识化、假名化所掌握的不同尺度。