公安数据开放场景下个人信息匿名化处理法律标准探究*

张　寒，张　宁
（中国人民公安大学　公安管理学院，北京　100032）

摘　要：在公安数据开放的背景下，个人信息匿名化处理逐渐成为平衡公安数据开放与个人信息保护的黄金分割点。通过国际横向比较分析，揭示了我国现行“无法识别特定个人且不能复原”的匿名化处理法律标准在操作层面的局限性，并选取我国17省市公安数据开放平台发布的《行政处罚决定书》为样本，对数据开放的类、量、质进行系统性评估。研究发现，公安数据匿名化开放尚处于初步阶段，存在顶层设计缺乏刚性约束、数据过度保密、格式不规范和处理标准不统一等现象。基于此，我国可以确立操作方法、风险检验及效果评估三维协同的匿名化处理法律标准：区分处理直接标识符与准标识符的操作方法标准，引入蓄意侵入者角色的再识别风险检验标准，以及去识别化效果评估标准。通过三重维度的协同作用，助推公安匿名数据最终实现“无法识别特定个人且不能复原”的法律效果。

关键词：公安数据开放；个人信息；法律标准；匿名化处理；标识符；再识别风险检验
中图分类号：D912　　**文献标识码：**A　　**DOI：**10.19358/j.issn.2097-1788.2024.11.016
引用格式：张寒，张宁. 公安数据开放场景下个人信息匿名化处理法律标准探究［J］. 网络安全与数据治理，2024，43(11)：101-109.

Exploration of legal standards for anonymization of personal information in public security data openness scenarios

Zhang Han, Zhang Ning
(Public Security Management College, People's Public Security University of China, Beijing 100032, China)

Abstract: Under the backdrop of public security data opening, the anonymization of personal information has gradually become the golden section point for balancing the opening of public security data and the protection of personal information. This study, through international comparative analysis, reveals the operational limitations of China's current legal standard for anonymization processing, which is "unable to identify specific individuals and cannot be reverted". It selects the Administrative Penalty Decision published by the public security data opening platforms of 17 provinces and cities in China as samples to systematically evaluate the type, quantity, and quality of data opening. The study finds that due to the lack of rigid constraints in top-level design, excessive data secrecy, non-standard formats, and non-uniform processing standards, the opening of public security data anonymization is still in initial stage. China can establish a three-dimensional coordinated legal standard for anonymization processing: operational method standards for distinguishing direct identifiers from quasi-identifiers, re-identification risk inspection standards for introducing the role of motivated intruders, and de-identification effect evaluation standards. Through the synergistic effect of the three dimensions, it promotes the final realization of the legal effect of " unable to identify specific individuals and cannot be reverted" for public security anonymous data.

Key words: public security data disclosure; personal information; legal standards; anonymization; identifiers; re-identification risk assessment

*基金项目：国家社科基金项目“社会治理现代化视域下的警务数据开放研究”（23BZZ071）

0 引言

伴随大数据、人工智能、区块链等前沿信息技术的迭代更新，人类社会正逐步向“数字时代”转型，数据作为这一转型过程的核心资产，在社会治安治理、市场经济运行以及科研教育等领域中的价值日益凸显。为充分利用政府持有的海量数据资源，推进社会对公开数据的深入挖掘和创新应用，中共中央于2022年12月颁布的《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》中指出，数据治理在数据要素功能发挥中的核心地位。政府数据作为一种公共资源，根据数据生命周期理论，数据开放是数据生命周期中的重要一环[1]。截至2023年8月，我国已有226个省市地方政府上线了数据开放平台[2]。公安机关作为我国政府领导下的行政职能部门，同样掌握着海量极具商业价值和社会价值的个人信息，鉴于公安数据存在涉密、隐私保护等缘故无法直接向社会公众开放，致使资源得不到有效利用而只能躺在数据库内“休眠”。

为释放数据多元潜力，北京、上海、重庆、山东等省市在地方数据条例中已规定相应的数据开放规制思路，即个人信息经过匿名化处理达至法律标准后，方可有条件或无条件开放。但值得商榷的是，公安数据开放过程中对个人信息的匿名化处理并非一劳永逸，其剩余再识别风险总是会如影随形地伴随着匿名化数据。当社会公众结合其他相关数据从匿名化数据中重新识别出特定数据主体时，这些数据便重新具有个人属性恢复为个人数据，并且重新受到《个人信息保护法》的规范和制约。由此采用何种法律标准界定个人信息在经过处理后是否达到匿名化状态，成为公安数据开放的重中之重。故此，本文在公安数据开放的宏观背景下，依据《个人信息保护法》第七十三条第四款、《网络安全法》第七十六条第五款所确立的法律基准，融会贯通本土情境和域外经验，对我国公安数据匿名化制度构建进行创造性探讨，探寻公安数据匿名化的法律标准、实践方法。

1 匿名化：平衡公安数据开放与个人信息保护之利益

1.1 利益冲突与现实困境

公安数据中所包含的个人信息，不仅与个人的人格尊严、人身自由息息相关，其后续的流通与利用亦关涉到他人和整个社会的利益，因而具有社会性和公共性。这种兼具公益与私益双重属性的信息，在公安数据开放与个人信息保护之间存在较难调和的利益冲突，即公众的数据利用利益（公益）与个人的信息保护利益（私益）之间的冲突[3]。一方面，传统的个人信息“个人控制论”过分强调个人信息的私益属性，对信息发挥其所蕴含的公益属性将产生阻碍。鉴于个人信息所固有的社会公共属性，学界已逐渐对个人信息的个人绝对控制论产生反思，并提出个人信息“社会控制论”的主张[4]。另一方面，公安数据开放不仅可以促进公安机关向数字化转型，提升公安工作的透明度与人民群众的认可度，还能释放数据的多重价值，满足社会对公共数据流通与利用的需求。

在个人信息保护领域，传统的“告知同意规则”在面对数据发展的多元需求时日益显得捉襟见肘[5]，已逐渐成为制约公安数据开放的最大桎梏。公安机关若按照相关法定程序开放个人信息，不仅在信息的收集、处理、利用等阶段需要征得个人的事前同意，而且事后也必须接受个人的查询、更正或者删除等要求[3]。在大数据时代，公安机关若想试图在不同应用场景中逐一获取个人信息主体的同意，将面临较高成本和较低的可行性。

1.2 个人信息匿名化

作为化解上述利益冲突的技术成果，个人信息匿名化以一种“隐私友好（privacy-friendly）”的方式满足社会对信息的需求[6]，尽管匿名化在一定程度上折损了信息的后续利用价值[7]。然而，技术的发展与应用离不开法律标准的保驾护航，匿名化技术亦是如此[8]。我国有关匿名化处理的法律制度尚处于初级阶段。相较之下，国外研究百呈现出多元化、精细化发展态势[9]。当下，匿名化数据的法律标准已成为国际范围内政府数据开放领域的重要议题，我国有必要立足本土情境，汲取域外在匿名数据处理、法律规则制定、再识别风险检验、去识别化效果评估等领域的有益经验，探索兼具实用性与有效性的匿名化处理法律标准。

2 法阈镜鉴：国际视域下匿名化处理法律标准的经验及启示

尽管不同法阈对匿名化信息的法律认定标准不一，但各国立法机构普遍对匿名化信息的法律标准设定了极为严苛的条件。从国际比较视域下，我国信息匿名化处理采用的是《个人信息保护法》中的“无法识别特定个人且不能复原”标准，呈现出“静态绝对”的特性；欧盟则适用《通用数据保护条例》（General Data Protection Regulation，GDPR）中的“所有合理可能性”标准，与我国“静态绝对”的法律标准相比更具动态性[10]；而美国采取的是更为开放的“专家判定法”与“安全港方法”标准，以促进数据的流通与再利用；英国信息专员办公室（Information Commissioner’s Office，ICO）于2012年发布《匿名化：数据保护风险管理实践准则》，该指引在进行匿名化剩余再识别风险评估时，更倾向选择“蓄意侵

入者检验”（Motivated Intruder Test）标准。

针对数据治理领域，个人数据匿名化处理面临着匿名数据动态管理风险被忽视、具体实践操作性不强这一现状[10]，我国可在立足于本国国情的基础上从域外汲取有益元素，平衡数据自由使用与个人信息安全之间的关系，健全我国匿名化数据治理法律标准。

2.1 中国：“无法识别特定个人且不能复原”标准

2.1.1 个人信息匿名化处理之法律标准涵义

关于个人信息匿名化处理规范，我国《民法典》第一百一十一条、《网络安全法》第四十四条、《个人信息保护法》第十条以及《数据安全法》第三十条指出，数据控制者在处理个人信息时必须承担相应的法律义务，维护个人隐私和数据安全。进一步地，《民法典》第一千零三十八条第一款、《网络安全法》第四十二条第一款、《关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》第三条第二款对个人信息的处理活动作出补充性说明[11]。并且《个人信息保护法》第四条明确指出“个人信息不包括匿名化处理后的信息”，匿名化信息就相应地被排除在外。若个人信息经过匿名化处理达到法律标准后便脱离上述规范性文件的限制，从而允许其进入市场自由流通，不受个人信息保护机制的约束。

我国《个人信息保护法》第七十三条第四款明确规定，“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”。纵观上述文件可见，我国现行的匿名化处理法律标准为“无法识别特定个人且不能复原”。其一，识别性（Identifiability）和关联性（Linkability）构成“个人信息”核心特征[12]，姓名、公民身份证号、详细地址等能直接指向特定个人的直接标识符（Direct Identifier）与性别、年龄、婚姻状况等能结合其他属性识别个人的准标识符（Quasi-identifier）均被列入个人信息再识别的范畴。其二，“无法识别”这一概念应涵盖直接识别与间接识别两种情形，即个人信息经过匿名化处理后，不能直接或结合其他信息识别出特定主体。其三，“不能复原”则要求经匿名化后的信息在任何场景下都不能被重新复原成个人信息。

2.1.2 法律适用困境及启示

这一“静态绝对”的匿名化处理法律标准，使得实践中数据控制者在数据处理时陷入进退两难的困境：首先，若匿名化处理后的信息仍存在被重新识别或复原的潜在可能性，那么其仍属于法律中个人信息的范畴；反之，数据控制者若真正使得匿名化处理达到“无法识别特定个人且不能复原”的法律效果，很可能会导致匿名化信息的社会科研价值、商用价值丧失，这与数据再利用的目的背道而驰[13]。美国学者 Paul Ohm 指出，在数据匿名化实践中，匿名化本身就是一个“破碎的隐私承诺”[14]。若要达到绝对的不可逆转或不可追溯，往往会不可避免地遭遇难以逾越的技术与法律鸿沟[15]。

鉴于匿名化技术在不同应用场景下的效用差异，以及法律规范与技术实践之间的鸿沟，意图通过抽象凝练的法律条文来规范复杂多变的匿名化处理行为存在一定的实际操作难度。特别是在数据治理与隐私权保护领域，个人信息匿名化追求的是风险最小化，而非百分百消除风险[16]。面对匿名化法定概念与实际技术效果脱节的问题，我国宜正视匿名化处理后残存的再识别风险，并明令禁止对匿名化信息实施非法再识别行为。

2.2 欧盟：“所有合理可能性”标准

欧盟个人信息匿名化处理的规范性文件主要体现在《通用数据保护条例》（GDPR）、“108 公约 +（Convention 108 +）”以及 WP29《匿名化技术》的意见书中。GDPR 于前言第 26 段界定了匿名化信息的识别认定标准，即在评估信息主体再识别风险时，应当考虑“所有合理可能性”[17]的客观情形，其中涵盖数据控制者或其他任何人所能直接或间接识别到特定信息主体的方式。

2.2.1 识别主体标准：数据控制者或其他任何人

对于匿名化数据再识别的主体标准具有“客观说”与“主观说”之分[18]。欧盟则采纳“客观说”作为匿名化数据处理的基本原则，并以“任一主体说”为核心，明确将数据控制者或其他任何人作为匿名化数据再识别的主体标准，即匿名化处理亟需达到任一主体都无法进行再识别的高度。这一标准与欧盟对个人数据所赋予的“具有广泛延伸性”[17]理念相吻合，旨在更全面地保障数据主体的合法权益。在评估匿名化数据的再识别风险时，可从识别主体的视角出发，鉴于不同主体所支配的外部信息量各不相同，从外界获取额外辅助信息以识别匿名数据中特定主体身份的能力也存在差异，由此要求在匿名化评估时应将所有可能获得的外部信息均纳入考虑范围。

2.2.2 识别方式标准：所有合理可能之方法

欧盟于 GDPR 前言第 26 条中明确规定在对匿名信息进行特定身份关联时采用“所有合理可能性”标准，识别方式是否具有合理可能性，需要结合资金、时间、技术、人力等要素在不同场景下进行动态衡量[19]。倘若识别所耗成本远超识别所获收益，即被认为采取不合理的方式进行身份识别，不属于欧盟“所有合理可能性”的范围，仍被认为是匿名化信息，其数据的流通与利用不受相关个人信息保护机制的限制。鉴于“个人信息”与“匿名信息”会因“所有合理可能性”而动态转换，GDPR 更加注重匿名化数据的再识别风险，并将“风险管

理”理念引入个人信息匿名化过程中[20]。从法律目的与识别技术的视角出发，匿名并非是绝对的，而是一个可识别程度问题。

2.3 英国：“蓄意侵入者检验”标准

英国有关个人信息的处理工作主要由 ICO 负责，ICO 作为国家层面独立的数据监管机构，在评估“匿名信息再识别风险”及“个体再识别动机”时则适用“蓄意侵入者检验”标准，即假设一位具备识别动机（积极侵权）和一定识别能力的侵入者，评估其能否将公开的匿名化信息识别至特定个人，进而判断其成功识别的概率[21]。

2.3.1 侵入者的识别动机

动机是匿名数据具备再识别风险的首要因素，若无动机驱动，再识别行为便无从谈起。ICO 所提出的侵入者在主观上要具有充分的再识别动机，即意图通过公开披露的匿名化信息重新识别到特定主体——“蓄意（Motivated）”[8]。针对再识别动机，El Emam 等学者设计了更为精细化的侵入者动机模型，分为“动机温和的侵入者”和“动机强烈的侵入者”，后者会持续识别并尽可能多地验证潜在匹配信息，但这种识别仍需考虑经济、时间等实际条件的限制[22]。

2.3.2 侵入者的识别能力

蓄意侵入者的识别能力是评估匿名化数据再识别风险的核心指标。ICO 假设侵入者具备一定的识别能力，能利用可公开获取的资源（如图书馆、网络），但不具专业知识或设备（如黑客技术或设备），也不会为再识别匿名化数据而诉诸犯罪行为[6,22]。该标准将侵入者识别能力的区间设定为，高于普通社会公众又低于专家，因而具有合理性。并且 ICO 强调，蓄意侵入者在尝试对匿名数据进行再识别之前，并不掌握与匿名主体有关的任何先验知识。这种既不过于宽松以致无法有效保护个人数据，也不过于严苛而阻碍数据合理利用，因而是有效的。同时，鉴于专业知识者（如警察、医生、律师等）在履行其职责时需遵守严格的保密义务，任何企图通过非法手段进行再识别的行为都将受到法律的严厉制裁。这种规定旨在强化数据保护的法律框架，使个人隐私得到尊重和保护。

2.4 美国：个人信息去识别化法律标准

美国国家标准与技术研究院（National Institute of Standards and Technology，NIST）于 2015 年发布《个人信息去识别化》指南，在为结构化数据的去识别方法进行梳理的同时，还扩展了对非结构化数据处理的指导，为不同应用场景下进行去标识化提供指引。在美国《健康保险流通与责任法案》（The Health Insurance Portability and Accountability Act of 1996，HIPAA）的隐私规则中描述了“专家判定法”标准、“安全港方法”标准以及“区分处理标识符”策略，以帮助组织和个人降低数据泄露风险并增强数据隐私保护。

2.4.1 “专家判定法”标准

“专家判定法”标准规定由专家审视数据并确定合适的去识别化方法，以有效降低匿名化数据再识别的风险。根据 HIPAA 第 164.514 条第 b 款的规定，“专家”是指在统计科学领域具备一定去识别化知识与经验的人士。专家运用这些原则和方法，判断经去识别化的信息单独或与其他合理可用的信息结合使用时，被预期接收者识别信息主体的风险能否非常小，进而判定相关信息能否构成匿名化信息。可见，“专家判定法”标准是一种基于对匿名化信息可识别性进行评估的标准，它旨在衡量个体被重新识别的概率，并据此评估信息主体隐私泄露的风险。

2.4.2 “安全港方法”标准

“安全港方法”标准中所详细列举的识别符在世界隐私法规中堪称最全[14]。HIPAA 第 164.514 条第 b 款第 2 项详细地列举了一系列用于识别个体的特定信息类别，包括姓名、地理区域、日期、电话号码、电子邮件地址、网络通用资源定位符（URL）、IP 地址等 18 项识别符。“安全港方法”标准规定数据控制者在删除“个人或其亲属、雇主、家庭成员”的上述 18 项识别符后，且并无合理依据相信该数据可用于识别个人身份，其即可被视为去识别化数据。此外，该条于第 e 款还规定，数据控制者若与预期接收者签订数据使用协议，则有权使用或披露被删除 16 种特定识别符的有限数据集。

但“安全港方法”标准可能并不安全，HIPAA 所列举的识别符未能全面囊括所有潜在场景，也无法列举周延。若不断根据再识别技术的进步而无限扩展需删除的识别符清单，则将会如“打地鼠游戏”一般陷入无尽循环[21]。

2.4.3 个人信息去识别化策略

NIST 在《个人信息去识别化》规定，直接标识符（Direct Identifiers）是指直接识别单个个人的数据，包括姓名、社会保险号码和电子邮件地址等。在去识别化的实践中，数据控制者必须删除或者以其他方式转换直接标识符。具体方法包括：第一，直接删除，在物理或逻辑上消除数据的存在；第二，替换为通用类别名称，如“某市”“某民族”等，以降低数据的可识别性；第三，替换为符号，对敏感信息用占位符加以遮盖以保护隐私，如“×××”“***”等[23]。

与直接标识符不同，准标识符（Quasi-Identifiers），是指单独不能识别特定个人，但可以被聚合并与其他信

息“链接”以识别数据主体的标识符[24]，包括生日、邮政编码、性别等。它通常蕴含对后续分析可能至关重要的信息，因而必须审视评估其所带来的再识别风险与潜在价值的关系。处理方法通常有以下几种：第一，抑制（Suppression），即删除准标识符；第二，泛化（Generalization），将准标识符的精确值扩展为宽泛的范围或集合中的一员，例如将年龄45岁调整为到40到50岁区间，泛化可应用于整个数据集或特定记录，在增加隐私性的同时又保留了数据分析价值；第三，干扰（Perturbation），对每个个体在给定的泛化水平内，以一致方式将个体数据进行随机或有规则的调整，例如年龄的微调或犯罪日期的系统性变化，在增强数据的隐私性的同时而不显著影响其统计特性[25]。在不牺牲数据集分析价值的情况下实现对个人信息的隐私保护。

对比不同法域下数据匿名化处理法律标准，中国、欧盟、美国总体呈现出由“静态绝对”“动态绝对”到“动态开放”的递进趋势[10]。我国在匿名化处理上的“双重要求”，不仅使得匿名数据在当前技术环节下无法与特定个人产生直接关联，并且在不可预见的未来，亦无法通过任何技术手段恢复其原始状态。在数据开放领域，我国公安机关可在个人信息处理所处的具体场景中进行动态的风险控制[26]，并创新数据治理策略，以有效应对数据交易市场的“负外部性”问题以及再识别风险的“动态性”挑战。

3 匿名化处理法律标准的三维重构

鉴于我国公安机关尚未形成独属的公安数据匿名化处理法律标准，目前仍需遵循《个人信息保护法》《网络安全法》等相关法律法规的指导。在当前公安行政处罚文书公开过程中，需关注重点问题——匿名信息公开与个人隐私权保护的平衡之争。根据《中华人民共和国行政处罚法》第59条之规定，对于行政处罚决定书中涉所及的信息，其中大部分既属于隐私，亦属于个人信息，因而不得不面对隐私与个人信息的关系争议[27]。需要认识到，这些数据的隐私属性是不可剥夺的，并不因要进入行政程序而性质有所改变[28]。

在公开行政处罚决定文书时，公安机关仍应对涉事公民的个人信息进行匿名化处理。为进一步提升匿名化规范的实用性，本文借鉴美国HIPAA中的“区分处理直接标识符与准标识符”策略与英国“蓄意侵入者检验”标准，并结合我国个人信息去标识化效果评估指南，构建起一套集合操作方法、再识别风险检验与去识别化效果评估三维一体的匿名化处理规范，以最终达到匿名化处理的法律效果。

3.1 再识别风险检验：“蓄意侵入者检验”标准

在推进我国公安数据匿名化开放进程中，宜采用“蓄意侵入者检验”标准来衡量数据的再识别风险。在识别认定主体方面，欧盟所采用的“任一主体说”将数据控制者及其他任何人均视为潜在的识别认定主体，即匿名化处理亟需达到任一主体都无法进行再识别的程度，其做法由于对数据控制者的要求过高而在客观上限制匿名数据的流通再利用，这与制度设计的初衷相悖；尽管“专家判定法”标准相对较为科学合理性，但其实施过程往往缺乏与专家资质认证、操作流程及责任追究等配套制度的协同，可能在实际操作中介入主观不确定性，进而影响判定结果的客观一致性。

“蓄意侵入者”标准则通过假设一位具备识别动机和一定识别能力的蓄意侵入者，这种侵入者仅为具有一定数据搜集能力的普通人，而非拥有特殊计算机技能的专家（如黑客破解技术）或犯罪分子（如非法获取数据）。该标准不仅在法律层面被予以明确界定，还对侵入者的识别能力进行现实适应性调整，使其与实际情景相符。此外，该标准还通过与职业伦理等原则融合，进一步强化了对匿名主体合法权益的保护。同时，在评估匿名化数据的再识别风险时，应遵循合理可能性原则。若识别匿名化数据的行为被现行法律所禁止，或其实施所需的成本显著超出合理范围（包括但不限于金钱、时间、技术成本），则应认定为不可识别，以维持隐私保护措施的合理性与可行性。

3.2 匿名化操作方法：区分处理直接标识符与准标识符

不同标识符的识别能力差异显著，其蕴含的科研、经济价值也千差万别。在匿名化处理过程中，本文借鉴美国HIPAA隐私规则中的“区分处理标识符”策略，参照《信息安全技术 个人信息去标识化效果评估指南》（GB/T 42460—2023）中的评估标准，并结合公安机关行政处罚文书开放实践，对个人信息中涉及的标识符进行分类梳理。具体而言，即按标准将公安《行政处罚决定书》中含有的个人信息标识符总量暂设定为14项，其中直接标识符细分为：姓名、居民身份证号码、户籍地、现住址、违法地址，准标识符细列为：性别、年龄、出生日期、受教育水平、政治面貌、工作单位、违法经历、违法事实、违法日期。针对这些标识符的匿名化处理，主要侧重于以牺牲数据准确性为限，模糊或间接淡化特定个人对应的可识别要素[13]。但与HIPAA中的规定不同，对上述标识符进行匿名化处理，绝不意味着上述所列举的标识符经处理后信息即实现匿名。公安机关还需依据特定的应用场景和预期目标，采用多样化的技术方法来调节数据的精度与真实性，在保持数据的匿名状态

的同时维持其后续利用价值。

3.2.1 个人信息中直接标识符的脱敏策略

对于直接标识符，鉴于其与个人信息主体的直接关联性，因而具有明确的个人识别属性。基于保护人格尊严之目的价值应当优先于利用个人信息之工具价值的理念[29]，公安机关必须采取严格的脱敏措施（如表 1 所示），使匿名化数据进入社会公众视野后，尽可能避免对信息主体合法权益的不当损害。

3.2.2 个人信息中准标识符的脱敏策略

准标识符个人信息主体的直接关联性相对较弱，其单独识别个人信息主体的能力有限。然而，当这些准标识符与其他数据结合使用时，会显著增加个体识别的风险。此外，通过某些准识别符还可以推知其他可识别信息[11]，例如通过居民身份证号码前六位可推测其户籍地信息，通过工作单位可推测收入水平等。因而公安机关需采取适度的数据脱敏策略（参见表 2），以降低信息的再识别潜力，并保持数据的分析和研究价值。

3.3 我国公安匿名数据去识别化效果评估

在对个人信息进行匿名化处理后，对其进行相应的去标识化效果评估是确保其符合数据开放标准的必经之路。鉴于信息在“可识别”与“不可识别”之间存在阈值界限[30]，本研究参考《信息安全技术 个人信息去标识化效果评估指南》（GB/T 42460—2023）中的评估框架，并结合美国“专家判定法”标准与我国公安工作实际，对评估流程进行系统性优化（如图 1 所示）。基于此，针对公示文书中的个人信息，公安机关必须实施分级分类脱敏处理。研究中，拟公开的个人信息须经历三阶段的严格评估流程：安全评估、定性评估、定量评估。在确保公示信息为非涉密数据后，再依评估结果将去识别化效果划分为四级：第一级指数据内有嵌直接标识符，此类数据在特定场景下足以直接识别信息主体的身份，须先行脱敏处理所有直接标识符后再进行后续操作；第二级涉及数据仅含有准标识符，其重标识风险（Re-identification Risk）高于或等于可接受风险阈值，鉴于此等级数

表 1 公安行政处罚决定书中个人信息直接标识符的脱敏处理操作

直接标识符	公安数据处理措施
姓名	部分脱敏处理，仅保留姓氏，其余部分使用占位符代替，以隐藏个人的具体身份信息。例如，将“李四”转换为“李＊”
居民身份证号码	脱敏处理，可视情况选择其中一种：（1）完全脱敏处理，如“＊＊＊＊＊＊”；（2）部分脱敏处理，仅公示身份证号码前四位，如“1100＊＊＊＊＊＊”，保护隐私的同时允许开展相应统计分析
户籍地/现住址	部分脱敏处理，仅公示到地级市一级的户籍地信息，如“浙江省宁波市＊＊＊＊＊”
违法地址	部分脱敏处理，仅公示到县级市一级，如“山东省济南市历下区＊＊＊＊＊”，以维护涉案地点后续开展合法活动

表 2 公安行政处罚决定书中个人信息准标识符的脱敏处理操作

准标识符	公安数据处理措施
性别	不脱敏处理，性别信息对于数据使用目的至关重要，并且一般不会直接导致隐私泄露或歧视
年龄	不脱敏或完全脱敏处理，根据具体情况和隐私保护需求，可以选择直接公示个人的年龄
出生日期	脱敏处理，可视情况选择其中一种：（1）完全脱敏处理，如“＊＊年＊＊月＊＊日”；（2）部分脱敏处理，通过泛化仅保留出生年份或更粗略的时间范围，如“1981 年”或“1980－1985”
受教育水平	不脱敏或完全脱敏处理，可根据保留数据分析价值或隐藏具体教育背景的需要进行选择
政治面貌	不脱敏或完全脱敏处理，政治面貌信息可以视情况采取完全脱敏，以保护个人的政治面貌隐私；或直接公示，以满足特定的研究或分析需求
工作单位	完全脱敏处理，以帮助受处罚人在认识到错误并改过自新后，减少社会歧视和就业障碍
违法经历	部分脱敏或完全脱敏处理，违法经历可以进行完全脱敏处理，以保护个人的隐私权益；或部分脱敏处理，以震慑潜在违法动机
违法事实	不脱敏或部分脱敏处理，违法事实若非涉密可直接公示，但应确保符合法律法规和伦理标准
违法日期	不脱敏或部分脱敏处理，公开违法日期可不经匿名化处理。若采用部分脱敏处理，可采用泛化处理或占位符进行替换，例如“2024 年 3 月”或“2024 年 3 月＊日”

据的再识别风险显著，须采用额外的匿名技术进行再脱敏处理，以有效削减重标识风险；第三级为数据重标识风险低于可接受风险阈值，这表明匿名化数据已达到安全开放标准，是否开放匿名化信息需根据各地区的具体法规和政策审慎决策；第四级指数据中未包含任何形式的标识符，实现了个人数据的最大程度匿名性，此时匿名化数据完全符合安全开放的标准。随着去标识化等级的递减，重识别风险也相应增加。在此过程中，所有评估活动皆应被详细记录，有关专家的沟通与协商应贯穿整个评估流程。在评估流程的最后环节，匿名化数据一旦通过上述评估标准，并符合《个人信息保护法》《网络安全法》等相关法律法规与政策要求后，方可启动公安匿名化数据公示工作。

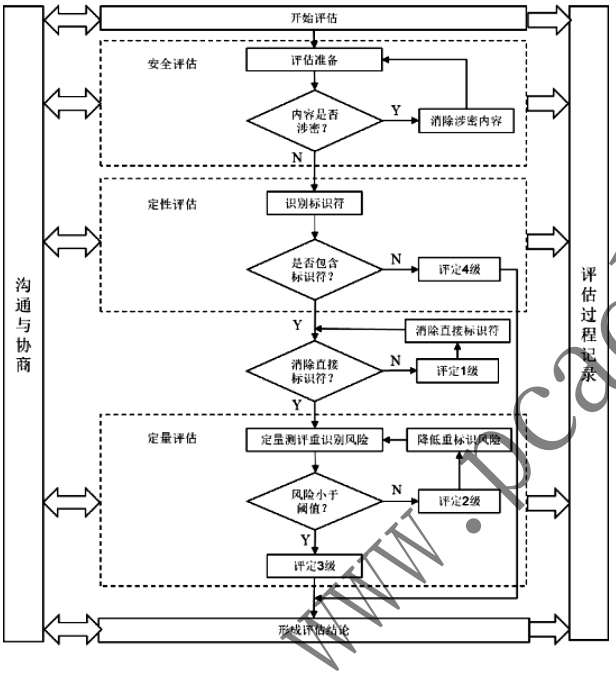

图 1　个人信息去识别化效果评定流程

面对与日俱增的数据开放总量，公安机关还应持续进行动态评估，以判断匿名数据是否仍然不可识别，同时考虑新的技术或新的可获取信息是否可用于再识别个人[11]。一旦监测到存在或可能存在个人信息泄露、毁损或丢失的风险时，应依据《网络安全法》第四十二条第二款的规定，立即采取补救措施并及时向有关主管部门报告。

4　现状检视：我国公安机关匿名化处理实践示例

4.1　我国公安机关匿名化处理现状

为进一步探究公安数据开放的未来发展趋势，本文以 2024 年 6 月 1 日为节点，选取了 17 个省市地区公安机关在当地政府数据开放平台上发布的《行政处罚决定书》作为分析样本，并针对其中部分地区的匿名化处理工作进行质效评估。研究中，将《行政处罚决定书》中涉及的个人信息标识符总量暂设定为 14 项，作为衡量匿名化处理效果的关键指标。研究发现，各地区文书在个人信息标识符的公示方面呈现出显著的地域性差异（如图 2 所示），其中山东省、天津市以 13 项标识符的开放数量并列位居第一，远高于平均约 8 项的标识符公开水平。

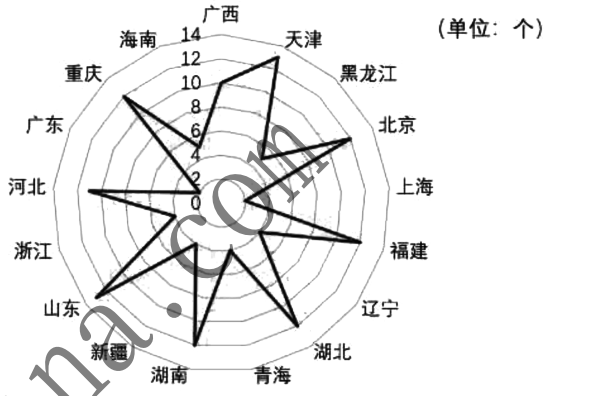

图 2　各地区公安行政处罚决定书中标识符定量分析

尽管北京、重庆、上海、山东等 17 省市地区已开始在行政处罚决定书的公示过程中对个人信息实施匿名化处理，但各地在匿名化处理的深度、方法以及公示内容上存在明显差异（如表 3 所示）。建议对现行行政处罚公示标准进行综合评估，并根据评估结果实施必要的优化和调整，进而推进全国范围内实现公安数据公示标准的一致性，提升个人信息保护的效率效果。

4.2　我国公安机关匿名化处理的现实路径分析

与 HIPAA 中所规定的“安全港”标准不同，我国目前并未明确规定所必须去除的隐私标识符，也未根据案件具体情况制定差异化的信息匿名公示策略，这在一定程度上影响了文书内容公示的标准性和灵活性。

（1）辽宁省公安机关

辽宁省公安机关在公示行政处罚决定书时，存在两个亟待解决的问题：其一，受处罚者的姓名信息未经脱敏处理即被公之于众。对受处罚者个人信息的脱敏处理需要进一步细致的考量、审查和优化（如表 1、表 2 所示）。其二，个别行政处罚文书在信息录入时的规范性存在偏差，如日期年份输入错误等。对此，宜制定详尽的信息录入标准和操作指南，使所有录入工作遵循统一的规范，并对录入数据实施双重或多重验证机制；同时引入先进信息技术，如自动化数据校验工具，以辅助识别和纠正录入过程中的潜在错误。

表 3　不同地区公安行政处罚决定书对个人信息的匿名化程度

序号	地区	姓名	性别	民族	年龄	文化程度	出生日期	居民身份证号码	户籍地	现住址	工作单位	违法经历	违法日期
1	北京市	/	×	×	×		√	√	/	/	√	×	×
2	天津市	/	×	√	√		√	√	/	/	√	√	×
3	上海市	/											
4	重庆市	/	×		×		×	√	√	√	√	/	×
5	山东省	/	×		√	×	√	√	√	√	√	√	√
6	河北省	/	×	√			√	√	/	/		/	×
7	浙江省	/											×
8	辽宁省	/											×
9	黑龙江省	/						/					×
10	福建省	/	×	×		√	/	/	/	/		/	×
11	湖北省	/	×		√		√	√	/	/	√	/	×
12	湖南省	/	×	√		√	√	√	/	/	√		×
13	广东省	/							/				×
14	海南省	/											×
15	青海省	/	×										×
16	广西壮族自治区	/	×				×	√	/	/	√	×	
17	新疆维吾尔自治区	/											×

注：表中 × 表示未脱敏；/表示部分脱敏，√表示完全脱敏

（2）湖北省、四川省公安机关

湖北省公安机关对行政处罚决定书的访问实施频率限制，这一措施为个人信息的不当再识别设置了壁垒，但也对学术界为该省行政处罚信息脱敏效果的研究与改进工作形成制约。四川省公安机关通过设置用户登录机制，要求用户在访问相关数据前进行个人信息的登录验证，在提高信息访问安全性的同时，增强对访问者身份的有效监管，从而保障匿名化信息安全。

（3）广西壮族自治区公安机关

广西壮族自治区公安机关在公示行政处罚信息时，由于未对出生日期等敏感个人信息实施必要的脱敏处理，这不仅使信息的再识别风险显著增加，而且当这些信息与其公示的户籍地信息相结合时，即可直接推断出受处罚人身份证号码的前 14 位信息。同时，性别信息的公示进一步限定了身份证号码中第 17 位（性别码）的可能值，从而将性别码的数量缩小至 5 个。在居民身份证号码中，第 15 至 17 位作为顺序码，其中第 15 至 16 位的排列组合总量为 100 种，通过对这两位数字进行解析后，并结合第 17 位的性别码，可衍生出 500 种可能的身份证号码序列。一旦前 17 位号码得以确认，第 18 位（校验码）即可通过特定加权算法计算得出，这进一步使得从 500 种潜在组合中准确识别出真实居民身份证号码成为可能。再通过将这些序列与已公开匿名主体的姓氏进行交叉验证，可迅速缩小潜在匹配范围。对此，建议有关部门对相应的个人敏感信息进行脱敏处理（具体措施参见表 1、表 2）。

5　结论

数据警务是未来警务发展的方向[31]，匿名化的应用正逐渐成为实现公安数据开放与个人数据保护平衡的“不二法门”。本文从国际视域下探讨欧盟、英国以及美国在匿名化处理法律标准领域的可行之处，并结合我国 17 省市地区公安机关数据开放实践，构建起集操作方法、再识别风险检验与去识别化效果评估于一体的匿名化处理规范。同时，这一规范还综合考虑数据保护的法律要求、技术实现的可行性以及社会对数据开放的需求，在保障数据安全与个人隐私的同时，最大化地释放数据的社会、经济价值与应用潜力。然而，再识别风险如影随形地伴随着匿名数据，一旦发生再识别事件，公安机关应立即开展应急响应工作，实施紧急干预措施以减轻信息泄露带来的损失，并最大限度地保障公民的信息隐私安全。

参考文献

[1] 丁红发，孟秋晴，王祥，等. 面向数据生命周期的政府数

据开放的数据安全与隐私保护对策分析［J］. 情报杂志，2019，38（7）：151－159.
［2］复旦大学数字与移动治理实验室. 中国地方公共数据开放利用报告——城市（2023 年度）［R/OL］.（2023－11－01）［2023－12－02］. http：//ifopendata. fudan. edu. cn/report.
［3］张艳，王璐瑶. 政府数据开放场景下匿名化数据的再识别风险防范［J］. 电子政务，2024（5）：64－76.
［4］高富平. 个人信息保护：从个人控制到社会控制［J］. 法学研究，2018，40（3）：84－101.
［5］高颖，杜娟. 大数据时代数据匿名化的法律规制［J］. 情报理论与实践，2021，44（10）：50－56.
［6］UK. Information commissioner’s office. anoymisation：managing data protection risk code of practice［EB/OL］.［2024－04－28］. https：//ico. org. uk/media/1061/anonymisation-code. pdf.
［7］齐英程. 我国个人信息匿名化规则的检视与替代选择［J］. 环球法律评论，2021，43（3）：52－66.
［8］程海玲. 个人信息匿名化处理法律标准探究［J］. 科技与法律（中英文），2021（3）：26－35.
［9］ALEXIN Z. Does fair anonymization exist?［J］. International Review of Law，Computers & Technology，2014，28（1）：21－44.
［10］衲钦，张慧春. 数智环境下匿名数据治理创新对策研究［J］. 科学管理研究，2022，40（2）：124－130.
［11］张晨原. 数据匿名化处理的法律规制［J］. 重庆邮电大学学报（社会科学版），2017，29（6）：52－58.
［12］韩旭至. 大数据时代下匿名信息的法律规制［J］. 大连理工大学学报（社会科学版），2018，39（4）：64－75.
［13］赵精武. 个人信息匿名化的理论基础与制度建构［J］. 中外法学，2024，36（2）：326－345.
［14］OHM P. Broken promises of privacy：responding to the surprising failure of anonymization［J］. UCLA Law Review，2010，57（6）：1701－1778.
［15］王立梅. 大数据视角下的个人信息匿名化规则构建［J］. 云南民族大学学报（哲学社会科学版），2021，38（5）：142－150.
［16］张涛. 欧盟个人数据匿名化的立法经验与启示［J］. 图书馆建设，2019（3）：58－64.
［17］Article 29 Data Protection Working Party. Opinion 04/2007 on the concept of personal data［EB/OL］.（2007－06－20）［2024－04－05］. http：//ec. europa. eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp 136_ en. pdf.
［18］程海玲. 个人信息匿名化处理的私法规制［D］. 重庆：西南政法大学，2020.
［19］BALBONI P，MACENAITE M. Privacy by design and anonymisation techniques in action：case study of Ma3tch technology［J］. Computer Law & Security Review，2013，29（4）：330－340.
［20］张建文，高悦. 我国个人信息匿名化的法律标准与规则重塑［J］. 河北法学，2020，38（1）：43－56.
［21］STALLA-BOURDILLON S，KNIGHT A. Anonymous data v. personal data-false debate：an EU perspective on anonymization，pseudonymization and personal data［J］. Wis. Int’l LJ，2016，34：284.
［22］EL EMAM K，DANKAR F K，NEISA A，et al. Evaluating the risk of patient re-identification from adverse drug event reports［J］. BMC Medical Informatics and Decision Making，2013（13）：1－14.
［23］National Institute of Standards and Technology. De-identification of personal information［R/OL］.（2015－10－20）［2024－05－13］. https：//nvlpubs. nist. gov/nistpubs/ir/2015/NIST. IR. 8053. pdf.
［24］DALENIUS T. Finding a needle in a haystack or identifying anonymous census records［J］. Journal of Official Statistics，1986，2（3）：329.
［25］EL EMAM K. Methods for the de-identification of electronic health records for genomic research［J］. Genome Medicine，2011（3）：1－9.
［26］范为. 大数据时代个人信息保护的路径重构［J］. 环球法律评论，2016，38（5）：92－115.
［27］许可，孙铭溪. 个人私密信息的再厘清——从隐私和个人信息的关系切入［J］. 中国应用法学，2021（1）：3－19.
［28］孔祥稳. 行政处罚决定公开的功能与界限［J］. 中外法学，2021，33（6）：1619－1637.
［29］胡文涛. 我国个人敏感信息界定之构想［J］. 中国法学，2018（5）：235－254.
［30］张丽，许多奇. 风险控制理念下我国个人信息匿名化处理的法律规制［J］. 重庆大学学报（社会科学版），2023，29（2）：220－231.
［31］张宁. 我国公安数据开放的政策过程、现状评估与拓展路径分析［J］. 情报杂志，2022，41（02）：160－168.

（收稿日期：2024－08－03）

作者简介：

张寒（2001－），男，硕士研究生，主要研究方向：公安数据开放、数据匿名化。

张宁（1980－），男，博士，副教授，主要研究方向：数字公安、警务数据开放。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部