重庆大学学报(社会科学版) 2023年第29卷第2期
JOURNAL OF CHONGQING UNIVERSITY(Social Science Edition) Vol. 29 No. 2 2023

Doi:10.11835/j.issn.1008-5831.fx.2021.09.003

欢迎按以下格式引用: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(2):220-231. Doi:10.11835/j.issn.1008-5831.fx.2021.09.003.

Citation Format: ZHANG Li, XU Duoqi. On the legal regulation of personal information anonymization in China under the risk control concept [J]. Journal of Chongqing University (Social Science Edition),2023(2):220-231. Doi:10.11835/ j.issn.1008-5831.fx.2021.09.003.

风险控制理念下我国个人信息匿名化处理的法律规制

张 丽[1],许多奇[2]

(1.上海交通大学 凯原法学院,上海 200030;2.复旦大学 法学院,上海 200438)

摘要:数据匿名化为数据的流通和共享提供了重要的技术助力,技术的易变性同时也对数据匿名化处理的法律规制带来诸多障碍。在当前的技术背景下,我国个人信息①匿名化面临身份识别标准不确定、身份再识别可逆转等风险,个人信息匿名化处理过程中所产生的技术风险给个人隐私利益带来极大挑战。目前以结果导向为目标的规制手段在个人信息匿名化规制方面缺乏灵活性,难以缓释技术带来的不确定性风险。平衡好个人隐私利益、企业经济利益以及社会公共利益之间的冲突是个人信息匿名化处理的终极目标。风险控制导向理念的优位在理念层面上摆脱了数据保护的现实诉求与理想价值判断之间的束缚,替代结果导向理念,为信息处理者内源性的数据合规与自律监管、信息主体外向性的权利保护与数据使用提供了较为有效的弥合思维进路,为个人信息匿名化处理的法律规制提供新的可能。以风险控制理念为核心的个人信息匿名化法律规制架构以实现数据的有效性与实用性之间的动态平衡为目标,围绕降低个人信息匿名化处理过程中的风险进行法律机制设计,通过课以信息处理者相应的信息处理风险评估义务实现对个人信息匿名化规制的良善治理。在无规范性文件作理据支撑的前提下,个人信息匿名化的风险评估及评估标准于实践中难以稳定量化,故而在个人信息匿名化法律规制方面,应当提倡将保障个人信息主体权利作为个人信息匿名化处理的核心,通过赋予信息主体在信息处理过程中相应的数据权利以实现信息主体自身的权利。个人信息匿名化以个人信息可识别为前提,在个人信息概念界定上应当依据具体场景加以个案判断,因此,对于匿名数据的认定也应当采取动态场景化的方式进行理解。在个人信息匿名化处理风险控制手段的实现上,通过确立相关隐私风险评估机制为信息处理者提供明确的数据利用指引,规范信息处理者的行为。

关键词:个人信息;匿名处理;风险评估;风险控制

中图分类号:D923;D922.16 **文献标志码**:A **文章编号**:1008-5831(2023)02-0220-12

基金项目:国家社会科学基金项目"我国互联网金融市场准入与监管法制重大理论与实践问题研究"(16BFX098)

作者简介:张丽,上海交通大学凯原法学院,Email:zhanglily0419@163.com。

①本文中,笔者将"个人信息"与"个人数据"作同一概念理解,不作具体区分。

一、问题之提出

在全球发达国家对应法域较之以前更加倡导数据开放、数据共享的浪潮下,全球范围内的多层次、多类型主体对于个人信息匿名化的需求亦与日俱增。匿名化技术在于切断原始数据集中数据所有者和敏感信息之间的一一对应关系,产生既满足隐私保护需求又保证数据可用的匿名数据集[1]。从原理上看,该等技术有效解决了数据利用过程中所带来的隐私泄露问题,极大促进了数据的自由流通,提升了数据共享和利用效率。换言之,匿名化技术已成为数据利用环节的重要前提和保障。与此同时,由于技术的负外部性,匿名化技术也面临明显的现实困境。首先,数据的形式多样,作用各异,很难运用统一的标准达到完全的匿名化[2]。其次,匿名化处理的再识别对个人信息匿名化法律标准的确立提出了挑战:匿名化与再识别技术的天然对抗性使匿名化处理面临从传统的完全匿名化困境到目前的可逆转风险的嬗变,无疑扩大了匿名化处理过程中的隐私风险。技术的易变性增加了个人信息匿名化法律规制的不确定性,如何确立个人信息匿名化的法律标准,以及如何规制成为亟待解决的现实问题。

通过梳理现有文献可以看出,相关学者的理论著述均不同程度吸收和借鉴了风险控制理念。有学者从如何实现真正的匿名化入手,提出构建事前、事中、事后的风险评估机制[3]。有学者从匿名化的实现方式上切入,建议进行功能性匿名化,并将比例原则引入个人信息匿名化的法律标准重塑中[4]。还有学者从匿名化的再识别风险出发,提出在事前、事中、事后分阶段构建基于再识别风险的匿名信息分级披露制[5]。通过梳理前期研究成果可知,学者主要是从风险评估及技术视角探讨如何实现信息处理者在个人信息匿名化处理时的合规行为,而从法律治理视角对个人信息匿名化的研究则有待进一步探讨。本文试从风险控制理念的视角切入,通过对我国个人信息匿名化规则的缺陷进行检视,在此基础上提出优化个人信息匿名化规则的方案,以期为我国个人信息匿名化的法律规制提供有益借鉴。

二、个人信息匿名化处理之技术风险的法律透视

信息技术发展为社会经济发展带来便利的同时,也带来极大的隐私侵权风险。法律长期发挥着规范秩序和控制风险的重要作用,当人工智能等信息科技带来诸多社会风险时,法律介入规制应当确立必要的风险社会理念,进行有效的风险控制[6]。数据保护自一开始就是一种风险监管机制[7],特别是针对技术所引发的隐私风险进行规制。以信息技术为代表的数据挖掘、数据分析等技术为社会经济发展带来诸多数据红利,各国也均纷纷致力于探究如何有效平衡数据隐私保护与数据利用之间的冲突。在此方面,匿名化技术通过元组泛化、抑制等数据处理的 K 系列匿名方案实现数据的匿名化,从而保证数据集发布的隐私安全[8],是信息处理者排除适用个人信息保护法规制的重要手段,是降低数据隐私侵权风险的技术保障。

匿名化技术主要包括两个不同的目标,主要目标在于对信息进行实际的身份识别,第二个目标是降低数据的敏感属性,换言之,匿名化减少了将信息与特定数据主体相关联的能力。在一定程度

上,隐含的风险特性涉及身份信息和敏感属性,假设可行,匿名化可以减少隐私风险[9]。然而,在信息技术不断更新迭代的当下,匿名化处理所固有的技术风险以及匿名数据法律标准的不确定性使以降低隐私风险为目标的匿名化仍面临挑战。

(一)身份识别标准的不确定性

如前文所述,匿名化首要目标在于对信息进行实际的身份识别。所谓"识别"是指个人信息与信息主体存在某一客观确定的可能性,简单说是通过这些个人信息能够把信息主体直接或间接"认出来"[10]。根据定义,匿名数据是"与识别或可识别自然人无关的信息或以数据主体不能或不再可识别的方式匿名提供的个人信息"②。由此可知,匿名数据的界定是以明确的个人信息边界为前提。在个人信息界定方面,我国现行立法对个人信息的认定标准采用身份识别标准(identification),此种标准也是目前世界范围内占据主流地位的认定标准之一[11]。根据欧盟《一般数据保护条例》(也称《通用数据保护条例》,系指General Data Protection Regulation,缩写为"GDPR")第4条规定,"个人数据是指与已识别或可识别的自然人(数据主体)相关的任何数据"③。对于"已识别"而言,通常是指可以通过直观的判断而无需借助任何技术或价值判断即可直接识别特定数据主体,如通过自然人的姓名、出生日期、身份证件号等数据可以直接识别特定数据主体。反之,对于"可识别"而言,也即所谓的间接识别,则往往需要借助特定技术手段甚至需要通过价值判断来实现,如位置信息、消费者行为信息甚至是网页浏览记录等信息。

对于可识别的判断标准,不少国家或地区的法律实践对其界定时,事实上采取的是一种预测判断,并且这种预测判断需要合理和实践可行[11]。即便在特定场景中,一笔信息(集)能否识别或关联特定个人,仍无法作"是或非"的二元化界定,只能作"程度化"考量,即评估构成个人信息的"可能性"[12],如欧盟采用的判断标准为"所有合理可能的方法"④。在是否构成个人信息方面,欧盟虽然强调要结合上下文语境进行判断,但"合理可能"本身就隐含了大量的技术判断及价值判断。同一条数据在某一方手中可能是个人数据,但是在另一方手中就不是个人数据[13]。因此,在个人信息界定上存在诸多不确定性。

此外,身份识别的标准也会随着信息处理者(GDPR中表述为数据控制者)的识别技术水平甚至是被识别主体对于隐私被侵犯的接受程度因人因情景而异。身份识别判断标准的不确定直接导致个人信息与非个人信息之间的界限模糊。更为甚者,通过信息技术,几乎所有数据都可纳入个人信息范畴,这会直接导致身份识别标准在界定个人信息方面的作用失效。数据流通、共享是以排除适用个人信息保护法为前提,而身份识别法律标准的模糊性以及不确定性无法为信息处理者提供明确的行为指引,在个人信息匿名化处理中信息处理者很难找到可参照的行为标准来对信息处理行为加以约束,这无疑会增加个人信息匿名化处理中的隐私侵权风险。

---

②GDPR序言第26条。

③GDPR第4条第1款。

④GDPR序言第26条已明确:为确定自然人是否具有可识别性,必须考虑可能使用的所有方法,例如数据控制者或其他任何人为直接或间接地识别自然人而采取的筛选方法。为确定某一方法是否可合理地用于识别自然人,必须考虑所有客观因素,例如识别所需的成本和时间,还需考虑处理和技术开发过程中可用的技术。

(二)身份再识别的可逆转性

匿名化的另一目标在于降低数据的敏感属性。从技术角度来看,其实现方式主要通过从数据中永久和完全删除个人标识符,如将个人身份信息转换为汇总数据。而匿名数据则是不能再以任何方式与个人关联的数据,一旦剥离了此数据中的个人识别元素,这些元素就永远无法与数据或底层个人重新关联。对此,欧盟第 29 小组特别强调匿名化的处理必须是“不可逆转的”(irreversible)[14]。实践中,在大数据技术的作用下,完全且彻底的匿名化已不再可能。有研究表明,特定个人被重新识别的可能性很高,即使匿名数据集严重不完整,也可以保证其重新识别的准确性,如在任何数据集中使用 15 个受众特征都会正确地重新识别 99.98%的美国人[15]。身份再识别带来的匿名数据可逆转的风险,给以降低隐私风险为目的的匿名化带来威胁。信息处理者使用、共享数据的合法性源于所用数据已切断与信息主体之间的可识别性,意味着信息处理者对数据安全保护义务的完成。然而在再识别技术下,该识别性再次被重新连结,攻击者可以通过收集的辅助信息来实现去匿名化,一方面加重了信息处理者的责任负担,另一方面直接导致隐私侵权风险加大。

三、风险控制理念对个人信息匿名化处理的回应

(一)风险控制目标:平衡数据的有效性与实用性

匿名化作为平衡个人信息隐私保护与利用之间冲突的技术手段,在一定程度上缓解了二者之间的矛盾冲突。同时,对于个人信息匿名化技术本身而言,在实现二者之间的平衡目标时,也面临数据有效性与实用性之间的冲突。具体而言,在匿名化的实现方式上,主要依靠去除能够识别信息主体标识符的方式达到隐私保护的目的。然而,单纯将原始数据中能够标识数据主体的标识符去除的方法并不能有效实现匿名保护,实现匿名保护通常要对数据在准标识符上的属性值作概化处理(generalization)方能实现数据的匿名化,而此举往往会在很大程度上降低数据的精确性继而导致降低共享数据的可用性⑤。笔者认为,过度的匿名化虽有利于数据隐私保护,但该保护以牺牲数据的利用价值为前提,有悖于个人信息匿名化提升数据利用价值的初衷。

从技术角度看,匿名化和去匿名化属于两个对立的概念,但二者之间的界限也并非非黑即白。实践中,通过一定的技术手段可以基本实现完全的匿名化,将匿名化所产生的隐私风险降至最低甚至消除;对于去匿名化而言,尽管已有诸多实例证明已经匿名化的数据有被再识别的风险,对此有学者提出,已匿名化的数据被重新识别的风险主要源于不良的匿名化,并已试图从技术角度提出如何实现充分的匿名化[16]。二者之间的界限虽然在很大程度上能够从技术角度加以区分,但考虑到数据的有效性和实用性,追求完全的匿名化或许并不具备特别积极的现实意义。而若默许非完全匿名化,抑或意味着为再识别留下一定空间,加大了再识别信息主体的隐私风险。笔者认为,之所以二者之间形成对立,在于研究主体对二者的判断是基于结果导向方法所致,并未充分考虑二者之

⑤所谓概化处理,即用较为抽象概括的属性值来代替原本具体的属性值。参见:王智慧、许俭、汪卫、施伯乐《一种基于聚类的数据匿名方法》,(《软件学报》,2010 年第 4 期 680-693 页)。

间存在的数据实用性空间。为有效平衡数据有效性和实用性之间的冲突,应当将重点转向关注降低匿名化处理过程中的风险。

(二)风险控制手段:从结果导向到风险控制

信息科技的发展使掌握核心技术的信息处理者与信息主体之间的地位出现严重失衡,私权力地位不断上升,信息主体弱势地位愈发凸显。越来越多的研究表明,以信息主体"知情—同意"为数据保护核心原则的传统在实践运用中受阻⑥。由于前述双方在技术理解、运用上存在明显的"知识沟壑",信息处理者掌握绝对的话语权,大量个人信息由信息处理者掌控,信息主体仅仅依靠日常经验及其对信息技术的一般理解,容易丧失对其自身数据掌控的能力。"知情—同意"制度进而可能成为纸上谈兵,信息主体合法权益难以获得有效保障。信息处理者责任承担的触发往往建立在信息主体合法权益已遭受侵权的基础上。然而,大数据时代的隐私侵权行为方式变得更加隐秘,性质更加模糊,后果呈现形式多样且损害程度更加严重,行为与结果之间的因果关系更加松散,致使信息主体在维权方面面临极大的障碍[17]。不仅如此,当前一些业内领先的大型企业已经比政府掌握了更多的公民信息,相当一部分公权力部门也不得不依赖它们,久而久之,在未建立明确约束机制而存在长期勾稽互通的语境下,前述依赖将模糊公权力与私权力之间的边界,使本来应当由政府监管的对象成为政府的合作伙伴乃至实际控制者[18]。信息处理者的"数据权力"缺乏制衡将导致信息主体只能被动承受数据被分析、使用甚至泄露等一系列后果而无力反抗[19]。对于个人信息匿名化而言,现实的诉求往往聚焦于追求匿名化的最终目标,以匿名化的结果是否实现来判断信息处理者是否达到保障隐私安全的义务。然而,如上文所述,匿名化的结果充满诸多不确定性,以结果导向的规制方式使匿名化处理过程中的隐私风险未能进行有效及时准确地识别,在侵权救济方面也面临极大障碍,无法有效保障信息主体的合法利益。有鉴于此,笔者认为,与其在价值判断与现实诉求上趋于理想化地专注于匿名化的最终目标,不如围绕降低风险的必要流程对法律进行机制设计,关注重新识别和敏感属性公开的规范方式[20],转变个人信息匿名化的规制方式,从结果导向转变为风险控制。

现代社会之所以强化个人信息保护,原因在于个人不易对个人信息流通中的风险进行有效管理。面对越来越复杂的信息收集方式和信息的不规范流转,个人也很难对相关风险加以判断和防范[21]。风险控制理念以风险监管为核心。不同于结果导向的事后救济,风险控制的方法更加强调技术对个人信息侵害的潜在的以及未知的影响,是基于事先的防范措施而不是以危害结果为导向的规制方式。事先防范措施意味着在保障信息主体合法权益方面更加强调信息处理者的义务及责任,将风险置于一定的可控范围内。风险控制的逻辑起点在于将风险分析工具嵌入信息处理全流程,其目的是评估每个处理操作的利弊,并以此为基础管理风险[22]。具言之,风险导向的个人信息保护方法将规制重点转向以信息处理者处理数据行为规范为中心,通过划分不同风险级别对信息处理者相应的义务及责任作出类型化规定,使信息处理者在个人信息处理活动中可形成自律监管

⑥知情同意所面临的困境学术界已存在诸多讨论,观点详情可参见: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3期84-101页);万方《隐私政策中的告知同意原则及其异化》,(《法律科学(西北政法大学学报)》,2019年第2期61-68页)。

的模式,从而达到保障信息主体合法权益的目的。

(三)风险控制结果:课以信息处理者相应的义务

风险评估是实现风险控制理念的核心。个人信息匿名化处理过程中面临身份识别标准不确定以及身份再识别的风险,两种风险贯穿个人信息匿名化处理的全流程。信息处理者作为风险评估的义务主体,应当承担降低隐私侵权风险的责任。评估匿名化处理过程中的风险是信息处理者的应有义务,通过信息处理者自律监管模式的数据评估,能够提升信息主体对信息处理者合法合理利用数据的信心,对于促进数据产业发展大有裨益。在数据处理中,赋予信息处理者一定的自由裁量权,由风险管理人员运用科学方法,对个人信息匿名化有关的风险进行系统分析与研究,确定各项风险的频度和强度,为选择适当的风险处理方法提供依据[23]。《中华人民共和国数据安全法》(以下简称《数据安全法》)提出,国家建立集中统一、高效权威的数据安全评估、报告、信息共享、监测预警机制,加强数据安全风险信息的获取、分析、研判、预警工作⑦。与《数据安全法》相呼应,笔者认为,通过风险评估并根据风险等级确定数据匿名化的程度,进一步依据风险程度的高低课以数据控制者相应的义务,这一风险导向的制度逻辑闭环比结果导向的制度逻辑闭环更具实践性。

四、风险控制理念下个人信息匿名化的制度因应

数据开放与共享是当前世界各国制定大数据发展战略重点关注的问题,《促进大数据发展行动纲要》明确指出当下我国数据市场面临数据开放不足的问题。在法律规范层面,《中华人民共和国个人信息保护法》(以下简称《个保法》)以及《中华人民共和国网络安全法》(以下简称《网络安全法》)第42条的但书条款构成了我国当前法律层面关于个人信息匿名化处理的主要规定⑧。以上规定作为基本法的规范,在文字表述上虽然相对抽象、笼统,但从原则与观念上,依然确立了我国个人信息匿名化处理不可识别不可复原的标准。可惜的是,由于具体法律标准缺失,在具体适用上缺乏具体的法律性规范解释,前述规范已几乎沦为宣誓性条款。此外,《信息安全技术 个人信息安全规范》中对于匿名化、去标识化以及再识别等相应内容也有所提及,可整体看来,同样缺乏对个人信息匿名化的细则说明⑨。在实践中,以上海和贵州为代表的大数据交易中心已经通过开展数据交易的模式来实现数据的流通、共享,并针对数据匿名化形成了一定标准规范⑩,囿于一些外部客观原因,亦多体现为原则性规定,并无针对个人信息匿名化操作的具体技术性、强制性标准规范。

总体来看,我国关于个人信息匿名化的规制碎片化现象突出,规范性文件位阶偏低,高位阶的规范性文件由于无法较好衔接细化规则,容易流于形式或者沦为宣誓性规定,缺乏可操作、可对接

⑦《中华人民共和国数据安全法》第22条。

⑧《中华人民共和国个人信息保护法》第4条规定:个人信息是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,不包括匿名化处理后的信息。《中华人民共和国网络安全法》第42条规定:网络运营者不得泄露、篡改、毁损其收集的个人信息;未经被收集者同意,不得向他人提供个人信息。但是,经过处理无法识别特定个人且不能复原的除外。

⑨中华人民共和国国家标准《信息安全技术 个人信息安全规范》GB/T35273-2020。

⑩以上海大数据交易中心《流通数据处理准则》为例,其明确可直接识别特定个人身份的标识与其他个人数据分别存管和处理的隔离原则,并进一步明确在任何情形下均不得擅自公开、向第三人提供带有身份标识个人数据的禁止公开原则等。

的具体机制。鉴于数据开放、共享日益重要,在立法层面亟待对个人信息的匿名化进行规范。个人信息匿名化本质上属于技术范畴,对于技术治理需要通过一定的程序或者机制让不同的利益相关者参与到相应的技术过程之中,充分考虑权利、资源和利益分配等问题,以实现解决冲突和理性决策的目标[24]。考虑到基于风险控制的个人信息匿名化仍具有诸多不确定性,在具体规制上要以保障信息主体权利为核心、以限制信息处理者权力为目的、赋予监管主体权力为手段的制度规范。

(一)以个人信息权利保护为核心

风险管理不只是一种技术分析方法,它还体现了重要的价值观和理想,尤其在问责制和责任层面[25]。从上文来看,基于风险的规制方法相较传统个人信息保护规制方法来说具有一定的灵活性,有效适应了大数据时代数据规制的需求。但其具体适用效果以及规制防范是否降低甚至排除适用以个人信息人权保障为核心的保护方式,存在较大争议:不同于其他领域的风险评估可以通过具体量化的方式进行,隐私风险评估不仅仅涉及技术规范,还隐含大量价值判断,很难通过量化的标准对其进行评估。目前,以风险控制为核心的欧洲数据保护系统亦欠缺统一的框架来衡量和评估已识别的隐私风险。虽然欧盟《一般数据保护条例》在《数据保护指令》的基础上更加明确了相关风险的具体类型,但仍然以灵活抽象的方式定义风险,并且需要由数据控制者根据每个数据处理案例的特殊性来指定和评估[26]。对此,有学者提出评估的决定权掌握在数据控制者手中是否会加强数据控制者的权力[27]。这种新兴的执法范式主要依靠数据控制者的自律监管,可能加剧破坏数据保护机构的作用,还可能会进一步减弱其有效数据保护的能力以及有损数据主体的基本权利[28]。还有学者提出,基于风险的规制方法与基于权利保护的方法相悖,基于权利的保护方法更加强调公平,将个人数据保护的范围公正平等地覆盖到每个数据主体,而基于风险的规制方法是有选择的,显然在保障公平性上有所欠缺[29]。

笔者认为,基于风险的规制方法虽然给予信息处理者在处理数据上评估风险的空间,但并非意味着欧盟摒弃了以个人信息人权保障为核心的保护方式。欧盟第29条工作小组(WP29)一直支持在欧盟数据保护法律框架中纳入基于风险的方法,同时工作小组也指出基于风险的方法并非是替代已确立的数据保护权利和原则,其实质上是一种数据控制者处理数据的合规方法[30]。进一步而言,信息处理者仍需以保障信息主体的人格利益为目标,只是在规制方法上采用更灵活的基于风险的监管方式,其目的仍是为保障信息主体的合法权益,对于匿名化而言,其本身的数据处理行为仍需要受到个人信息保护基本原则的限制,如相关机制的构建并未豁免知情同意原则、数据最小化原则以及目的限制原则等。在个人信息主体权益保护方面,我国《个保法》更是专章对个人信息主体在个人信息处理活动中享有的权利进行了规定,形成了相对全面的个人信息权利体系。当然,个人信息主体权利的实现有赖于个人信息处理者履行相应的行为义务,对信息主体权利的实现提供全面保障[31]。需要指出的是,风险控制导向理念的优位,是在理念层面上摆脱数据保护现实诉求与理想价值判断的束缚,替代结果导向理念,以此在理解技术发展与技术壁垒、平衡数据保护与数据共享的同时,从强化可行性的实践视阈,为信息处理者内源性的数据合规与自律监管、信息主体外向性的数据使用与权利保护提供较为有效的机制弥合的思维进路,而任何导向及语境下的思维进路

均未以规范原则与规范手段完全替代的实践模式作为前述机制弥合的优化模式,亦即风险控制导向理念侧重于为解释既有立法的执行方向与未来立法的规范方向提供框架性理据支撑。

(二)完善个人信息分类保护制度

个人信息匿名化建立于个人信息可识别的基础上,系个人信息分类保护的大前提。值得注意的是,在个人信息界定上,各国将"场景"理念嵌入个人信息界定中,即对个人信息的界定需要依据具体的场景加以个案判断。同样,对于匿名数据的认定也采取动态场景化的方式进行理解[32]。动态场景化的界定方式从个案判断出发,根据数据所处上下文语境进行判断,能够对是否构成个人信息进行客观评价,但动态场景化的界定方式仍无法突破信息技术判断标准因人而异所带来的差异。场景一词作为研究中的变量往往千差万别,去匿名化过程中所存在的风险更是因场景不同而所有差异[33]。学者 Paul Ohm 提出匿名化是"破碎的隐私承诺",提议取消传统的个人数据与非个人数据的界分[34]。应当看到,取消个人信息与非个人信息的方式并不可取,原因在于对个人信息进行准确厘定是清晰权利与义务的前提,既是信息主体确认其基本权利的起点,亦属于对信息处理者课以相应义务的起点。应当摒弃个人信息与非个人信息的绝对化区分,根据具体场景与制度功能对个人信息的范围予以厘定并加以规制,为应对场景化理论存在的不确定性问题,可以建立模块化的个人信息分类保护制度[35]。

如上文所述,身份识别标准存在诸多不确定性风险,个人信息范围的抽象性和不确定性无法为企业等数据利用者提供明确的行为预期[36]。现有关于个人信息的界定已经无法适应信息技术背景下个人信息隐私保护和利用需求,个人信息分类保护成为当前学界普遍认为合理可行的方式。在具体类型划分上,有学者提出可参考保罗·施瓦茨与丹尼尔·索洛夫所提出的"个人信息 2.0"的概念,将可识别的个人信息分为三类⑪:已识别个人的信息、可识别个人的信息、不可识别的个人信息。还有学者提出根据能否直接识别信息主体、社会性强弱以及是否具有敏感性这三项要素对个人信息进行类型化构建[37]。另有学者提出将个人信息的识别性和相关性进行程度上的区分,即从识别性上可分为已识别信息、可识别信息、匿名信息,从相关性程度上可分为个人敏感信息、个人一般信息、完全无关的信息[38]。从以上类型划分来看,虽然体现了差异化的个人数据分析,但无法脱离"可识别"与"不可识别"这一二分制的界定方式。笔者认为,该分类标准存在局限性,无论是"已识别"还是"可识别"抑或是"身份"本身都有其相对性,不能直接帮助规范的制定者或争议的裁判者了解相关信息在生活实践中的应用价值[39]。在数字技术背景下,数字技术的发展更是改变了信息识别个人的能力和方式,"识别"与"可识别"之间并非非此即彼[40]。

事实上,"可识别"与"不可识别"之间存在可识别性的连续性,可在可识别性的连续性上定义一个阈值。如果数据集的可识别性高于阈值,则将其视为个人数据,反之,则为非个人数据[41]。根据个人数据被识别的难易程度,学者 Emma 提出了可识别性的五级模型,据该类型划分,从可明确识别的数据到汇总数据的再识别,重新识别需要付出的精力、成本、时间以及技巧越高,数据被重新识别

⑪观点详情参见:丁晓东《用户画像、个性化推荐与个人信息保护》,(《环球法律评论》,2019 年第 5 期 82-96 页);金耀《个人信息去身份的法理基础与规范重塑》,(《法学评论》,2017 年第 3 期 120-130 页)。

的风险愈小。根据以上标准划分为:(1)可明确识别的数据(Readily identifiable data)⑫;(2)掩码数据(Masked data)⑬;(3)暴露数据(Exposed data)⑭;(4)托管数据(Managed data)⑮;(5)汇总数据(Aggregate data)⑯。该数据类型划分突破了可识别和已识别之间的二元制界限,除可明显识别的数据和完全无法识别的数据外,对处于中间状态的数据更加强调数据控制者对数据的管理能力,如通过匿名或者去标识符的方式来对数据进行安全管理[42]。我国2021年4月发布的《信息安全技术 个人信息去标识化效果分级评估规范》(征求意见稿)亦对个人信息的去标识化分级进行了有益尝试,为去标识化效果评估提供了国家标准⑰。笔者认为,该划分方式较好地以信息处理者的责任和义务为核心,根据可识别风险的大小来课以其义务,能够敦促信息处理者合法合规的处理数据。

(三)确立相关隐私风险评估机制

风险评估的目的在于更好地对风险进行控制和管理。隐私风险评估是对组织机构所收集、储存、管理、利用、开放的数据是否对隐私产生影响所进行的生命周期的、系统的评估过程和结果[43]。其目的在于为信息处理者提供明确的数据利用指引,规范信息处理者的行为。隐私风险评估对少数国家的政府机关来说是一项强制性义务,但大多数国家主要还是将其作为一种风险防控的商业手段,多体现在行业制度、企业内部规范之中[44]。我国《数据安全法》明确将风险评估确定为信息处理者的一项强制性义务。笔者认为,通过隐私风险评估可以实现两方面的目标,一是明确信息处理者的义务,二是可以形成相对透明的问责机制。个人信息匿名化最终目的是发布无涉个人隐私的数据用于流通、共享。鉴于匿名化处理中存在的诸多风险,应当通过隐私风险评估识别可能的隐私侵权风险,将匿名化处理过程中的隐私风险评估作为数据控制者一项强制义务予以规范。此外,隐私风险评估本质上是一种工具、方法,也需要通过制定相应的操作性规范将隐私风险评估机制落到实处。

五、结语

匿名化作为平衡数据隐私保护与数据利用冲突的重要技术手段,为数据的流通、共享提供了可能。考虑到个人信息匿名化处理过程中所面临的诸多风险,应当将风险控制理念嵌入个人信息匿名化的法律制度构建中。从个人信息匿名化标准确立到个人信息匿名化风险识别再到“匿名”数据发布,风险控制理念能够与存在诸多不确定性风险的个人信息匿名化形成有效契合。在具体制度

---

⑫可明确识别的数据可通过社会安全号码(SSN)及生物特征和出生日期或其他识别信息直接识别到数据主体,该级别需要最小的努力来重新识别到个人。

⑬处于该级别的数据根据随机化和创建可逆或不可逆的方式操纵识别变量,其主要作用是防止个人身份信息、敏感个人数据以及商业敏感数据被暴露给未经授权的用户。

⑭该级别数据是指除屏蔽标识符(如姓名和出生日期)外,还屏蔽了被视为准标识符(如日期、年龄和性别)的变量,但由于数据的可识别性不可确定,因此,该级别的数据代表了托管人的高风险暴露。

⑮该级别数据可以是微数据或以表格形式出现,并且仅在此级别上,数据可以从个人信息转移到非个人信息,数据托管人可以管理重新识别的风险。

⑯特指明显无法识别的数据。

⑰《信息安全技术 个人信息去标识化效果分级评估规范》将个人信息标识度分为四级:能直接识别主体的数据、消除直接标识符的数据、重标识风险可接受数据以及聚合数据。

构建上,应当紧紧围绕个人信息权利保护,通过完善个人信息分类制度以及隐私风险评估制度将匿名化的不确定性风险置于可控范围内,为信息处理者处理数据提供清晰明确的指引。

参考文献:

[1]刘湘雯,王良民.数据发布匿名技术进展[J].江苏大学学报(自然科学版),2016(5):562-571.

[2]孙广中,魏燊,谢幸.大数据时代中的去匿名化技术及应用[J].信息通信技术,2013(6):52-57.

[3]王融.数据匿名化的法律规制[J].信息通信技术,2016(4):38-44.

[4]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(1):43-56.

[5]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(6):52-58.

[6]马长山.人工智能的社会风险及其法律规制[J].法律科学(西北政法大学学报),2018(6):47-55.

[7]GELLERT R. Data protection:A risk regulation? Between the risk management of everything and the precautionary alternative [J]. International Data Privacy Law,2015,5(1):3-19.

[8]宋健,许国艳,夭荣朋.基于差分隐私的数据匿名化隐私保护方法[J].计算机应用,2016(10):2753-2757.

[9]SHAPIRO S S. Situating anonymization within a privacy risk model[C]//2012 IEEE International Systems Conference SysCon 2012. Vancouver,BC:IEEE,2012:1-6.

[10]齐爱民.界定法律意义上的信息[J].社会科学家,2009(3):6-10.

[11]苏宇,高文英.个人信息的身份识别标准:源流、实践与反思[J].交大法学,2019(4):54-71.

[12]范为.大数据时代个人信息定义的再审视[J].信息安全与通信保密,2016(10):70-80.

[13]Information Commissioner's Office. Determining what is personal data[R/OL].(2012-12-12)[2023-02-24]. https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf.

[14]Article 29 Data Protection Working Party. Opinion 05/2014 on anoymisation techniques[EB/OL].(2014-04-10)[2023-02-24]. https://www.pdpjournals.com/docs/88197.pdf.

[15]ROCHER L,HENDRICKX J M,DE MONTJOYE Y A. Estimating the success of re-identifications in incomplete datasets using generative models[J]. Nature Communications,2019,10:1-9.

[16]SANCHEZ D,MARTINEZ S,DOMINGO-FERRER J. Comment on"Unique in the shopping mall: On the reidentifiability of credit card metadata"[J]. Science,2016,351(6279):1274.

[17]徐明.大数据时代的隐私危机及其侵权法应对[J].中国法学,2017(1):130-149.

[18]郑戈.人工智能与法律的未来[J].探索与争鸣,2017(10):78-84.

[19]冯果,薛亦飒.从"权利规范模式"走向"行为控制模式"的数据信托:数据主体权利保护机制构建的另一种思路[J].法学评论,2020(3):70-82.

[20]RUBINSTEIN I S,HARTZOG W. Anonymization and risk[J]. Washington Law Review,2016,91:703-760.

[21]丁晓东.个人信息私法保护的困境与出路[J].法学研究,2018(6):194-206.

[22]GELLERT R. We have always managed risks in data protection law: Understanding the similarities and differences between the rights-based and the risk-based approaches to data protection[J]. European Data Protection Law Review,2016,4(2):481-492.

[23]张涛.大数据时代个人信息匿名化的规制治理[J].华中科技大学学报(社会科学版),2019(2):76-85.

[24]程海东,王以梁,侯沐辰.人工智能的不确定性及其治理探究[J].自然辩证法研究,2020(2):36-41.

[25]POWER M. The risk management of everything:Rethinking the politics of uncertainty[J]. The Journal of Risk Finance,2004,5(3):58-65.

[26]MACENAITE M. The"riskification"of European data protection law through a two-fold shift[J]. European Journal of Risk Regulation,2017,8(3):506-540.

[27]DIJKA N V,GELLERT R,ROMMETVEIT K. A risk to a right? Beyond data protection risk assessments[J]. Computer Law & Security Review,2016,32(2):286-306.

[28]GONÇALVES M E. The risk-based approach under the new EU data protection regulation: A critical perspective[J]. Journal of Risk Research,2020,23(2):139-152.

[29]LYNSKEY O. The foundations of EU data protection law[M]. Oxford:Oxford University Press,2015:84.

[30]Article 29 Data Protection Working Party. Statement on the role of a risk-based approach in data protection legal frameworks [EB/OL]. (2014-05-30) [2023-02-24]. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf.

[31]姚佳. 个人信息主体权利的实现困境及其保护救济[J]. 中国法律评论,2022(6):132-142.

[32]STALLA-BOURDILLON S, KNIGHT A. Anonymous data v. personal data-a false debate: An EU perspective on anonymization,pseudonymization and personal data[J]. Wisconsin International Law Journal,2016,34(2):284-322.

[33]郑佳宁. 数据匿名化的体系规范构建[J]. 政法论丛,2022(4):61-71.

[34]OHM P. Broken promises of privacy:Responding to the surprising failure of anonymization[J]. UCLA Law Review,2010,57:1701-1777.

[35]丁晓东. 论个人信息概念的不确定性及其法律应对[J]. 比较法研究,2022(5):46-60.

[36]齐爱民,张哲. 识别与再识别:个人信息的概念界定与立法选择[J]. 重庆大学学报(社会科学版),2018(2):119-131.

[37]项定宜. 个人信息的类型化分析及区分保护[J]. 重庆邮电大学学报(社会科学版),2017(1):31-38.

[38]谢琳. 大数据时代个人信息边界的界定[J]. 学术研究,2019(3):69-75.

[39]岳林. 个人信息的身份识别标准[J]. 上海大学学报(社会科学版),2017(6):28-41.

[40]高富平. 个人信息流通利用的制度基础:以信息识别性为视角[J]. 环球法律评论,2022(1):84-99.

[41]EL EMAM K. Risk-based de-identification of health data[J]. IEEE Security & Privacy,2010,8(3):64-67.

[42]NELSON G S. Practical implications of sharing data: A primer on data privacy, anonymization, and de-identification[EB/OL]. (2015-04-26)[2023-02-24]. https://support.sas.com/resources/papers/proceedings15/1884-2015.pdf.

[43]迪莉娅. 大数据环境下隐私泄露影响评估研究[J]. 情报杂志,2016(4):141-146.

[44]肖冬梅,谭礼格. 欧盟数据保护影响评估制度及其启示[J]. 中国图书馆学报,2018(5):76-86.

On the legal regulation of personal information anonymization in China under the risk control concept

ZHANG Li[1], XU Duoqi[2]

(*1. Koguan School of Law, Shanghai Jiao Tong University, Shanghai 200030, P. R. China;*
2. Law School, Fudan University, Shanghai 200438, P. R. China)

Abstract: Data anonymization provides essential technical support for the circulation and sharing of data. Technology variability also brings many obstacles to the legal regulation of data anonymization. Under the current technical background, China's anonymization of personal information faces risks such as uncertain identification standards and reversible identity re-identification. The technical risks generated by anonymizing personal information significantly challenge personal privacy interests. The current result-oriented regulatory means lack flexibility in the regulation of personal information anonymization, and it is difficult to mitigate the uncertain risks brought about by technology. Balancing the conflict between individual privacy, corporate economic, and public social interests is the ultimate goal of anonymizing personal information. The superiority of the risk control-oriented concept eliminates the shackles between the realistic demands of data protection and

ideal value judgments at the conceptual level, replacing the result-oriented concept. It provides a more practical approach to bridging thinking for the endogenous data compliance and self-regulation of information processors, the protection of rights and the use of data of information subject, and offers new possibilities for the legal regulation of the anonymization of personal information. The legal framework of personal information anonymization with the concept of risk control as the core aims to achieve a dynamic balance between the validity and practicability of data and designs legal mechanisms around reducing the risk in the process of anonymization of personal information. The excellent governance of personal information anonymization regulation can be realized by imposing corresponding information processing risk assessment obligation on information processors. Without the support of normative documents, the risk assessment and evaluation standards of anonymizing personal information are difficult to quantify stably in practice. Therefore, in terms of legal regulation of personal information anonymization, it should be advocated that the protection of the rights of personal information subjects should be the core of anonymization processing of personal information, and the rights of the information subject should be realized by giving the information subject corresponding data rights in the process of information processing. The anonymization of personal information is based on the premise that personal information can be identified, and the definition of personal information should be judged on a case-by-case basis based on specific scenarios. Therefore, the identification of anonymous data should also be understood in a dynamic and scenario－based manner. In terms of implementing risk control means for anonymizing personal information, relevant privacy risk assessment mechanism should be established to provide straightforward data utilization guidelines for information processors and regulate the behavior of information processors.

Key words: personal information; anonymous processing; risk assessment; risk control

(责任编辑　袁　虹)