主题研讨

个人信息流通利用的制度基础

——以信息识别性为视角

高富平

内容提要:《个人信息保护法》以信息主体同意为基础,构筑了个人控制的个人信息直接利用制度,但其是否为流通利用提供了通道仍存疑问。信息因其识别性能的差异,可区分为直接标识符、间接标识符和准标识符,三者给个人权益带来的危害风险不同。《个人信息保护法》规定的匿名化和去标识化本质上是针对特定数据集中信息识别风险的制度安排,能消除因信息本身识别性产生的风险,而很难消除基于识别分析的识别性产生的风险。因此,缺失针对"基于识别分析的识别性产生的风险"的措施,现行关于匿名化和去标识化的规范均不能支撑个人信息流通利用。去标识化需要改造成为"去直接标识符+识别控制"的受控去标识化制度,在防控个人信息识别风险的前提下,为个人信息流通利用提供制度保障,以最大化实现个人信息的社会价值。

关键词:个人可识别　去标识化　匿名化　个人信息流通利用

高富平,华东政法大学法律学院教授。

进入数据化时代,各种硬件设备形成的海量数据记录与特定个体关联,可被用于匹配与分析。而人不应被当作客体而随意处理(分析),由此建立起个人数据(个人信息,本文在相同意义上使用)保护制度,以保护信息主体的权益。我国已形成以《个人信息保护法》为核心的信息主体权益保护的法律体系。信息因具有"识别性"而成为个人信息,受《个人信息保护法》调整。识别性是个人信息的效用或价值所在,也是个人信息的风险所在。针对交往(交易)和服务(治理)对象开展识别分析(profiling),[1]有助于更加快速和精准地作出判断、预测和决定,大大提升社会治理和经济运营的效率,提升社会整体生产力。但是,对个人的识别分析可能侵害隐私利益,并危害人的尊严、自治甚或平等等权益。

〔1〕 本文所谓"识别分析"与"画像"(profiling)同义,其作用的结果是了解特定对象的个性特征,并非识别特定自然人的身份。于是,本文中"识别"的目标是身份;"识别分析"的目标是个性特征。

因此,《个人信息保护法》旨在通过个人信息处理规范来控制处理行为给信息主体带来的侵害风险,“促进个人信息合理利用”。这里的“合理”的前提是能够利用,保持信息具有识别分析的功能。如果缺失识别分析的功能,个人信息就不再具有效用。“合理”意味着个人信息的使用对信息主体权益的侵害风险保持在社会或信息主体可接受的范围内。为控制风险,《个人信息保护法》建立了个人信息正当处理的基本原则,同时赋予信息主体对处理行为以一定的控制权。

为实现这样的控制,《个人信息保护法》不得不将个人信息的使用维持在特定处理者在特定目的范围内的直接使用关系,超出特定目的或者对外提供(即流通利用)则需要经信息主体同意。可以说,《个人信息保护法》基本没有触及流通利用。[2] 流通利用是否一定会给个人带来额外风险?如果有,那么我们应当在排查该风险的基础上探寻有效的制度,将流通利用风险控制在合理的范围内,从而构建安全可信的流通利用秩序。这不仅关系《个人信息保护法》的实施,而且也关系未来数据要素市场的建设。保持个人信息的效用,同时消除信息主体权益受损的风险,就成为“促进个人信息合理利用”需要解决的基本问题。本文从识别含义和识别方式入手,寻找个人信息识别侵害信息主体权益的风险点及现实的规制路径,检讨《个人信息保护法》中匿名化和去标识化制度作为识别分析和流通利用制度的可行性,构建我国个人信息效用和权益侵害风险相平衡的制度。

一 个人信息流通利用的制度缺失:《个人信息保护法》的缺憾

《个人信息保护法》并未完全禁止信息处理者对外提供,只是受到信息主体意志的左右。该法的宗旨是控制个人信息的滥用或不当处理的风险,而实现这一目的的手段被认为是维护信息主体对个人信息使用的控制。这种控制包括处理前的同意、处理中的拒绝以及处理后的删除等。在这方面,《个人信息保护法》以信息主体对个人信息的控制为本位,赋予信息主体较强的控制力。在个人信息处理前,只要没有法定事由,收集和使用个人信息就应当取得信息主体同意。同意作为处理的合法性基础,仅在于允许特定信息处理者在同意范围内或特定目的的必要范围内利用个人信息,并没有允许信息处理者自由利用个人信息,更未授予信息处理者向他人提供信息的权利。即使初始收集时的同意包括向业务合作伙伴等第三方提供信息的内容,《个人信息保护法》仍然要求信息处理者须获得信息主体的单独同意才能向外提供个人信息(第 23 条)。另外,该法第 45 条还规定了“可携权”,个人信息处理者应配合移转,提供移转途径。

可以说,从个人信息保护相关立法来看,我国立法已将同意工具用到了极致。[3] 为

〔2〕《个人信息保护法(草案一审稿)》第 1 条明确地将保障个人信息依法有序自由流动作为立法目的之一,但二审稿之后该表述被删除。

〔3〕除了明确同意要件(《个人信息保护法》第 14 条)外,《个人信息保护法》还有三种情形需要重新同意,五种情形需要单独同意。一般情形下,个人信息的处理目的、处理方式和处理的个人信息种类发生变更的即应当重新取得信息主体同意(第 14 条第 2 款);在组织变更、对外提供两种特殊情形下,接收方变更原先的处理目的、处理方式的,应重新取得信息主体同意(第 22 条和第 23 条)。单独同意的五种情形是:对外提供其处理的个人信息(第 23 条);公开个人信息(第 25 条);公共场所采集信息,用于公共安全以外的目的(第 26 条);敏感个人信息的处理(第 29 条);向境外提供(第 39 条)。

实现信息主体的控制,其需要与信息处理者之间建立意定或法定目的的使用关系。这样的制度决定了它不考虑赋予信息处理者以对外提供信息的权利,不允许个人信息脱离信息主体的“控制”而不断流动。其直接法效果是信息处理者不享有任何向他人提供其所控制个人信息的权利,第三人没有从信息处理者手中间接取得个人信息的合法渠道(当然,具有《个人信息保护法》第 13 条第 1 款第 2 – 7 项合法性基础的除外)。这样,《个人信息保护法》的“利用”就不包括再利用或流通利用,也就没有合理再利用的空间。

在《个人信息保护法》之前,《网络安全法》第 42 条的但书条款规定“经处理不能识别特定信息主体且不能复原的”可以不经信息主体同意而向他人提供。《民法典》第 1038 条也作了类似规定。这曾经被解释为个人信息非经信息主体同意不得流通的法律依据。《个人信息保护法》开始启用“匿名化”一词,且将匿名化定义为“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”,将匿名化的个人信息排除在了个人信息范畴之外。这样,匿名化处理后的信息流通就不再属于“个人”信息流通利用,即匿名化不能作为个人信息流通利用的法律依据。《个人信息保护法》在规定匿名化的同时,还引入了去标识化制度,并将去标识化视为个人信息处理的安全措施(《个人信息保护法》第 51 条第 3 项、第 73 条),但未明确去标识化个人信息是否可以非经信息主体同意而流通利用。由此,《个人信息保护法》并未在同意之外建立个人信息流通利用的规则。

综上,《个人信息保护法》没有为个人信息流通利用提供法律依据,然而数据要素市场建设要求探寻个人信息社会化、市场化利用的可行制度方案。在这方面,去标识化信息成为平衡个人信息利用与信息主体权益侵害风险、实现对外提供的合法化的唯一突破口。

二 个人识别理论:信息的识别性分析

根据《个人信息保护法》,与已识别和可识别的自然人有关的信息均属个人信息。但是对于信息缘何能够“识别”个人这一问题,尚无细致深入的研究。而对此的正确理解,既关系着《个人信息保护法》的正确适用,也关系着该法是否给个人信息的流通利用留下空间。

(一)识别个人的社会实践和分类:“已识别”与“可识别”

个人信息的基本效用是识别个人。从识别的目的角度,分为识别身份和识别个性特征:[4]识别身份是将信息与特定个人关联起来,使该信息归属于该个人或让其承担该信息的后果;识别个性特征是描述信息主体的个性特征或预测其行为倾向。不过,这两种识别都以信息具有识别性(identifiability)为前提。

识别的基本含义就是辨识某人的身份,使其被识别出来。达姆曼(Ulrich Damman)认为,只有当清楚地表示数据是关于此人而非他人时,此人才能被认为是已识别的。[5] 简(Rosemary Jay)认为某一信息主体是已被识别的,在于存在充足的信息能够联系或认知

〔4〕 参见高富平:《个人信息保护:从个人控制到社会控制》,《法学研究》2018 年第 5 期,第 93 – 94 页。

〔5〕 Vgl. Dammann, in: Simitis, Bundesdatenschutzegesetz, 7. Aufl. 2011, § 3 Rn. 297, 310.

该人,通过一定方式将该人与其他人区分开来,并知道该人是谁。[6] 欧盟立法者也坚持这样的观点,认为只有当某人从一群人中被区分出来,该人才被认为是已被识别的。[7] 因此,识别的基本含义是从信息辨识出个人,并且往往是以姓名为代表的特定个人。但有时候,当信息指向或具象到某个个体但不知其是谁时,我们称其为"可识别的自然人"。可识别的自然人不是社会身份明确的具体个人,而是社会中的某个个体。在身份识别之前,其为一个抽象的存在。显然,当我们认为"身份"指一个具体个人的社会身份时,可识别的自然人仍然身份不明。这也意味着,当我们与可识别的人打交道时,只需要收集该人的信息进行分析,了解对方是一个什么样的人,很多时候不需要再行识别身份。这样的结论符合社会常识。这时,便涉及识别的另一个含义——识别具体个人的个性特征。所运用素材是有关该个人的过往事实或行为,至于分析什么特征则取决于交往目的。识别个性特征的价值至少可以归纳为两方面:一是分析信用,防范不必要的商业风险和安全风险;二是分析偏好、行为倾向、需求等,以决定是否和如何通信或采取行动等。

数字技术给个性识别创造了无限便利和可能性。识别个性特征对信息的基本要求是信息之间可链接,只要能够将关于某信息主体的信息都链接在一起,就能够在不识别身份的前提下勾勒出该信息主体的形象或特征,甚至实现活动目的。在万物互联的泛在网络环境中,存在着许多可以指向某个体的网络和设备ID(亦被称为数字ID),这些数字ID的背后是用户,而用户对应社会中的某个主体(这里仅讨论自然人)。但是,数字ID是一种匿名个体,在识别其真实身份之前,只是可识别的人(identifiable person)。数字ID的出现和大量应用使识别个性特征能够方便地开展,这是因为计算机的应用使收集、存储和获取关于某信息主体的信息成为可能,且使不断累积和重复使用所掌握的信息主体数据并识别分析成为可能。[8] 在此种技术环境下,识别信息主体的真实身份不再是个性特征识别或用户画像的前提。综上,数字技术大大改变了信息识别个人的能力和方式。因此,信息的识别性理解和研究成为解开个人信息保护法谜团的钥匙。

(二)信息的识别性:直接标识符、间接标识符与准标识符

利用信息进行识别分析是通过计算处理实现的。计算机语言使用标识符来标识变量、函数或属性的字符序列,由此可以关联或结合相关信息,形成关于某对象的文档或数据集。作为计算机语言的用语,标识符一定是在一个处理域内具有唯一性。标识符被应用于识别分析中,在一个数据集中区分出标识符和非标识符性质信息。标识符的本质是

---

〔6〕 See Rosemary Jay, *Data Protection Law and Practice*, 4th ed., Sweet & Maxwell, 2012, p. 172.

〔7〕 See Article 29 Data Protection Working Party, Opinion 4/2007 on the Concept of Personal Data (01248/07/EN WP 136), p. 12.

〔8〕 正因为如此,长期存储的关于信息主体数据的文档是《个人信息保护法》的立法起点:欧洲委员会的1981年《有关个人数据自动化处理中的个体保护公约》(*Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*)第2条明确定义了"自动化数据文档"(automated data file)和"文档的控制者"(controller of the file),将关于信息主体的文档控制和利用作为规制主线;欧洲议会与欧洲理事会1995年《关于涉及信息主体数据处理的信息主体保护以及此类数据自由流通的第95/46/EC号指令》(*Directive 95/46/EC on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data*)第2条定义条款定义了信息主体数据存档系统(personal data filing system):指可按照特定标准访问(或获取)的结构化的信息主体数据集合。

具有指示或指向个人作用(不一定能表彰身份),为更多信息的识别分析提供“根基”。标识符也进入了个人信息保护法。欧盟《一般数据保护条例》在个人数据的定义中以列举方式使用了标识符,将姓名、身份证号、定位数据、网络标识符等视为标识符,并包括在线标识符(因特网协议地址、cookies 或者射频标识别标签等)。美国一些特殊立法及行业实践中广泛使用“个人可识别信息”(PII),其实就指标识符。我国《个人信息保护法》中的“去标识”也是针对标识符而言。

标识符是个人信息中直接指向或关联到个人的信息,且这种指向或关联具有唯一性。任何人不作进一步分析,仅凭常识即可以判断该信息属于或指向特定个人。简言之,标识符具有直接识别出个人的能力,只是这里的个人时常被解释为不明身份的个体。标识符与信息主体的关联性不完全相同。荷兰蒂尔堡大学的莱恩斯教授(Ronald E. Leenes)发现了标识符内部不同部分发挥不同作用,并提出查找型识别和认知型识别标识符的分类。〔9〕 受这一分类的启发,本文将标识符分为直接标识符和间接标识符两类,划分标准是标识符是否对应到现实中具体的个人身份。直接标识符是指那些唯一能够关联信息主体身份的标识符,如姓名、身份证号(及其他类似唯一身份的权威编码)以及指纹、面部轮廓等生物识别信息等。在某种意义上,含直接标识符的信息约等于与已识别个人有关的信息。间接标识符是指不能单独识别具体信息主体身份,但是结合其他信息可以识别具体信息主体的部分标识符,其身份仍然是隐匿的状态。间接标识符以数字 ID 与在线标识符为代表。大量网络和设备 ID 的存在使得识别分析可以很方便地开展,并进行网络通信(如精准推送)和科学研究。〔10〕

自从标识符进入法律后,其范畴一直存在争议,尤其是唯一关联个体的数字标识符,不同的法域或机构有不同认知。〔11〕 例如,MAC 地址、广告 ID、像素标签、账户别名(用户名)、设备指纹等也具有唯一性,是否应当作为标识符并列入去标识处理范畴就有截然不同的做法。〔12〕 这些标识符的使用可能会留下痕迹,当与其他信息相结合时,可用于创建

〔9〕 莱恩斯教授从信息在识别中的作用的角度,对识别符作进行类型化分析,将标识符区分查找型识别(look-up identifiability,简称 L 型)和认知型识别 recognition identifiability,简称 R 型)标识符。前者用于个人身份的识别,识别性体现于标识符与有姓有名的个人之间联系,如身份证号码、住址以及手机号码等;后者是在未与显名个人产生联系的情况下对其进行认知,它需要被识别人出现或活动,利用特定场景下事实信息的关联。参见 Ronald Leenes, Do They Know Me? Deconstructing Identifiability, 4 *University of Ottawa Law & Technology Journal* 135 (2007)。

〔10〕 比如,一家企业使用 WiFi 分析数据来计算不同零售店每小时的访客数量。这需要使用移动设备的“媒体访问控制”(MAC)地址向其公共 WiFi 热点广播探测请求。MAC 地址对于设备来说是唯一的,即使企业不知道设备使用人的姓名,使用 MAC 地址(或其他唯一标识符)来跟踪设备,就是对潜在可识别个体的识别,也达到了识别出可识别个人的效果。

〔11〕 尽管就去标识化的必要性和价值达成了广泛共识,但关于数据是否以及何时可以真正去标识化的争论似乎仍无休止。学术界、监管机构和其他利益相关者多年来一直在寻求建立去标识化的通用标准,但迄今为止,甚至没有形成一个通用术语,see Jules Polonetsky, Omer Tene and Kelsey Finch, Shades of Gray: Seeing the Full Spectrum of Practical Data De-identification, 56 *Santa Clara Law Review* 593, 596 - 599 (2016). 例如,最近美国统一州法委员会制定的《统一个人数据保护法》(*Uniform Personal Data Protection Act*)定义的直接标识符就非常狭义,与本文的直接标识符相似:直接标识符是指通常(commonly)用于识别数据主体的信息,包括姓名、物理地址、电子邮件地址、可辨识(recognizable)的照片、电话号码和社会保障号码等。

〔12〕 例如,美国国家标准研究院对直接标识符限于直接与个人关联的数据,广告行业也采取这样的做法,而在健康领域,则将直接标识符扩展到设备 ID。

个人数据集并识别相关主体。除此而外,网络或系统应用会形成大量的所谓间接个人数据(indirect personal data)[13],亦可以用来进行关联分析,但未进入标识符清单。

本文将大量存在的可以将个体区分开来、但又不能直接识别身份的网络和设备 ID 归为间接标识符,是基于识别分析实践,旨在建立更清晰的个人信息处理规范。间接标识符实际上只是发挥实现信息关联或链接的作用,需要"额外信息或通过公共领域的其他信息进行交叉链接(cross-linking)才可用于识别个人",[14]是否属于个人信息取决于它是否真实地与特定个人相关联。[15] 因此,我们应当考虑将其独立于直接关联到人的标识符来对待。网络广告促进会(NAI)在其最新的《2020 年行为准则》中更新了术语分类体系,反映了这样的趋势。该行为准则将标识符区分三类:识别个人的信息(personally-identified information, PII),即用来或旨在直接识别特定个人数据;识别设备的信息(device-identified information, DII),即与浏览器、设备或一组设备关联的且不用于直接识别某个人的数据;去识别的信息(de-identified information),即与个人与设备均不发生关联的数据。[16] 显然,这样划分的标准是信息直接关联的对象,识别设备的信息是直接关联到物的,就是本文所称的间接标识符。本文的数字 ID 与 DII 基本上可以等同。将 DII 独立出来就是为了不识别特定个人的个人信息利用。[17]

法律时常使用单独识别与结合识别来描述身份的识别方式,如《网络安全法》第 76

〔13〕 这里的间接个人数据类似于本文所说的间接标识符。行业报道显示,在系统或系统架构中对个人数据进行加密编码时可以创建这种唯一性的数据。这组数字本身没有任何意义,但当它在注册表中用于识别一个人时,该密钥在所有场景中都成为个人数据。有两种情形可以生成这样的识别符:其一,从用户管理中派生出来的注册表项,它们是自动创建的,以便在使用信息系统时确保可追溯性。这样的字段在设计良好的系统中随处可见。第二种类型是用户活动在系统中呈现的,行业术语称为个人外键(personal foreign keys),包括系统各部分之间对个人主数据的引用,例如对销售订单联系人或产品信息中的产品所有者的引用。这些外键通常是用户活动的结果。See Indirect Personal Data Findings Cause Surprises in GDPR Analyses, INEO News, https://www.ineo.fi/en/indirect-personal-data-findings-cause-surprises-gdpr-analyses/,最近访问时间[2022-01-02]。

〔14〕 International Organization for Standardization, ISO/TS 25237:2008 Health informatics - Pseudonymization.

〔15〕 比如,IP 地址可能是一个简单的收集点(point of collection),没有附加任何其他信息,则如果公开危害会更小;因此,它不应被视为个人信息(如果未链接到任何传记或联系信息)。如果指向设备的 IP 地址不透露"私密"细节(例如,它仅用于记住其网站用户的首选语言,而不会以其他方式提供给第三方),则该 IP 地址也可能不被视为个人信息。与传记信息(biographic information)相结合的 IP 地址变得更加"敏感",尤其是当与具有"亲密"性质的传记信息相关联时;如果该信息尚未"可用",则会变得更加"敏感"。在这种情况下,IP 地址(连同与其链接的信息)显然可以作为个人信息。当与联系点(例如电子邮件地址、用户账户或实际地址)相关联时,该 IP 地址将变得更具潜在危害性。See Eloïse Gratton, If Personal Information is Privacy's Gatekeeper, then Risk of Harm is the Key: A Proposed Method for Determining What Counts as Personal Information, 24 *Albany Law Journal of Science and Technology* 105, 176-177 (2014).

〔16〕 2020 NAI Code of Conduct, https://thenai.org/wp-content/uploads/2021/07/nai_code2020.pdf,最近访问时间[2022-01-02]。网络广告促进会之前的行为准则遵循的是二分法,建议其成员在可能的情形下避免收集和保存直接识别个人数据(directly identifying data),如姓名、email 和邮政编码、电话或社会保障号。结果许多公司收集和保留了不直接识别个人的标识符,如 cookie、移动广告标识等。显然,这样的标识符认定与现实社会对隐私认知是不一致的。于是,三分法旨在指引成员对待和保持数据联结或关联的标识符,而不是继续维系虚幻的数据可识别性与现实社会对隐私危害的分裂状态。

〔17〕 网络广告促进会《2020 年行为准则》认为,借助 DII 可以收集和以假名方式使用数据,将某种利益联结到某个特定但不知身份的用户,或者对其作出推断或预测。就定向广告而言,以下收集和使用 DII 数据的方式是符合标准的:(1)作为一种确保所收集和接受的数据不被用于识别特定个人的措施,如仅使用随机产生的数字标识符而不是姓名或电子邮箱地址;(2)公开承诺将数据保存为 DII;(3)采取合理措施如合同,防范获取数据的任何公司试图将数据与 PII 结合或者采取将数据用来识别特定个人的行为(除非 DII 是专属于接受者)以发布定制广告。

条与欧盟《一般数据保护条例》第 4 条的规定。所谓单独识别,指单一信息能直接辨识出特定个人身份,显然,只有直接识别符(如姓名、身份证号、指纹和面部轮廓等生物识别等)才具有这性的性能。在没有直接标识符时,通常需要结合识别方式来识别信息关联的主体的身份。有了结合识别就会扩大用于身份识别的信息类型和数量。除标识符外,还有大量信息均属于描述信息主体特征或某方面特性的属性信息,可以在特定场景下"还原"或对应现实中的具体个人,因而也具有识别人的身份的功能。在当今的智能分析技术下,从大量人口收集数据并从在该数据收集中出现的属性来识别个人,已被大量应用(被称为间接识别分析)。这些属性信息因可以结合识别身份,所以在行业实践和学术研究中,也被称为"间接标识符"。[18] 但本文认为其属于"准标识符":准标识符并不是标识符,[19]因为它不具备标识符的唯一性,并不唯一指向特定主体或者不存在直接的关联。准标识符范围广泛,几乎没有边界,性别、婚姻状况、种族、民族、出生国、邮政编码或其他地理位置信息、出生日期或年龄、职业、语言、公民状态、受教育情况、奖惩记录、总收入、宗教信仰等均可能成为准标识符。某个准标识符在多大程度上可以清晰地识别身份,取决于特定的场景。比如,有时性别、职位、职业等几个准识别符就足以将人识别出来。仅含有准标识符的信息本身不具有识别性,但信息间的可链接性使其通过识别分析过程仍然可以具有识别性。

从上述分析,我们可以得出三点结论:第一,信息本身的识别性是一条连续的光谱,一端是姓名等直接指认出身份的直接标识符,另一端无法识别个人,而在这两端之间有间接标识符、准标识符及各种可以通过结合分析而匹配到某信息主体的大量信息。第二,信息的识别性不仅取决于其本身,也取决于分析行为。根据信息本身是否有指向或关联能力,我们区分出标识符信息;从标识符与社会身份的关联性,又区分出直接标识符和间接标识符。而纳入个人信息范畴的大量其他信息则没有与主体之间的直接关联属性,仅依赖信息的可链接性而归属于或关联某个标识符(尤其是数字 ID),正因为信息汇集和结合分析使很多信息具有了识别性。第三,标识符是识别分析的工具。如果说个人信息最主要的价值是个性特征分析,以便于作出各种决策或采取某种行动,那么实现这一目标所需要的恰恰是非标识符信息(属性信息和行为信息),而标识符的作用仅在于使这种识别分析成为可能,或者说,标识符主要充当识别分析的手段。

(三)《个人信息保护法》中"识别"和"有关"的含义

《个人信息保护法》第 4 条使用了"识别"和"有关"两个核心术语,以"识别"修饰识别的对象即已识别或者可识别的自然人,以"有关"来限定个人信息的范围;同时用"无法识别"而不是"无关"来界定非个人信息。这里的"识别""有关"及"无法识别"的含义及其关系值得探讨。

---

〔18〕 See Jules Polonetsky, Omer Tene and Kelsey Finch, Shades of Gray: Seeing the Full Spectrum of Practical Data De - identification, 56 *Santa Clara Law Review* 593, 605 (2016).

〔19〕 信息主体数据中具有指向属性的信息也被纳入标识符体系,突破了计算机语言对标识符要求在特定场域的唯一性要求。本文认为,只有唯一性、指向作用的信息才是标识符,只有标识符才有直接与间接区分。本质上,准标识符不是标识符,而是识别信息主体的因子。

根据前文论述,已识别和可识别的自然人在识别分析语境中表现为直接标识符和间接标识符。按照《个人信息保护法》第4条的字面理解,“有关”就是这些指向信息、描述信息,或者归属于或以某主体为主题的信息;而是否有关则需要判断、分析和确定。这里的问题在于如何将散布于社会中的信息收集和归纳起来,形成关于某个人的信息(在计算环境中称为文档、数据集)。在万物互联的泛在网络环境下,个人识别分析基本上依赖计算机和网络产生的数据,而网络环境下的数据均有记录可查,均可追溯来源,均可归属于某网址或某设备。不管是人利用计算机或网络工具生成的,或在网络互动通信中形成的,还是服务器、传感器、智能设备记录的人们从事的行为或事实,都具备这些特点。这样,“有关”的判断就是基于网络数据可溯源性、可链接性来实现的,是通过数据的结合分析来实现的。

一般而言,一个信息可链接的信息越多,个性识别的机会就越大、准确度就越高。而一旦与标识符结合起来,就成为可以识别特定主体的信息。计算机科学家加芬克尔(Simson L. Garfinkel)以可链接(linkable)来定义个人信息:个人信息是“与特定个人的信息存在逻辑联系的信息”。[20] 另一位计算机领域的乔丹教授(Scott Jordan)则增加了“合理”标准,将个人信息定义为“与关联特定个人或家庭有关的信息有合理可能的逻辑联系的信息”。[21] 该定义虽然并未得到普遍采纳或适用,但是它揭示了信息上的个人可识别性源自于信息与信息的逻辑联系,只是所联系的信息与个人是否关联或关联程度存在不同。由此可见,“识别”包括信息本身可识别和识别分析后可识别,这样广义的识别约等于“有关”,其基本要求是信息具有可链接性。该结论对于分析匿名化和去标识化具有重要指导意义。

三 信息无法识别:《个人信息保护法》的匿名化和去标识化规范检讨

为实现个人信息利用与信息主体权益的平衡,《个人信息保护法》使用了匿名化和去标识化两种处理技术,并为其赋予两种不同法律效果:匿名化信息彻底转化为非个人信息,而去标识化只是个人信息的合规和安全措施。这样规定背后的逻辑,需要检讨和分析。

(一)“无法识别”的信息与匿名化和去标识化技术

根据前述识别理论,信息的个人识别性存在差异。我们需要探讨匿名化和去标识化如何实现“无法识别”。首先需要指出的是,这两种处理措施都是针对特定数据集而言的,主要看处理后的数据集的信息是否还具有识别个人的能力。就此而言,是否能够识别个人不能只看信息本身的识别性,而需要考察这些信息间是否因可链接而具有识别性。对于信息本身是否能够指向个人(标识符)是可以判断的;而对于信息之间是否可链接,在进行结合分析之前并不能直观地判断出来。

于是,要使一个数据集达到“无法识别”个人的效果,就只能集中于对直接标识符的

〔20〕 Simson L. Garfinkel, De-identification of Personal Information, National Institute of Standards and Technology Report NISTIR 8053 (October 2015) at 40, http://dx.doi.org/10.6028/NIST.IR.8053,最近访问时间[2022-01-02]。

〔21〕 Scott Jordan, Aligning Legal Definitions of Personal Information with the Computer Science of Identifiability (July 26, 2021), SSRN: https://ssrn.com/abstract=3893833,最近访问时间[2022-01-02]。

处理。这一点在国际社会对匿名化的定义中能够得到充分验证。国际标准化组织(ISO)《隐私框架标准》(ISO 29100:2011)对匿名化的定义较有代表性。匿名化是"信息主体可识别信息"(PII)以不可逆转的方式改变,使得 PII 控制者不能单独或与其他主体合作直接或间接识别 PII 主体。[22] 欧盟第 29 条工作组对匿名化的定义与国际标准化组织的定义相似。总之,匿名化处理的核心是"处理"PII,使数据集中本来与信息主体关联的信息不再与个人关联。因为匿名化与去标识化在技术上存在相同之处,美国一开始就没有刻意区分匿名化和去标识化,而统一称为去标识化。美国国家标准与技术研究院(NIST)发布的《保护信息主体身份信息(PII)机密性的指南》将"去标识化信息"定义为"删除或隐藏(又称屏蔽或混淆)足够多的 PII 的记录,使得剩余信息无法识别信息主体身份,并且没有合理依据相信该信息可用于识别信息主体。"[23]

之所以匿名化和去标识化在技术角度不存在根本性差异,主要因为标识符是可去除或可处理的,而信息的可链接性是难以消除的,甚至在技术上是不可实现的。因此,匿名化中的"无法识别",能够做的也只是去除直接标识符,而不是消除信息之间的可链接性。[24] 在本文看来,只要数据保持可计算分析的原始状态,就具有识别分析的能力(至少可以进行个性识别分析,是否能够识别出身份则取决于拥有的数据量、场景等)。处理后的无法识别只能是在边界相对固定的数据集中去除直接标识符,达到身份无法识别。

(二)《个人信息保护法》匿名化的法效果检讨

《个人信息保护法》第 4 条明确,匿名化信息属于非个人信息。这意味着匿名化信息属于可自由处理(包括对外提供)的信息,不受《个人信息保护法》调整。第 73 条对匿名化的定义,强调匿名化的无法识别是"不能复原"的。结合前文论述,"无法识别"至少有两种理解,一种理解是,其针对一个数据集中的直接标识符,即不能单纯从信息识别出个人;另一种理解是,数据集中的信息不仅无直接标识符,而且无可链接性,即无法通过结合分析而识别出个人。由于匿名化的信息被置于个人信息之外,因而匿名化的无法识别应理解为后者。[25] 但是,这几乎无法实现。如前所述,信息即使没有标识符,也存在链接更多信息对应到某个体的可能性。在当下,数据或信息的可链接性无限增强,给结合分析提供了无限可能性。深度学习领域的权威沃登(Pete Warden)认为,通过交叉引用不同的信息源,重新识别通常是可能的,匿名化过程是一种错觉。[26] K 匿名和差分隐私(differenti-

〔22〕 ISO/IEC 29100:2011 Information technology - Security techniques - Privacy framework.

〔23〕 National Institute of Standards and Technology, Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) (April 2010), http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf,最近访问时间[2022-01-02]。NIST 的 PII 中不包括 cookie ID 和设备 ID,因此许多 AdTech 公司、广告商和发布商将它们视为非 PII。但是在欧盟《一般数据保护条例》及其所影响下的所有立法中,这些均属于个人数据(个人信息)。

〔24〕 本文将信息与主体之间的可关联性称可识别(identifiable);而将关于某个人的信息之间可关联性称为可链接性(linkable)。在有些文献中不加区分使用二者,泛化了二者区别。在本文作者看来,前者是人们可辨识的,而后者只有在特定技术环境下才能实现。

〔25〕 See Scott Jordan, Aligning Legal Definitions of Personal Information with the Computer Science of Identifiability (July 26, 2021) at 21, SSRN: https://ssrn.com/abstract=3893833,最近访问时间[2022-01-02]。

〔26〕 Pete Warden, Why You Can't Really Anonymize Your Data (17 May 2011), https://www.oreilly.com/content/anonymize-data-limits,最近访问时间[2022-01-02]。

al privacy)被认为是实现匿名化的两项技术,但是有报告认为,经二者处理的信息也存在重新识别风险。[27]

《个人信息保护法》的相关规定也许是受欧盟《一般数据保护条例》影响。该条例立法说明第26条规定:"数据保护的原则不适用于匿名信息,即与已识别或者可识别的自然人无关的信息,或者以某种导致信息主体不可识别或不再可识别的方式匿名提供的信息。"这可以理解为匿名信息不适用个人信息保护法。但第29条工作组认为,匿名化可以为信息主体和整个社会带来"开放数据"的好处,同时减轻相关信息主体的风险。[28] 这说明在其看来,匿名数据的可链接性并未丧失,而只是从中不能辨识出信息主体的身份。如果《一般数据保护条例》对其不适用,那么将使虚假的"匿名化"信息不受个人信息保护法管控。[29]

在《个人信息保护法》第4条已经规定匿名信息并非个人信息情形下,我们只能在解释上寻找合适的出路。"无法识别"只有在知识或规律总结层面(规律判断、抽象信息或知识)才是能够确证的。如果处理后的信息仍然是源自于网络的事实信息或数据,而对后续的利用行为不加管控,那么将匿名化信息视为非个人信息就有脱法之嫌疑,将给信息主体权益带来极大风险。因此,对于是否达到《个人信息保护法》规定效果的匿名化必须作严格解释,即处理者必须证明,匿名化的信息已经丧失了信息间的可链接性。

(三)《个人信息保护法》去标识化的法效果分析

《个人信息保护法》将去标识化定义为"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"。这一定义的核心是"不借助额外信息"的情况下"无法识别特定自然人"。这意味着去标识化承认技术处理的有限性,只有在信息处理者不再获取新的额外信息并利用额外信息进行识别的条件下处于"无法识别"的状态。显然,这是对信息处理者的行为约束,而不是技术约束。也就是说,此处的无法识别是建立在"如果"信息处理者不识别的基础上。在立法者看来,一旦信息处理者做到去标识化后不识别,那么就不会给信息主体权益造成危害,其处理行为也就合规。但是,我国对信息处理者去标识化程度和去标识化后的信息利用并没有明确的规定。关于去标识化的法效果,《个人信息保护法》仅笼统地规定为一种安全措施,即"确保个人信息处理活动符合法律、行政法规的规定"和防止泄露的安全风险,并没有明确可以确保哪些处理活动的合规和安全。这是去标识化应用亟待明确和解释之处。不清楚去标识化后的信息可以作哪些处理,即去标识化真正的法效果,就不能确定去标识化的标准是什么。

国家推荐标准《信息安全技术 个人信息安全规范》(GB/T 35273—2020)对去标识化信息的可处理行为作了明确规定,包括两类场景:一是展示个人信息宜采取去标识化处理

---

〔27〕 See Marc Dautlich et al., Introduction to Anonymisation, https://go.privitar.com/rs/588-MYA-374/images/2021-07-Privitar-Bristows-Intro_to_Anonymisation.pdf,最近访问时间[2022-01-02]。

〔28〕 See Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques (0829/14/EN WP216).

〔29〕 需要指出,欧盟的个人数据采纳的是与"已经识别和可识别的个人"关联标准,因而其范畴远大于美国的可识别个人信息(PII),美国的PII仅相当欧盟法中的"已识别和可识别的个人"数据。因此,在美国法的语境下,去标识数据就可以被认为是非个人数据,但是在欧盟语境下就不能得出这样的结论。正因此,美国统一州法委员会2021年发布的《统一个人数据保护法》就明确规定"个人数据"为通过直接标识符识别或描述数据主体的记录,包括假名化数据(pseudonymized data),不包括去标识数据;参见Section 2(10)。在欧盟立法语境下,将匿名数据排除于个人数据之外是不合逻辑的。这一结论同样适用于我国的《个人信息保护法》。

等措施,降低个人信息在展示环节的泄露风险(7.2);二是个人信息经去标识化处理且确保数据接收方无法重新识别或者关联信息主体的,可以不经信息主体同意进行共享、转让(9.2b)。这两种处理均建立在去标识化的信息仍是个人信息,而仍然可以被用于识别分析个性特征的基础上。也正因此,该规范在 3.15 条将去标识化定义为"通过对个人信息的技术处理,使其在不借助额外信息的情况下,无法识别或者关联信息主体的过程",并以注释方式明确"去标识化建立在个体基础之上,保留了个体颗粒度,采用假名、加密、哈希函数等技术手段替代对个人信息的标识"。

这样的定位应当成为对《个人信息保护法》去标识化规则的合理解释方案。相对于《个人信息保护法》,上述标准的去标识化明确了"去什么"和去标识化信息可作怎样的处理。去标识化去除的是信息与信息主体的关联(标识符),而去标识化的信息仍然可以进行识别分析个性特征和对外提供,只是去除信息与主体关联的风险。正如前文论述匿名化时已经指出的,只有与主体存在直接关联标识符的信息才是可以进行技术处理的,而信息之间的可链接性则不易明辨,无法去除。只要信息存在可链接性,就可以进行结合分析,用于识别分析。因此,去标识化的数据集是保留了信息识别分析价值和去除信息与主体直接关联所引发风险的一种制度安排。

这样的定位与欧美去标识化规范是一致的。美国奉行单一去标识化体制,将去标识化定位为在保护个人信息效用(识别分析)前提下减少个人信息直接关联风险的制度。《个人信息保护法》的去标识化规定与欧盟《一般数据保护条例》第 4 条的假名化相似,只是表述上存在差异。[30] 欧盟 29 条数据保护工作组认为,假名只是对信息主体身份进行伪装,形成的假名信息可被复原,假名化处理仅仅是减少了数据与可识别信息主体之间的关联能力,是一种有用的安全措施。[31]《一般数据保护条例》接受这一观点,将假名化作为保护信息主体权利、降低数据安全风险的一种措施(该条例立法说明第 28 条)。假名化机制下通过使用附加信息可以关联到某个自然人的数据,则属于可识别的自然人的信息(该条例立法说明第 26 条)。假名化被视为信息主体数据合法处理考虑的因素(第 6 条、第 89 条),也是许多场景下数据安全考虑的因素(第 32 条)。只要采取有效的假名化措施,应当允许对经假名化处理的信息主体数据进行一般分析(该条例立法说明第 29 条)。保持去标识化数据的识别分析个性特征功能,就是保留个人信息的识别性,也就是保留个人信息的基本效用。

去标识化作为一项制度的约束条件应当有二:一是标识符管理措施,避免信息处理者复原或再关联;二是信息处理者不得进行再识别。但是,《个人信息保护法》对去标识化的定义太过简单,并不足以支撑去标识化信息的公开展示和对外提供。本质上,去标识化不应当理解为一种技术,而是一种制度措施。本文将之表述为"受控去标识化"制度。

---

〔30〕 欧盟《一般数据保护条例》第 4 条 5 项对假名化作了清晰的定义:"假名化(pseudonymisation)是指以如下方式处理信息主体数据,即:除非使用额外信息,否则无法将信息主体数据归属于某个具体的信息主体,且上述额外信息应当被独立存储并受制于适当的技术和组织措施,以确保信息主体数据不会连结到某个已识别或可识别的自然人。"如果认为我国法上的"识别"限定于身份识别,那么在去除哪些信息方面,我国的去标识化与欧洲的假名化就不存在实质差异。

〔31〕 See Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques (0829/14/EN WP216).

四　受控去标识化:个人信息流通的合法基础

《个人信息保护法》仅将去标识化定义为一种安全措施,而没有明确为对外提供(流通)和公开展示的安全措施。如果要将去标识化打造为确保个人信息流通利用的安全措施,那么还需要匹配相应的制度措施,确保去标识化信息的流通利用合法和安全。

(一)受控去标识化制度的提出

前文已述,信息可识别性是一个连续变量,因而在去除什么标识符或去除到什么程度等问题上一直存在争议。越来越多的学者和专家认识到个人信息不是"非黑即白",无论是美国法语境下的"PII"和"non-PII"还是欧盟的个人数据和非个人数据二分法,都不能反映识别分析的实践。美国著名的两位隐私学者施瓦茨(Paul Schwartz)和索洛夫(Dan Solove)认为,在许多情况下,非 PII 可以与个人联系起来,并且可以重新识别未识别的数据。实际上,个人识别性是风险连续体(continuum of risk)。他们提出信息的个人识别性三分法理论(即 PII 2.0),将信息分为"已识别、可识别或不可识别"。[32] 之后有学者进一步指出应以各种灰色阴影看待数据,提出根据利益和产生的风险对数据进行适当的控制;他们认为,应根据直接标识符、间接标识符(指准标识符)以及访问和使用的保障和控制三个主要变量的相互作用对数据进行分类,为数据的收集、使用和控制制定适当的规则。[33] 最近还有学者提出更复杂的信息识别性分类,认为可以匿名信息、去标识化信息、不可追踪信息、可追踪信息和合理可识别信息的定义,替代个人可识别信息与去标识化信息之间过于简单的区别,以根据每种类型的信息的特征定制告知和同意要求。[34]

所有这些研究均说明,单纯去除标识符并不能去除信息的识别性、达到不能识别的效果。尤其是在万物互联的数字环境下,要消除信息的识别性是非常困难的。因此,"无法识别"不能单纯依靠去标识符实现,需要辅之以"控制识别"(不允许再识别或重新关联)的制度规范才能实现规范目标。在《个人信息保护法》区别规范且将匿名化信息定位于非个人信息的情形下,我们只能通过对去标识化制度进行解释,在保护个人信息权益且保证个人信息效用前提下实现个人信息流通利用的目的。

将去标识化理解为一种制度,既是承认技术有限性的结果,也是实现信息(数据)资源有效利用的制度需要。鉴于数据的可链接性这一事实,我们只能依赖去除或隐藏信息与主体之间的直接关联(标识符)来降低个人信息在利用或遭意外披露时的风险。去除标识符的作用主要体现在两个方面:一是降低侵犯隐私的可能性。当数据集与特定信息主体不存在直接关联时,隐私就不归属于特定信息主体,就不会直接引发隐私侵害。二是

---

〔32〕 Paul Schwartz & Dan Solove, The PII Problem: Privacy and a New Concept of Personally Identifiable Information, 86 *New York University Law Review* 1814, 1877 - 1879 (2011).

〔33〕 See Jules Polonetsky, Omer Tene and Kelsey Finch, Shades of Gray: Seeing the Full Spectrum of Practical Data De-identification, 56 *Santa Clara Law Review* 593, 607 - 608 (2016).

〔34〕 Scott Jordan, Aligning Legal Definitions of Personal Information with the Computer Science of Identifiability (July 26, 2021), SSRN: https://ssrn.com/abstract=3893833,最近访问时间[2022 - 01 - 02]。

减少信息或数据集在泄露或提供给他人时,给信息主体带来的不必要安全风险和社会危害。包含直接标识符信息的公开或流动会使个人信息滥用的风险增大,例如被利用以实施电信诈骗等违法犯罪行为。从这两个方面看,去标识化有助于信息主体权益保护。

但是,识别并不以存在直接标识符为前提;存在间接标识符即已足够。甚至在信息具有普遍链接性特征的条件下,准标识符亦促进了信息间的链接。信息链接之后的识别分析既是数字技术带给社会的红利,也不是单纯依赖技术可以阻止的情形。因此,符合当今现实的做法是通过技术消除直接显性风险,而辅之以控制识别制度,使识别分析以社会可接受的合理方式进行。早在 2012 年,美国联邦贸易委员会(FTC)发布的《在快速变化的时代保护消费者隐私:对企业和政策制定者的建议》即按照这样思路设计去标识化制度。该建议认为,联邦贸易委员会隐私框架仅适用于与消费者"合理关联"的数据,并认为公司采取以下措施,即满足数据不构成"合理关联":(1)采取合理措施确保数据去标识化;(2)公开承诺不尝试重新识别数据;(3)合同上禁止下游接收者尝试重新识别数据。[35] 至于去标识化的处理方法,一般是删除直接标识符,只是在美国并没有统一清单,而是根据风险来确定。美国标准研究院(NIST)认为直接标识符是"直接识别单个个体的数据",因而仅列举姓名、社会安全号码和电子邮件地址作为示范。[36] 而健康领域的特别立法《健康保险可携性和责任法案》(*Health Insurance Portability and Accountability Act*,"HIPAA")将直接标识符扩展为 18 种。[37] 不过,再全面的直接标识符清单,也只是消除数据集中有限的直接与个人或设备关联的信息,而不排除仍然可以进行识别分析的可能。去标识信息的利用必须运用合同,控制去标识化个人信息之处理者的使用行为。

因此,去标识化技术结合控制识别的后续措施,构成受控去标识化。这种制度设计确保了数据流通利用的安全,实现了信息主体权益与信息利用的平衡。要使《个人信息保护法》的去标识化承担去标识化信息的发布和分享利用的合规和安全保障制度,就必须引入受控去标识化制度作为我国个人信息流通利用的安全保障措施。

(二)作为防控个人信息流通利用风险的受控去标识化制度

在我国明确提出大数据战略、发展数字经济的今天,数据被作为生产要素,包含个人信息在内的数据资源的社会利用成为数据要素市场建设的必然要求。《个人信息保护法》本应考虑国家数字经济发展的制度诉求,为个人信息的流通利用开辟通道,但遗憾地留下了制度空白。在现阶段,我们只能依赖合理地解释该法第 51 条上的去标识化,在丰满和细化其制度条件的同时,将去标识化制度作为个人信息流通利用的制度基础。这意

〔35〕 Federal Trade Commission, Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers (March 2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf, 最近访问时间[2022-01-02]。

〔36〕 Simson L. Garfinkel, De-identification of Personal Information, National Institute of Standards and Technology Report NISTIR 8053 (October 2015), http://dx.doi.org/10.6028/NIST.IR.8053,最近访问时间[2022-01-02]。

〔37〕 See US Department of Health & Human, Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard,最近访问时间[2022-01-02]。

味着要将去标识化明确为去除数据集中的直接标识符，在防范直接关联风险的前提下，实现个人信息合理利用（包括流通利用）制度，将去标识化作为合规和安全地实现个人信息价值的制度。为实现这样的目标，去标识化制度尚需要补齐以下制度短板。

首先，从技术的角度来看，应对去标识化的范围作出指引性规范。由于去标识化本身就是因行业和场景不同而存在较大差异，因而应当鼓励行业主管部门、行业协会等发展出适合本行业的去标识化清单。该清单一定是基于风险考量的产物，而不是逻辑推演的适用于一切情形的“大一统”指引。

其次，建立去标识化信息的使用限制，形成去标识化信息应用的内部约束。我国的相关立法仅仅规定了不获取额外信息情况下的无法识别，但没有对标识符的安全管理作出要求。缺失对去除的标识符的安全管理措施，信息处理者就可以很方便地重新关联，识别信息对应的身份。例如，欧盟《一般数据保护条例》第 4 条所规定的假名化包含相应管理措施，为假名化设置了前提条件，从数据中删除直接标识符，将这些标识符与数据分开，并使它们受技术和组织控制。在实践中，这可能意味着：（1）标记或加密直接标识符，实质上是将其从数据中删除；（2）将令牌映射或加密密钥与数据分开保存；（3）保护令牌值或加密密钥，这需要技术措施（例如网络安全控制）和组织措施（例如禁止未经授权使用令牌库或加密密钥的政策或标准操作程序）。

最后，在去标识化信息分享或流通利用时，应当对接收和使用数据的处理行为进行控制。在对外提供数据时，宜由第三方为数据提供者与接受者提供去标识化服务，以便形成双方均不可能随意复原或关联的数据匹配和计算分析的安全环境，使所有处理过程可溯源、可审计，使去标识化及其程度可证明。同时，数据提供者要通过协议明确去标识化信息的使用条件，并约束接受者在一定范围内、在非直接关联信息主体的条件下开展识别分析。如果接受者重新识别信息主体、针对信息主体作出决策，应当向去标识化服务的第三方提供其使用行为遵守《个人信息保护法》、具有正当合法性基础的证明（这往往意味着获得信息主体的同意）。这样就使得去标识化个人信息得以在遵循《个人信息保护法》的框架下被利用。因此，本文建议法律能够确立以下制度，即信息处理者对所控制的数据集进行去标识化处理并采取防范接受者重新识别的使用措施的，可以向他人提供数据。

上述三方面构成完整的受控去标识化制度。显然，这超出了法律解释和适用的范围，需要负责个人信息保护的主管部门推进行业探索、创制符合个人信息保护法精神的实施细则才能实现。控制去标识化信息的识别应用是受控去标识化制度有效运行的核心，需要有可操作、可检验的一套制度规则。《数据去标识化共享指南》地方标准率先提出“商业驱动、技术支撑和法律保障”三位一体的数据共享框架，通过“规范共享数据内容、控制过程的安全有序，约束数据的有限使用”来保障去标识化数据的流通利用合规和安全，这也许可以作为一种解决方案。〔38〕 在域外也有行业专家提出“受控可链接数据”（controlled linkable data），以在遵守欧盟《一般数据保护条例》、增强个人数据主体隐私保护的前提下，实现数据使用和数据价值的释放。受控制去标识化的核心是构建更现实的去标识化

〔38〕 2021 年 7 月 27 日，上海市市场监管局发布了地方标准《数据去标识化共享指南》（DB31/T 1311－2021）。

形式,通过技术和组织措施,在数据的整个生命周期内保护数据主体权利。[39] 这些说明了本文提出的受控去标识化并不是纯理论的构想,而是可操作和可实现的制度规则。

五 结 论

信息识别性是一个连续的变量,可以从中区分出直接与个人关联的直接标识符、直接与设备关联的间接标识符、可结合分析关联到个人属性的信息(准标识符)。具有识别性的信息在识别分析中被称为标识符,其中个人可控制和第三人可识别的限于直接标识符;间接标识符不是个人可直接控制的数据,也不是第三人可以直接判断关联到特定个人的数据,其本身是匿名的。其他用于识别分析的信息属于非识别性信息,仅因为与可识别性信息的关联而与个人关联,成为个人信息。对于此类信息个人不能事先控制,第三人也不能够清晰判断,只是被用于特定场景和特定目的的分析。[40]

《个人信息保护法》应当根据不同类型信息在识别分析中的作用和个人控制或预防不当识别的可能性,配置处理前的同意和处理中的控制性权利。直接标识符是任何行业或领域去标识化须清除的,事前的同意亦应仅限于直接标识符。对于间接标识符,可能的路径是允许使用,但是允许个人拒绝(区别于处理前的同意),同时以法律或行业自律规范限定其用途或方式。作为间接标识符的设备识别标识或数字 ID,是保持识别分析不直接识别出具体个人又能够实现个人信息社会效用的工具。其是否进入去标识的清单,取决于其触及或识别出个人的效果。对于其他非识别性信息,则只能通过对识别分析行为的控制,预防侵害结果的发生并给个人以救济保护。

在《个人信息保护法》将匿名化信息定义成为非个人信息的情形下,为防止脱法行为,只能对是否达到《个人信息保护法》规定效果的匿名化作严格解释;同时,将去标识化打造为与国际社会接轨的、确保个人信息流通利用安全的法律制度,而不只是将其作为一种技术或一种内部安全措施。从技术的角度来看,信息间的可链接性是不可能去除的。要对基于可链接信息的结合识别进行规制,本质上依赖于对识别行为的控制,而不是对信息的技术处理。这种识别控制并不是要绝对禁止接受者进行识别分析,而是要禁止通过识别分析识别出身份。如果进行识别信息主体身份的使用,那么就必须接受是否符合《个人信息保护法》合法性基础和是否保障个人信息权益的审查。应在《个人信息保护法》实施中采取“去标识符 + 识别控制”的受控去标识化制度,允许各行业发展出适合各自风险的受控去标识化机制,并探索出符合个人信息保护法精神、确保个人信息流通利用的安全保障措施。确立受控制去标识化制度、打造可信去标识化信息流通环境,才能真正在保护个人信息效用的前提下,在信息主体权益侵害可控范围内实现个人信息的社会化利用。

〔39〕 Mike Hintze and Gary LaFever, Meeting Upcoming GDPR Requirements While Maximizing the Full Value of Data Analytics (January 2017), SSRN: https://ssrn.com/abstract=2927540,最近访问时间[2022-01-02]。Anonos 公司将受控可链接数据开发成为一种合规解决方案(命名为 BigPrivacy),对外提供服务。

〔40〕 参见李群涛、高富平:《信息主体同意的适用边界》,《财经法学》2022 年第 1 期,第 7-8 页。

［**Abstract**］ The Personal Information Protection Law (PIPL) establishes a system of direct utilization of personal information under the control of individuals, but it is questionable whether it provides a channel for the sharing of personal information. Information has different identifiers: direct identifiers are directly linked to the identity of the information subject, indirect identifiers can identify individuals but are not directly linked to their identity, and quasi-identifiers can profile individuals by combining two or more linkable information. The three types of identifiers bring different risks of harm to the rights and interests of individuals. Both anonymization and de-identification provided for in the PIPL are essentially institutional arrangements against the personal identifiable information risk within a particular dataset, which can eliminate the risk arising from the identification of the information itself, but not that arising from profiling. Therefore, in the absence of measures for addressing the risk from profiling, neither the anonymization nor the de-identification under the current PIPL can support the sharing of personal information. Now that the PIPL defines anonymized information as non-personal information, to prevent deregulation, it is necessary to interpret stringently as to whether anonymization meets the requirements of the PIPL, and to make de-identification a legal system that is in line with the requirement of the international community and can ensure the safe sharing and use of personal information. The PIPL should arrange for the consent before processing and the controlling rights during processing according to the role of different types of information in the identification analysis and the possibility for individuals to control or prevent improper identification. What must be removed for de-identification is the direct identifiers, and the consent before processing should also be limited to direct identifiers. For indirect identifiers, a possible path is to allow their use, but at the same time allow individuals to refuse their use (as opposed to consent before processing), and limit their use or the manner of their use by law or industry self-regulation. For other non-identifiable information, the only way to prevent infringement and provide remedies for individuals is to control the identifying and analyzing conduct. This identification control is not intended to absolutely prohibit the processor from performing profiling, but rather to prohibit identification through profiling. In short, de-identification needs to be reconstituted into a controlled de-identification system of "de-direct identifier + identification control" to allow each industry to develop a controlled de-identification mechanism suitable for its own risk and explore security measures for ensuring the sharing and use of personal information in accordance with the spirit of the personal protection law. Only in this way can we, on the premise of controlling the personal identifiable risk, provide an institutional foundation for the sharing of personal information and maximize the social utilities of personal information .

（责任编辑:余佳楠）