■市场经济法治 文章编号:1001-2397(2022)01-0125-19

政府数据开放中个人信息保护的范式转变

张 涛

(清华大学 法学院,北京 100084)

摘 要:政府数据开放并非静态的单一行为,而是动态的系统过程。借助数据生命周期理论,可以将政府数据开放解构为数据收集、转换、存储、公开和使用五个阶段。根据《个人信息保护法》和《数据安全法》确立的最新规则,个人信息保护风险可能同时存在于政府数据开放生命周期的各个阶段。然而,政府数据开放中现有的个人信息保护范式主要采取“基于结果的方法”,重点关注政府数据在公开时的状态,依靠技术性匿名化手段,难以有效应对政府数据开放中的个人信息保护风险。与此相对应,“基于过程的方法”与政府数据生命周期、个人信息保护的程序化和数据安全全流程管理相契合,可以弥补“基于结果的方法”的不足。通过将风险预防原则和程序、技术、经济、教育和法律等手段分散放置在政府数据开放生命周期的每个阶段,能够最大限度减少个人信息保护风险,在个人信息保护与政府数据开放之间实现动态平衡。

关键词:政府数据开放;个人信息保护;基于过程的方法;匿名化

中图分类号:DF36 **文献标志码**:A

DOI:10.3969/j.issn.1001-2397.2022.01.09 **开放科学(资源服务)标识码(OSID)**:

一、问题的提出

随着“数据”成为第五类生产要素,数据开放共享对于数字经济发展和数字社会建设至关重要。政府部门维护着大量数据,这些数据已经成为一种宝贵的资源,可以被从个人到企业甚至其他政府机构的各种实体所利用。开放数据是释放政府数据价值的重要方法,它意味着任何人都可以从任何渠道获取以公开形式存在,并且满足一些特定条件的政府数据。[①]事实证明,政府数据开放不仅具有经济价值,如促进产业转型、助推大众创业等,而且还具有社会价值和政治价值,如提升公众生活

收稿日期:2021-12-01

基金项目:2019 年国家社科基金重大项目“大数据、人工智能背景下的公安法治建设研究”(19ZDA165);2021 年度教育部人文社会科学研究青年基金项目“政府数据开放利用机制研究”(21YJC820035)

作者简介:张涛(1991),男,贵州铜仁人,清华大学法学院助理研究员、博士后,法学博士。

① [美]乔尔·古林:《开放数据——如何从无处不在的免费数据中发掘创意和商机》,张尚轩译,中信出版社 2015 年版,第 6 页。

品质、增强政府透明度、优化公共决策水平等。[①] 政府数据开放的重要性也获得了我国有关政策文件和立法的肯认,正逐渐向法制化方向发展。[②] 2015 年 8 月,国务院印发的《促进大数据发展行动纲要》将"加快政府数据开放共享,推动资源整合,提升治理能力"确立为促进大数据发展的"主要任务";2020 年 3 月,中共中央、国务院印发的《关于构建更加完善的要素市场化配置体制机制的意见》将"推进政府数据开放共享"作为"加快培育数据要素市场"重要举措。贵州、浙江、上海等地通过地方性法规、地方政府规章的形式对政府数据开放共享予以专门规范,如《贵州省政府数据共享开放条例》《上海市公共数据开放暂行办法》等。

然而,政府数据开放在创造巨大价值的同时,也带来潜在风险,主要包括:一是开放数据本身有可能泄露国家秘密、商业机密和个人隐私;二是开放数据被误用或滥用后会损害公共利益及第三方利益;三是开放数据由于质量问题会对数据使用者和社会造成损失。[③] 如何平衡政府数据开放的效用与风险,已经成为一个亟待解决的重要课题。[④] 实证研究表明,与泄露机密信息(如个人信息和商业秘密)相关的风险已经成为政府部门拒绝开放共享其数据的主要理由。[⑤] 随着我国《个人信息保护法》和《数据安全法》的出台,与个人信息保护相关的原则、规则以及机制得以确立,这意味着政府部门在处理数据开放共享与个人信息保护的关系时将面临新的、更高的要求。[⑥] 现有的研究成果主要聚焦于政府数据开放中的隐私保护,倡导将隐私法的一些原则和规则运用到政府数据开放中,以求在隐私利益和其他利益之间维持平衡。[⑦] 然而,传统的隐私法过于关注个人信息的收集、使用或者披露的信息的性质,当具体的损害难以表述,甚至难以定位时,事后的、个性化的补救措施常常会失去效用。[⑧]

此外,个人信息与个人隐私之间存在明显的区别[⑨],而现有的隐私法规范也未给政府数据开放提供明确的指引,这意味着我们需要一个更为复杂和精细的数据保护方法来为个人数据提供强有力的保护,并提高政府数据开放的效用。[⑩] 本文便以此作为重点,探讨相关问题,密切关注法学、数据科学、统计学等学科在个人信息保护方面的最新进展,并通过如下结构展开论证:首先,以数据生命周期理论为指引对政府数据开放进行阶段性解构,并以《个人信息保护法》和《数据安全法》的有关规定,检视不同阶段可能存在的个人信息保护风险。其次,对政府数据开放中现有的个人信息保护模式进行评析,在此基础上,尝试以"基于过程的方法"来建构新的个人信息保护范式。最后,从

---

① 张涛:《藏智于民:开放政府数据的法理基础与规范重塑》,载《电子政务》2019 年第 8 期,第 78-80 页。

② 张涛:《开放政府数据法制化的地方实践与制度完善——以浙江等 9 个省市为分析样本》,载《贵州大学学报(社会科学版)》2019 年第 5 期,第 78 页。

③ 郑磊:《开放的数林:政府数据开放的中国故事》,上海人民出版社 2018 年版,第 21 页。

④ 蒋冰晶、李少军:《包容与合作:大数据时代政府数据开放的行政法治理念》,载《河北法学》2019 年第 12 期,第 103 页。

⑤ *See* Bernd W. Wirtz, Robert Piehler & Marc-Julian Thomas et al., *Resistance of Public Personnel to Open Government: A cognitive theory view of implementation barriers towards open government data*, 18 Public Management Review 1335, 1356(2016).

⑥ 商希雪、韩海庭:《政府数据开放中个人信息保护路径研究》,载《电子政务》2021 年第 6 期,第 113 页。

⑦ 邹东升:《政府开放数据和个人隐私保护:加拿大的例证》,载《中国行政管理》2018 年第 6 期,第 75 页。

⑧ 王学辉、赵昕:《隐私权之公私法整合保护探索——以"大数据时代"个人信息隐私为分析视点》,载《河北法学》2015 年第 5 期,第 65 页。

⑨ 王利明:《论个人信息权的法律保护——以个人信息权与隐私权的界分为中心》,载《现代法学》2013 年第 4 期,第 66-68 页。

⑩ *See* Ira S. Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 Washington Law Review 703, 720-721(2016).

不同角度就如何在政府数据开放中落实“基于过程的”个人信息保护机制提出建议。

二、政府数据开放中个人信息保护面临的风险

从广义上看,政府数据开放是一个动态的系统过程,而非静态的单一行为。[①] 在数据科学中,学者们用“数据生命周期”理论来描述数据从“产生”到“解释”、从原始比特转化为终端用户价值的过程,并且强调在生命周期的每个阶段都需要考虑数据隐私和数据伦理。[②] 为了能够对政府数据开放中个人信息保护面临的风险有一个较为准确的把握,我们有必要以数据生命周期理论为指引对政府数据开放过程进行阶段性解构,并具体分析每个阶段可能存在的风险。

(一)政府数据开放过程的阶段性解释

在数据科学中,一般认为,“数据生命周期”概念的目标是“提供一种结构来组织与项目或者组织内的数据管理有关的任务和活动”。[③] 这个概念被具体化为各种数据生命周期模型,涵盖了数据从生产到归档(或者删除)的整个生命周期,并将整个过程视为同一过程的下一次迭代,使其形成一个循环。这样一个概念对于政府数据开放而言至关重要,因为它“提供了一个结构来考虑数据记录在整个生命周期中需要执行的诸多操作”。[④]

不过,数据生命周期理论仍面临挑战,那就是没有统一的数据生命周期模型,因为数据生命周期在每个领域、场域或组织中都存在差异。尽管如此,仍然有一些学者尝试开发出应用于政府数据开放的数据生命周期模型。德国波恩大学学者朱迪·阿塔德(Judie Attard)等人将政府数据开放生命周期(open government data life-cycle)划分为“3个板块9个阶段”:预处理(pre-processing)板块是准备要发布的数据,包括创建、选择、协调、发布等4个阶段;开采(exploitation)板块是使用已发布的数据,包括互联、发现、探索、挖掘等4个阶段;维护(maintenance)板块是维护已发布的数据,使其具有可持续性,包括管护。[⑤] 希腊爱琴海大学学者亚尼斯·查拉比迪斯(Yannis Charalabidis)等人在开放数据支持工作组提出的链接OGD(Open Government Data)生命周期的基础上,开发了一个扩展的政府数据开放生命周期,它由9个阶段构成,包括创建、预处理、策划、存储/获得(store/obtain)、发布、检索/购得(retrieve/acquire)、处理、使用和用户协作。[⑥] 国内学者鲍静等人将政府数据开放生命周期划分为6个阶段,包括数据生成和发布、权限配置管理、网上流转、数据呈现、利用管理和更新管理。[⑦] 黄如花等人则将政府数据开放生命周期划分为5个阶段,包括创建与采集、组织与处理、存

---

① *See* Judie Attard, Fabrizio Orlandi & Simon Scerri et al., *A Systematic Review of Open Government Data Initiatives*, 32 Government Information Quarterly 399, 399(2015).

② *See* Jeannette M. Wing, *The Data Life Cycle*, 1 Harvard Data Science Review 1, 1(2019).

③ Line Pouchard, *Revisiting the Data Lifecycle with Big Data Curation*, 10 International Journal of Digital Curation 176, 180(2015).

④ Alex Ball, *Review of Data Management Lifecycle Models* (*version* 1.0), University of Bath, 2012, p.4.

⑤ *See* JSee Judie Attard, Fabrizio Orlandi & Simon Scerri et al., *A Systematic Review of Open Government Data Initiatives*, 32 Government Information Quarterly 399, 403(2015).

⑥ *See* Yannis Charalabidis, Charalampos Alexopoulos & Euripidis Loukis, *A Taxonomy of Open Government Data Research Areas and Topics*, 26 Journal of Organizational Computing and Electronic Commerce 41, 56(2016).

⑦ 鲍静、张勇进、董占广:《我国政府数据开放管理若干基本问题研究》,载《行政论坛》2017年第1期,第28页。

储与发布、发现与获取、增值与评价。[①]

事实上,数据的不同用途可能对应不同的生命周期。例如,数据可能只是为了保存记录而被归档,也可能被用于一次性的法律决策,还有可能在决策支持系统中被持续处理。当我们在开发一个特定的数据生命周期模型时,面临着在通用性和复杂性之间进行权衡:数据生命周期模型越复杂,它就越能描述简单个案,但在描述其他大数据使用案例时却可能缺乏通用性。[②] 为了在通用性与复杂性之间取得平衡,以现有的研究成果为基础,本文采用的政府数据开放生命周期模型主要有两个目标:一是模型足够简单,以概括其他法律领域中许多大数据用例的共同特征;二是模型足够具体,以捕获政府数据开放过程中有意义的、独特的"阶段"。在这个数据生命周期模型中,主要分为 5 个不同的阶段:数据收集、数据转换、数据存储、数据公开和数据使用,下文将分别对这 5 个阶段进行简单描述,同时分析可能存在的个人信息保护风险。

(二)数据收集阶段侵害个人信息权益

政府数据开放生命周期的第一个阶段始于数据收集。本文在广义上使用"收集"一词,包括接受、提取或者获取数据。数据无处不在,正以不易察觉又显而易见的方式嵌入"我们日常生活的结构"中,而技术正在改变数据产生、收集、维护和利用的方式。[③] 在大数据时代,政府数据的收集主要呈以下特点:(1)政府数据收集的主体越来越多元化。除了传统负责交通运输、环境保护、治安管理、教育卫生、文化旅游等业务的行政机关在履行职责的过程中会收集各种数据外,代行政府管理职能的组织在履行职责的过程中也会采集各类数据。(2)政府数据收集的类型越来越多样化。除了个人信息以外,环境数据、气象数据、税务数据、交通数据等也在政府数据收集的范围之内。(3)政府数据收集的方式越来越隐蔽和便捷。在传统的行政管理或者政务服务中,行政机关往往通过线下访问以纸质文件等形式来采集数据。随着数字政府建设的不断推进,很多行政任务由线下转为线上办理,行政机关可以借助移动应用程序、生物识别设备等快速、无接触地采集数据。(4)政府数据收集的途径更加多元。行政机关除了可以通过在履行法定职责时直接收集数据外,还可能从第三方数据中介组织或者其他政府部门收集数据。

大数据技术虽然给政府数据收集带来了诸多革命性变革,但也存在违反《个人信息保护法》的风险,主要体现在以下几个方面:

1. 过度收集个人信息。《个人信息保护法》第 34 条规定,国家机关为履行法定职责收集个人信息,不得超出履行法定职责所必需的范围和限度。实践中,一些政府部门在收集个人信息时,往往采取"应采尽采、应归尽归"的方法,超越职责和权限收集个人信息。2020 年 12 月,APP 违法违规收集使用个人信息治理工作组发布了 35 款存在个人信息收集使用问题的 APP,其中,安徽省数据资源管理局负责运营的"皖事通"存在"未明示收集用户详细地址、支付宝账号、社保账号等个人信息

① 黄如花、赖彤:《数据生命周期视角下我国政府数据开放的障碍研究》,载《情报理论与实践》2018 年第 2 期,第 8 页。

② *See* Simon Vydra, Andrei Poama & Sarah Giest et al., *Big Data Ethics: A Life Cycle Perspective*, Erasmus Law Review, Issue 1, 2021, http://www.erasmuslawreview.nl/tijdschrift/ELR/2021/1%20(incomplete)/ELR-D-20-00036.pdf(Last visited on July 11, 2021).

③ *See* Barbara L. Cohn, *Data Governance: A Quality Imperative in the Era of Big Data, Open Data and Beyond*, 10 A Journal of Law and Policy for the Information Society 811, 811(2015).

的目的、方式和范围”。①

2. 未履行告知义务收集个人信息。《个人信息保护法》第 35 条规定,国家机关为履行法定职责处理个人信息,应当履行告知义务。在 APP 违法违规收集使用个人信息治理工作组发布的 35 款存在个人信息收集使用问题的 APP 中,由湖北省人民政府主办、湖北省楚天云有限公司运营的“鄂汇办”存在“未明示收集的人脸特征等个人信息的目的、方式和范围,且收集时未同步告知用户其目的”。②

3. 未经同意收集个人信息。根据《个人信息保护法》第 13 条的规定,国家机关不需取得个人同意而处理个人信息的条件是“为履行法定职责或者法定义务所必需”,这意味着若收集的个人信息并非履行法定职责所必需,则仍然需要取得个人同意。在前面提到的“皖事通”和“鄂汇办”都存在未经同意收集个人信息的问题,如“鄂汇办”在“用户明确表示不同意打开位置权限后,仍频繁征求用户同意,干扰用户正常使用”。③

(三)数据转换阶段侵害个人信息权益

数据转换(data transformation)是将一种格式或者状态的数据转换为对另一种目的有用的格式或状态的过程。④ 数据转换的目的是使数据可使用、可管理,并符合现行的数据治理标准,一旦数据被转换,数据使用者就可以使用分析方法,从中获得可信赖的、可操作的信息。数据转换是“提取、转换和加载”过程的中间行动,为分析准备数据,而“提取、转换和加载”是一个数据管道,用于从各种来源收集数据,根据业务规则转换数据,并将其加载到一个目标数据存储。⑤ 数据转换通常涉及各种操作,如过滤、排序、聚合、连接数据、清洗数据、重复数据和验证数据等;数据转换通常也包括一系列的改变,转换可能是结构性的或者语义性的,也可能是有损的或者无损的。⑥

就政府数据开放而言,数据转换至关重要,原因主要有两点:第一,数据转换是政府数据达到可机读、非专属等数据标准的必经阶段。为了最大限度发挥政府数据的效用,各国政府或者国际组织在制定政府数据开放政策、战略、立法或者倡议时均对数据标准作出规定,要求政府数据在公开时应当满足可机读、非专属、以接口形式提供等标准。⑦ 政府数据来源广泛、类型众多,结构性数据与非结构性数据并存,要达到预期的数据标准,需要利用各种技术手段,进行数据转换。第二,数据转换是政府数据由收集阶段迈入存储阶段的中间桥梁。如前所述,政府部门借助各种信息技术、媒介或者平台从各种来源收集数据,如官民互动数据、行政文书、数据库或者由机器与传感器产生的流

① 申佳平:《35 款 APP 存在个人信息收集问题 “鄂汇办”等多个政务平台被通报》,载人民网 2020 年 11 月 13 日,https://baijiahao.baidu.com/s?id=1683240493240792290&wfr=spider&for=pc。

② 申佳平:《35 款 APP 存在个人信息收集问题 “鄂汇办”等多个政务平台被通报》,载人民网 2020 年 11 月 13 日,https://baijiahao.baidu.com/s?id=1683240493240792290&wfr=spider&for=pc。

③ 申佳平:《35 款 APP 存在个人信息收集问题 “鄂汇办”等多个政务平台被通报》,载人民网 2020 年 11 月 13 日,https://baijiahao.baidu.com/s?id=1683240493240792290&wfr=spider&for=pc。

④ 朱东妹:《数据仓库与数据挖掘概念、方法及图书馆应用》,安徽师范大学出版社 2017 年版,第 38-39 页。

⑤ 罗会兰:《数据提取、转换和装载技术研究》,载《计算机工程与设计》2004 年第 5 期,第 764 页。

⑥ *See* Micah Altman, Alexandra Wood & David R. O'Brien et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkeley Technology Law Journal 1967, 2020(2015).

⑦ *See* Ashit Talukder, *Big Data Open Standards and Benchmarking to Foster Innovation*, 10 A Journal of Law and Policy for the Information Society 799, 802-803(2015).

媒体数据,这些数据只有经过适当转换后,才能加载到云数据存储库。

与数据收集相比,数据转换虽然并不直接与数据主体打交道,但也可能因为一些主观或客观因素的影响,侵害个人信息权益,其中最主要的风险是可能违反准确性义务(原则)。从国内外个人信息保护立法的现状与趋势来看,准确性都被视为一项重要的原则或者义务。欧盟《通用数据保护条例》第 5 条第 1 款第(d)项规定,个人数据必须准确、及时、保持更新。新加坡《个人数据保护法》第 23 条规定,机构应当作出合理努力以确保由机构或者代表机构收集的个人数据是准确且完整的。我国《个人信息保护法》第 8 条规定:"处理个人信息应当保证个人信息的质量,避免因个人信息不准确、不完整对个人权益造成不利影响。"在实践中,大部分与数据打交道的人都知道,利用数据转化歪曲事实是有可能的。达莱尔·哈夫的经典之作《统计数字会撒谎》描述了数据可以被歪曲的事实,同时创造一个事实的虚假表象,方法主要包括主观的数据选择、范围的操控、部分数据点遗漏,这些方法直到今天还在使用。① 2021 年 5 月,江苏南通一位市民在查询个人征信时发现,其征信报告"工作单位"一栏被写上了"专业做×十年",该事件引发社会公众对征信机构公信力的质疑,其主要原因是征信机构在处理个人信息时未履行准确性义务。②

(四)数据存储阶段侵害个人信息权益

数据存储是指记录和保存数字信息,如应用程序、网络协议、文档、媒体、地址簿、用户偏好等背后的比特和字节,用于未来的操作。数据存储是大数据的核心组成部分,也是政府数据开放生命周期中的重要阶段。③ 在某种程度上,数据的创造以及大数据概念的诞生,正是计算机的发展、数字数据取代模拟数据的进步,以及处理和存储数据的速率提高等因素的结果。④ 在大数据技术发展的早期,数据存储的成本十分高昂,通常以模拟数据的方式进行代替,如缩微拍摄、摄影以及纸媒等。随着技术的进步,计算环境解决了存储模拟数据方面的诸多难题,目前,常见的数据存储类型包括软件定义存储、云存储、网络附加存储、对象存储、文件存储、块存储等。数据存储技术的发展与广泛运用,也对政府数据存储产生了深远影响,使得计算环境中的政府数据具有以下特征:第一,数据可以得到完整存储;第二,数据变得易于复制;第三,数据有高度的可访问性;第四,较之物理存储,数据存储的成本显著降低。⑤

尽管政府数据存储为后续的数据公开、使用奠定了基础,但仍然可能对个人信息权益造成侵害,主要体现在以下几个方面。

1. 无限期存储数据可能违反存储限制。从国内外个人信息保护立法的现状来看,存储限制是一项重要的原则。欧盟《通用数据保护条例》第 5 条第 1 款第(e)项对"存储限制"进行了规定,个人数据允许以数据主体可识别的形式保存,保存时间不得超过处理个人数据所需的时间。我国《个人

① [美]DAMA 国际(Data Management International):《DAMA 数据管理知识体系指南》,DAMA 中国分会翻译组译,机械工业出版社 2020 年版,第 34 页。

② 《如此儿戏! 女子个人征信报告被写"专业做鸡十年"》,载澎湃新闻 2021 年 5 月 25 日,http://m.thepaper.cn/baijiahao_12853184。

③ *See* Gherardo Carullo & Christian Ernst, *Data Storage by Public Administrations*, 26 European Public Law 545, 545(2020).

④ [美]道恩·霍尔姆斯:《大数据》,李德俊、洪艳青译,译林出版社 2020 年版,第 14 页。

⑤ [美]拉塞尔·沃克:《从大数据到巨额利润》,王正林译,广东人民出版社 2019 年版,第 8 页。

信息保护法》第19条规定:“除法律、行政法规另有规定外,个人信息的保存期限应当为实现处理目的所必要的最短时间。”在实践中,各种类型的数据库不断建立,如社会保障信息数据库、个人身份信息数据库、公共信用信息数据库、车辆识别信息数据库等,而不同类别的数据所需的保存期限不同。在大数据时代,各种基于数据驱动的预测性技术其背后的运行逻辑便是“从过去预测未来”①,在这种思维的主导下有些个人信息的保存期限可能会被有意或者无意地被延长,这样就可能会违反存储限制。

2. 可能造成数据泄露,违反安全义务。从国内外的个人信息保护立法现状来看,数据泄露问题成为重要的规范内容,而数据安全义务也得到不断强化。我国《数据安全法》对“数据安全”进行了系统规定,明确了许多义务和机制。我国《个人信息保护法》第9条将“安全”作为一项重要原则,要求个人信息处理者应当采取必要措施保障个人信息的安全,第36条和第40条进一步规定了“安全评估”机制。在实践中,政府数据泄露的风险可以大致分为三个方面:一是意外数据泄露,即由于某些内部或外部原因,数据库中的数据被他人获取,导致某些敏感数据被公布,造成隐私侵犯;二是无意中的数据泄露,即因为内部人员操作失误,违反安全政策,引发数据泄露,导致敏感数据被非法盗取、修改、复制等;三是恶意数据泄露,即外部有针对性的攻击,如黑客攻击、计算机病毒、“信息间谍”等,导致数据库信息泄露。② 数据泄露已经成为近年来导致社会提升对个人信息保护问题关注度的重要诱因,原因在于数据泄露造成的危害可能在短期内不会立即显现,但随着时间的推移却可能“积少成多,积重难返”,引发“身份盗窃”、欺诈或者名誉受损等风险,而这些风险会进一步引发社会公众的焦虑情绪,甚至可能引发公众对政府收集数据的“寒蝉效应”。③

(五)数据公开阶段侵害个人信息权益

数据公开是政府数据开放的核心阶段,在某种意义上也可以称为狭义的政府数据开放,这也是研究政府数据开放的学者最为关注的阶段,诸如数据标准、数据质量、数据门户等问题都与数据公开有关。从广义上看,数据公开实际上是双向的,既包括政府部门主动或者依申请发布各种政府数据,也包括数据主体或者其他个人及组织访问与之相关的政府数据。

在政府数据公开阶段,也存在侵害个人信息权益的风险,主要体现在以下几个方面。

1. 敏感个人信息被过度披露。从国内外个人信息保护立法内容来看,“识别”都是个人信息的核心要素。④ 在信息隐私监管中,个人可识别信息(Personally Identifiable Information,简称PII)是核心概念之一,隐私法的范围通常取决于是否涉及PII,适用法律背后的基本假设是,如果不涉及PII,就不可能有隐私损害。⑤ 我国《个人信息保护法》专门对“敏感个人信息的处理规则”予以规定,其背后的主要考量是敏感个人信息一旦泄露或者非法使用,容易导致自然人的人格尊严受到侵害或

① [英]维克托·迈尔-舍恩伯格、肯尼思·库克耶:《大数据时代》,盛杨燕、周涛译,浙江人民出版社2013年版,第16页。

② *See* A. Michael Froomkin, *Government Data Breaches*, 24 Berkeley Technology Law Journal 1019, 1019(2009).

③ *See* Daniel J. Solove & Danielle Keats Citron, *Risk and Anxiety: A Theory of Data-Breach Harms*, 96 Texas Law Review 737, 737(2018).

④ 程德理、赵丽丽:《个人信息保护中的“识别”要素研究》,载《河北法学》2020年第9期,第45页。

⑤ *See* Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 New York University Law Review 1814, 1814(2011).

者人身、财产安全受到危害。在实践中,政府数据公开阶段经常出现过度披露敏感个人信息的现象,侵犯个人信息权益。2021 年 4 月,最高人民检察院发公布了 11 件个人信息保护公益诉讼典型案例,其中,在“江西省某县人民检察院督促规范政府信息公开行政公益诉讼案”中,江西省某县农业农村局在官方网站公布农机购置补贴情况的政府信息时,有 1044 人的身份证号码、家庭住址、银行账户、手机号码等个人信息被完整公开。①

2. 过于关注个人而忽略群体隐私保护。从个人隐私保护到个人信息保护,存在一种“原子化”(atomistic)的本体论,即单独照顾每个成员,群体也会自动没事。换言之,只要我们保护好可识别个人的个人信息,对群体的保护就会自行解决。② 这从国内外个人信息保护立法对“个人信息”或者“个人数据”的界定以及与之相关的一系列保护机制可以得到印证。我国《个人信息保护法》第 4 条规定,个人信息是与“已识别或者可识别的自然人有关的各种信息,不包括匿名化处理的信息”。然而,我们应该看到,开放数据的友好和非友好用户可能并不关心具体的“张三”或者“李四”,而是关心这些人(不管他们是谁)是否属于高(低)收入、是否经常去高(低)消费地方消费或者是否属于患有某种疾病等。在先进的数据收集技术和分析技术的加持下,“群体”的概念比以往任何时候都更有意义:首先,更多关于现有群体的信息得以发现和披露;其次,可以在个人不知情的情况下,通过“提取”数据中的群体特征,将他们推断为某一群体的成员;最后,可以在数据分析阶段发生不为人知的“分组”过程,而分析者本人并不知情。③ 在这种情况下,个人隐私有可能得到有效保护,但群体本身却得不到充分保护。

(六)数据使用阶段侵害个人信息权益

政府数据开放的重要目的之一是通过政府数据的“再利用”进而激发政府数据所蕴含的各种价值。不过,一旦政府数据从正式的数据库系统中释放后,可以有效应用的一系列控制措施就可能发生变化,个人信息保护风险也会随之发生演变。具体而言,主要体现在以下几个方面:

1. 数据使用违反目的限制。目的限制原则是个人信息保护的基本原则之一,被称为个人信息保护法的“帝王条款”,我国《个人信息保护法》第 6 条专门对该原则进行了规定:“处理个人信息应当具有明确、合理的目的,并应当与处理目的直接相关,采取对个人权益影响最小的方式。”在政府数据开放中,政府部门在收集个人信息时,可能是严格按照“目的限制原则”的规定,具有明确、合理的目的。然而,一旦包含个人信息的政府数据随着政府数据公开而进入使用阶段,则有可能违反目的限制原则,原因在于使用政府数据的主体可能性质各异,既有企业,也有个人,还有其他政府机构或者研究机构,而这些使用者处理数据的目的也不尽相同,使用数据有可能违背政府最初收集数据时所宣示的目的。例如,包含个人信息的医疗保健数据,其最初处理目的可能在于改进与个人相关的医疗保健服务,但如果保险公司获得了这些数据,那么这些数据就很有可能被用来改进和支撑保

① 《检察机关个人信息保护公益诉讼典型案例》,载最高人民检察院官网,https://www.spp.gov.cn/spp/xwfbh/wsfbt/202104/t20210422_516357.shtml#2

② *See* Luciano Floridi, *Open Data*, *Data Protection*, *and Group Privacy*, 27 Philosophy & Technology 1, 2(2014).

③ See Lanah Kammourieh, Thomas Baar & Jos Berens, et al., *Group Privacy in the Age of Big Data*, in Linnet Taylor, Luciano Floridi & Bart van der Sloot(eds.), Group Privacy: New Challenges of Data Technologies, Springer, 2017, p. 38.

险公司的市场营销策略。①

2. 匿名化数据被“去匿名化”,导致个人信息被“再识别”。为了避免政府数据公开给个人隐私造成的侵害,很多政府数据开放倡议或者立法都要求政府部门在公布包含个人信息在内的政府数据时,应当进行匿名化处理或者去标识化处理。② 尽管匿名化处理使个人信息的“再识别”变得异常困难,但却并非完全不可再识别,原因主要有两个方面。一方面,匿名化技术本身可能存在缺陷,导致个人信息的匿名化并不彻底,出现“伪匿名化数据”;另一方面,尽管匿名化技术在不断改进,但是,去匿名化技术亦在不断发展,曾经或现在成功实现匿名化的个人信息,也有可能因为去匿名化技术的使用而再次被识别。③

3. 不同数据的聚合,引发新的隐私风险。政府数据的使用者(消费者)获得的政府数据可能来自不同的政府部门,属于不同的数据类型,孤立地看这些单一的数据集,可能并不存在侵犯隐私的问题。然而,当多个不同来源的数据被整合到一个数据集时,政府数据使用者便可以作出新的推断:数据的“组合”可以创建关于个人的新数据。在预测模型和自主学习算法的帮助下,数据使用者可以在个人并未主动提供的情况下生成个人信息,而这些信息可能准确预测某人未来生活的细节。④

三、以“基于过程的方法”应对个人信息保护风险

面对政府数据开放中可能存在的个人信息保护风险,现有的法律框架采取了“基于结果的方法”,即重点关注政府数据在公开之时的“状态”或者“性质”,最典型的便是要求对个人信息进行“脱敏处理”,使之变为“匿名数据”或者“假名数据”。本文认为,结果保护范式下以“技术性匿名化”为核心的保护手段并不足以应对政府数据开放中存在的或隐藏的个人信息保护风险,应当以风险预防原则重新确立政府数据开放中个人信息的保护理念,在此基础上,以“基于过程的方法”重塑政府数据开放中个人信息保护范式。

(一)结果保护范式下“技术性匿名化”之不足

现有的法律框架在对政府数据开放中的个人信息保护问题进行规范时,主要受到传统隐私法的影响,重点关注行为是否会产生某种可见的损害后果,其背后的逻辑是:“可识别性”是个人信息的核心标准,要想避免侵犯个人信息权益,那么通过技术手段“消除”政府数据中个人信息的“可识别性”就可以解决问题。⑤ 在这种结果主义思维的主导下,“匿名化”或者“去标识化”等技术手段逐渐获得法律规范的肯认⑥,并成为解决政府数据开放中个人信息保护问题最为重要的手段之一,逐

① *See* Liane Colonna, *Privacy*, *Risk*, *Anonymization and Data Sharing in the Internet of Health Things*, 20 Pittsburgh Journal of Technology Law and Policy 148, 148(2020).

② 张涛:《大数据时代个人信息匿名化的规制治理》,载《华中科技大学学报(社会科学版)》2019 年第 2 期,第 76 页。

③ *See* Paul Ohm, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, 57 UCLA Law Review 1701, 1701 (2010).

④ [荷兰]玛农·奥斯特芬:《数据的边界——隐私与个人数据保护》,曹博译,上海人民出版社 2020 年版,第 49 页。

⑤ *See* Ira S. Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 Washington Law Review 703, 726-727 (2016).

⑥ 张涛:《欧盟个人数据匿名化治理:法律、技术与风险》,载《图书馆论坛》2019 年第 12 期,第 91 页。

渐演化为“发布即遗忘模式”,即去标识化的个人信息可能会通过互联网向公众发布,一旦以这种方式发布,一个组织可能很难或者不可能召回这些信息。①

尽管匿名化技术在政府数据开放共享中确实为保护个人信息提供了重要支撑,但理论界与实务界围绕匿名化展开的论战也一直在持续。不看好匿名化的学者认为,长期以来,我们一直认为匿名化能够“拯救我们”,然而,科学研究已经表明,隐藏在匿名数据中的个人能够被“再识别”或者“去匿名化”,因此,监管机构必须迅速而有力地应对这种颠覆性的技术变革,以恢复法律的平衡,保护我们免受迫在眉睫的重大伤害。② 支持匿名化的学者认为,一方面,匿名化的批判者对“数据公地”的社会效用存在误解,并大大低估了其价值,如果政策制定者终止或者限制公开发布非识别数据集,社会将遭受新的数据“公地悲剧”;另一方面,匿名化的批判者错误地解释了计算机科学文献,过度强调了匿名化的无用性,事实上“去匿名化”或者“再识别”风险主要是理论上的,数据共享所带来的现实风险可以忽略不计。③

本文认为,上述观点均有一定的合理性,但却是相反的两个极端,并且几乎都将分析完全限制在个人信息在“公开”这个时点是否处于“匿名”状态,低估了政府数据开放中个人信息保护风险及其应对机制的复杂性,导致“匿名化技术”必须承担平衡个人信息保护和使用的全部“重量”,造成功能过载。总体而言,本文认为,以“技术性匿名化”为中心的结果保护范式难以有效应对政府数据开放生命周期不同阶段的个人信息保护风险,也难以担负平衡个人信息保护与使用的大任,具体理由如下。

1. 匿名化与数据再利用之间存在冲突,无法实现预期的平衡目标。如前所述,匿名化已经成为世界主要国家个人信息保护法治用以解决平衡个人信息保护与使用问题的重要方法。这种方法在理论上运行良好,但前提是来自数据的输出潜力仍然保持其效用,而实际情况却并非如此。这是因为通过使用自动化算法软件寻找模式(即链接数据点之间的关系),可以使从分析数据集中获得的价值或者知识最大化。然而,匿名化的目的是解除这种数据点之间的联系,因为它们与可以收集到的关于特定个人及其身份的信息有关。④ 这便引发了一个问题:政府部门如何确保对自己拥有的数据进行有效的匿名化,同时保留这些数据的效用,以便将来可能向第三方披露,并由第三方进一步处理?

2. 匿名化只关注数据本身的状态,忽视数据保护的过程。如前所述,以匿名化技术为核心的结果保护范式主要受到传统隐私法的影响,但是传统隐私法的许多策略并不真正适合于与政府数据开放有关的具体场景,因为大多数现有的隐私法都是采取一种“原子主义”方法论,将着力点放在特定的信息主体和离散的信息类型,而不是整个数据集。⑤ 这种“原子主义”方法论也影响到现有的个

---

① *See* Paul Ohm, Broken Promises of Privacy: *Responding to the Surprising Failure of Anonymization*, 57 UCLA Law Review 1701, 1711-1712 (2010).

② *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA Law Review 1701, 1723 (2010).

③ *See* Jane Yakowitz, *Tragedy of the Data Commons*, 25 Harvard Journal of Law & Technology 1, 1(2011).

④ *See* Sophie Stalla-Bourdillon & Alison Knight, *Anonymous Data v. Personal Data - False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data*, 34 Wisconsin International Law Journal 284, 285(2016).

⑤ *See* Ira S. Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 Washington Law Review 703, 726(2016).

人信息保护立法,导致与“去标识化”制度或者数据披露有关的法律规则都与数据本身的“输出”有关。我国《个人信息保护法》和欧盟《通用数据保护条例》都将“匿名数据”排除在个人信息保护范围之外,美国《健康保险可携性和问责法》(*The Health Insurance Portabicity and Accountability Act*,HIPAA)的去标识化制度也取决于数据是否缺乏某些属性。这种保护方法忽视了信息运行的周期性、过程性特征,将关注点聚焦于数据本身在公开这一时点的状态,无法满足个人信息保护的程序性要求。

3. 匿名化强调“技术制衡技术”,忽视技术与其他工具的协同。面对新兴技术给法律规范体系带来的冲击,囿于法律本身的滞后性,一些学者主张转变监管思路,采用“技术制衡技术”的方法。[①] 匿名化技术在个人信息保护中的重要地位得以确立,也受“技术制衡技术”思维的影响。然而,这也带来一个弊端,那就是保护方法上的“一刀切”,引发的问题便是过度保护或者保护不足。此外,要求政府部门在开放数据之前进行匿名化处理,容易营造两种引人误解的表象:一是匿名数据肯定是已经满足适当保护措施的数据;二是匿名化处理仅仅是政府部门的责任。事实上,由于政府数据开放本身是一个系统性工程,影响因素、利益相关者众多,这就意味着我们在应对个人信息保护风险时,除了要发挥技术的作用外,还要充分激发法律、教育、经济等手段的作用;除了政府部门要承担责任外,信息主体、社会公众、政府数据的再使用者等主体也应当积极参与个人信息保护。

需要说明的是,本文对以技术性匿名化为核心手段的结果保护范式提出批评意见,并非是对匿名化技术的全盘否定,而是不赞同政府部门在政府数据开放过程中将保护个人信息的重任全部放置在匿名化技术上,忽视了在其他阶段采用不同的个人信息保护手段,以降低个人信息保护不足的风险。

(二)以风险预防原则确立个人信息保护理念

近年来,“风险”概念开始在个人信息保护理论与实践中得以扩散,风险管理已经成为确保数据得到适当处理和个人基本权利得到有效保护的重要工具。[②] 面对技术风险,法律规范通常可以区分为两种面向:一种是压制性的(repressive),即在事故发生后才开始行动,所关注的问题是:是否有人应当对该事故负责?是否需要有人支付费用(民法)?是否有人会受到惩罚(刑法)?另一种是预防性的(preventive),即试图规范风险活动,以避免事故发生,这主要是行政法的目的,它包含成百上千的法律和条例,倾向于限制特定技术活动的风险。[③] 就个人信息保护领域而言,预防性法律规范呈现爆炸式增长,尤其是欧盟《通用数据保护条例》制定出台以后,个人信息保护法制的“风险化”已经成为一种趋势。在欧盟,个人数据保护法的“风险化”主要体现在两个方面:一是在实践层面,转向基于风险的数据保护执法和合规;二是在更广泛而坚实的规制层面,转向了风险规制。[④] 在我国,“风险”一词在《个人信息保护法》中出现了 4 次,在《数据安全法》中出现了 10 次,《个人信息保护

① 季卫东:《数据、隐私以及人工智能时代的宪法创新》,载《南大法学》2020 年第 1 期,第 1 页。

② *See* Christopher Kuner, Fred H. Cate & Christopher Millard et al., *Risk Management in Data Protection*, 5 International Data Privacy Law 95, 95(2015).

③ *See* Hansjörg Seiler, *Harmonised Risk Based Regulation — A Legal Viewpoint*, 40 Safety Science 31, 31(2002).

④ *See* Milda Macenaite, *The ‘Riskification’ of European Data Protection Law through a two-fold Shift*, 8 European Journal of Risk Regulation 506, 506(2017).

法》第 11 条确立了“风险预防”原则,《数据安全法》第 22 条要求建立集中统一、高效权威的“数据安全风险评估机制”。

需要指出的是,尽管风险管理已经成为遵守个人信息保护法、确保个人信息得到适当处理和个人基本权益得到有效保护的关键工具,并且已经初步实现了法制化。然而,在实践中,诸多风险管理过程或者工具往往还是非正式的、非结构化的,未能充分利用其他领域已经广泛接受的风险管理的诸多原则和工具,因此基于风险的方法(risk-based approach)仍然没有为个人信息保护实践或者法律提供广泛而坚实的基础。① 对于我国现阶段的个人信息保护而言,这既是机遇,也是挑战。一方面,我们可以充分吸收借鉴其他领域长期积累的风险管理经验,开发一个现代化的、有效的风险管理框架;另一方面,我们需要积极采取行动,以跟上新兴技术的巨大变化。一般认为,风险管理主要涉及三个关键要素:一是识别和评估危害和其他负面影响的系统过程;二是避免或者减轻那些不能用利益和其他积极影响来证明的危害;三是接受和管理剩余风险。② 本文认为,政府部门应当在政府数据开放过程中建构一个现代化的风险管理框架,同时开发有效的风险管理工具,以“风险预防”来重新确立政府数据开放中个人信息保护的理念与目标。一方面,基于风险的政府数据开放制度契合当前个人信息保护的趋势;另一方面,基于风险的政府数据开放制度将帮助我们超越关于匿名化有效或者无效的争论。

1. 在政府数据开放过程中应当考虑不同的风险因素。风险管理从本质上讲是一种平衡测试,它需要考虑诸多因素,包括个人的基本权益、拟议的处理将损害个人的可能性、发生损害的严重程度、可用来降低风险的措施、数据控制者的权益等。③ 因此,在建构风险管理框架时,政府部门也需要确定不同的风险因素,以确定在公开政府数据时需要采取何种保护措施。参照美国国家标准与技术研究院(National Institute of standards and Technology,NIST)发布的《个人信息去标识化》(*De-Identification of Personal Information*)报告④,本文认为,在政府数据开放过程中至少应当考虑以下风险因素:一是政府数据的数量,信息量会影响重新识别和敏感属性披露的风险;二是政府数据的敏感性,生物识别信息、行动轨迹信息、银行账户信息等是比较敏感的,也更有可能成为攻击目标;三是政府数据的接收者,至少包括内部接收者、受信任的接收者和普通公众三种不同类型的数据接收者,并且风险呈递增趋势;四是政府数据的用途,数据的不同用途可能会给攻击者带来不同的再识别动机;五是数据的公开方式,不同的公开方式引发的风险不同,受控制的公开比无条件开放的风险低。

2. 在政府数据开放过程中应区分不同的风险等级。与其他领域的风险管理相似,个人信息保护中的风险管理也应当根据不同的风险因素、管理目标,确定不同的风险等级。欧盟《通用数据保

---

① *See* Maria Eduarda Gonçalves, *The Risk-Based Approach under the New EU Data Protection Regulation: A Critical Perspective*, 23 Journal of Risk Research 139, 139(2020).

② *See* Christopher Kuner, Fred H. Cate & Christopher Millard et al., *Risk Management in Data Protection*, 5 International Data Privacy Law 95, 96-97(2015).

③ *See* Christopher Kuner, Fred H. Cate & Christopher Millard et al., *Risk Management in Data Protection*, 5 International Data Privacy Law 95, 98(2015).

④ See National Institute of Standards and Technology, *De-Identification of Personal Information*, https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf(Last visited on July 11, 2021).

护条例》采用了基于风险的数据保护方法,鼓励控制个人数据的组织实施与其数据处理活动的风险水平相适应的保护措施:首先是高风险,《通用数据保护条例》对从事“高风险”活动的数据控制者提出了更高的要求。具体而言,在从事这种活动之前,数据控制者可能被要求咨询数据保护机构并进行详细的隐私影响评估;在发生数据泄露的情况下,数据控制者可能被要求通知可能受影响的个人。其次是风险,对于没有被标记为“高风险”的活动,数据控制者仍然必须采取与该活动的风险水平相适应的措施。例如,数据控制者被要求“确保与风险相适应的数据安全水平”,并实施基于风险的措施以遵守一般法律义务。最后是低风险,如果对数据主体的风险很小,数据控制者可以免于向数据保护机构通报数据泄露的要求。尽管《通用数据保护条例》没有提及数据控制者应当如何评估和量化风险,但是这种基于风险等级采取不同控制措施的思路仍然值得我国政府部门在建构政府数据开放的风险管理框架时予以借鉴。

(三)以“基于过程的方法”重塑个人信息保护模式

“基于过程的方法”(process-based approach)是以数据生命周期为基础,全面、动态地考察政府数据开放整个过程的一种保护模式,它为我们观察政府数据开放提供了一种全新的视角,改变了传统结果保护范式的认知与适用局限。从逻辑与事实的关联性来看,“基于过程的方法”与政府数据开放中个人信息保护之间具有内在关联和外在契合。

1.“基于过程的方法”与政府数据开放生命周期之间存在内在契合。如前所述,政府数据开放本身就是一个动态的过程,遵循了数据生命周期的一般规律,可以分为收集、转换、存储、公开和使用等5个阶段,这与“基于过程的方法”本身所主张的全面、动态的观察视角相吻合。

2.“基于过程的方法”与政府数据开放的风险管理相契合。“基于风险的方法”已经逐渐成为个人信息保护领域的重要方法,风险管理也成为关键工具,而风险管理本身就包括风险识别、风险量度、风险评估、风险应对等阶段,这与“基于过程的方法”存在外在关联。

3.“基于过程的方法”与当前个人信息保护的程序主义进路相吻合。从国内外个人信息保护立法的现状来看,无论是信息主体享有的权利,还是信息处理者应当承担的义务,以及最终的权利救济机制,都出现了大量的程序性规则。[①] 个人信息保护已经成为一个不断识别、预防、降低风险的过程,即使信息处理者没有实际的违规行为,只要其未能采取行业内普遍认可的措施来充分降低风险,其就有可能承担法律责任[②],这与“基于过程的方法”的核心理念是相吻合的。

4.“基于过程的方法”与数据安全全流程管理制度是相吻合的。我国《数据安全法》第27条规定,开展数据处理应当“建立健全全流程数据安全管理制度”,这意味着数据安全管理应当融入整个数据生命周期,这与“基于过程的方法”的基本主张是一致的。

对于政府数据开放中的个人信息保护而言,一个更加具有可持续性的方法是将重点侧重于保护所需的前提条件和过程。政府部门在制定政府数据开放策略时,应当以更为全面的、动态的视角来看待政府数据开放过程,而不是仅仅关注政府数据在公开这个时点所处的状态。本文认为,“基

① *See* Antonella Galetta & Paul De Hert, *The Proceduralisation of Data Protection Remedies under EU Data Protection Law: Towards a More Effective and Data Subject-oriented Remedial System?* 8 Review of European Administrative Law 125, 125(2015).

② *See* Ira S. Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 Washington Law Review 703, 730(2016).

于过程的方法”可以为政府数据开放与个人信息保护提供一个平衡框架,它可以考虑不同的信息类型和保护手段,最终实现全生命周期的协同保护。

第一,区分不同的信息类型。在政府数据开放中,根据信息类型不同,会产生不同程度的个人信息保护风险。① 为了平衡个人信息保护与政府数据效用之间的关系,可以区分四种不同的信息类型:(1)原始个人信息,即未经处理或者简化的信息,包括姓名、身份证号码、电话号码等。目前,绝大多数国家或地区的政府数据开放政策都规定,开放数据通常不应当包含原始个人信息。当然,在某些特定情况下,原始个人信息也可以公开,条件是公开的公共利益超过了个人利益。(2)假名化信息,这意味着一个人的身份识别信息被随机的唯一标识符所取代。(3)匿名化信息,这是不再可识别的信息,从数据集中删除了所有个人可识别信息,并将可识别信息转化为匿名。不过,随着大量数据的积累,数据挖掘者可以从数据集中发现隐藏的个人信息,导致个人信息被再识别。(4)非个人信息,主要是指不包含个人信息的数据集,如公共交通时间、天气状况、公共部门预算、环境污染等。② 不过,需要说明的是,尽管从理论上可以对政府数据中不同的信息类型进行相对独立的区分,但是,在实践中这四类信息的边界仍然是比较模糊的,匿名化信息可以再识别或者去匿名化,而非个人信息也有可能提供关于个人的信息。

第二,采用不同的保护手段。随着信息技术的快速发展,个人信息保护变得越发复杂,仅仅依靠单一的主体、工具均无法为个人和社会提供有效的保护,“合作规制”逐渐成为一种趋势,要求整合多元治理主体和多元治理手段③,我国《个人信息保护法》第 11 条亦对此进行了规定。在政府数据开放过程中,为了保护个人信息,可以采取的保护手段大致有以下五类:(1)程序手段,广义上是指在组织内部采用各种程序,如执行通知、制定数据清单、审查数据库访问等。(2)技术手段,广义上是指包括统计方法、计算方法(如加密)和人为因素分析(如隐私政策的可读性分析)等在内的技术手段。(3)教育手段,广义上是指包括任何旨在告知信息主体、信息处理者、信息接收者或者广大公众有关个人信息保护规则与风险的干预措施。(4)经济手段,广义上是指任何旨在改变利益相关者经济动机的干预措施,如征税、收费、罚款或者提供保险等。(5)法律手段,广义上是指任何旨在改变利益相关者的法律权利义务或者法律关系的干预措施,如私人诉讼、行政问责、公益诉讼等。④

四、完善政府数据开放中基于过程的个人信息保护机制

前文阐述了政府数据开放生命周期不同阶段可能存在的个人信息保护风险,并对以技术性匿名化为核心手段的结果保护范式的不足进行了分析,在此基础上,本文提出以“基于过程的方法”重

① 宋烁:《论政府数据开放中个人信息保护的制度构建》,载《行政法学研究》2021 年第 6 期,第 85 页。

② *See* Frederik Zuiderveen Borgesius, Jonathan Gray & Mireille Van Eechoud, *Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework*, 30 Berkeley Technology Law Journal 2073, 2120(2015).

③ *See* Irene Kamara, *Co-regulation in EU Personal Data Protection: The Case of Technical Standards and the Privacy by Design Standardisation 'Mandate'*, 8 European Journal of Law and Technology 1, 1(2017).

④ *See* Micah Altman, Alexandra Wood & David R. O'Brien et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkeley Technology Law Journal 1967, 2016-2017(2015).

塑政府数据开放中个人信息保护模式,并且就该保护模式的基本内容与核心措施进行了初步分析。为了进一步加强政府数据开放中的个人信息保护,有针对性地就不同阶段的个人信息保护风险采取措施,还需要对程序、技术、教育、经济、法律等手段的范围作进一步的说明,以及对它们在政府数据开放生命周期不同阶段如何发挥作用进行阐述。为此,本文以本·格林(Ben Green)等人在《开放数据隐私》报告中提出的"在数据生命周期的每个阶段考虑隐私问题"为基础,[①]参照迈卡·奥特曼(Micah Altman)等人提出的分析框架[②],结合我国《个人信息保护法》《数据安全法》和《信息安全技术 个人信息安全规范》(GB/T 35273—2020)的有关规定,尝试提出一个将不同的个人信息保护手段与政府数据开放生命周期不同阶段协调衔接的控制网络。

(一)数据收集阶段的个人信息保护手段

一旦个人信息被收集,它就有可能作为开放数据或者通过回应公共记录请求而被公开,因此,限制数据收集往往是限制未来披露的最佳方式。为了应对政府数据收集阶段可能存在的个人信息保护风险,政府部门可以采取如下手段。

1. 确保收集个人信息的合法性。根据《个人信息保护法》第 5 条的规定,合法原则是个人信息保护的重要原则。政府部门在收集包含个人信息的数据时,应当依照法律、行政法规规定的权限与程序进行,不得通过其他非法手段和途径收集个人信息。

2. 确保收集个人信息的最小必要。根据《个人信息保护法》第 5 条的规定,必要原则是个人信息保护的基本原则。该法第 6 条进一步明确"收集个人信息,应当限于实现处理目的的最小范围,不得过度收集个人信息"。该法第 34 条则规定:"国家机关为履行法定职责处理个人信息,应当依照法律、行政法规的权限、程序进行,不得超越履行法定职责所必需的范围和限度。"政府部门在收集个人信息时,所收集的个人信息类型应当与履行法定职责或者提供公共服务的业务功能直接关联。所谓直接关联,就是没有上述个人信息的参与,履行职责或提供服务的功能就无法实现。

3. 确保收集个人信息的诚实守信。根据《个人信息保护法》第 5 条的规定,诚信原则是个人信息保护的基本原则,这对政府部门在收集个人信息时提出了"正反"两方面的要求。一方面,政府部门要严格落实"告知-同意"要求。在国内外个人信息保护法中,"告知-同意"是公平信息实践原则的基石,也是保护个人信息的常用工具。就"通知"而言,除了常规的隐私政策外,还可以采取公共教育措施,让公民了解收集的数据类型、如何使用这些数据以及与之相关的风险[③];就"同意"而言,虽然履行法定职责、紧急情况下的公共利益可以成为豁免条件,但是,政府部门仍然需要制定更为有效的同意计划,原因在于同意的效力并不仅仅局限于收集阶段,而是辐射到整个数据生命周期。另一方面,这意味着政府部门不得以欺诈、诱骗、误导等方式收集个人信息,也不得隐瞒各类政务 APP 所具有的收集个人信息的功能。

---

① *See* Ben Green, Gabe Cunningham & Ariel Ekblaw et al., *Open Data Privacy* (2017), Berkman Klein Center Research Publication, https://dash.harvard.edu/bitstream/handle/1/30340010/OpenDataPrivacy.pdf? sequence=5(Last visited on July 11, 2021).

② *See* Micah Altman, Alexandra Wood & David R. O'Brien et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkeley Technology Law Journal 1967, 2049-2058(2015).

③ *See* Micah Altman, Alexandra Wood & David R. O'Brien et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkeley Technology Law Journal 1967, 2017(2015).

4. 完善个人信息的内部管理机制。政府数据开放共享是政府数据治理的重要内容。随着政府数据治理体系的不断完善,越来越多的政府部门开始任命专门的数据治理官员或者组建专门的数据治理机构,如政府首席数据官,为政府数据治理提供组织保障。① 我国《个人信息保护法》第 52 条规定,处理个人信息达到一定规模的个人信息处理者应当指定个人信息保护负责人。就政府部门而言,应当以建立和完善政府首席数据官制度为契机,强化政府首席数据官对政府部门处理个人信息的监督职责。

5. 完善个人信息保护影响评估机制。在其他风险管理领域,风险评估是最为重要的工具之一,如环境保护影响评估。在隐私和个人信息保护领域,传统的隐私影响评估工具经常被作为平衡效用和隐私以及选择适当隐私保障措施的工具。如今,在隐私影响评估机制的基础上,个人信息保护影响评估机制成为个人信息保护的重要工具。尽管个人信息保护影响评估在不同的机构之间可能略有不同,但通常涉及个人信息的处理目的、处理方式是否合法、正当、必要,个人信息的预期用途和接收者,对个人权益的影响及安全风险,采取的保护措施是否合法有效等。

(二)数据转换阶段的个人信息保护手段

如前所述,尽管数据转换阶段与数据主体并不直接相关,但却可能因为违反准确性、完整性等法律要求,产生侵害个人信息权益的风险。为了有效应对政府数据转换阶段存在的个人信息保护风险,结合相关理论研究成果和实践经验,政府部门可以采取如下手段。

1. 完善内部的数据转换制度和操作规程,强化专业培训教育。数据转换对于整个政府数据开放至关重要,要求确保在数据转换过程中既符合个人信息保护要求,也能达到未来政府数据公开的标准,政府部门应当制定内部的数据转换操作规程,同时加强对相关工作人员的专业培训,既包括先进处理技术的培训,也包括个人信息保护法规、数据安全法规、信息伦理要求等方面的培训教育。

2. 积极采取加密等技术手段,确保数据转换过程安全。在数据转换过程中,政府部门还应当积极采取一些先进的“隐私增强技术”或者“数据安全技术”,预防和减少数据转换阶段可能出现的信息泄露、篡改、丢失等风险,其中比较常见的便是采用加密技术,包括公钥加密和私钥加密。

3. 保障信息主体的访问权和更正权,增强数据转换的透明度。为了确保数据安全,数据转换的具体过程一般具有一定的机密性,但是数据转换的方法及结果应当遵循透明度要求。政府部门应当通过适当的途径,将数据转换的方法和结果告知信息主体,保障信息主体能够有意义地行使访问权和更正权。

(三)数据存储阶段的个人信息保护手段

随着数字政府建设的不断向前推进,政府部门维护着众多的数据集,并且这些数据集通常分布在不同的部门或者机构,使得政府部门很难跟踪其众多的数据资源。如果没有对现有数据集的全面了解和掌控,政府部门就有可能做出错误的数据管理决策或者进行多余的数据收集、超期的数据保存。此外,未知的和未被充分监测、评估的数据集可能会带来风险,再加上人员流动和信息管理系统定期升级增加了评估旧数据集所涉隐私及个人信息风险的难度。为了有效应对政府数据存储

① 张涛:《数据治理的组织法构造:以政府首席数据官制度为视角》,载《电子政务》2021 年第 9 期,第 58 页。

阶段的个人信息保护风险,政府部门可以采取如下手段。

1. 制定数据清单及目录,并确立相应的隐私与个人信息风险等级。《数据安全法》第 42 条规定,国家应当制定政务数据开放目录;《个人信息保护法》第 51 条规定,个人信息处理者应当对个人信息实行分类管理。前文对政府数据的分类进行了初步分析,政府部门在编制数据清单时,还应当重点解决如下问题:数据是如何产生的、什么记录管理系统产生了这些数据、这些数据是什么时候收集的、这些数据是如何收集的、是否还有与这些数据相关的其他补充信息、哪些部门和个人负责这些数据、这些数据规定的保存期限是多长等。数据清单的另一个关键部分是确定相应的隐私与个人信息风险等级,政府部门应当开发一个分级模式,将每个数据集的隐私与个人信息风险划分为数个类别,如高、中、低风险,提供一个风险概览图,帮助政府部门将有限的资源用于降低最紧急的风险。

2. 限制个人信息存储期限,避免无限期保存。如前所述,存储限制是《个人信息保护法》提出的明确要求,政府部门也应当积极遵守。一方面,个人信息存储期限应为实现政府部门处理信息目的或者信息主体授权目的所必需的最短时间;另一方面,当个人信息超出上述存储期限后,应当对个人信息进行删除或者匿名化处理。

3. 完善数据安全风险评估制度,及时监测存在的安全风险。《数据安全法》第 29 条规定“开展数据处理活动应当加强风险监测”;第 30 条规定:“重要数据的处理者应当按照规定对其数据处理活动定期开展风险评估”。拥有众多大型数据库的政府部门面临的数据安全风险并不天然地低于私营部门。因此,上述规定同样也适用于政府部门,其更应当积极完善数据安全风险评估制度,降低数据泄露风险。一般而言,数据安全风险评估通常包括所存储的数据类型与数量、可能存在的风险、可能采取的应对措施。

4. 编制数据安全事件应急预案,完善数据泄露通知制度。尽管各种保护措施可以将数据安全风险降至最低,但是,数据泄露事件仍然无法避免。为此,《数据安全法》第 23 条规定:“国家建立数据安全应急处置机制”;《个人信息保护法》第 51 条规定,个人信息处理者应当“制定并组织实施个人信息安全事件应急预案”。政府部门应当借鉴其他领域的应急管理经验,充分听取数据安全领域的技术专家、法律学者的意见,制定本部门的数据安全事件应急预案。当发生大规模数据泄露事件时,政府部门应当按照《个人信息保护法》第 57 条的规定,应当“立即采取补救措施,并通知履行个人信息保护职责的部门和个人”,让其能够积极采取防护措施,避免数据泄露引发的次生伤害。

(四)数据公开阶段的个人信息保护手段

如何确定要公开的数据集是政府数据开放所面临的最常见的挑战之一,因为政府部门并不总是十分清楚哪些政府数据在公开后会带来隐私与个人信息风险,稍有不慎,就可能过度披露敏感信息。为了应对在政府数据公开阶段可能存在的个人信息保护风险,政府部门可以采取如下几种手段。

1. 区分不同的公开方式,采用不同的政府数据访问机制。政府部门应当清楚地认识到,并非所有的数据集都适合以“开放数据”格式公开,如数据的敏感性和颗粒度很高时,以不可控的开放数据格式公开,被再识别的风险就会很高。因此,不同类型的政府数据、不同风险级别的数据集应当设

置与其风险水平相适应的公开方式和数据访问机制。一般认为,在组织内部和组织之间共享和访问数据,可以采用的访问机制大致包括以下几类:一是限制性共享,政府部门可以根据预期的数据接收者类型及其相应的风险等级,来决定如何公开数据。二是管理性访问,政府部门可以确定到底是谁在访问数据集,同时保持对其传播效果的控制。三是互动方法,比较典型的是“差分隐私”(differential privacy),即在数据集中添加一定量的“噪音”或者只提供关于底层数据集的“统计结果”。四是混合方法(hybrid),即政府部门可以将一个包含有可能再识别的个人信息的数据库进行拆分,再融入前述三种方法。①

2. 广泛使用去标识化技术,降低政府数据的被再识别风险。从技术的角度看,尽管并不存在完美无瑕的去标识化技术,但不可否认的是各类去标识化技术仍然在个人信息保护中发挥着重要的作用。政府部门在准备要公开的政府数据时,应当广泛使用去标识化技术,既可以由人工执行,也可以通过自动化程序执行,或者两者兼备。一旦认定数据集已经完成去标识化后,应当手动审查或者以其他方式进行数据审计,以判定是否还有任何可识别信息,或者审查是否有可能通过与其他数据源的关联来恢复被删除的敏感属性。

(五)数据使用阶段的个人信息保护手段

一旦政府数据从数据库中公开进入使用者手中,前面几个阶段的保护手段就可能失去效用,再加上可能存在的数据滥用、数据聚合等风险,从而导致个人信息权益受到侵犯。为了应对政府数据使用阶段存在的个人信息保护风险,政府部门应当积极开发新的保护手段或者机制来消解“发布即遗忘模式”的弊端,比较常见的方法主要有以下两种。

1. 通过信用监管强化对政府数据使用者的事前审查。尽管各国在制定政府数据开放政策时,都要求政府部门不应当对政府数据使用者设置歧视性条件,但这并不意味着政府数据使用者不需要满足任何条件或者达到一定要求。为了避免政府数据使用者在使用政府数据时产生新的隐私与个人信息保护风险,在实践中,政府部门可以要求政府数据使用者在数据存储、数据处理和数据安全保护能力等条件方面应当达到相应的信用等级。

2. 通过数据使用协议强化对政府数据使用者的事后监管。为了应对政府数据滥用可能带来的隐私与个人信息保护风险,数据使用协议成为常见的保护手段。在数据使用协议中,政府部门可以就政府数据再使用的目的、用途等进行约定,如禁止重新识别信息或者联系个人,同时,要求政府数据使用者应当依法履行相应的数据保护职责,并接受政府部门的监督检查,承担相应的法律责任。

五、结论

政府数据开放被认为有助于实现各种社会、经济及行政目标,如提升公众生活品质、促进产业转型、增强政府透明度等。然而,由于政府部门维护的数据集中含有大量的个人信息,因此,政府数

① See Bendert Zevenbergen, Ian Brown & Joss Wright et al., *Ethical Privacy Guidelines for Mobile Connectivity Measurements*, Oxford Internet Institute, https://www.oii.ox.ac.uk/archive/downloads/research/files/Ethical_Privacy_Guidelines_for_Mobile_Connectivity_Measurements.pdf(Last visited on July 11, 2021).

据开放可能会对个人信息权益造成侵害。这种侵害不局限于政府数据公开阶段,而且可能存在于政府数据收集、转换、存储、使用等多个阶段。在我国《个人信息保护法》和《数据安全法》通过并实施后,如何在个人信息保护与政府数据开放之间维持平衡,成为一项重要议题。传统的以技术性去匿名化为核心手段的结果保护范式由于过度关注政府数据在公开这一时点的状态,忽视了政府数据开放本身所具有的动态性、周期性、系统性特征,难以有效应对政府数据开放全生命周期各个阶段存在的个人信息保护风险。对此,本文提出以“基于过程的方法”重塑政府数据开放中的个人信息保护范式,以风险预防原则重新确立个人信息保护理念与目标。为了构建政府数据开放中基于过程的个人信息保护机制,需要将程序、技术、经济、教育、法律等多种手段置于在政府数据开放生命周期的不同阶段,有针对性地减少隐私与个人信息风险,形成全生命周期的协同保护机制。ML

Paradigm Shift of Personal Information Protection in Open Government Data

ZHANG Tao

(Law School, Tsinghua University, Beijing 100084, China)

Abstract: Open government data (OGD) is not a static single behavior, but a dynamic system process. With the help of data life cycle theory, open government data can be deconstructed into five stages of data collection, transformation, storage, release, and use. According to the latest rules established by the Personal Information Protection Law and the Data Security Law, personal information protection risks may exist at all stages of the OGD life cycle. However, the existing personal information protection paradigm in OGD mainly adopts an "outcome-based approach", focusing on the state of government data at the time of release, and relying on technical anonymization methods, it is difficult to effectively deal with personal information risks in OGD life cycle. Correspondingly, the "process-based approach" is compatible with the OGD life cycle, the proceduralization of personal information protection, and the full-process management of data security, and can make up for the shortcomings of the "outcome-based approach". By decentralizing risk precautionary principles and procedural, technical, economic, educational, and legal tools at each stage of the OGD life cycle, the risk of personal information protection can be minimized, and a dynamic balance can be achieved between personal information protection and open government data.

Key Words: open government data; personal information protection; process-based approach; anonymization

本文责任编辑:林士平

青年学术编辑:孙　莹