交大法学 SJTU Law Review
No. 4 (2019)

个人信息的身份识别标准：源流、实践与反思

苏　宇　高文英*

目次

摘要　我国现行立法对个人信息的认定标准是能够识别特定自然人身份的“身份识别标准”。这种标准涵盖范围甚广，但却存在着以下问题：“可识别性”概念边界模糊、信息的识别区分度未被有效考虑、信息的结合性识别无法预判。在建立个人信息分类保护制度的前提下，针对不同的信息类型，单一的身份识别标准亦应当发展成一个差异化、动态调整的标准体系。应当建立个人信息分类保护制度，对不同类别的个人信息实行不同强度的外延式保护。对于目录内不同类型的个人信息，应当建立有一定差异性的认定程序，最终由主管部门发布个人信息分类目录，并定期进行动态调整；应当结合外延式保护及内涵式保护的进路，对不同类型的个人信息提供有层次的界定方式；应当通过执法和司法实践的不断归纳总结，结合技术专家的意见，形成更具操作性的界定标准，或者进一步形成临时性的参考清单。

关键词　个人信息　身份识别标准　可识别性　信息法学

随着信息产业与信息经济的迅猛发展，人被不断卷入信息社会的洪流之中，人的一举一动都进入了信息技术的视野。通过政府的信息采集工作、移动互联网应用和终端的信息收集过程、监控摄像设备及各种物联网设备的自动化拍摄或记录等途径，政府与企业逐渐掌握了大量

* 苏宇，中国人民公安大学法学院讲师、法学博士；高文英，中国人民公安大学法学院教授、法学博士。本文系中国法学会 2017 年度部级法学研究课题“大数据时代背景下的警察执法信息公开范围研究”（项目编号：CLS2017C07）及 2018 年国家社会科学基金重大项目“大数据时代个人数据保护与数据权利体系研究”（项目编号：18ZDA145）的阶段性研究成果。

个人信息。移动互联网、大数据与云计算等技术的发展既为数字经济增添了强劲的动力，也给个人信息的不当获取与利用带来了显著的风险。〔1〕现实生活中，个人信息泄露的风险无处不在，个人信息的法律保护需求亦随之不断增强。《民法总则》第111条规定："自然人的个人信息受法律保护。任何组织和个人需要获取他人个人信息的，应当依法取得并确保信息安全，不得非法收集、使用、加工、传输他人个人信息，不得非法买卖、提供或者公开他人个人信息。"这一规定明确承认了个人信息的法律地位，表明个人信息独立于其他民事权利，为个人信息的私法保护提供了重要的规范依据。〔2〕同时，《网络安全法》《电子商务法》《消费者权益保护法》乃至《刑法》等一系列法律规范也对个人信息从不同角度进行了保护。类似地，西方国家也非常重视个人信息安全问题，通过一系列立法与司法实践加强了对个人信息的保护。尽管部分国家和地区更倾向于使用"个人数据"（personal data）之概念，但相关的数据保护法律制度与我国的个人信息保护法制实属异曲同工。

然而，在林林总总的个人信息保护规范中，一个根本的问题正变得日益关键：什么是个人信息？易言之，个人信息应当建立何种认定标准？这一问题已经变成横亘在世界各国个人信息保护法律实践面前的一个基础性理论问题。我国现行立法对个人信息的认定标准是"身份识别标准"，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。对身份识别标准的理解与应用，对于信息时代的个人权利保护可谓举足轻重。

一、身份识别标准的形成及基本内涵

（一）身份识别标准的形成

如果我们将个人信息的学理脉络追溯到隐私权，则早在隐私权概念被提出之时，此方面的权利保护即已隐含了身份识别的要素。沃伦（Samuel D. Warren）和布兰代斯（Louis D. Brandeis）在1890年发表的著名论文《隐私权》中就指出，事物的排他性所有权必须经过识别（identification）才能确认，也只有经过这种识别才能适用美国宪法与法律中个人专属权利（property，在原文中非"财产权"之意）的保护。〔3〕个人信息的身份识别标准大约形成于20世纪60年代。最初从隐私权角度提出身份识别标准的法律学者是瑞布豪森（Oscar M. Ruebhausen）与布林（Orville G. Jr. Brim），他们在《隐私与行为研究》一文中针对研究数据泄露个人隐私的情况进行了分析，坚持个人身份能被外界识别则必须取得当事人同意的研究准则，主张研究数据的对外提供必须遵守匿名化条件，避免受访者被识别，并提出了对研究数据进行匿名化处理与避免个人身份识别的系列建议。〔4〕此时，通过信息技术收集和处理数据的活动尚未广泛兴起，但离大数据时代的来临也已为时不远。

1972年美国的"水门事件"促使美国政府成立了自动化个人数据系统咨询委员会（Advisory

〔1〕 参见高秦伟：《个人信息保护中的企业隐私政策及政府规制》，载《法商研究》2019年第2期，第16页。

〔2〕 参见张凌寒、杜婧：《基于隐私权的个人信息保护路径研究——以美国为研究视角》，载《网络法律评论》2016年第1期，第43页。

〔3〕 "Property"在此处指"任何人特有或固有的、排他性地归属于其个体的权利"，作者援引库迪斯（Curtis）的学说，强调"property"来源于拉丁语"proprius"一词及"个人自有"（one's own）的含义。See S. Warren & L. Brandeis, "The Right to Privacy", 4 Harvard Law Review 206 (1890).

〔4〕 See O. Ruebhausen & O. Brim, "Privacy and Behavioral Research", 65 Columbia Law Review 1184-1211 (1965).

Committee on Automated Personal Data Systems)以调查信息科技产生的潜在有害后果及提出相应权益保障建议,该委员会发布了名为《记录、计算机与公民权利》(Records, Computers and the Rights of Citizens)的报告(下称《1973年报告》),并促成美国国会制定了1974年的《隐私权法》。[5]《1973年报告》明确使用身份识别标准界定个人隐私的边界,并主张"自然个体的个人隐私受到一份记录中与其相关之可识别性信息的披露与运用的直接影响。一份记录如包含可识别形式之有关个人的信息,就必须受这样一种程序约束:赋予该个体参与程序以决定记录内容为何、如何披露与使用其中的可识别性信息。可识别性个人信息的任何记录、披露与使用若不受此种程序约束,则必须被认为是一种不公正的信息活动,除非此种记录、披露或使用获得法律的专门授权"。这一文件标志着身份识别标准的正式形成。美国1974年制定的《隐私权法》沿用了身份识别标准并加以细化,但其具体范围是以直接可识别的信息为主。[6] 以《隐私权法》为基础,美国在个人信息权益保护方面采取了隐私权保护进路,非法取得个人识别信息(personal identifiable information,简称PII)只是触发隐私权保护机制的前提条件,美国法也并没有对个人识别信息给出统一的界定。[7] 在具体法律实践中,美国的个人识别信息基本上仅限于当前条件下可以直接识别的信息,[8]范围尚属有限。在当时的技术和社会条件下,这一标准并未受到充分关注。由于美国对个人信息的保护是以消费者权利及隐私权保护的进路为主,尽管其后仍有若干与个人信息相关的重要立法(如1984年的《联邦有限通讯政策法》),个人信息仍未泛化为私法上的权利,[9]其个人信息界定标准也未产生重大的后续影响。

伴随信息科技与信息经济的发展,身份识别标准逐渐在全球范围内获得广泛认可,其范围与功能也发生了变化。1977年,联邦德国已经开始在立法层面正式承认身份识别标准。[10] 1980年,世界经济合作与发展组织(Organization for Economic Co-operation and Development,OECD)制定了《隐私保护和个人数据跨境流通指南》(Guidelines on the Protection of Privacy and Transborder Flows of Personal Data,以下简称《1980年OECD指南》)。《1980年OECD指南》在附录第一部分第1条将"个人数据"界定为"与一个被识别或可被识别的人(即数据主体)有关的任何信息",这一定义使身份识别标准明确包括了"被识别"(即当前能够直接识别)以及"可能被识别"(即在可预见的未来可能被识别)两层含义,扩展了身份识别标准的内涵,使之能够应对新的信息技术条件下个人信息保护的需求。1995年10月24日欧洲共同体发布的一项数据保护指令第26段[11]延续了这一界定方式,确定数据保护的原则适用于"被识别的或可被识别的个人"(an identified or identifiable person),而不适用于具备匿名化效果的、已不再可识别个人身份的数据。2012年11月新加坡制定的《个人数据保护法》(Personal Data Protection Act 2012)同样采取了身份识别

---

〔5〕 See M. Bharwaney & A. Marwah, "Personal Data Privacy in the Digital Age", 43 Hong Kong Law Journal 806 (2013).

〔6〕 See 5 U.S.C. §552(a).

〔7〕 See P. Schwartz & D. Solove, "The PII Problem: Privacy and a New Concept of Personally Identifiable Information", 86 New York University Law Review 1816 (2011).

〔8〕 See P. Schwartz & D. Solove, "Reconciling Personal Information in the United States and European Union", 102 California Law Review 880 (2014).

〔9〕 参见丁晓东:《个人信息私法保护的困境与出路》,载《法学研究》2018年第6期,第198页。

〔10〕 See P. Schwartz & D. Solove, *supra* note〔7〕, at 1874.

〔11〕 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 (on the protection of individuals with regard to the processing of personal data and on the free movement of such data), para. 26.

标准，并且规定了可识别性的扩展范围——“从这一数据或其他信息中可以或有可能掌握识别（个人身份）的途径”。〔12〕2018 年开始实施的欧盟《通用数据保护条例》（General Data Protection Regulation，即“GDPR”）亦采取了“被识别或可被识别”的身份识别标准。〔13〕以欧盟与新加坡等为代表的国家及地区对个人信息或个人数据采取了不必借助隐私权概念的直接保护方式，身份识别标准亦在这些国家及地区成为统一的、通用的权益界定标准，在信息法领域日益发挥着基础性的作用。尽管也有地区（例如美国加州）提出了“可关联性”（linkable）标准，即以相关信息是否能够直接或通过设备关联到消费者或住户作为判断构成个人信息的标准，〔14〕这种标准在突出场景导向、把握个人信息辨识个人的方式与实质、增强法律适用确定性方面有积极的意义，〔15〕但尚未对身份识别标准的主导地位造成显著冲击。

在我国，个人信息保护的主要法律规范及司法解释也采取了身份识别标准。例如 2013 年制定的《电信和互联网用户个人信息保护规定》第 4 条规定：“本规定所称用户个人信息，是指电信业务经营者和互联网信息服务提供者在提供服务的过程中收集的用户姓名、出生日期、身份证件号码、住址、电话号码、账号和密码等能够单独或者与其他信息结合识别用户的信息以及用户使用服务的时间、地点等信息。”2014 年通过的《最高人民法院关于审理利用信息网络侵害人身权益民事纠纷案件适用法律若干问题的规定》也采取了身份识别标准，并进一步明确个人信息的范围包括“自然人基因信息、病历资料、健康检查资料、犯罪记录、家庭住址、私人活动等个人隐私和其他个人信息”。2016 年制定的《网络安全法》第 76 条第（5）项又在此基础上明确增加了“个人生物识别信息”。2017 年通过的《最高人民法院、最高人民检察院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 1 条亦有类似规定，并将财产状况和行踪轨迹也纳入个人信息的范围。2017 年发布的国家标准《信息安全技术 个人信息安全规范》（GB/T 35273—2017）中 3.1 条也确认了类似的界定标准。总而言之，无论是“硬法”还是“软法”层面，个人信息的身份识别标准也已被我国法律实践所认可。

（二）身份识别标准的基本内涵

身份识别标准包括“被识别”与“可被识别”两层内涵。在法律实践中，“被识别”指的是事实上已借由此种信息识别特定自然人，或者一般人能够不借助技术分析而直接识别出特定自然人；而“可被识别”实际上指向的是一种相对的概率，即识别某种信息与特定自然人相关的相对可能性（relevant possibility），〔16〕特别是作为一个人最常用的主要标识——姓名与其相关数据相联系的可能性。〔17〕因此，不少国家或地区的法律实践对“可被识别”的界定事实上是一种预测性的判断，并且此种预测需要“合理”（reasonable）和“实践可行”（practicable）的基础，〔18〕或者如《通用数据保护条例》所要求，需要具备“合理倾向性”（reasonable likelihood），又或者如加拿大的法律实践所

〔12〕 Personal Data Protection Act 2012 (Singapore), 2(1).

〔13〕 General Data Protection Regulation, art. 4(1).

〔14〕 See California Consumer Privacy Act of 2018 § 1798.140.

〔15〕 参见范为：《大数据时代个人信息保护的路径重构》，载《环球法律评论》2016 年第 5 期，第 100～101 页。

〔16〕 See N. Purtova, “The law of everything. Broad concept of personal data and future of EU data protection law”, 10 Innovation and Technology 44 (2018).

〔17〕 See Stacy-Ann Elvyp, “Paying for Privacy and the Personal Data Economy,” 117 Columbia Law Review 1415 (2017).

〔18〕 See W. Chik & J. Pang, “The Meaning and Scope of Personal Data under the Singapore Personal Data Protection Act”, 26 Singapore Academy of Law Journal 367 (2014).

要求，需要有“严肃的可能性”(serious possibility)。[19] 事实上，“可被识别”的可能性存在从零风险到必然被识别的模糊范围，对“可被识别”的标准也就存在不同程度的把握。[20] 况且，由于“合理”与“实践可行”等辅助性标准的存在，单纯理论上可被识别的可能性并不必然满足“可被识别”的标准。例如有一种相当有影响力的观点还主张考虑识别信息的成本、去识别技术的防御作用、识别工具的可得性等种种条件，[21]避免过度扩展个人信息保护法制的适用范围。因此，“可被识别”的范围实际上存在不确定性，成为引人关注的焦点问题。

一旦“可被识别”的标准内部被添加了“合理可行”或“严肃的可能性”之类的内涵，“可识别性”就不再是单纯的技术判断，而隐含了相当程度的价值权衡：不将纯粹理论上具备身份识别潜能或识别可能性极低的信息划入“个人信息”之范围，是避免个人信息保护法制给信息占有者带来过分苛刻的遵从义务，影响它们对其所占有信息的使用权与利用能力。[22] 此种价值权衡目前尚不存在一种明确的界线。随着技术发展，再识别(re-identification)的风险日益超出了数学和统计学模型的预测能力，规制者正在尝试“从数学到社会学”的思维转换，放弃单纯从技术层面分析身份识别的可能性，而是追问谁有可能从数据中获取个人信息、成功的可能性如何，特定的行业或经济部门是否存在有关侵害隐私之倾向性的某种历史、惯例或传统，以及管理者采取的保护措施是否足以降低此种风险等等。[23] 更多地考虑综合性的因素去判断匿名化的数据是否可能被发现个人身份信息，使得身份识别标准日益脱离字面上的简单含义、日益承载复杂的价值与技术判断。

不仅如此，面对信息技术的不断变革和发展，“可被识别”的标准还是动态变化的，同一个数据集可能在不同时间、不同技术条件甚至技术人员手中产生不同的识别结果，[24]而此种识别结果的发生与否、何时发生并非全然可预测。鉴于此种原因，有人认为“可被识别”的标准呈现“语境依赖”(context-dependent)的性质，[25]但即使给出了特定的语境，由于对信息技术的理解与把握存在差异，不同人对同一数据集的身份识别能力也可以做出不同的判断。这几乎使得对身份识别标准的具体理解成为“罗生门”式的图景。既然统一的判断尺度难以掌握，许多国家与地区的法律实践也就没有积极尝试严谨、精确地规定身份识别标准的内涵，而是在对身份识别标准的内涵进行一定程度抽象阐释的同时，从外延方面着手界定个人信息保护的具体范围。若要认识身份识别标准下个人信息的具体保护范围，还需要观察立法与司法的实践状况。

在不同国家与地区的法律实践中，对于何种信息属于“被识别或可被识别”的问题，具体法律实践非常复杂，时常需要辅之以外延性的界定。例如，欧盟《通用数据保护条例》第4条第(1)款除明确将自然人的姓名、身份证明的号码、位置以及网络地址列为个人数据外，还将某一自然人特有的生理、心理、基因、精神、经济、文化或社会标识(identity)列为个人数据。易言之，只要通过一个信息能够产生身份上的区分性，都有可能被列为个人信息(数据)；至于具体的区分程度或识别的可能程度，身份识别标准的法律实践至今尚未形成明确、统一的界定；在个人信息保护制度中，可

---

〔19〕 See W. Chik & J. Pang, *supra* note〔18〕, at 386.

〔20〕 See N. Purtova, *supra* note〔16〕, at 42.

〔21〕 See N. Purtova, *supra* note〔16〕, at 46－47.

〔22〕 See W. Chik & J. Pang, *supra* note〔18〕, at 388－389.

〔23〕 See Paul Ohm, “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization”, 57 UCLA Law Review 1761－1762 (2010).

〔24〕 See N. Purtova, *supra* note〔16〕, at 47.

〔25〕 See N. Purtova, *supra* note〔16〕, at 48.

识别性(identifiability)仍然是一个边界含糊的概念。〔26〕这一状况给各国的个人信息保护法律实践带来了实质性的挑战。

二、身份识别标准下个人信息保护范围之展开：以我国的法律实践为例

我国的法律实践同样存在着“可识别性”概念边界模糊的问题，不仅立法上的规定较为宽泛，司法实践中的个人信息内容亦相当丰富。笔者自中国裁判文书网上选取了近年来刑事、民事与行政案件中在个人信息认定方面包含不同个人信息内容的若干案例(见表1)：

表1　我国生效裁判文书中认定个人信息范围的若干实例

性质	案　　例	被认定属于“个人信息”的内容
刑事	何×、肖×侵犯公民个人信息案〔27〕	公民姓名、住址、联系方式、女性化妆品类信息
刑事	李××、赖××侵犯公民个人信息案〔28〕	姓名、电话号码、身份证件号码、财产信息
刑事	安徽聚思文化传播有限公司、王××侵犯公民个人信息案〔29〕	公民姓名、所属公司名称、联系电话
刑事	谭××侵犯公民个人信息案〔30〕	机主姓名、手机号码、所在企业名称、固定电话号码、联系地址等
刑事	金×与曹×等侵犯公民个人信息案〔31〕	公司名称、法人姓名、职务、联系方式、部分车牌信息
刑事	张××、赵××侵犯公民个人信息案〔32〕	邮箱账号、密码、信用卡信息
民事	丁芝玲诉汪锡奎案〔33〕	身份证上的姓名、性别、民族、出生日期、住址和身份证号码(张贴裁判文书)
民事	黄立红诉付丽丽案〔34〕	微信二维码、姓名、联系电话、私密照片
民事	王静诉王茹香、李春香等案〔35〕	姓名、身份证号码、银行卡号
民事	周容北诉巫蓝霞案〔36〕	姓名、职务、照片
民事	重庆市足下软件职业培训学院诉毛志刚案〔37〕	毕业生姓名、性别、电话号码、毕业院校

〔26〕 See N. Purtova, *supra* note〔16〕, at 44－47.
〔27〕 随州市曾都区人民法院(2017)鄂1303刑初206号刑事判决书。
〔28〕 中山市第一人民法院(2017)粤2071刑初1804号刑事判决书。
〔29〕 合肥市蜀山区人民法院(2018)皖0104刑初268号刑事判决书。
〔30〕 合肥市包河区人民法院(2018)皖0111刑初377号刑事判决书。
〔31〕 江苏省南京市中级人民法院(2018)苏01刑终279号刑事裁定书。
〔32〕 江西省萍乡市中级人民法院(2018)赣03刑终158号刑事判决书。
〔33〕 德阳市旌阳区人民法院(2017)川0603民初4743号民事判决书。
〔34〕 苏州市姑苏区人民法院(2018)苏0508民初1860号民事判决书。
〔35〕 北京市门头沟区人民法院(2017)京0109民初4623号民事判决书。
〔36〕 珠海市中级人民法院(2018)粤04民终541号民事判决书。
〔37〕 重庆市高级人民法院(2018)渝民申853号民事裁定书。

续表

性质	案　　例	被认定属于“个人信息”的内容
民事	庞理鹏诉北京趣拿信息技术有限公司案〔38〕	姓名、手机号、行程信息
行政	汪华斌诉合肥市蜀山区井岗镇人民政府案〔39〕	姓名、住址、财产状况
行政	杨永芳诉兰州市西固区人民政府政府信息公开案〔40〕	特定自然人的《补偿安置协议》及《评估报告表》(含个人财产状况)
行政	赵迎香等诉西安市公安局浐灞生态区分局案〔41〕	某村人口花名册(具体内容不详)

目前,我国司法实践中对个人信息的认定与保护已经逐渐从个人隐私中分离出来,明确以身份识别标准界定个人信息的保护范围。例如,在“丁芝玲诉汪锡奎案”中,人民法院以详细的论证明确区分了公民个人隐私权与个人信息权之间的区别,〔42〕明确了个人信息保护范围只需要是“与特定自然人相关”以及“可以据此将该自然人特定化”,而不需要确定信息是否敏感、是否超出一般人的容忍度等,判决被告张贴未经处理的生效裁判文书侵犯了公民的个人信息权。此种立场在近年来的裁判文书中并不鲜见,如“孟凡勇诉来安县金点子广告传媒有限公司案”的判决书亦认为擅自传播未经处理的生效裁判文书属于侵权。〔43〕这一趋势对于加强个人信息保护具有相当积极的意义,但脱离隐私权的保护路径后,个人信息的识别标准与保护范围问题就更加凸显。

事实上,与“可被识别”的标准相类似,由于“能与其他信息结合识别特定自然人身份的信息”之具体范围并不明确,除法条中明确规定的用户姓名、出生日期、身份证件号码、住址、电话号码、账号和密码等标识性极强的信息外,其余何种信息能够被列入个人信息之范围,尚不无疑问。在表1所列案例中,除公民姓名为普遍出现的内容外,联系方式、住址及身份证号码的出现频次亦较高,其余如女性化妆品类信息、公民所在单位、邮箱账号、信用卡号码、财产信息、行程信息、车牌号码、职务、照片、微信二维码甚至性别及毕业院校等亦在裁判文书中被列入“个人信息”之列。当

〔38〕 北京市第一中级人民法院(2017)京01民终509号民事判决书。

〔39〕 安徽省合肥市中级人民法院(2018)皖01行终22号行政判决书。

〔40〕 甘肃省高级人民法院(2018)甘行终379号行政判决书。

〔41〕 西安铁路运输法院(2017)陕7102行初1402号行政判决书。

〔42〕 “……2. 关于自然人个人信息及自然人信息权。自然人个人信息主要是指据以识别特定自然人身份的任何生物性、物理性的数据、文件、档案等资料,其范围不仅包括自然人的身份证信息、户籍信息、家庭构成、职业情况、社会交往、电子数据等物理性数据。任何与特定自然人相关的,可以据此将该自然人特定化的信息均属个人信息。自然人的个人信息涉及自然人的身份和地址,具有人身属性,属于人格权范畴。公民个人享有自然人信息权,依法受法律保护。3. 关于公民个人隐私权。公民个人即自然人享有其生活中不愿被他人知晓的信息的权利,未经本人同意任何对该类信息的获取都是非法的,都是对自然人隐私权的侵犯。公民个人的隐私信息是公民个人的自然人信息的一部分,但并不完全就是公民个人的自然人信息。自然人隐私信息是自然人独有的、不愿意公之于众的资料。而一般性的个人基本信息(如居民身份证、护照上载明的信息),一般情况下属于自然人个人身份信息,不属于严格的法律意义上的个人隐私。其界限和区别是:① 不超过一个‘一般人’的‘社会容忍度’;② 不涉及敏感的信息;③ 已经公开的个人信息。只要具备这三个特征之一即不再具有隐私的特点,不属于个人隐私。但若未经本人同意而利用其个人身份信息进行非法活动,则构成对公民个人信息权的侵犯。”判决书原文,请参见德阳市旌阳区人民法院(2017)川0603民初4743号民事判决书。

〔43〕 来安县人民法院(2018)皖1122民初77号民事判决书。

然，仔细研读相关裁判文书，性别、职务与毕业院校等信息并未被单独列入个人信息范围，而是被归属于“能与其他信息结合识别特定自然人身份的信息”。这部分信息的范围相当宽泛，就目前的法律实践而言，只要泄露或买卖个人信息的内容包括公民姓名以及任何其他与姓名对应的信息，都有可能被认定为侵犯个人信息之法益。需要注意的是，法律规定中的“身份”并不必然需要以姓名或身份证号码等表示。有学者认为“识别”包含身份识别与个体特征识别，步态、语音等应当属于个体特征而非身份。〔44〕此种学说在逻辑上是正确的，但难以解释目前的法律实践，否则除姓名和身份证号码等极少数信息外，其他信息都不能被单独列为“个人信息”。应当认为，目前我国的法律实践对“身份”的理解是泛化的，只要某一特定个体能够通过某一信息被与其他个体区分，不论其呈现何种外在形式，该信息就有可能被划入个人信息之列。这一理解实际上与“关联性”标准的精神已经有所相通。

值得注意的是，司法实践中对个人信息的认识已经开始出现一些重要分类，例如司法解释及国家标准中已经明确了识别的对象包括“特定自然人身份”或者“特定自然人活动情况”，此种分类对于个人信息保护范围的进一步明确富有积极意义。

表 1 中“庞理鹏诉北京趣拿信息技术有限公司案”可谓个人信息保护范围界定方面的经典案例。在该案中，趣拿公司与东航涉嫌泄露了庞理鹏的姓名、手机号码及航班信息。该案的判决书确认“公民个人信息包括身份识别信息和活动情况信息”，并由此将机票上的行程信息列入个人信息范围。不仅如此，针对手机号码是否属于个人信息，该判决书认为，虽然单个的、孤立的甚至在民事生活中可以（部分）公示的信息，如姓名和手机号码，并不必然需要保密，但这些信息“一旦被收集、提取和综合”，就可以“与特定的个人相匹配，从而形成某一特定个人的详细而准确的整体信息”，〔45〕进而导致“个人的隐私将遭受巨大威胁”，在社会生活中将有可能发生针对性的诈骗（如本案中有人向庞理鹏发送航班取消的诈骗信息）等危害后果。这一论述至少揭示了个人信息范围界定的两条具体准则：一是相关信息是否有能力与其他信息相结合，形成与特定自然人相匹配的整体信息（以下简称“整体匹配准则”）；二是形成整体信息有可能引致更有针对性的侵权风险，因此这些单个的、孤立的甚至部分范围内公示的信息就有受法律保护的价值（以下简称“侵权风险准则”）。这两个要点正是从技术层面与价值层面对个人信息识别标准进行把握的基础，也与表 1 中其他案例的范围界定结果及裁判文书的论证是一致的。

但是，对于“整体匹配准则”与“侵权风险准则”，可以存在两种不同的理解，导致个人信息保护范围的认定存在分歧：如果两条准则是相互独立的，则任何来源于自然人的信息都有可能被识别出与特定个人的关联性而成为构造整体信息的一部分，都满足“整体匹配准则”，也就都应当被纳入个人信息的保护范围；如果两条准则是紧密结合的，则在信息占有者所持有的全部信息不足以导致进一步侵权风险的前提下，即使个别信息在理论上有助于构造特定个人的整体信息，也会被排除在个人信息的范围外。如果缺失了姓名或身份证号码这些关键的、直接的身份识别信息，余下的其他信息是否也能够构成个人信息呢？目前的法律实践倾向于否定这一点。例如，一些企业使用了匿名化处理后的网络行为痕迹与标签信息作为商业决策的参考（如淘宝的“生意参谋”数据产品），这种数据产品未被法律实践认为包含个人信息。在“淘宝（中国）软件有限公司诉安徽美景信息科技有限公司案”中，人民法院采取了“个人信息”与“非个人信息”二分的标准，并认定匿名化处理后的行为痕迹与标签信息不具备单独或者与其他信息结合识别个人身份的能力，“对网络用

〔44〕 参见高富平：《个人信息保护：从个人控制到社会控制》，载《法学研究》2018 年第 3 期，第 93 页。

〔45〕 北京市第一中级人民法院（2017）京 01 民终 509 号民事判决书。

户信息提供者不会产生不利影响”,〔46〕因而不属于个人信息。类似地,早在2015年判决的“北京百度网讯科技有限公司与朱烨隐私权纠纷案”中,与个人身份信息相分离的网络行为痕迹,即已经不被界定为个人信息;判决书认为这些信息已经“无法确定具体的信息归属主体”,亦不会导致对用户隐私权的侵犯。〔47〕在此,“侵权风险准则”实际上占据了主导地位,法律实践容许经过匿名化处理的信息可以流通以及被第三方利用。

在目前的法律实践中,即便是对于健康医疗信息等与个人密切相关的生活信息,也并不必然排斥合法进入数据市场及其他数据传播过程的可能性。例如,福州市人民政府2017年制定了《福州市健康医疗大数据资源管理暂行办法》,随后福州市人民政府办公厅又印发了《福州市健康医疗大数据资源管理实施细则》,其中就有关于个人敏感信息脱密脱敏处理后开放的规定。〔48〕其他一些地方政府在制定大数据方面的发展政策时,也做出了类似的政策安排。例如,2016年发布的《浙江省促进大数据发展实施计划》提出“……对经过脱敏等安全处理的数据资源进行充分的挖掘使用,开展数据增值服务”。大数据产业和信息经济的勃兴促使法律与政策实践在社会经济发展与个人信息保护之间尽可能争取最优的平衡状态,也使得个人信息的认定标准需要考虑更多的社会经济因素。

个人信息保护范围的确定,从信息经济与信息产业的发展上看亦不失其合理之处。一旦任何来源于自然人的信息都被划入个人信息的范围,不仅企业数据交易与政府数据共享将面临重大法律困境,〔49〕需要大数据作为训练集的人工智能也将只能在棋类、作曲、无人机驾驶等少数领域发挥作用,而无法为自然人提供各类精准的工作与生活服务。面对个人信息被侵犯的风险,企业和政府采取了匿名化手段以确保数据集在传播与扩散的过程中不至于发生被侵权的情形。

不过,新的问题也随之而来:在这一系列法律实践中,身份识别标准逐渐开始出现浮动的状

---

〔46〕 该案涉及的主要法律问题之一是“生意参谋”数据产品是否侵犯了公民的个人信息。在判决书中,人民法院认为:“用户信息包括个人信息和非个人信息。前者指向单独或与其他信息结合识别自然人个人身份的各种信息和敏感信息,后者包括无法识别到特定个人的诸如网络活动记录等数据信息”。随后进一步指出:“从……‘生意参谋’数据产品所涉网络用户信息主要表现为网络用户浏览、搜索、收藏、加购、交易等行为痕迹信息以及由行为痕迹信息推测所得出的行为人的性别、职业、所在区域、个人偏好等标签信息。这些行为痕迹信息与标签信息并不具备能够单独或者与其他信息结合识别自然人个人身份的可能性,故其不属于网络安全法中的网络用户个人信息,而属于网络用户非个人信息。”判决书原文,请参见浙江省杭州市中级人民法院(2018)浙01民终7312号民事判决书。

〔47〕 本案中,一审原告朱烨在使用百度网站时发现自身的搜索痕迹被记录并用于网络推广,遂起诉百度公司。一审判决百度公司的行为侵犯了朱烨的隐私权,但被二审判决否定。二审判决书中指出:“网络活动轨迹及上网偏好一旦与网络用户身份相分离,便无法确定具体的信息归属主体,不再属于个人信息范畴。”参见南京市中级人民法院(2014)宁民终字第5028号民事判决书。

〔48〕 《福州市健康医疗大数据资源管理暂行办法》(2017年通过)第16条规定:“市数字办会同市卫计委等相关行业主管部门及数据运营单位建立健康医疗大数据开放开发机制,规范健康医疗大数据应用领域的准入标准,建立大数据应用诚信机制和退出机制,严格规范大数据挖掘、应用和开发行为。除法律、法规另有规定外,涉及商业秘密、个人隐私的健康医疗大数据应当进行脱密脱敏处理后开放。”第19条规定:“高校或者科研院所获得的数据只限用于科研教育等非营利性活动。数据使用单位确需使用可识别个人身份和隐私内容的个案信息的,应向数据运营单位提出应用服务申请,在征得个人同意或经脱敏脱密后,方可实行。”《福州市健康医疗大数据资源管理实施细则》(2017年通过)第21条第(4)项规定:“在数据开放开发过程中,应根据不同的应用场景对数据进行脱敏脱密处理,采用动态脱敏和静态脱敏两种技术手段。”

〔49〕 参见韩旭至:《个人信息的法律界定及类型化研究》,法律出版社2018年版,第35页。

态。一组信息是否能够识别特定自然人不再完全取决于其内容种类，而需要检视匿名化处理及脱敏处理所能达到的标准与效果，即是否满足侵权风险准则。对此，信息技术的发展使得身份识别标准逐渐呈现扑朔迷离的状态：一方面，常规匿名化处理操作、差分隐私（differential privacy）、零知识证明（zero-knowledge proof）、全同态加密（fully homomorphic encryption）等处理技术力求切断数据集的整体数学特征与个人信息、个人身份信息或公开活动信息与其他相关匿名化信息之间的关联性；另一方面，数据挖掘与数据分析技术又不断突破去识别技术的防御，信息技术的迅速发展和不断交锋使得对身份识别标准的把握变得更为复杂和专业，身份识别标准也开始面临前所未有的挑战。

三、身份识别标准的缺陷与反思

识别意味着区分和辨认；身份识别标准意味着存在区分和指向特定自然人的可能性。〔50〕在身份识别标准的范围内涵盖的个人信息非常丰富，“可识别性”（identifiability）之下似乎无所不包，但却暗含着两个难以解决的关键问题：一是信息的识别区分度，二是信息的结合性识别。这两个问题导致了身份识别标准日益面临法律上的重大挑战。

（一）信息的识别区分度问题

理论上，所有源于自然人的信息都包含着一定程度的身份识别概率，但不同类型信息的区分度有所差异，不同技术条件下能够识别的信息也有差异，导致身份识别概率呈现一种参差不齐、动态变化的态势，从而影响身份识别标准的判断结果。

随着数据挖掘与数据分析技术的发展，尤其是伴随深度学习（deep learning）技术的不断成熟，原先区分度不足的信息有可能在一定条件下成为具有身份识别能力的信息。易言之，可识别性的范围是动态发展的。〔51〕例如交通工具、地铁出入口及公共道路上的公共摄像设备所拍摄的视频本身并不必然属于个人信息，但如果运用人脸识别技术和数据挖掘技术，通过对足够多自然人单独或共同出入某些地点、乘坐交通工具的情况进行分析，就有可能识别出特定自然人的行动轨迹、人与人之间的许多潜在联系，进而推断出部分个体的工作单位、婚姻家庭状态、亲子关系甚至感情历史等等。又如，网络用户发帖中部分匿名化叙事的内容目前不属于个人信息，但随着文本作者身份识别（authorship identification）技术的发展，作者的身份一旦被确认，其叙事的内容就有可能涉及本人或他人的个人信息。当数据库中包含的信息量足够庞大，而这些信息又直接或间接来源于特定自然人，即便这些信息不属于个人信息，经由大数据挖掘与分析，也很有可能识别出特定自然人，而且难以通过简单的去识别化技术加以防范。〔52〕不仅如此，随着生物识别和痕迹检验等技术的发展，所有自然人的行为轨迹及相关的社会联系都有可能被破解，个人在公共场合产生的许多信息（如步态、文本、笔迹、驾驶习惯、语音特征等）都将可能成为识别特定自然人身份的关键信息，这给个人信息保护带来了棘手的难题。

这些难题本质上是信息识别区分度的程度高低与动态变化导致的。自技术原理层面观之，许多信息在应然上都具备区分能力，只是在当前的技术条件下未必能够准确区分出特定个体而已。但是，随着技术条件的发展，信息的区分度也会发生相应的变化，即使某些信息不能完全准确地定

〔50〕 见前注〔44〕，高富平文，第93页。

〔51〕 See N. Purtova, *supra* note〔16〕, at 47.

〔52〕 参见孙南翔：《论作为消费者的数据主体及其数据保护机制》，载《政治与法律》2018年第7期，第22页。

位到特定个体，也有可能将识别范围缩小到足可推测或跟踪特定个体的范围。例如，仅有IP地址的信息不能识别特定自然人，因为可能存在多人共用IP地址等情形，但IP地址信息往往伴随着网络活动时间等附随信息，有了IP地址的信息已经不难推测或追踪到特定自然人。〔53〕对于此种区分度尚未达到直接识别程度的信息（可以暂称为“部分可识别信息”），随着数学理论及数据分析技术的发展，区分度有可能取得实质性的突破。在当代，每一种信息的区分能力都是很难被精确评估的。因为从信息识别身份的能力高度依赖于技术条件，而数据挖掘与分析技术正处于迅速的发展变化之中，这给识别区分度的具体判断带来了显著的困难。

一旦数据分析技术产生了识别区分度方面的重大突破，从合法或公开途径挖掘与分析特定自然人的部分可识别信息将变得更加容易，坚持识别到特定自然人的身份识别标准就面临一个艰难的抉择：如果将部分可识别信息也纳入个人信息的范围，则个人信息保护的范围会进一步扩展到公共场合或网络公共空间产生的大量信息，信息保护制度有可能扩展至一个无所不包的范围，〔54〕个人信息保护成本将大大增加，且行政及司法机关很难及时对识别特定自然人的具体风险做出准确的评估或预测；如果坚持原有的标准，则个人信息保护的力度与效果将会大打折扣，并且产生了界定标准不一致的问题——如果自然人的财产状况或行程信息属于能够识别特定自然人的信息，那么其步态、笔迹、语音特征、行文习惯等难道不能够产生同样的区分性吗？这些在社会生活中难免会部分公开的信息，法律上又应当如何处理呢？身份识别标准面临的此种困惑，本质上是由于个人信息的保护包含着复杂的价值冲突，即个人信息背后隐私权、安宁权、人身与财产安全等价值与社会生活必不可少的公共性与便利性之间的利益冲突，远远超出了技术上的考量。

（二）信息的结合性识别问题

身份识别标准的范围不仅包括直接识别特定自然人的信息，还包括所谓的“间接识别信息”，例如我国立法中规定的“能够与其他信息结合识别特定自然人”的信息。随着信息技术的发展，许多数据分析方法可以通过匿名化信息与公开信息相结合的手段识别出个人信息，“结合性识别”也越来越成为身份识别标准的关键疑难问题。

如前所述，在我国的法律实践中，脱离个人身份或匿名化处理后的行为痕迹或活动轨迹被排除在个人信息的范围之外，因为仅有此种信息并不足以识别特定自然人。然而，此种个人信息保护范围及方式并非绝对能避免个人身份相关信息的泄露。在所谓的“去识别”（de-identification）之后，通过一定的手段，特别是依赖公开来源信息，数据挖掘者可以实现“再识别”（re-identification）。〔55〕个人在进行社会交往及社会活动时往往有一部分信息是在一定范围内公开的。例如个人在社交媒体上发布的私人生活、学者及科技工作者的部分参会信息、作家的签售会及读者见面会信息、娱乐工作者的部分商演行程及商业活动信息等等，这些人员的外形、衣着、车辆型号、语言习惯甚至所用手机型号等信息都有可能被相关参与者接触到，上述部分暴露信息与特定的行为痕迹或活动轨迹数据相结合，即有可能识别特定自然人的敏感信息。例如，单纯通过匿名的车辆轨迹信息（包括出租车轨迹），只要借助公开途径获得明星活动信息及含有出租车及时间戳的娱乐新闻照片，利用MD5加密算法的漏洞，一些社会名流的行动轨迹甚至住

〔53〕 See P. Schwartz & D. Solove, *supra* note〔7〕, at 1839.

〔54〕 See N. Purtova, *supra* note〔16〕, at 49.

〔55〕 See P. Schwartz & D. Solove, *supra* note〔7〕, at 1845.

址都被外界通过猜解的方法成功推测。〔56〕又如，如同生理指纹一样，每个人的行为都会留下数据指纹（data fingerprint），研究者们早已深入了解导致数据指纹形成的人类直觉（intuition）基础，只要从更容易获取的公开来源信息中发现相应的数据指纹，即可识别特定自然人。〔57〕

不仅如此，如果一部分企业已经掌握了一些经过用户同意收集或者通过公开渠道获得的个人信息，而又从另一部分企业或其他合法来源获取了匿名化的行为痕迹或活动轨迹数据，通过数据匹配（data matching）等方法即有可能获悉相关个体的更多个人信息甚至个人敏感信息，此种可能性早已具备现实基础，〔58〕也已为我国学者所洞见。〔59〕企业通过不同来源的数据合成大数据库，可以获得更充分的个人信息；甚至有时单独凭借搜索记录都可以从匿名化的行为痕迹信息中推断出特定自然人的身份。〔60〕代理商（data brokers）从各种渠道合法获取数据然后叠加于单个可检索的数据库中以对每一个用户获得更充分的了解，在美国也已经蔚然成风，行政机构也经常从数据代理商购买数据以丰富对相对人的认识。〔61〕一般情况下，信息掌握者积累的信息越多，就越有可能通过聚合操作（aggregation）或其他技术手段获得个人信息。〔62〕鉴于“商业机构在将数据去识别化之后的再应用已不可控制”，〔63〕如无深度介入的规制手段，此种个人信息的获取途径就很难受到实质性的限制。

易言之，识别能力的判断需要确定信息掌握者实际拥有或能够获取的其他数据与信息，它是一个高度依赖于语境（context，或译“场景”）的概念，个人信息在部分条件下不具备可识别性，并不意味着在所有条件下都不具备可识别性；而信息掌握者的数据资源及识别特定自然人的潜在技术可能性又是难以判断的。显而易见，目前排除匿名化处理信息的身份识别标准尽管一定程度上实现了对个人信息的保护，但它更有利于已经掌握了相当数据资源的大企业，因为它们可以较为容易地利用市场优势获取个人信息，也容易绕过身份识别标准交易数据并挖掘个人信息。对此，有学者主张，应当基于场景和风险导向的理念，“舍弃传统路径中全有全无的‘二元化’判断，转而进行‘程度性’评估”，〔64〕突破身份识别标准的局限性。

不仅如此，由于“结合性识别”问题的存在，身份识别标准还加剧了信息获取与整合能力的不对等性：企业可以通过大数据分析从并不属于个人信息的数据中挖掘出个人信息，而相比拥有海

---

〔56〕 See N. Purtova, *supra* note〔16〕, at 47 (2018); J. Trotter, “Public NYC Taxicab Database Lets You See How Celebrities Tip” (gawker.com/the-public-nyc-taxicab-database-that-accidentally-track-1646724546, last access 2019-04-15).

〔57〕 See Paul Ohm, *supra* note〔23〕, at 1723.

〔58〕 See Vagelis Papakonstantinou, “A Data Protection Approach to Data Matching Operations among Public Bodies”, 9 International Journal of Law and Information Technology 40 (2001); David C. Vladeck, “Consumer Protection in an Era of Big Data Analytics”, 42 Ohio Northern University Law Review 497-498 (2016); Paul M. Schwartz & Daniel J. Solove, “The PII Problem: Privacy and a New Concept of Personally Identifiable Information”, 86 New York University Law Review 1842 (2011).

〔59〕 参见岳林：《个人信息的身份识别标准》，载《上海大学学报（社会科学版）》2017年第6期，第35页。

〔60〕 见前注〔52〕，孙南翔文，第22页。

〔61〕 See D. Vladeck, “Consumer Protection in an Era of Big Data Analytics”, 42 Ohio Northern University Law Review 497-498 (2016).

〔62〕 See P. Schwartz & D. Solove, *supra* note〔7〕, at 1842.

〔63〕 张平：《大数据时代个人信息保护的立法选择》，载《北京大学学报（哲学社会科学版）》2017年第3期，第149页。

〔64〕 见前注〔15〕，范为文，第99页。

量数据资源及强大信息技术支持的大企业，绝大多数个体在防范数据挖掘和用户画像等方面既无专业经验亦无技术基础，对通过数据交易与信息整合形成的个人信息侵权风险缺乏清晰的认识，甚至对此种风险的具体细节可能一无所知，信息获取的告知与同意机制沦为摆设。〔65〕但是，如果对身份识别标准中的“能够与其他信息结合识别特定自然人”之范围作深入追问，则目前大数据开发利用的许多商业行为将面临法律上的困境，甚至对一些大型信息技术企业的核心盈利模式产生冲击。此外，如果一味仿照欧美国家，对企业施行严格的个人数据保护义务，包括对结合性识别进行严密的技术防范，更容易加重中小企业的责任和负担，伤害其发展。〔66〕

因此，如果仅仅规定了个人信息的身份识别标准，而缺乏针对特定信息类型的防御规则与机制，将很可能陷入一种两难处境：若对结合性识别采取严格的判定标准，则将使大数据产业的发展进退维谷；若对此采取宽松的判定标准，则将使个人信息保护制度在许多场合下形同虚设。

（三）个人信息保护标准的深层价值难题

个人信息保护标准面临的这两个难题，不仅仅是技术层面的，更是价值层面的。这两个难题之所以存在，本质上是因为侵权风险的高度不确定性，导致概念内涵的高度不确定性。自一般法理层面观之，“可被识别”本身就是一个不确定法律概念（unbestimmter rechtsbegriff）；由于它的判断包括对盖然性的预测，它是不确定法律概念中的“倾向概念”（dispositionsbegriff）。对它的解释不仅依赖于事实判断，也隐含着价值判断。〔67〕在法律规定不尽清晰之际，这种价值判断将决定个人信息保护范围的边界，身份识别标准要在每一个模糊个案中处理复杂的价值判断，这才是该标准所面临的深层问题。

回顾世界范围内的相关法律实践史，早在隐私权兴起之时，它与政府信息公开、言论和新闻自由等就发生过价值冲突，法院的权衡与考量并不必然固定地支持某一方，而是需要综合各种条件做出细致的判断。例如，对于有公共身份的人物或公共意义的事件之相关信息，就存在一个保密与公开的变化梯度：低级公务员可以享受更多的隐私保护，而信息公开的公共利益随着政府组织在等级阶梯（hierarchical ladder）中的上升而增加，政府雇员的等级、地位与其行为的重要性等均属于影响隐私保护之价值权衡的因素。〔68〕同样地，在美国，言论和新闻自由与隐私权之间的价值权衡也因其所涉及的对象和所牵涉公共利益的重要性而有所区分，如果被报道或评论的对象属于公众人物，并且该言论涉及公共利益，则言论和新闻自由一般处于优势地位；〔69〕但当此种言论涉及弱势群体成员的隐私权或者生活安宁时，即使与公共利益有关，言论和新闻自由亦不一定能够占据价值优势。〔70〕可见，在隐私权保护领域，价值权衡的方法早已不可或缺，它已经发展成为一种需要平衡多重价值因素的规范思维，广泛地影响隐私权保护的法律判断。

个人信息的身份识别标准亦处于类似的处境之中，它所隐含的价值冲突问题甚至更加复杂。

---

〔65〕 见前注〔1〕，高秦伟文，第21页。

〔66〕 参见邓明理：《大数据背景下个人数据的监管保护》，载《北京邮电大学学报（社会科学版）》2019年第1期，第24页。

〔67〕 参见王天华：《行政法上的不确定法律概念》，载《中国法学》2016年第3期，第77页。

〔68〕 See *Jesse C. Trentadue v. Integrity Committee*, 501 F.3d 1215 (2007); *Larry Tomscha v. General Services Administration*, 158 Fed. Appx. 329 (2005).

〔69〕 See *Alvarado, et al v. KOB-TV, LLC*, 493 F.3d 1210 (2007).

〔70〕 See *National Association of Retired Federal Employees v. Constance Horner, Director, Office of Personnel Management*, 879 F.2d 873 (1989).

对于一般自然人而言，单独的某些个人信息本身并不必然具有确定的实质性价值，它更多地体现为一种价值载体，其他主体有可能通过这一载体侵害个人的安宁权、隐私权、人身与财产安全等实质性价值目标。识别特定自然人的后果，并非必然使其遭受明确的、固定的损害，而是存在通过识别特定自然人而使用该信息造成侵害不特定法益的盖然性，而不同信息的识别对于各种法益所产生的侵害盖然性又存在较大差异。个人信息（包括隐私权保护）及其他合法权益之冲突，亦已为欧盟明确认识。2014 年，欧盟一个专门梳理此方面权益冲突的工作组发布了关于合法利益的指导性意见，认为表达和信息自由、艺术和科学自由、访问资料权、人身自由与安全权、思想信仰和宗教自由、从商自由、财产权、获得有效救济和公正审判权等都有可能与个人信息及隐私权保护之要求相冲突。〔71〕 在当前的中国，个人信息的法律保护不仅同样需要考量言论自由、政府信息公开等法律价值，更需要考虑数字经济（包括一些战略性新兴产业）、信息社会与信息科技的发展，考虑特定自然人的数据权利（例如我国《网络安全法》已确认的删除权、更正权及欧盟《通用数据保护条例》中的数据携带权、被遗忘权等），甚至考虑国际上的“长臂管辖”（long-arm jurisdiction）及法律冲突之情形，等等。

种种复杂的价值因素使得个人信息的界定标准面临选择上的难题：如果身份识别标准界定太窄，面对强大的现代数据技术，个人隐私将有较大的受侵害之风险；界定太宽，则有可能使个人信息法律制度的保护范围扩展到无所不包的信息，转变为一种麻烦且不可行（cumbersome and unworkable）的规制。〔72〕 事实上，大部分掌握信息的主体并不一定有足够的能力挖掘和分析出身份信息，即使有此种能力完成身份识别，也未必有动机耗费高昂的人力物力成本去进行识别，更未必会进一步导致对其他法益的侵害。对此，身份识别标准的既有学说中已经出现“客观说”“主观说”和“任一主体说”的争论，即究竟应当以一般公众、特定信息控制者还是社会中任何法律主体的识别能力作为判定可识别性的基准。〔73〕 虽然存在 Github、Stack Overflow 等共享信息技术发展的知识社区及专业人员之间的其他交流渠道，但信息技术的发展仍然存在技术能力差异大、各方信息不对称的态势，如果要求绝对避免身份识别的“零风险”标准，就会面临一个几乎不可能的任务：要证明某一时刻之前所有专业人士所掌握的所有信息技术都无法从一个数据集中识别特定自然人的身份是极为困难的。对此，身份识别标准势必要在“零风险”到“百分之百确定能被识别的风险”之间做出价值平衡与选择。“侵权风险准则”的法律实践也表明，个人信息保护范围的确定实际上是一个包含复杂价值判断的风险评价问题。

面对此种价值判断难题，法律需要首先确认一个关键前提：个人信息被识别的风险并非绝对确定、绝对严重的，个人信息被识别的系列相关风险与信息被识别的可能性有关，但更与信息或数据的类型及内容相关；因此，将不同内容的信息笼统地塞进同一个类型中进行保护的做法并不明智。〔74〕 更何况，与部分有统计学基础支持的风险规制及风险治理问题不同，个人信息保护标准的考量因素正在发生“从数学到社会学”的转变，正是因为从数学与统计学层面已经难以确定身份识别的可能性、难以预估风险发生的概率，才需要转换思路，从社会学的风险分析进路对这些风险进行评估。对于难以预测潜在危害和无法把握侵害概率的风险领域，认识它们的社会建构（social

〔71〕 参见谢琳：《大数据时代个人信息使用的合法利益豁免》，载《政法论坛》2019 年第 1 期，第 76 页。

〔72〕 See P. Schwartz & D. Solove, *supra* note〔7〕, at 1827.

〔73〕 见前注〔49〕，韩旭至书，第 41～43 页。

〔74〕 See P. Schwartz & D. Solove, *supra* note〔7〕, at 1876.

construction)本质并进而建立风险社会学或文化理论之分析视角,是很有必要的。〔75〕

在风险的文化理论和社会学认识中,风险背后的众多价值冲突不断随着社会生活的发展而动态变化,不同主体对于不同价值的考量亦截然有别,我们不可能对此确立统一的、客观的衡量准则,但可以从风险认知的主观层面着手破解风险治理的难题。如果我们放弃直接从林林总总的价值关系中直接确定某种价值秩序或衡量基准的思路,从文化理论和社会学进路上认识个人信息被识别的风险,可以考虑雷纳(Steve Rayner)教授提出的多重定义"公式"——风险(R)=概率(P)×后果大小(M)+信任(T)·责任(L)·认同(C),后三者不一定是乘积关系,但对风险的总体评价结果都有着某种正相关性。在风险的工程学一端,这个"公式"变为R=P×M;在社会学一端,"公式"就变为R=T·L·C,它主要考虑的因素是对风险制造和治理者的信任程度、受风险影响者的可接受程度和责任承担者(利益相对方)的承担能力。〔76〕这一公式对信息技术背景下的风险治理有重要启发:即使某些风险无法在客观层面上被充分计算、彻底消除,也可以在主观层面上,通过价值可认同、过程可信任、责任可追究的风险治理机制,与前述"从数学到社会学"的个人信息保护思路相契合,使风险在主观层面上被消融。

一般情况下,个人信息所包含的不同内容足以引起"公式"在社会学一端对三种要素的不同考量。例如,在信任机制方面,收集与特定商业应用或行政职能无关的活动信息或身份信息更难以赢得用户或相对人的信任;在认同或可接受程度方面,一般公民对于个人敏感信息被泄露的后果将更难以接受,也更难以认同企业在无充分相关性的前提下利用个人敏感信息;在责任的可追溯性方面,个人信息的外延越是明确、具体,就越容易保证责任追究的清晰性;在责任承担能力方面,如果要求企业防止未经用户同意采集、传输和利用可被直接识别的身份信息,企业是完全可以遵守此种义务的,因为风险预防成本相对较低、可以控制在合理的可承受范围;反之,如果所有信息和数据的传输与利用都要经过严格的逐一授权和再识别风险监测,企业将很可能面临过度沉重的负担。这些因素都将成为个人信息保护的重要考量因素。不仅如此,类似于风险治理过程,受影响的个人及组织也需要有充分的机会参与到个人信息的界定与保护过程,使得个人信息的界定能够符合利害相关者对风险的认知与评价,实现动态的、多元的价值平衡。

因此,个人信息的法律保护不仅需要引入能够平衡多种价值的衡量方法与过程,更需要注重风险的社会建构内涵,根据个人信息的不同类型及范围,发展不同的法律调整机制及保护方式。我们无法一劳永逸地解决衡量信息识别区分度及结合性识别问题,但可以分门别类、循序渐进地对不同范围内的个人信息给予不同程度的保护,对个人信息保护逐步建立价值共识,尽可能降低与个人信息相关的各种法益受重大侵害的可能性。这就需要形成个人信息分类界定的一系列标准,同时建立与之相应的个人信息保护机制,从而超越目前简单而含混的可识别性判断。

四、身份识别标准的变革与完善

个人信息保护标准需要变革与完善。单有笼统的可识别性不足以完成与其他复杂法益或法律价值之间的平衡与整合,所有个人信息也不必然具备同等的保护理由基础,不必然需要相

〔75〕 See K. Tierney, "Toward a Critical Sociology of Risk", 14 Sociology Forum 221-222 (1999).

〔76〕 参见[英]史蒂夫·雷纳:《文化理论与风险分析》,载[英]克里姆斯基、[英]戈尔丁编著:《风险的社会理论学说》,徐元玲等译,北京出版社2005年版,第106页。

同的标准予以界定及保护。面对日益精致而复杂的信息社会及科学技术条件，立法者若不考虑偏离目前仍在世界范围内占主导地位的身份识别标准，则可以在身份识别标准的基础上，根据不同的风险评估与价值考量之结果，发展出一系列有关个人信息的分类保护标准。将单一的身份识别标准发展成一个差异化的界定标准体系，是信息法学面临的必然任务。这些标准的确立与发展不可能经由理论建构之努力而一蹴而就，但形成这些标准所需要遵循的基本要点却已初见端倪：

首先，应当建立个人信息分类保护制度，对不同类别的个人信息实行不同强度的外延式保护。根据目前的法律和政策实践，从个人信息泄露的后果出发，个人信息可以分为（常规）个人信息与个人敏感信息；〔77〕同时，个人信息又可以根据识别条件与识别内容的不同作进一步细分，例如前文所述的“身份识别信息”和“活动情况信息”；又如可以将直接识别特定自然人的数据或信息称为“直接识别信息”，将只能与其他信息或数据结合而识别特定自然人的信息称为“间接识别信息”；此外，还可以根据需要特别保护的群体专门界定部分受特别保护的个人信息，例如2019年国家互联网信息办公室起草的《儿童个人信息网络保护规定》中的“儿童个人信息”。〔78〕在可识别性这一总体标准之下根据识别条件、识别内容及泄露后果作进一步的分类，有利于针对不同信息类型建立更加精细的界定标准。其中，对信息主体更加关切、更难以接受被泄露、更可能产生较大识别风险及危害后果的直接识别信息与个人敏感信息，必须采取更严格的法律保护措施。例如，对于个人直接识别信息，法律可以要求明确的、专门的用户同意或授权，禁止未经再次特别同意的数据传输、处理与流通；对于个人敏感信息，则更要针对所有相关应用场景，建立充分的信任机制：获取和使用个人信息不仅需要遵循“多重因素检测”的考量，〔79〕还需要设置安全、合理、可验证的匿名化利用规则，保证处理的程度达到当前已公开可知的信息技术不能实现再识别的严格标准（例如欧盟《通用数据保护条例》第26条所规定的“所有合理可能使用方式”之标准），并且明确要求匿名化处理标准随技术条件的改变而动态调整。在不同的界定标准与分类维度之下，法律规则可以叠加适用，力争取得最佳的个人信息保护效果。

其次，对于目录内不同类型的个人信息，应当建立起有一定差异性的认定程序，最终由主管部门发布个人信息分类目录，并定期进行动态调整。目录内个人信息的认定程序可以有一个共通的基础框架。例如，目录的初始草案可以由行政主管部门、数据代理商、技术与法律专家以及第三方机构等共同起草；所有法律规范中明确规定的各类个人信息、实践中已出现侵权后果或侵权风险的信息以及技术专家认为很可能造成侵权风险的信息，都可以按照此前所述的分类列入目录，并根据法律实践的经验和信息技术的发展由主管部门对目录进行定期修订，避免身份识别标准在实践中的僵化与停滞。同时，个人信息目录的确定和调整应当广泛吸收各方专业人士与社会公众的意见，凝聚有关个人信息利用与风险方面的社会共识，逐步形成相对清晰、稳定的个人信息保护范

〔77〕 根据全国信息安全标准化技术委员会组织制定和归口管理的国家标准《信息安全技术　个人信息安全规范》（GB/T 35273—2017）（2017年通过），个人信息是指“以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人身份或者反映特定自然人活动情况的各种信息”。个人敏感信息是指“一旦泄露、非法提供或滥用可能危害人身和财产安全，极易导致个人名誉、身心健康受到损害或歧视性待遇等的个人信息”。

〔78〕 直接与间接、一般与敏感的个人信息类型划分相对接受度较高，亦多见于其他学者的理论主张之中，参见项定宜：《个人信息的类型化分析及区分保护》，载《重庆邮电大学学报（社会科学版）》2017年第1期，第33～35页；陶盈：《我国网络信息化进程中新型个人信息的合理利用与法律规制》，载《山东大学学报（哲学社会科学版）》2016年第2期，第56页等。

〔79〕 参见胡文涛：《我国个人敏感信息界定之构想》，载《中国法学》2018年第5期，第249页。

围。目录内若干类型信息的社会风险触角更加敏感，也就需要更为充分、更具利益代表色彩的多元互动过程。例如，对于个人敏感信息和特殊群体的个人信息，在形成目录的过程中还应当更加注意融合风险社会学的思考，增强用户及易受影响人群（例如病人、未成年人）的参与比例及程序影响力；对于特殊群体的个人敏感信息，还可以赋予相关群体更灵活的程序启动权甚至一定比例的决定权，使易受影响的群体获得更强的公众充能（civil empowerment）。由此，多元参与、动态调整的个人信息目录制度将为公民维护个人信息权利提供更具正当性、更易清晰理解的依据，为企业利用数据资源提供更具体可行的标准，也为执法和司法机关认定个人信息提供更富操作性的实务指引。

再者，应当结合外延式保护及内涵式保护的双重进路，对不同类型的个人信息提供有层次的界定方式。目前法律实践中已被普遍确认的、可以直接识别特定自然人的数据或信息，应当在法律或个人信息目录中通过列举的方式直接明确其法律地位。除日常用于识别个人的姓名及照片外，对于诸如手机号码、电子邮箱、基因序列、身份证号码、即时通讯软件的用户名或号码等可以一一对应地直接识别个体的信息，有关部门应尽可能在立法或法定机构发布的个人信息目录中直接明确列示为个人信息，从外延入手确立其法律地位，防止执法与司法者基于对个人信息内涵的不同认识而产生各种各样的分歧与矛盾。同理，对于需要严格保护的个人敏感信息等亦应尽可能通过立法或个人信息目录的方式直接列示，使得此类个人信息的保护有更充分的法律依据。相反，对于一些一时不便列入或未明确是否列入目录的间接识别信息及常规个人信息，可以更多地通过抽象性规定的方式加以界定，或者通过更灵活、更频繁的调整程序，由执法与司法机关在法律实践中因时制宜、因事制宜，进行具体风险评估与价值权衡。无论是具体的信息目录还是抽象的界定标准均是动态、可调整的，任意一项信息都可随科学技术及社会风险认知的发展而被移入或移出保护范围。

最后，对于目录外需要一定专业技术基础方可认定的间接识别信息，应当通过执法和司法实践的不断归纳总结，结合技术专家的意见，形成更具操作性的界定标准，或者进一步形成临时性的参考清单。间接识别信息的界定颇为复杂，我们固然应当注意吸纳技术专家对于识别特定自然人身份及活动轨迹的风险分析，但更要注重将侵权风险、侵害结果与规制此种信息传播与利用的社会经济成本以及规制的其他不利影响作综合权衡。在信息社会中，鉴于自由信息流（free information flow）具有重要价值，〔80〕对间接识别信息的认定与限制应当慎重。〔81〕如果采取临时清单的措施，主管部门可以先根据实践中已出现个人信息侵权风险及技术专家认为很可能造成侵权风险的间接识别信息建立专门的清单，对采集、传输与处理清单内信息的活动实行备案、报告或留存操作记录等干预程度较低的措施，再根据具体技术条件判断这些信息是否具备可识别性，进而决定是否需要纳入正式的目录，并为信息来源主体提供异议与救济途径。对于可能涉及大量个人信息的数据交易等行为，还应当引入事中的风险评估机制，确保其中的个人信息已经被匿名化处理且不能被再识别。〔82〕此外，法律还应当明确禁止数据的占有人和使用人不得进行个人身份的再识别，并要求参与数据交易及大数据开发利用的企业明确声明不从事个人身份再识别活动，

〔80〕 See Paul Ohm, *supra* note〔23〕, at 1769.

〔81〕 间接可识别信息在大数据应用中有重要价值，不能被一概排除在信息经济之外，参见金耀：《个人信息去身份的法理基础与规范重塑》，载《法学评论》2017年第3期，第129页。

〔82〕 参见齐爱民、张哲：《识别与再识别：个人信息的概念界定与立法选择》，载《重庆大学学报（社会科学版）》2018年第2期，第129页。

甚至形成“去身份化”的专门标准。[83] 通过以上手段，政府对间接识别信息的规制能够尽可能满足比例原则的要求，避免给信息技术产业造成过度的负担，同时也有力地促进个人信息保护目标的实现。

结　　语

在信息社会中，信息已经成为经济建设与社会发展的重要战略资源，个人信息的法律地位亦不断提升。确定个人信息的识别标准与保护范围，已经成为信息法学的最关键议题之一。个人信息保护任重而道远，本文无意于直接提出一整套个人信息界定标准，但求获抛砖引玉之效，与学界及实务界同仁一道探求和增进个人信息保护的社会共识，为信息时代的基础性法律命题共同勉力。

Abstract　China's current legislation defines personal information by the "identification standard", which means someone is able to identify the identity of a particular natural person from such information. This standard has a broad meaning and covers a wide range, but faces the following problems: the concept of "identifiability" is blurred, the degree of identification is not well-considered, and the binding identification of information cannot be forecast. Under the premise of establishing a classified personal information protection system, the single identification standard should be developed into a differentiated and dynamically adjusted standard system for different categories of information. The system should be established to carry out denotative definition and protection of different types of personal information with different intensity. For different types of personal information in the directory, differentiated procedures of defining different personal pieces of information should be established and finally published by the competent departments as a personal information directory with regular dynamic adjustment. It is necessary to provide a hierarchical definition of different types of personal information in the light of denotative and connotative recognition. More operational criteria or even further a temporary reference list should be formed with the help of continuous reflection of relevant law enforcement, judicial practice and the views of technical experts.

Keywords　Personal Information, Identification Standards, Identifiaibility, Jurisprudence of Information Law

（责任编辑：徐彦冰）

〔83〕 见前注〔81〕，金耀文，第129～130页。