· 研究综述与前沿进展 ·

国外公共数据资源开放共享中的隐私风险控制研究综述

苏君华 杜 念*
（上海大学文化遗产与信息管理学院，上海 200444）

摘 要：［目的/意义］通过阐述国外公共数据资源开放共享中的隐私风险控制研究进展，为国内相关理论研究与实践提供参考。［方法/过程］运用内容分析方法，从公共数据资源开放共享中的隐私风险控制教育手段、法律手段、技术手段和程序手段 4 个方面解析国外研究进展。［结果/结论］研究发现，国外对公共数据资源开放共享中的隐私风险控制主题进行了广泛探索，主要挑战是平衡数据隐私和数据效用，但现有研究整体上还不成熟，需构建隐私风险控制体系，加强公共数据资源开放共享中的隐私控制技术与实践研究，以便推进公共数据资源开放共享的发展。

关键词：公共数据资源；开放共享；隐私风险控制；数据生命周期；综述
DOI：10.3969/j.issn.1008-0821.2024.03.015
〔中图分类号〕TP309；G203 〔文献标识码〕A 〔文章编号〕1008-0821（2024）03-0164-14

Review of Foreign Research on Privacy Risk Control Under Open Sharing of Public Data Resources

Su Junhua Du Nian*
（School of Cultural Heritage and Information Management，Shanghai University，Shanghai 200444，China）

Abstract：［Purpose/Significance］Through illustrating the research progress of foreign privacy risk control under the open sharing of public data resources，to provide reference for relevant theoretical research and practice in China.［Method/Process］Using the content analysis method，this article analysed foreign research progress of privacy risk control from four aspects：educational means，legal means，technical means and procedural means，in the context of open sharing of public data resources.［Result/Conclusion］This study finds that the topic of privacy risk control under the open sharing of public data resources has been extensively explored abroad，and the main challenge is to balance data privacy and data utility. However，the existing research is still immature as a whole，and there is a need to construct a privacy risk control system and strengthen the research on privacy control technology and practice under the open sharing of public data resources，in order to promote the development of public data resources open sharing.

Key words：public data resource；open sharing；privacy risk control；data life cycle；review

2021 年，十三届全国人大四次会议通过的《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》中明确指出，“完善公共数据开放共享机制”，强调了基于隐私和安全的公共数据开放的重要性。学者们就如何平衡公共数据隐私和效用进行了深入的研究，提出了许多控制方法，如数据信任[1]、差分隐私[2]、去识别化[3]、区块链[4]等。同时，有关隐私风险控制的理论、实践研究逐渐走向深入，产生了一系列相关研究成果。尽管国内已发表隐私计算[5]、开放数据[6]等研究述评，但未见相关综述全面反映国外公共数据资源开放共享中的隐私风险控制研究成果。因此，

收稿日期：2023-06-02
基金项目：国家社会科学基金项目“国家大数据战略背景下档案数据质量优化控制研究”（项目编号：21BTQ016）。
作者简介：苏君华（1979-），女，教授，博士，博士生导师，研究方向：信息资源管理、数据治理。
通讯作者：杜念（1999-），女，硕士研究生，研究方向：信息资源管理、隐私保护。

本文采用内容分析法来梳理、总结国外公共数据资源开放共享中的隐私风险控制研究进展。

本文的主要贡献是从教育、法律、技术和程序手段 4 个方面梳理当前公共数据开放中的隐私风险控制研究进展，并分类阐述各个方面的研究现状及主要控制措施。最后，总结并展望了当前公共数据开放中的隐私风险控制现状和发展方向，以期为后续研究人员及中国公共数据开放的建设提供参考和思路。

1 数据样本与研究方法

《上海市公共数据和一网通办管理办法》第三条规定，公共数据指本市各级行政机关以及履行公共管理和服务职能的事业单位在依法履职过程中，采集和产生的各类数据资源。其中，公共数据的共享、开放、授权经营亦或交易，必须有利于促成公共利益，而不是谋取个人或少数群体的利益。根据公共数据定义，2023 年 3 月 30 日，笔者以 TS = (Open and Data and Privacy and Risk and (Government OR Public)) 为检索式，对国外主要全文数据库(含 Web of Science、EBSCO、Elsevier、Springer、Taylor & Francis、Emerald、SAGE Journal、Wiley、Scopus)进行检索，排除社论材料、专利、信函、新闻等文献类型。为了使检索更全面，以 TI = ("Privacy Risk") 为检索式补充检索了相关重要文献。手动剔除显著不相关(如非公共数据、涉及隐私风险而无隐私控制的文献)与重复文献后，最终得到 99 篇文献。

借助 VOSview 对文献进行关键词聚类分析。首先获取文献关键词，文献关键词整合作者关键词和来源数据库提供的关键词，包括 Web of Science 关键词、IEEE 关键词、Scopus 检索关键词及主题词。其次预处理关键词，将外文文献中的关键词翻译为中文，合并同义词与相近词，删除无实际意义关键词等。如将"数据采集"与"数据收集"合并为"数据收集"；"法律和立法"与"法律"合并为"法律/立法"等；删除"文章""程序""调查"等关键词。选取关键词频率大于 3 的关键词(91)，运用 VOSviewer 软件生成国外公共数据资源开放共享中的隐私风险控制主题关键词共现网络图谱，如图 1 所示。

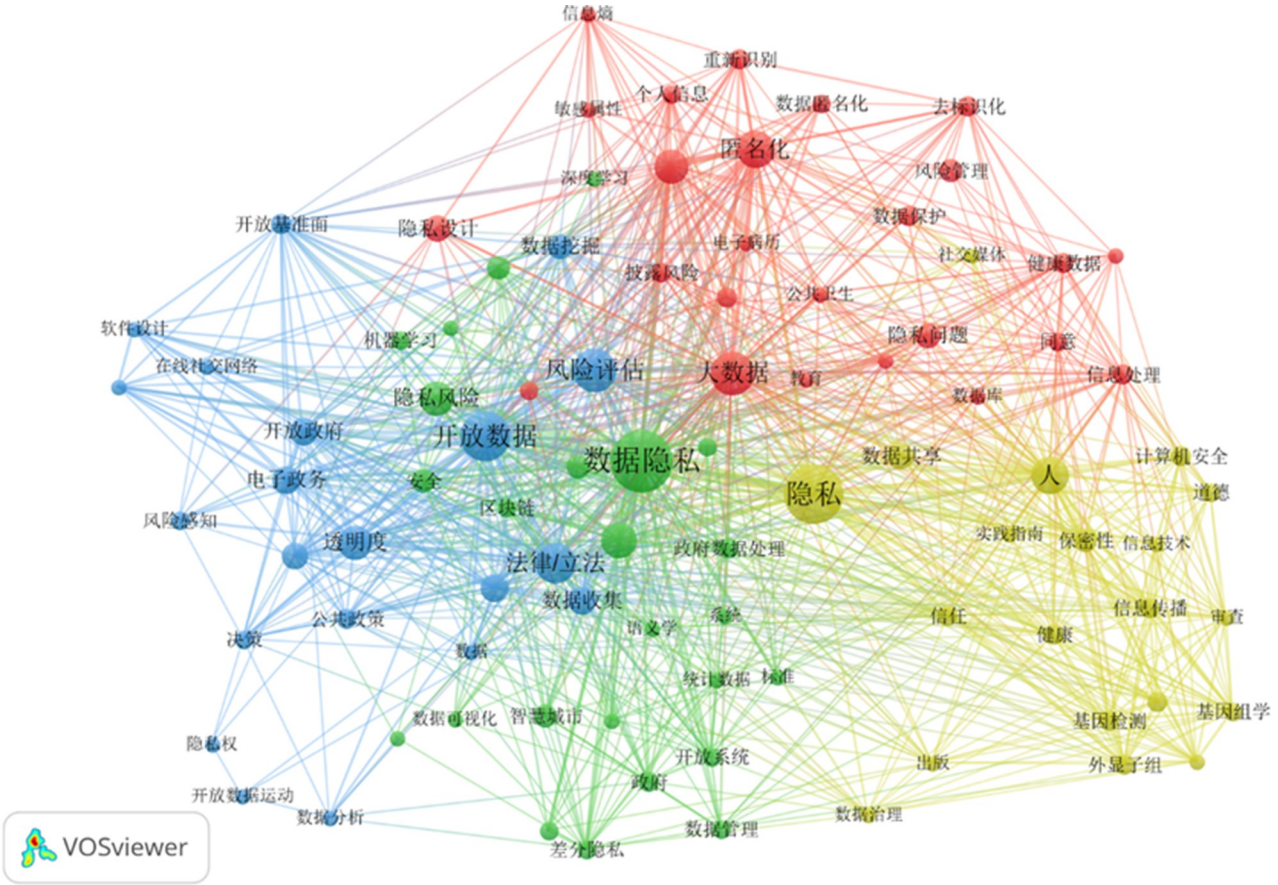

图 1 国外公共数据资源开放共享中的隐私控制主题关键词共现网络图谱

Fig. 1 Keyword Co-occurrence Networks of Foreign Public Data Resources Privacy Control Under the Open Sharing

根据图 1，国外公共数据资源开放共享中的隐私控制主题研究形成了 4 个主要类团。第一个类团(黄色部分)主要涉及公共数据资源开放共享中的隐私风险控制教育手段；第二个类团(蓝色部分)主要覆盖公共数据资源开放共享中的隐私保护法律手段；第三个类团(红色部分)主要涉及公共数据资源开放共享中的隐私风险控制技术手段；第四个类团(绿色部分)主要涉及公共数据资源开放共享中的隐私风险控制程序手段。各类团包含的研究主题、主要关键词及其频次、关键词数量、总频次信息如表 1 所示。

表 1 国外公共数据资源开放共享中的隐私风险控制主题关键词类团信息一览

Tab. 1 List of Keyword Groups Information of Foreign Privacy Risk Control Under the Open Sharing of Public Data Resources

类团名称	研究主题	主要关键词及其频次	关键词数量	总频次
公共数据资源开放共享中的隐私风险控制教育手段		隐私(38)；人(16)；数据共享(8)；保密性(6)；信任(5)；健康(5)；生物样本库(5)；道德(4)等	20	132
公共数据资源开放共享中的隐私保护法律手段	隐私保护主要法律法规	法律/立法(15)；透明度(13)；数据收集(9)；隐私权(3)等	21	169
	隐私保护主要标准	开放数据(26)；风险评估(21)；开放政府(10)；电子政务(9)；公共政策(4)；开放数据运动(3)等		
公共数据资源开放共享中的隐私风险控制技术手段	去标识化	隐私设计(8)；去标识化(5)；信息处理(5)；人工智能(3)；敏感信息(5)；个人信息(4)等	25	145
	匿名化	大数据(21)；匿名化(15)；k-anonymity(13)；重新识别(5)；数据匿名化(4)；披露风险(4)等		
公共数据资源开放共享中的隐私风险控制程序手段		数据隐私(43)；隐私保护(14)；隐私风险(13)；智慧城市(6)；安全(6)；人工智能(6)；政府数据处理(6)；政府(5)；开放系统(4)；智能手机(4)；数据管理(4)；差分隐私(4)；区块链(4)等	25	159

综合图 1 与表 1 可以发现：①目前国外对公共数据资源开放共享中的隐私风险控制主题研究主要集中在隐私风险控制教育、法律法规和标准、技术、程序 4 个方面，其中法律手段类团的关键词数量和总频次较高，说明当前公共数据中的隐私风险控制正处于发展阶段，法律正在不断完善中，其中隐私与开放的平衡是研究重点；②当前公共数据资源开放共享中的隐私风险控制教育手段研究成果相对较少，研究者开始注重道德、公众信任对隐私的影响，这对于推进公共数据资源开放共享具有重要作用；③在技术、程序手段方面，隐私增强技术广泛应用于各个领域，特别是政府服务、城市管理方面。并且区块链、差分隐私、人工智能等新兴技术被应用于隐私控制实践中；④4 个类团的关键词分布存在交叉现象，说明国外公共数据资源开放共享中的隐私风险控制主题研究热点之间的界限还比较模糊。

2 公共数据资源开放共享中的隐私风险控制研究分析

基于关键词共现与聚类分析，本文从收集和接受、转换、保留、发布和访问、访问后 5 个方面出发，将公共数据资源开放共享中的隐私风险控制实践情况分为程序、经济、教育、法律、技术手段，如表 2 所示。其中，程序手段指采用组织内部的程序；技术手段包括统计方法、计算方法和人为因素分析；教育手段包括旨在通知与系统交互的数据主体、数据控制者和数据接收者的任何干预措施，及一般数据主体、控制者、接收者或广大公众对隐私惯例和风险的看法；经济手段包括旨在改变利益相关者的经济激励的任何干预措施；法律手段旨在改变利益攸关方的合法权利或利益相关者间关系的干预措施。

表 2 各数据生命周期阶段中隐私风险控制手段一览表

Tab. 2 List of Privacy Risk Control Methods in Each Data Life Cycle Stage

	教育	法律	技术	程序
收集/接受	同意教育[7]；透明度[8]；通知[9]；公共教育[10]；数据信任/信托[11]；动态社会契约[12]	数据最小化(PIPA)；通知和同意(PIPA)；目的规范[13]；数据信任/信托[11]；动态社会契约[12]	数据信任/信托[1]；动态社会契约[12]	道德框架[8]；收集限制[13]；数据最小化[13]；目的规范[13]；通知和同意[7,9,14-15]；适当使用规则[1]；数据保护官(GDPR)；机构审查委员会[9,16-17]；风险评估[16,18]；道德委员会审查[9]
转换	透明度[19]；数据信任/信托[11]；动态社会契约[12]	更正或修改的权利[13]；数据信任/信托[11]；动态社会契约[12]	加密[1]；删除标识符[16]；数据抑制[11,16]；泛化[20]；汇总统计数据[21]；差异私有数据摘要[11]；合成数据[11,22-23]；差分隐私[11,24]；微聚合[22]；数据信任/信托[11]；动态社会契约[12]	去识别化[18]；匿名化[25]；清点数据资产(如元数据)[19]；技术匿名化小组[26]
保留(存储)	数据资产寄存器[8]；透明度[14]；数据信任/信托[11]；动态社会契约[12]	违规报告要求[1]；数据保留和销毁要求[27]；完整性和准确性要求[13]；数据信任/信托[11]；动态社会契约[12]	加密[1]；密钥管理[1]；个人数据存储[12]；隐私仪表盘[28]；数据信任/信托[11]；动态社会契约[12]	审计[29]；受控备份[29]；目的规范[13]
发布和访问	数据资产登记簿[19]；通知(PIPA)；透明度(PIPA)；数据信任/信托[11]；动态社会契约[12]	完整性和准确性要求[13]；数据使用协议(与数据接收方签订合同)/服务条款认证[11,19]；数据信任/信托[11]；动态社会契约[12]	k－anonymous[18,30]；差分隐私[31]；加密(功能、同态)[32]；入侵者测试[18]；交互式查询系统(数据安全站点)[21,33]；抽象编程接口或数据集下载点[1,34]；安全的多方计算[35]；贝叶斯信念网络[36]；身份验证和授权[10]；数据信任/信托[11]；动态社会契约[12]	访问控制[13,16]；同意(PIPA)；专家小组[37]；注册[33]；风险评估[16,38]；数据访问委员会[7]；数据避风港[32]
访问后	隐私仪表板[39]；透明度[1]；许可证[13]；数据信任/信托[11]；动态社会契约[12]	民事和刑事处罚[40]；数据使用协议/服务条款[9,41]；数据信任/信托[1]；动态社会契约[12]	个人数据存储[12]；区块链[8]；日志[34]；数据信任/信托[11]；动态社会契约[12]	审计程序[1]；重用限制[13]；问责制[8,13,42]；机器智能委员会[8]

2.1 公共数据开放中的隐私风险控制教育手段

Watson H 等[43]提出，健康数据环境中，需要转变思维方式。他强调优先考虑以患者为中心的研究，并减轻需要量化的患者实际隐私风险，同时必须从患者本身开始自下而上地采取激励措施。这强调了公众信任的重要性。在开放公共数据过程中，日益增长的背景知识导致现有的数据发布隐私保护模式大多无法抵御攻击[44]；通过使用人工智能，数据隐私风险及得出有偏见或错误结论也变得更加突出[45]。这极大地削减了公众对政府、组织机构

的信任，同时阻碍了公共数据的开放与共享。然而，当人们意识到他们的隐私受到尊重和保护时，他们会更加自信地参与社会和经济活动。隐私仪表板和个人数据存储是个人用于表达有关保留和使用其数据的详细权限的工具，它帮助提供透明度和对个人数据的控制，有助于提升公众的信任。个人可以使用基于 Web 的隐私仪表板向选定方或特定用途授予对其数据的精细访问权限。还可在用于监控和自动评估的“仪表板”中查阅个人层面的现有数据源被链接情况[28]。个人数据存储使个人能够有效地对有关他们信息的存储位置和访问方式进行细粒度控制，从而选择在特定时间与特定方共享特定个人信息。个人数据存储不仅提供了增强的控制，而且作为用户控制的交互式系统，是开发更丰富的问责机制、在线聚合方法和高级安全机制的潜在基础。

对于信息或数据管理平台而言，公众的信任也至关重要。研究表明，社会规范、媒体代表和报道、对责任方的看法会影响人们对信息管理平台的信任和使用意愿[46]，而保持公众对信息平台的信任是减轻对数据安全性、隐私和功能效率的普遍担忧的关键。如 Shi M 等[47]从用户角度分析了影响医疗大数据安全和隐私泄露的关键指标，包括用户访问行为和信任度。并且在判断用户“非法行为”方面，将用户的信任值纳入风险评估指标，可以减少系统误判的可能性。

信任与隐私是彼此紧密联系、相互作用的两个重要因素，包括数据主体、控制者、接收者之间的信任。Ruotsalainen P 等[48]认为，可信度(即以合乎道德的方式处理健康和保健信息并保证隐私)是未来个人健康系统、无处不在的医疗保健和普遍健康的基石之一。基于无处不在的信息空间的普遍健康和风险分析框架模型，他们制定了实现可信信息共享的原则，包括数据主体应有权动态验证信任并控制其健康信息的使用，以及设置基于情境的上下文感知个人策略的权利；数据收集者和处理者的责任包括信息处理的透明度，以及利益、政策和环境特征的开放性。这些原则为自主管理健康领域的隐私和信息奠定了基础。基于信任的方法，Zuo Y J 等[49]构建了包括供应链成员信任评估、数据分类和基于信任的决策在内的框架，旨在控制和减轻参与者在供应链网络中面临的信息风险(如信息机密性、隐私和完整性的风险)。这充分表明建立数据主体、控制者、接收者及公众间的相互信任对于数据保护和防止个人信息泄露的重要性。

信任的建立依赖于透明和安全的数据处理措施。动态社会契约模型以一套商定的关于如何共享数据的合理期望为基础，对提供可接受的保证的治理以及谁对什么负责达成一致[50]。如 Open Mustard Seed Platform 是一个开放数据平台，它允许个人就其个人信息的使用进行社会契约谈判，通过基于同意的平台来管理数据，使人们在合法构成的“信任框架”内共享个人数据[12]。数据信托或数据受托人是另一种新兴方法。这种法律和政策框架方法考虑了第一方或第三方实体(数据主体除外)，受一组经批准的法律可执行义务的约束，以管理数据。Potoczny-Jones I 等[1]针对智慧城市协调数据敏感风险与预期收益问题，提出了“数据信任”解决方案。该技术框架强制实施个人身份数据最小化、访问控制以及灵活而精细的披露和编辑控制，并结合了法律上可执行的数据使用义务和责任；Young M 等[11]提出由第三方公私数据信托提供的综合法律技术方法，旨在透明度、所有权、隐私和研究目标之间取得平衡。基本成员资格允许公司和机构实现对合规性报告和核心方法研究数据的低风险访问，而模块化数据共享协议支持广泛的项目和用例。除非协议中另有明确规定，否则所有数据访问最初都是通过定制的合成数据集提供给最终用户。安全共享站点以安全和隐私保护的方式对数据进行计算，而无需发布原始数据，且所有数据共享都是透明且可审计的[33]。这种方案解决了对数据垄断的担忧，即没有人拥有数据。数据信托或数据受托人、动态社会契约模型等都涉及整个数据生命周期的隐私控制，在每个数据生命周期阶段通过“信任”框架或方案来管理数据，尽管他们提供“信任”方式不同，前者主要借助信托，后者基于谈判。此外，利用区块链的去中心化、透明化和可信度也有助于解决信任、安全和隐私问题。Kang H 等[51]开发了一种基于区块链的新型接触者追踪移动应用程序 BeepTrace，旨在缓解大流行并缓解接触者追踪的隐私问题，特别是解决了第三方信任问题。

2.2 公共数据开放中的隐私风险控制法律手段

2.2.1 隐私保护主要法律法规

当前，隐私风险控制主要的法律法规包括个人信息保护法(PIPA)、数据保护法(DPA)、通用数据保护条例(GDPR)、个人医疗信息保护法(HIPAA)等。

韩国《个人信息保护法》主要保护个人的自由和权利，并通过规定处理和保护个人信息来实现个人的尊严和价值。2020 年 2 月的修订允许未经信息主体同意，将“假名信息”处理为有限目的，这为私营公司和公共机构的公共大数据交付提供了更多动力。然而，修订后的法律侧重于利用大数据、企业间信息合并等，并引入了安全措施义务、罚款、刑事处罚等，但并没有为应对公共机构扩大公共数据开放而承担特别强化的风险管理责任[40]。

英国《2018 年数据保护法》法案规定了公共机构及其雇员如何处理与个人有关的数据。根据 DPA，参与处理数据的人员被称为维护和操作数据的数据处理者，负责做出有关数据以及是否可以共享数据决定的数据控制者。数据处理者和控制者必须遵守严格的数据保护原则，并确保个人数据得到合法、公平和透明的方式处理；收集和处理最低限度的必要数据，且仅用于特定目的、准确、保存时间不超过必要时间，并得到适当保障[14]。

欧盟《通用数据保护条例》是第一部通过明确定义欧盟内外个人数据处理和移动的背景来直接规范个人隐私的法律。GDPR 的颁布带来了巨大的变化[52]。关于受保护的信息，摆脱了传统的个人信息/匿名信息二分法，首次引入假名信息概念。在降低信息主体风险的同时，减轻个人信息处理者义务。关于同意，在法规全文中明确规定，对于用于研究的个人信息的使用，可以“广泛同意”，而不是信息主体的具体同意。关于同意豁免的原因可以确认，即使是敏感信息，信息主体对医疗目的、公共卫生和研究目的的同意也相对广泛地得到豁免。此外，GDPR 明确定义了个人(主体)的权利，即：①个人数据泄露通知；②访问收集的数据及其使用方式和目的；③删除数据的权利；④数据可移植性；⑤收集和处理过程中的数据保护。同时，引入并建立数据保护官，他们有义务将其数据处理活动通知当地数据保护机构。根据 1995 年的数据保护指令，欧盟委员会(2012 年)提出了对欧盟数据保护规则的全面改革。此外，ISO/IEC 29100 标准还定义了 11 项隐私原则(ISO/IEC-29100 2011)[53]。

美国有单独的个人医疗信息保护法，对非识别方法有具体规定，如“专家决策方式”和“保障港法”[52]。此外，还对“有限信息聚合”概念进行了单独规定，部分放松了管制。关于免除同意的理由，若机构审查委员会等批准同意豁免，即使未经个人同意，也有可能将个人医疗信息用于研究目的[52]。

2.2.2 隐私保护主要标准

人们可能一度认为公共机构收集的任何信息只会用于最初收集的目的，但开放数据的引入改变了这种情况。开放数据是任何人都可以免费下载、共享和重复使用的数据，除了可能需要引用来源之外，对重复使用或重新分发没有限制。开放政府旨在通过使数据易于获取来提高透明度和公民参与与协作。当前许多国家(地区)承诺开放政府，并将公共数据作为开放数据提供。这有助于让公民参与使用或重用政府数据，并且使数据分析提供商或其他政府组织具有通过促进更好的理解和加强决策来帮助政府改进其程序的潜力[36]。但在履行这一承诺时，公共机构需确保以开放格式发布的任何数据不包含个人或敏感数据，即识别或可用于帮助识别公民个人的数据[14]。因为这些潜在的风险因素会影响问责制，甚至降低公共机构的声誉。因此，需要权衡隐私风险和数据效用。①一个重要框架是公平信息原则(FIP)。FIP 的有影响力的版本是经济合作与发展组织(OECD)的《个人数据隐私和跨境流动保护指南》[13]。OECD 成员国于 1980 年通过了《隐私准则》，该指南强调，它们提供了“最低标准”，且没有“根据其性质和收集、存储、处理或传播的背景，对不同类别的个人数据实施不同的保护措施。其原则包括：收集限制原则、数据质量原则、目的说明原则、使用限制原则、保安保障原则、开放原则、个人参与原则、问责原则。现在，几乎每个 OECD 成员国都有以 FIP 为核心的数据隐私法；②平衡隐私和其他利益。为组织内决策提供信息的另一个重要方法是进行风险评估。美国国家标准与技术研究院(NIST)或国际标准组织(ISO)等公认的国

际机构制定了评估安全风险的指南(BS ISO 27000：2017；NIST，2012)。此外，为针对匿名性的要求并帮助组织实施数据去标识化流程以增强隐私，ISO 提出了一系列数据去标识化方法，如 ISO 20889 和 ISO 29100 系列。ISO 29100 和 ISO 29191 标准为大数据链接和开放数据提供了额外的保护，也减轻了公众和科研人员对隐私侵犯或无意中非法侵犯个人数据的担忧。ISO 29192-1 至 ISO 29192-5 用于少量信息安全的技术标准，包括分组密码、流密码和非对称加密等机制[54]。

2.3 公共数据开放中的隐私风险控制技术手段

2.3.1 去识别化技术

数据的发布可能会导致私人信息泄露。为防止泄露，应在全部或部分个人信息被删除或转换后发布数据，这些技术被称为去识别化[3]。其技术解决方案是从数据集中删除识别信息，同时保留数据的其余实用程序。表 3 总结了各种去标识化技术的概念和集成技术。

表 3 各种去标识化技术的概念和集成技术

Tab. 3 The Concepts and Integration Techniques of Various De-identification Technologies

加工技术	描 述	集成技术
假名化	将个人身份数据替换为无法直接识别的其他值	假名化、加密、交换
集合体	将统计值应用于个人信息，使其无法识别特定个人	数据平均法、微聚合
数据缩减	删除可用于识别个人信息的特定数据值	减少记录、减少标识符
数据抑制	通过将给定的识别信息转换为组的代表性值或预定义的范围来防止唯一信息跟踪和识别	替换、泛化
数据屏蔽	将个人标识信息转换为备用值，例如空格和“*”噪声	添加随机噪声、空白和插补

1）假名化指将个人身份数据替换为无法直接识别的其他值。如加密加盐方法，将虚假信息添加到隐私字段中并使用其他算法进行加密，以使恢复原始数据更加困难。Huang H H 等[54]通过加密加盐方法(Cryptographic Salting)对一组来自中国台湾地区的电子收费数据进行去识别。这种去识别技术提高了隐私字段的安全性，混淆了原始数据的内容。但没有改变原始去标识方法的一对一对应关系，其获得的结果与原始数据结果相同，但隐私字段的内容更复杂，更难观察。

2）集合体指将统计值应用于个人信息，使其无法识别特定个人。如隐私字段数据平均法，将信息的详细部分转换为简化的分类[54]。该方法提高了数据粒度，但不会导致扭曲和不准确；结合基于距离的记录链接与微聚合方法，通过记录链接对去识别化的开放政府数据进行数据挖掘[22]。该方法能够解决匿名和已经发布的开放政府数据的挖掘问题，支持异构数据挖掘以进行深入分析；Zouinina S 等[55]提出了两种通过微聚合实现 k-anonymity 的技术：k-CMVM 和 Constrained-CMVM。两者都使用拓扑协作聚类来获取 k-anonymity 数据，前者自动确定 k 个级别，后者通过探索来定义它。然而，集合体难以进行基于汇总数据的精确分析。并且当汇总数据量很小时，可以在数据合并过程中提取或预测个人信息。

3）数据缩减指删除可用于识别个人信息的特定数据值。删除直接标识符和准标识符是清理或去标识化的最常见方法，如删除敏感数据和隐私识别信息[16]；删除所有可能包含个人身份信息的自由文本数据字段[41]。但其可用信息数量有限，只能用于粗略的统计分析[54]，当涉及大型数据集时，删除标识符并不总是足以保护隐私，因为几个准标识符组合起来可以具有强大的识别能力。

4）数据抑制指通过将给定的识别信息转换为组的代表性值或预定义的范围来防止唯一信息跟踪和识别。抑制包括用一些特殊值替换原始数据，例如“*”。与之类似的，泛化指故意降低数据准确性(如将年龄转换为年龄组)。然而，数据抑制和泛化都难以进行精确数值分析。

5）数据屏蔽指通过隐藏准标识符的一部分将数据划分为多个组[3]。Templ M 等[35]通过向事件历史日期中添加噪声，发现即使在高噪音水平下，

也能保持高效用，与原始数据相比，保留了事件数据的基本属性；Badu-Marfo G 等[56]发现在两种地理随机扰动方法(地理不可区分性(Geo-indistinguishability)和甜甜圈地理掩码(Donut Geomask))中，实现的 k-estimate 匿名性随甜甜圈地理掩码所需的匿名性线性增加，而地理不可区分性高度依赖于其隐私预算因素，且在确保期望实现的 k-estimate 匿名性方面不是很有效。甜甜圈地理掩码是 k-anonymity 位置隐私保护机制的实现，通过使用点位置的基础邻域人口密度来确定混淆距离以实现隐私保护。地理不可区分性是位置数据差分隐私的实现。它保证受访者的位置在指定的保护距离内受到保护，增加的噪声水平随距离而降低，其速率取决于所需的隐私级别。

在实践中，即使通过上述技术执行了足够的去标识化措施，若数据没有通过与匿名化相关的充分性评估，它仍可以通过逆向工程将数据与补充信息相结合进行推断而被识别并视为个人信息。

2.3.2 匿名技术

匿名技术是隐私保护领域的重要手段[25]。通常，以下匿名化测试主要用于评估去标识化过程的充分性：k-anonymity、l-diversity 和 t-closeness 等。

k-anonymity 模型是最基本的评估技术之一，生成数据集时通常会检查 k-anonymity[41]，可以修改准标识符以避免任何数据链接[30]。Luthfi A 等[36]提出贝叶斯信念网络方法，该模型使用像 k-anonymity 这样的抑制技术来匿名化敏感属性，并构建决策过程的因果关系，以开放健康患者记录中的数据。此外，k-anonymity 也涉及数据隐私和效用的权衡，k-anonymity 原则是若无法将个人与公开发布的数据集中的 $k-1$ 个其他个人区分开来，则可以实现隐私。其中，k 值越高，重新识别风险就越低[33]。特别是 Santos W 等[30]对 ARX k 匿名算法的 k 值进行的敏感性分析表明，匿名化过程可能导致少数群体和社会人口弱势群体的代表性不足。因此，需根据需求情况决定 k 值。k-anonymity 模型的缺陷是易受到同质性攻击和背景知识攻击。因此，Tudor C 等[18]讨论了一种弱 k-anonymity 的替代方案，它要求仅在记录的一个子集中强制执行，这意味着那些不通过 k-anonymity 控制的变量有可能被用来识别某人。然而，当对这些变量的兴趣较低时，这种风险通常很小。因此，这可能是一个更实用的选择。

l-diversity 是一种降低泄露机密信息风险的技术[41]。l-diversity 将大于或等于 l 的良好表示敏感值分配给每个等价类，通过额外要求在每个匿名组中存在表示良好的值来扩展 k-anonymity。Ali S 等[57]采用 l-diversity 来保护敏感标签，避免攻击者利用这些标签来推测私人信息；疾病控制和预防中心的病例监测科将流行病学数据集与隐私保护算法相结合，通过自动化工作流和 R 统计软件实现和验证 k 匿名性的字段级抑制和 L 多样性[41]，并根据该流程生成了两个去识别化的公共数据集。然而，l-diversity 无法防止概率推理攻击和属性披露。

t-closeness 是 l-diversity 的进一步延伸。这种方法不仅保证敏感值的良好表示，还要求匿名组内每个敏感属性的分布与属性在整个数据集上的分布相同，取模阈值 t。然而，与 k-anonymity 和 l-diversity 一样，t-closeness 下的年龄、性别、种族甚至工作类型等受保护的属性仍然可以从加速度测量数据中推断出来。并且，这些传统的隐私保护模型对攻击模型和攻击者的背景知识做了过多的假设，各种匿名公共记录的传统方法已被证明存在隐私泄露风险。直到差分隐私技术的出现，这个问题才得到了很好的解决。

差分隐私方法是通过在原始数据或统计数据中添加噪声来处理数据信息和转换原始数据。该模型可降低最大后台攻击风险，并定义隐私保护等级的量化评估方法。与 k-anonymity 不同，差分隐私是基于概率的，它使用不同的机制来隐藏数据的真实价值以保护隐私，如引入噪声或虚假数据。差分隐私的缺点与 k-anonymity 的缺点相似：为了实现足够的隐私级别，必须添加一定量的噪声。添加噪声等效于有意向数据集添加错误。这可能导致从数据分析中得出一些错误的结论。Nahmias Y 等[2]建议监管机构应该使用差分隐私算法在准确性和隐私之间进行权衡，并提出基于雾计算的政府统计数据发布的差分隐私框架，开发了一种基于 MaxDiff 直方图的数据发布算法，可用于实现基于雾计算的用户隐私保护功能；Piao C H 等[44]也提出了一种基于

MaxDiff 直方图的数据发布算法，通过应用差分方法，将拉普拉斯噪声添加到原始数据集中，根据最大频率差，对相邻数据箱进行分组，构建平均误差最小的差分隐私直方图。该方法可以有效保护公民隐私，降低查询敏感度，提高发布数据的实用性。然而，差分隐私并不能解决所有隐私问题，也不会保护个人免受未经授权的信息收集、处理或防止安全漏洞。

安全多方计算方法，受密码学领域启发，信息泄漏量根据对手可访问的信息量来衡量[55]。它使两方或多方(彼此不完全信任)能执行涉及其两个数据集的计算，而不透露彼此的任何信息。其他高级加密方法可对数据进行计算，并限制对基础数据的学习。如功能或同态加密能够对加密数据进行计算[32]，而无需解密数据并将其暴露给攻击者。

除上述技术外，提供匿名数据的另一种方法是生成与原始数据具有相同特征的合成数据，可使用机器学习和统计建模方法。合成数据是从使用原始数据集开发的统计模型生成的。生成合成数据最初被用来填补缺失的条目，现在被广泛用于保护隐私，因为合成数据集不直接指向任何“真实”的人。如 Li W 等[58]基于深度学习的生成模型来解决敏感数据被开放发布问题，该模型生成模拟数据以掩盖原始数据。合成数据通常具有非常低的披露率，但当原始数据具有复杂的结构时，数据效用也相对较低。因此，Young M 等[11]在发布数据之前从数据集中删除不需要的偏见和专有信息，并将这些方法与差分隐私技术相结合，当合成数据集不足以进行分析时，调用由强大治理支持的结构化数据使用协议。Lee J S 等[22]将微聚合应用于合成数据生成器以链接和利用异构开放政府数据(微数据)，允许用户调整隐私阈值水平，以确定隐私披露风险和数据效用之间的适当平衡。这种将合成数据与对原始数据的强大法律保护结合使用，可在透明度、所有权、隐私和研究目标之间取得平衡。

2.4 公共数据开放中的隐私风险控制程序手段

通知和同意是数据收集和接受中常用的隐私保护工具，并且在管理个人数据处理的欧洲法律中，获得数据主体的同意是支持公平合法处理个人数据的主要程序机制。同时，通知和同意也有助于确保公共数据开放共享的透明度。在寻求适用的规范时，仅仅遵守法律和采用一次性同意程序不足以确保数据使用在道德上是合理的。因为人们通常希望在数据科学项目的所有阶段都具有透明度，并被告知何时和为什么收集有关他们的数据以及项目的结果是否实现[8]。因此，许多知情同意模型被提出以适应不同情景下的个人隐私保护。①分层或分类同意模型为研究参与者提供了如何以及何时使用数据的选择；②动态同意允许参与者随着时间的推移更新他们的同意偏好，并将结果返回给感兴趣的人；③一揽子或一般同意模型指参与者可选择同意未来对其数据的所有研究使用，而无需获得该研究可能需要的详细信息；④选择退出指参与者主动退出研究。通过“选择退出”，数据主体可以反对将数据用于次要目的[50]；⑤自动同意模型以高精度地预测用户的数据共享决策，来避免提示用户做出大多数决策。

尽管在信息生态中，同意可能是主要的，但同意绝不是合法处理个人数据的唯一机制[50]，它还与维护自治原则、隐私、透明度和不歧视有关。足够的透明度、对有害使用和商业化的控制、反对的能力，特别是反对任何被认为不适当或特别敏感的处理对用户接受具有较低个人控制水平的同意模式至关重要[7]。更友好的智能设备界面可能是一个好的方式以便用户能够控制数据的使用内容和方式[39]，同时也可以帮助用户更好地接收通知并控制同意选择。

透明度和滥用问责制对实现数据效用和个人隐私保护平衡至关重要。如数据资产登记册[19]、公开辩论[8](或相关记录[26])可告知公众，政府持有和发布哪些类型的信息，他们如何决定向公众发布或隐瞒哪些数据，并在特殊情况下记录在案；而数据资产清单，可帮助组织制定与部门活动相关的数据管理计划及治理结构，以处理出现的问题[59]。此外，算法问责制和透明度对确保数据安全非常重要[8]。机器智能委员会旨在确保随着新一代算法的开发，公共利益得到保护。而区块链可提供更大的问责制和安全性的承诺[8]。对于数据收集者和数据的个人主体而言，需了解数据的潜在和实际用途。为数据主体实现此类透明度的一种工具是隐私

仪表板，该仪表板向个人提供有关哪些实体正在访问其数据、他们如何使用数据以及他们因使用其数据而可能面临的任何隐私风险的通知。在问责制方面，对滥用数据的制裁很重要，包括违反保护或滥用的处罚或其他后果的信息[7]。滥用责任包括使个人能够了解其数据是如何被共享和使用，对侵犯隐私的行为进行民事和刑事处罚，以及因不当使用其数据而受到伤害的个人的私人诉讼权。

在发布信息时，机构必须平衡隐私和效用，包括广泛利益相关方的专家小组参与制定决策来确保合规性。同样重要的是评估重新识别风险，因为即使数据集严重不完整，也可能不符合匿名化的现代标准[30]。开放数据的关键性级别、开放性、攻击风险、信任和使用限制是隐私风险与收益权衡中的重要考虑因素，可通过决策引擎和评分矩阵进行评估[16]；也可通过基于熵的再识别风险来衡量开放数据中的隐私泄露风险，同时结合基于熵的数据实用模型，在保证隐私的同时保证数据的可用性[3]；也可通过贝叶斯信念网络模型分析打开数据时导致风险的因果机制[36]；情况目录通过列出在评估是否以及在何种条件下发布数据集时应考虑的情况或因素，及应如何重新发布数据集的不同选项来帮助作出决定[13]。此外，进行正式的入侵者测试有助于评估重新识别风险，涉及使用“友好的入侵者”来尝试查看他们是否可以重新识别数据集中的任何人并捕获入侵者可能链接到数据集的其他信息以发生泄露，其中适当选择入侵者对于获得准确结果非常重要[18]。

组织在通过信息系统共享数据时，使用访问控制来保护隐私。①这种系统可能要求所有用户注册并共享个人信息，并且使用系统配置文件进行身份验证[10]。如，在发布个人的犯罪历史之前，平台可以要求请求者提供其标识符、全名和动机。通过此功能，专业人员在获得必要的批准后才能够下载他们想要的数据[16]，并且这种请求可通过速率限制来防止快速触发请求[42]；②区分 3 种类型的用户，即管理、授权用户和运营商，来进行访问控制。如管理层负责向用户授予访问权限并批准运营商上传的数据集在平台上发布[16]；③借助分层访问系统进行访问控制，通过身份验证模块将对私人信息的访问限制为数据所有者，实现分层访问，从而保护隐私[20]。分层访问还可包含更高级的数据共享模型。如向公众提供列联表形式的汇总统计数据等；④向研究人员群体提供交互式查询系统，向通过仔细筛选程序获得批准的少数分析人员提供原始数据。交互机制可使用户能够提交有关数据集的查询并仅接收查询分析结果，分析结果可通过图表等可视化形式呈现。

对披露数据进行重复使用限制是在隐私和开放数据政策之间取得平衡的另一种方法。重复使用限制可以以许可证形式呈现[13]。许可证可能要求用户不要重新识别数据，或在发现个人可以或已被重新识别的情况下通知许可人。此外，在线提供数据的组织通常会提供服务条款或参考道德准则，这些准则描述了使用有关个人机密数据的准则和最佳实践。如科学用途数据集仅发布给签署数据使用协议的经批准的研究人员，且包含比公共使用数据集更多的变量[41]。数据使用协议通常涉及对数据的使用、共享和重用的限制，保护数据的义务，因使用或滥用数据而造成损害的责任，以及执行协议条款的机制。一般召集/管理数据的组织对其正确访问和使用负有法律责任[32]。然而，在实践中很难发现违反数据使用协议的行为并执行条款。因此，需要通过签署合同或协议等方式确保数据后续使用存在的隐私问题。如在与特定学生共享敏感数据前，可与其签订特殊的合同协议[16]；只有在证明道德批准、签署数据使用协议和明确的数据管理计划的情况下才授予大学教师访问权限[9]等。

此外，在用户访问数据后，还涉及审计系统，该系统包括法律和技术机制，用于检测信息滥用和防止个人违反数据使用政策。如 Tzermias Z 等[42]提出可确定负责个人身份信息泄漏的公务员，通过使用诱饵文件及“诱饵”信息来识别泄漏；Potoczny-Jones I 等[1]在试点平台中检测违反策略的行为，并向公众提供未发生此类违规行为的透明度以保护隐私；Lee J T 等[34]提出在发生数据泄露风险时通过日志查看以往请求，识别访问过数据的个人，并请求他们返回或销毁受损信息。此外，可能需要第三方审计每年审查数据隐私和安全程序，并且有权访问数据的承包商也可能需要进行此类审

计[29]。

3 结 语

本文对公共数据资源开放共享中的隐私风险控制研究进展进行了综述。综合分析发现，隐私风险控制研究基本覆盖教育、法律、技术、程序方面，涉及公共数据的收集和接受、转换、保留、发布和访问、访问后等阶段。总的来看，当前研究：①重视利益相关者对隐私的看法，倡导积极与利益相关者沟通、交流，特别是构建利益相关者间“信任”；②公共数据中的隐私风险控制法律政策正处于发展、完善阶段，主要聚焦于隐私与开放的平衡问题；③各个领域，特别是政府、城市方面，积极采用新兴技术来保护隐私，如区块链、差分隐私等。

但目前的研究存在：①缺少对公共数据隐私风险控制的讨论。现有研究集中于政府开放数据下的隐私风险控制，忽视了履行公共管理和服务职能的事业单位或企业的开放数据隐私问题；②隐私风险控制方法尚未形成体系。诸如“差分隐私”“数据信任”等隐私控制方法，由于对隐私保护和数据效用的目标或参数(这通常由数据发布者选择)不同，因此其方法、定义、名称存在差异。此外，研究多涉及技术或框架等隐私风险控制方法，而很少构建包括政治、经济、社会、技术在内的综合的隐私风险控制体系；③缺少公共数据资源开放共享中的隐私风险控制的实证研究。虽然内容分析法被广泛用来分析公共数据资源开放共享中的隐私风险控制政策法规，但缺乏相关的隐私风险控制调查与实证研究，未能有效揭示公共数据资源开放共享中的隐私风险控制现状。

这对数据开放和隐私保护的实践工作有一定启示：①深化公共数据资源开放共享中的隐私风险控制的理论研究，探索数据开放与隐私保护相统一的法律制度体系，针对不同时期出现的新问题，及时完善在数字化进程中的法律空白。第一，明晰公共数据与政府数据的区别与联系，考虑履行公共管理和服务职能的事业单位或企业的开放数据隐私问题；第二，面向采集、存储、传输、共享、开放、使用、销毁等公共数据全生命周期过程，探究有针对性的数据隐私保护措施，落实公共数据风险评估、分类分级、合规监管等要求，从而明确安全保障职责，强化安全运行管理，提升安全保障能力；第三，探索完善的公共数据管理组织机构制度，明确公共数据治理和数据保护人员、机构及其职责；第四，通过法律鼓励公共数据创新与应用。探究支持公共数据创新和应用的政策和法律，以推进公共数据驱动的创新，促进公共机构和企业利用公共数据解决社会问题和提供公共服务。②加强对隐私技术的研究，首先，综合运用区块链、隐私计算、数据安全沙箱、边缘计算、同态加密、多方安全计算等隐私增强技术，探索新型开发利用模式；其次，探索技术理念与实践相统一的隐私风险控制途径，通过实施适当的技术和组织安全措施，以确保个人数据的安全，防止未经授权的访问、使用、披露、更改或破坏；最后，构建“一中心一张网一平台”隐私风险控制体系，确保公共数据开放中的隐私风险得到有效控制。③坚持以公众意见为导向。首先，探究建立公众信任的方法。在为公共利益共享数据的情况下，更清楚地说明如何定义和判断公共利益。如增加公共数据资源开放共享中的隐私风险控制的实地调查与实证研究，引入公共数据资源开放共享中的隐私风险的量化分析，提高研究结果的可靠性、可操作性与适用性等；其次，个人的隐私权必须与其他公民权利以及更广泛的社区和社会的权利相权衡。明确公共部门处理数据共享项目过程以建立一致性和透明度；最后，关注人工智能对公众隐私的影响。从法律和监管角度为开发和部署人工智能创造合适的环境，包括将伦理原则纳入共识规范框架，确保整个社会了解其基本情况，并能够就其与人工智能技术的关系做出积极决策。

本文区别于已有综述文章，重点梳理了公共数据开放与隐私相互作用的研究进展。将公共数据开放中的隐私风险控制研究分为教育、法律、技术、程序手段 4 个方面进行阐述，有助于了解、把握当前开放与隐私保护研究发展现状。最后，文章总结了当前研究的现状与不足，并展望未来发展方向，为推动安全、隐私的公共数据开放提供了新的研究方向。

参 考 文 献

[1] Potoczny-Jones I，Kenneally E，Ruffing J，et al. Encrypted Data-

set Collaboration Intelligent Privacy for Smart Cities [C] //Assoc Computing Machinery. Proceedings of the 2nd ACM/EIGSCC Symposium on Smart Cities and Communities (SCC). Portland, OR, USA, 2019: 1-8.

[2] Nahmias Y, Perez O, Shlomo Y, et al. Privacy Preserving Social Norm Nudges [J]. Michigan Telecommunications & Technology Law Review, 2019, 26 (1): 43-91.

[3] Kim S H, Jung C, Lee Y J. An Entropy-based Analytic Model for the Privacy-Preserving in Open Data [C] //Ieee. Proceedings of the 4th IEEE International Conference on Big Data (Big Data). Washington, DC, 2016: 3676-3684.

[4] 周鑫, 张静, 谢津, 等. 区块链赋能突发公共卫生事件开放数据隐私保护研究 [J]. 现代情报, 2023, 43 (1): 141-150.

[5] 杨瑞仙, 李兴芳, 王栋, 等. 隐私计算的溯源、现状及展望 [J]. 情报理论与实践, 2023, 46 (7): 158-167.

[6] 盛小平, 毕畅畅, 唐筠杰. 国内外开放科学主题研究综述 [J]. 图书情报知识, 2022, 39 (4): 101-113.

[7] Kalkman S, Delden J, Banerjee A, et al. Patients' and Public Views and Attitudes Towards the Sharing of Health Data for Research: A Narrative Review of the Empirical Evidence [J]. Journal of Medical Ethics, 2022, 48 (1): 3-13.

[8] Drew C. Data Science Ethics in Government [J]. Philosophical Transactions of the Royal Society A Mathematical Physical & Engineering Sciences, 2016, 374 (2083): e20160119.

[9] Pisani A R, Kanuri N, Filbin B, et al. Protecting User Privacy and Rights in Academic Data-sharing Partnerships: Principles from a Pilot Program at Crisis Text Line [J]. Journal of Medical Internet Research, 2019, 21 (1): e11507.

[10] Moustaka V, Theodosiou Z, Vakali A, et al. Enhancing Social Networking in Smart Cities: Privacy and Security Borderlines [J]. Technological Forecasting and Social Change, 2019, 142: 285-300.

[11] Young M, Rodriguez L, Keller E, et al. Beyond Open vs. Closed: Balancing Individual Privacy and Public Accountability in Data Sharing [C] //Association for Computing Machinery. Proceedings of the FAT* '19: Conference on Fairness, Accountability, and Transparency. Atlanta GA USA, 2019: 191-200.

[12] Hardjono T, Deegan P, Clippinger J H. Social Use Cases for the ID3 Open Mustard Seed Platform [J]. IEEE Technology & Society Magazine, 2014, 33 (3): 48-54.

[13] Borgesius F Z, Gray J, Eechoud M. Open Data, Privacy, and Fair Information Principles: Towards A Balancing Framework [J]. Berkeley Technology Law Journal, 2015, 30 (3): 2073-2131.

[14] Henriksen-Bulmer J, Faily S, Jeary S. Privacy Risk Assessment in Context: A Meta-model Based on Contextual Integrity [J]. Computers & Security, 2019, 82: 270-283.

[15] Marshall Z, Brunger F, Welch V, et al. Open Availability of Patient Medical Photographs in Google Images Search Results: Cross-Sectional Study of Transgender Research [J]. Journal of Medical Internet Research, 2018, 20 (2): e70.

[16] Ali-Eldin A, Zuiderwijk A, Janssen M. A Privacy Risk Assessment Model for Open Data [C] //Springer International Publishing Ag. Proceedings of the 7th International Symposium on Business Modeling and Software Design (BMSD). Barcelona, SPAIN, 2018: 186-201.

[17] Zayatz L. Privacy and Confidentiality Resources [J]. Journal of Empirical Research on Human Research Ethics, 2009, 4 (3): 33-34.

[18] Tudor C, Lowthian P, Spicer K. Opening Up Government Data While Maintaining Data Privacy [C] //CEUR-WS. Proceedings of the Joint Workshops of the International Conference on Extending Database Technology and the International Conference on Database Theory, EDBT/ICDT-WS 2015. Ambato, Ecuador, 2015: 263-269.

[19] Whittington J, Calo R, Simon M, et al. Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government [J]. Berkeley Technology Law Journal, 2015, 30: 1899-1966.

[20] Rodriguez-Hoyos A, Estrada-Jimenez J, Urquiza-Aguiar L, et al. Digital Hyper-Transparency: Leading e-Government Against Privacy [C] //Ieee. Proceedings of the 5th International Conference on eDemocracy and eGovernment (ICEDEG). Quito, ECUADOR, 2018: 263-268.

[21] Kotwal V, Parsheera S, Kak A, et al. Open Data & Digital Identity: Lessons for Aadhaar [C] //Ieee. Proceedings of the ITU Kaleidoscope Conference-Challenges for a Data-Driven Society (ITU K). Nanjing, China. 2017: 1-8.

[22] Lee J S, Jun S P. Privacy-preserving Data Mining for Open Government Data from Heterogeneous Sources [J]. Government Information Quarterly, 2021, 38 (1): e101544.

[23] Giannouchos T V, Ferdinand A O, Ilangovan G, et al. Identifying and Prioritizing Benefits and Risks of Using Privacy-enhancing Software Through Participatory Design: A Nominal Group Technique Study with Patients Living with Chronic Conditions [J]. Journal of the American Medical Informatics, 2021, 28 (8): 1746-1755.

[24] Piao C H, Shi Y J, Yan J Q, et al. Privacy-preserving Governmental Data Publishing: A Fog-computing-based Differential Privacy Approach [J]. Future Generation Computer Systems, 2019, 90: 158-174.

[25] Liu L P, Piao C H, Cao H R. Clustering-Anonymity Method for Privacy Preserving Table Data Sharing [C] //Springer International Publishing Ag. Proceedings of the 16th IEEE International Conference on e-Business Engineering (ICEBE). Shanghai, China, 2020: 405-420.

[26] Minssen T, Rajam N, Bogers M. Clinical Trial Data Transparency and GDPR Compliance: Implications for Data Sharing and Open Innovation [J]. Science & Public Policy, 2020, 47 (5): 616-626.

[27] Rantala S, Swallow B, Paloniemi R, et al. Governance of Forests and Governance of Forest Information: Interlinkages in the Age of Open and Digital Data [J]. Forest Policy & Economics, 2020, 113: e102123.

[28] Bourgeois J, De Ridder L, Bogaert S, et al. Data-driven Change Towards Integrated Care [J]. International Journal of Integrated Care, 2018, 18 (S2): A267.

[29] Altman M, Wood A, O'brien D, et al. Towards a Modern Approach to Privacy-Aware Government Data Releases [J]. Berkeley Technology Law Journal, 2016, 30: 1899.

[30] Santos W, Sousa G, Prata P, et al. Data Anonymization: K-anonymity Sensitivity Analysis [C] //IEEE. Proceedings of the 15th Iberian Conference on Information Systems and Technologies (CISTI). Seville, Spain, 2020: 1-6.

[31] Ho D H, Lee Y Y. Big Data Analytics Framework for Predictive Analytics Using Public Data with Privacy Preserving [Z]. 2021 IEEE International Conference on Big Data (Big Data). Orlando, FL, USA; IEEE, 2021: 5395-5405. https://doi.org/10.1109/BigData52589.2021.9671997.

[32] Jones K, Daniels H, Heys S, et al. Toward a Risk-Utility Data Governance Framework for Research Using Genomic and Phenotypic Data in Safe Havens: Multifaceted Review [J]. Journal of Medical Internet Research, 2020, 22 (5): e16346.

[33] Austin L M, Lie D. Safe Sharing Sites [J]. N Y Univ Law Rev, 2019, 94 (4): 581-623.

[34] Lee J T, Freitas J, Ferrall I L, et al. Review and Perspectives on Data Sharing and Privacy in Expanding Electricity Access [J]. Proceedings of the IEEE, 2019, 107 (9): 1803-1819.

[35] Templ M, Kanjala C, Siems I. Privacy of Study Participants in Open-access Health and Demographic Surveillance System Data: Requirements Analysis for Data Anonymization [J]. JMIR Public Health Surveill, 2022, 8 (9): e34472-e34472.

[36] Luthfi A, Rukanova B, Molenhuis M, et al. Bayesian-belief Networks for Supporting Decision-making of the Opening Data By the Customs [C] //CEUR Workshop Proceedings. Proceedings of the 2020 Ongoing Research, Practitioners, Posters, Workshops, and Projects of the International Conference EGOV-CeDEM-ePart, EGOV-CeDEM-ePart 2020. Linköping, Sweden, 2020: 51-58.

[37] Madeyski L, Lewowski T, Kitchenham B. OECD Recommendation's Draft Concerning Access to Research Data from Public Funding: A Review [J]. Bulletin of the Polish Academy of Sciences Technical Sciences, 2021, 69 (1): e135401.

[38] Emam K, Arbuckle L, Koru G, et al. De-identification Methods for Open Health Data: The Case of the Heritage Health Prize Claims Dataset [J]. Journal of Medical Internet Research, 2012, 14 (1): e33.

[39] Harper S, Mehrnezhad M, Mace J. User Privacy Concerns in Commercial Smart Buildings [J]. Journal of Computer Security, 2022, 30 (3): 465-497.

[40] Kwon E. A Legal Study on Data Risk Management in the Public Sector Focused on Personal Information Risk Associated with Opening up Public Data [J]. Administrative Law Journal, 2020, 60: 165-190.

[41] Lee B, Dupervil B, Deputy N P, et al. Protecting Privacy and Transforming COVID-19 Case Surveillance Datasets for Public Use [J]. Public Health Rep, 2021, 136 (5): 554-561.

[42] Tzermias Z, Prevelakis V, Ioannidis S. Privacy Risks from Public Data Sources [C] //Springer-Verlag Berlin. Proceedings of the 29th IFIP International Information Security and Privacy Conference (SEC). Marrakech, Morocco, 2014: 156-168.

[43] Watson H, Gallifant J, Lai Y, et al. Delivering on NIH Data Sharing Requirements: Avoiding Open Data in Appearance Only [J]. BMJ Heal Care Inf, 2023, 30 (1): e100771.

[44] Piao C H, Shi Y J, Zhang Y Z, et al. Research on Government Data Publishing Based on Differential Privacy Model [Z]. 2017 IEEE 14th International Conference on e-Business Engineering (ICEBE). Shanghai, China; IEEE. 2017: 76-83. https://doi.org/10.1109/ICEBE.2017.21.

[45] Gao Y Y, Janssen M. Generating Value from Government Data Using AI: An Exploratory Study [C] //Springer International Publishing Ag. Proceedings of the 19th IFIP WG 85 International Conference on Electronic Government (EGOV). Linköping, Sweden, 2020: 319-331.

[46] Yeo V A, Mi Y, Kwok K T. Factors Affecting Adoption of Digital Contact Tracing During the COVID-19 Pandemic: A Literature Review [J]. Journal of Public Health and Emergency, 2022, 6: 23.

[47] Shi M, Jiang R, Zhou W, et al. A Privacy Risk Assessment Model for Medical Big Data Based on Adaptive Neuro-Fuzzy Theory [J]. Security & Communication Networks, 2020, 2020 (2020): e5610839.

[48] Ruotsalainen P, Blobel B, Nykänen P, et al. Framework Model and Principles for Trusted Information Sharing in Pervasive Health [C] //IOS Press. Proceedings of the Studies in Health Technology and Informatics, 2011: 497-501.

[49] Zuo Y J, Hu W C. Trust-based Information Risk Management in a Supply Chain Network [J]. Intemational Journal of Information Systems and Supply Chain Management, 2009, 2 (3): 19-34.

[50] Sexton A，Shepherd E，Duke-Williams O，et al. The Role and Nature of Consent in Government Administrative Data [J]. Big Data & Society，2018，5 (2)：1-17.

[51] Kang H，Zhang Z，Dong J，et al. BeepTrace for COVID-19 Pandemic：A Demo [C] //IEEE. Proceedings of the 2021 3rd Conference on Blockchain Research and Applications for Innovative Networks and Services，BRAINS 2021. Paris，France，2021：6-7.

[52] 최계영. Personal Information Protection in the Medical Field-Centering on the Legislative Systems of the European Union and the USA [J]. 경제규제와 법，2016，9 (2)：206-223.

[53] Ali-Eldin A M T，Zuiderwijk A，Janssen M. Opening More Data a New Privacy Risk Scoring Model For Open Data [C] //SciTePress. Proceedings of the 7th International Symposium on Business Modeling and Software Design，BMSD 2017. Barcelona，Spain，2017：146-154.

[54] Huang H H，Lin J W，Lin C H. Data Re-IdentificationA Case of Retrieving Masked Data from Electronic Toll Collection [J]. Symmetry-Basel，2019，11 (4)：550.

[55] Zouinina S，Bennani Y，Rogovschi N，et al. Data Anonymization Through Collaborative Multi-view Microaggregation [J]. Journal of Intelligent Systems，2021，30 (1)：327-345.

[56] Badu-Marfo G，Farooq B，Patterson Z. Perturbation Methods for Protection of Sensitive Location Data：Smartphone Travel Survey Case Study [J]. Transportation Research Record Journal of the Transportation Research Board，2019，2673 (12)：244-255.

[57] Ali S，Osman T，Mannan M，et al. On Privacy Risks of Public WiFi Captive Portals [Z] //Pérez-Solà C，Navarro-Arribas G，Biryukov A，et al. Data Privacy Management，Cryptocurrencies and Blockchain Technology：ESORICS 2019 International Workshops，DPM 2019 and CBT 2019. Luxembourg；Springer International Publishing，2019：80-98. https://doi.org/10.1007/978-3-030-31500-9_6.

[58] Li W，Meng P，Hong Y，et al. Using Deep Learning to Preserve Data Confidentiality [J]. Applied Intelligence，2020，50 (2)：341-353.

[59] Whittington J，Calo R，Simon M，et al. Push，Pull，and Spill：A Transdisciplinary Case Study in Municipal Open Government [J]. Berkeley Technology Law Journal，2015，30：1899.

（责任编辑：郭沫含）

（上接第 139 页）

[30] Liu J，Euan A. Five Challenges in Altmetrics：A Toolmaker's Perspective [J]. Bulletin of the Association for Information Science & Technology，2013，39 (4)：31-34.

[31] 荣国阳，李长玲，范晴晴，等. 主题热度加速度指数——学科研究热点识别新方法 [J]. 图书情报工作，2021，65 (20)：59-67.

[32] 杨建林，钱玲飞. 基于关键词对逆文档频率的主题新颖度度量方法 [J]. 情报理论与实践，2013，36 (3)：99-102.

[33] Luhn H P. The Automatic Creation of Literature Abstracts [J]. IBM Journal of Research and Development，1958，2 (2)：159-165.

[34] 李长玲，高峰，牌艳欣. 试论跨学科潜在知识生长点及其识别方法 [J]. 科学学研究，2021，39 (6)：1007-1014.

[35] 潘玮，牟冬梅，李茵，等. 关键词共现方法识别领域研究热点过程中的数据清洗方法 [J]. 图书情报工作，2017，61 (7)：111-117.

[36] 高楠，周庆山. 新兴技术概念辨析与识别方法研究进展 [J]. 现代情报，2023，43 (4)：150-164.

[37] 韩芳，张生太，冯凌子，等. 基于专利文献技术融合测度的突破性创新主题识别——以太阳能光伏领域为例 [J]. 数据分析与知识发现，2021，5 (12)：137-147.

[38] Hwang C，Yoon K. Methods for Multiple Attribute Decision Making [M]. Hwang C，Yoon K. Multiple Attribute Decision Making：Methods and Applications A State-of-the-Art Survey. Berlin，Heidelberg：Springer Berlin Heidelberg，1981：58-191.

[39] Nakamoto S. Bitcoin：A Peer-to-Peer Electronic Cash System [EB/OL]. https://bitcoin.org/bitcoin.pdf，2022-09-08.

[40] 袁勇，王飞跃. 区块链技术发展现状与展望 [J]. 自动化学报，2016，42 (4)：481-494.

[41] 杨晓晨，张明. 比特币：运行原理、典型特征与前景展望 [J]. 金融评论，2014，6 (1)：38-53.

[42] Garfield E. Citation Frequency as a Measure of Research Activity and Performance [J]. Essays of an Information Scientist，1973，1 (73)：406-408.

[43] 中华人民共和国国务院. 国务院关于印发“十三五”国家信息化规划的通知 [EB/OL]. https://www.gov.cn/zhengce/content/2016-12/27/content_5153411.htm，2022-09-27.

[44] 中华人民共和国国务院办公厅. 国务院办公厅关于支持国家级新区深化改革创新加快推动高质量发展的指导意见 [EB/OL]. http://www.gov.cn/zhengce/content/2020-01/17/content_5470203.htm，2022-09-27.

[45] 罗棋，闵超，颜嘉麒，等. 国际区块链研究主题挖掘及演化分析 [J]. 现代情报，2021，41 (9)：157-166.

（责任编辑：杨丰侨）