

全生命周期视阈下公共数据伦理准则框架研究

马海群^{1,2}, 李金玲¹, 于同同¹, 张涛^{1*}

(1.黑龙江大学 信息管理学院, 哈尔滨 150080; 2.黑龙江大学 信息资源管理研究中心, 哈尔滨 150080)

摘要: [目的/意义]公共数据作为国家基础战略资源,在完善社会治理、提高政府服务水平等方面的作用举足轻重,其伦理准则框架构建有助于公共数据参与者规范自己的行为,避免发生隐私泄露、数据过度收集和数字鸿沟等公共数据伦理问题。[方法/过程]本文结合以往研究和现有公共数据政策,全面总结公共数据的内涵、特征,在此基础上结合典型数据生命周期模型构建公共数据全生命周期的循环层次模型。[结果/结论]在探索了公共数据发展应遵循的伦理准则的前提下,最终以公共数据全生命周期模型的六大阶段和17个子阶段作为一、二级指标构建公共数据伦理框架,为规约公共数据伦理问题提供参考和契机,激发公共数据要素价值,助力推进公共管理和服务数字化进程,建立良好公共服务秩序。

关键词: 公共数据; 数据伦理; 全生命周期模型; 准则框架

中图分类号: G203; G82-05

文献标识码: A

文章编号: 1002-1248 (2023) 06-0029-14

引用本文: 马海群,李金玲,于同同,等.全生命周期视阈下公共数据伦理准则框架研究[J].农业图书情报学报,2023,35(6): 29-42.

1 引言

数据是国家基础性战略资源。2019年10月,十九届四中全会首次提出将数据作为新的生产要素参与分配,特别在公共数据的开放和开发利用方面。此后,地方政府对公共数据管理、开发利用方面的条例蔚然成风,但其伴随的安全风险和伦理问题受到了社会广泛关注。人们在享受着公共数据产品带来的便捷

之余,也对其带来的数据隐私安全问题产生担忧,亟需伦理准则框架约束。一方面,公共数据的开发利用能够激发数据要素价值,催生数据公共服务企业和产品;另一方面,公共数据被非法获取、损毁、篡改或者不当利用引起数据隐私泄露^[1]、数据垄断^[2]和数字信任危机^[3]等社会伦理道德问题,损害公共利益和公共部门的公信力。因此,构建完善的公共数据伦理准则框架成为重中之重,也为公共数据多维度研究和发展提供契机。

收稿日期: 2023-05-10

基金项目: 国家社科基金项目重大项目“面向数字化发展的公共数据开放利用体系与能力建设研究”(21&ZD336); 2021年度黑龙江大学学位与研究生教育教学改革研究项目“面向德才兼备研究生高层次人才培养的数据信用风险集成化治理研究”(JGXM-YJS-2021012); 黑龙江省教育科学“十四五”规划重点课题“后疫情时代高校线上教学效果影响因素实证分析”(GJB1422044)

作者简介: 马海群(1964-),男,教授,博士生导师,博士,研究方向为信息政策与法律。李金玲(1998-),女,硕士研究生,研究方向为信息计量与评价。于同同(1997-),男,硕士研究生,研究方向为文献计量

*通信作者: 张涛(1981-),男,副教授,硕士生导师,博士,研究方向为数据与算法安全治理。Email: zhangtao@hlju.edu.cn

目前,从数据伦理方面出发,社会各界对伦理问题的研究多集中于“大数据伦理”^[4]、“人工智能伦理”^[5]、“医疗数据伦理”^[6]和“教育数据伦理”^[7]等,鲜少有人从公共数据的角度对其伦理问题进行深入探究;从公共数据方面出发,其研究也多集中于“公共数据平台”^[8]、“公共数据开放共享”^[9]、“公共数据授权运营”^[10]等公共数据赋能数字化改革、公共数据安全的热点,鲜有学者从数据伦理角度对公共数据进行准则框架研究。

数据伦理是避免在数据流通中引起社会伦理失范问题(如数据隐私泄露、数据垄断等)而作出的规范,是基于问题现象的阐述。数据道德是指数据参与者进行数据决策时需合乎道德,是数据伦理的基础原则。数据隐私是指数据中包含很多敏感个人信息和事项,极易泄露,因此是数据伦理要求保护之一。数据安全是指以数据为底层逻辑形成的数字技术、制度等的安全状态,有事前预防和事后追述两方面,而数据伦理属于事前预防部分。本文主要为避免公共数据伦理问题的发生,进行公共数据伦理框架构建,因此当公共数据参与者进行数据决策时需要在合乎道德的基础上,在框架构建时主要考虑涉及到伦理问题的技术、原则、规范问题,而非纯数据安全的技术操作规范。

2 公共数据伦理研究现状

国内对公共数据伦理的研究主要包括各领域公共数据伦理研究和公共数据应用于各领域引起的伦理问题探究。^①各领域公共数据伦理框架和规范研究。张誉元聚焦于课堂教育场域下的教育数据,从数据生命周期的角度,分别映射隐存的伦理问题和风险并提出了教育数据伦理问题的4条规约之径^[7]。关键出台系列专题,深入分析医学科学数据共享面临的伦理挑战,并提出了伦理要求,以保护个体利益的前提下促进医学数据共享和再利用^[11]。刘亚金等将利己主义因素、功利主义因素、利他主义因素及实用主义因素作为数据伦理参照标准,对科研数据伦理进行了界定,构建了面向科研人员的科研数据伦理框架^[12]。^②公共数据

应用伦理研究。刘霞从信息伦理的角度出发,对G2C电子政务中影响用户接收行为的信息伦理因素进行探讨^[13]。王亚强就电子政务系统的负效应与伦理诉求话题进行探讨,提出了对网络行政伦理和制度性创新的诉求,对当时电子政务伦理的发展具有深远影响^[14]。吴爱荣则对政府出行交通数据在全数据生命周期中出现的数据伦理问题进行总结,运用演绎推理法分析其成因并提出了相应的对策^[15]。

国外对公共数据伦理的研究主要包括:WANG通过案例分析发现了存在于公共部门的隐私、财产、准确性和可访问性这4个信息伦理关键问题^[16]。2015年11月,英国皇家统计学会召集来自政府、学术界和私营部门的28名参与者成立英国数据伦理委员会,并协助讨论出一套的切实可行的数据决策遵循的伦理优先事项,在决策者的数据优先事项基础上,对于政府数据的伦理考量做了进一步规定,包括改进数据访问和开发数据基础设施,促进了政府数据伦理框架的发展^[17]。随后国外学者们对政府数据伦理框架的探索蔚然成风:DREW阐述了政府数据伦理框架的必要性,并提出可以采取一种开放的、循证的、以用户为中心的方法来创建一个伦理框架^[18]。JOSEPH分析了美国公共部门数据泄露事件,提出公共数据泄露的受害者不仅是个人企业信息,还包括公共部门和机构提供的公共服务,并发现后斯诺登时代人们对公共机构的数据安全和隐私协议的关注倍增^[19]。DREW阐述了将数据伦理框架用于政府数据,以提供更多实用指导的认可^[20]。同年,英国政府中央数字和数据办公室发布了《数据伦理框架》,用以指导公共部门组织在规划、实施和评估新政策或服务时适当和负责任地使用数据^[21]。ROSEN对公共卫生中大数据的新作用和伦理影响,讨论了这些与审判入狱者(JIPs)有关的问题,并提出了促进这一领域的伦理分析和指导方针制定的初步步骤^[22]。MOLLIDREM以美国艾滋病感染者网络为例,分析了大数据技术应用于公共卫生数据是如何挑战管理数据重用和公共利益的基本逻辑和监管范式,以及会紧急出现的伦理问题和风险,并提出利益相关者应该针对公共卫生数据的重复使用就监管框架、伦理规范和最佳实践进行一

系列新的以改革为导向的对话^[23]。

综上所述, 国外多采用“政府数据伦理”的概念, 起始于公共部门数据泄露侵犯个人隐私等伦理问题频发引起的思考。以 2013 年斯诺登事件为转折点, 此后各国开始重视公共数据伦理问题, 并形成了政府学者研究相辅相成、相互促进的局面。以英国为例, 2015 年英国成立数据伦理委员会对政府数据伦理进行的考量和规定为政府数据伦理的发展谋势, 促使部分各界学者对应用于各界的政府数据伦理框架进行针对性探索和研究, 从而反哺国家权威机构出台框架——2018 年英国发布针对公共部门的《数据伦理框架》。后学者们将其运用于各领域的政府数据, 政府数据伦理研究开始走上正轨。

不同于国外爆炸性事件起始, 中国对公共数据的研究是由和平式的公共数据开放需求拉动的, 对公共数据伦理的研究起步较晚, 其概念整体上经历了“政务信息 - 政务数据 - 公共数据”的扩展性演进历程^[24]。相较于国外对通用型和领域专用型公共数据伦理框架的研究双管齐下, 中国目前仅局限在各领域专用型数据伦理框架的研究, 而缺少整体性、普适性较高的通用型公共数据伦理框架研究。同时缺乏以公共数据全生命周期视阈下对公共数据伦理的研究。本文将以公共数据特点和公共数据全生命周期为基础构建公共数据伦理框架, 为规约公共数据伦理问题提供参考和契机。

3 公共数据的概念、特征、全生命周期

对公共数据的概念、特点以及数据全生命周期的探析是构建公共数据伦理框架的必要前提。公共数据概念的明晰是分析公共数据特征的前提, 公共数据特征决定了公共数据全生命周期阶段和伦理框架内容的准确性, 而公共数据生命周期阶段的探索能够为伦理框架形成提供阶段划分依据和正当性基石。

3.1 公共数据的概念

目前中央政策立法中尚未对公共数据做出明确的

概念解释, 但部分省市级地方对公共数据进行先行探索的立法时, 对公共数据进行了相关定义。据不完全统计, 山东、江苏、江西、浙江、广东、北京、湖北等 5 个省份及其在内的 5 个市级已经专门针对公共数据出台了管理办法或条例; 而山东、贵州、吉林、浙江、上海等 10 个省份(省级市、自治区)及其在内的 4 个市级则出台了针对公共数据安全、公共数据开放的管理办法或暂行办法, 且政策中均对公共数据做出了相关定义和解释。

结合 24 个管理办法和条例中的公共数据定义及其法理逻辑, 我们以主体、行为和内容要素最大化原则总结公共数据的定义。公共数据, 是指公共管理和服务机构在依法履行公共管理职责、提供公共服务过程中, 收集、产生和制作的以电子或其他形式记录和保存的各类数据及其衍生数据。其中, 主体要素包括行政机关、人大、司法等国家机关和法律法规授权的具有公共管理和公共服务职能的组织, 企业事业单位, 人民团体等; 行为要素中收集的数据包括以一定形式记录、存储和传输的文件、资料、文字、图像等各类可机器读取的数据, 法律另有规定或涉及国家利益、公共安全、商业秘密、个人隐私等的数据除外。

3.2 公共数据的特征

公共数据的概念源自于政府数据(本文中等同于“政务数据”), 但其范围高于政府数据。因此在拥有政府数据类型多样性、数据来源可靠性、数据内容公共性、数据分布“碎片化”等特点的基础上, 公共数据的数据主体更丰富, 数据开放范围更广泛, 且授权数据可用于盈利。

3.2.1 更丰富的数据主体

在梳理 24 个管理办法和条例中的公共数据主体及其公共数据流通方向后发现, 在现行关于公共数据的立法和政府规章中, 对公共管理和公共服务机构中的公共数据有序流通方向和层次大致采用“多级伞形”结构, 如图 1 所示。

从图 1 看出, 公共数据包含政府数据, 同时, 除政府数据之外还包括公共管理和公共服务的其他主体。

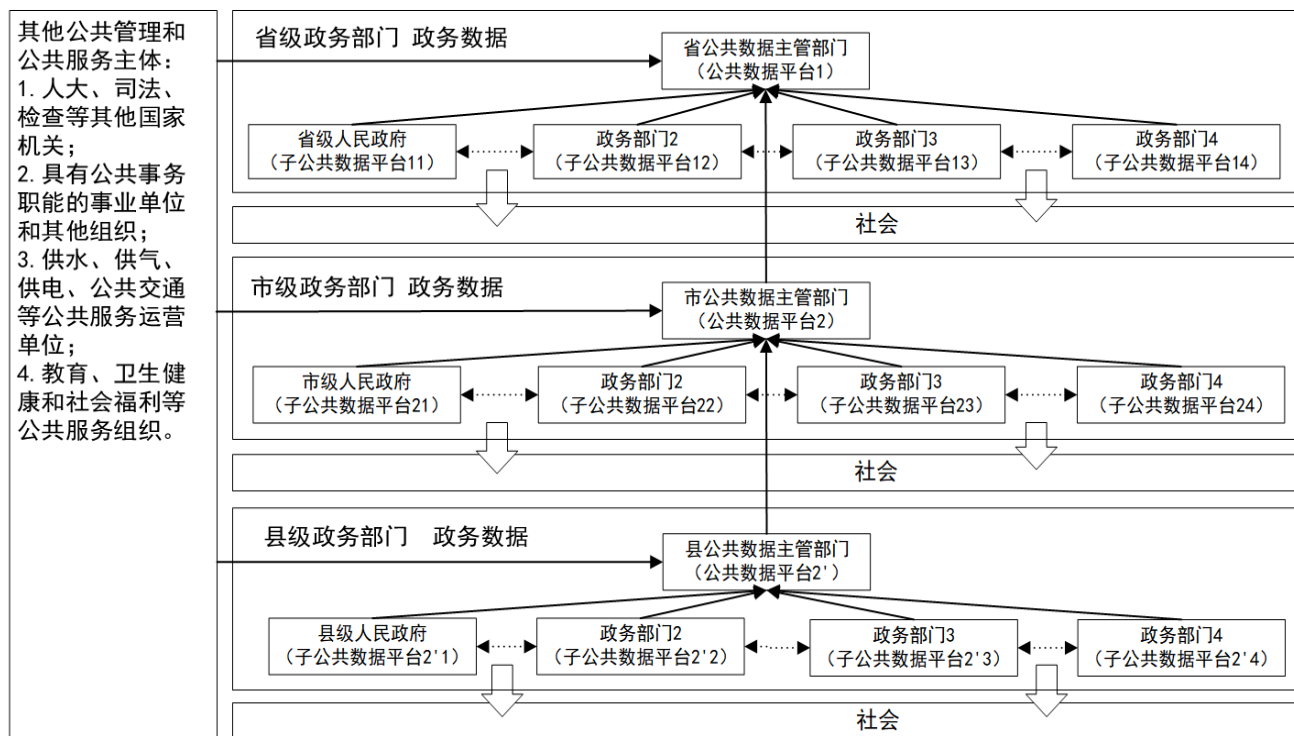


图1 公共数据各级有序流通方向和层次

Fig.1 Direction and level of orderly circulation of public data at different levels

其中，公共管理职能的主体包括：①人大、司法、监察、行政等国家机关；②具有公共管理事务职能的其他组织，如工会、妇联、共青团等组织。公共服务职能的主体主要包括：①供水、供电、供气、公共交通等公共服务运营单位；②教育、卫生健康和社会福利等公共服务组织。由此可见，公共数据的主体相较于政府数据主体来说更加丰富，毫无疑问扩大了公众获取公共数据资源的基数，进一步促进了公共数据的开放发展，保障公共利益。

3.2.2 更强劲的开放力度

公共数据通常比政府数据更具有开放性。据统计，仅贵州省、沈阳市等个别省市级政府出台政务数据开放政策，且执行力度较弱，甚至部分政策中未规定开放等级。同时，政府数据在最初原则上仅在行政机关内部共享，并不当然对外开放，随着数据价值的提升促进其开放；而公共数据概念的提出是由公共数据资源开放需求拉动，以数据开放为目的。在已有的公共数据管理政策中，目录化管理、开放等级清晰、数据开放安全责任等规定相对全面，同时规定以开放为原

则，不开放为例外。

3.2.3 公共授权数据可营利

公共数据具有巨大的经济价值。《浙江省公共数据条例》第三十五条规定：“授权运营单位应当依托公共数据平台对授权运营的公共数据进行加工，对加工形成的数据产品和服务，可以向用户提供并获取合理收益。^[25]”由此可见，授权运营单位以授权运营的公共数据形成的准公共产品具有交易性和营利性；政府数据仅能够向其他行政机关提供，因此数据本身通常是绝对不可交易的，不能用于其他目的。《数据安全法》第四十条规定，“国家机关委托他人建设、维护电子政务系统，存储、加工政务数据应当经过严格的批准程序，受托方不得擅自留存、使用、泄露、向他人提供。即使是其他行政机关获得政务数据，未经原机关允许亦不能与其他机构共享或对外开放。”^[26]

3.3 公共数据全生命周期

数据是人们用于记录事务的物理符号，本身没有很大的意义，只有经过收集、处理、分析等数据生命

周期阶段后, 才能充分激发出数据的价值^[7], 公共数据亦如此。本文在总结典型数据生命周期模型基础上, 结合公共数据特征探索公共数据全生命周期, 为公共数据伦理准则框架的指标构建奠定基础。

3.3.1 典型数据生命周期模型

在地方政府出台的公共数据管理办法中, 各条文已经呈现出公共数据全生命周期模型的雏形。但条文中更多规定的是以公共数据保存、共享和开放阶段的工作要求, 而缺乏对前期数据规划、处理、审计等的阶段规定, 因此本文梳理了部分典型的数据生命周期模型(表1), 并将对政策中未做出的数据阶段进行补充。

3.3.2 公共数据全生命周期模型

表1中可以看出, 典型的数据生命周期模型基本包括了数据规划、收集、处理、保存、发现与获取等阶段。而公共数据的流通具有循环性, 且因涉及的数

据主体范围大、开放力度更强、且具有盈利性等, 其安全问题备受关注。鉴于此, 本文增加包括数据产生与收集、数据汇交、数据分级分类、数据描述与评估在内的数据创建和包括数据审计和技术保障在内的数据监管阶段, 以强调保障公共数据安全的基础工作的重要性。故公共数据生命周期包括数据规划、数据创建、数据处理、数据保存、数据开放、数据监管6个阶段, 构成公共数据生命周期的循环层次模型, 具体如图2所示。其中, 每个阶段包括共17个更细粒度的子阶段, 以便形成伦理框架时提供更细致的指导。

(1) 数据规划。公共数据面向政府机构服务、公众使用, 其数据的生命周期活动需提前规划, 避免数据过度采集、数据滥用等公共数据伦理问题的发生, 因此设计数据规划阶段, 包括数据情景分析和规划管理设计。^①数据情景分析是指根据公共数据的适用范围和参与人员等构建数据场景, 规划不同场景下数据

表1 部分典型的数据生命周期模型

Table 1 Some typical data lifecycle models

名称	特征	阶段	说明
COSA-DLC 数据生命周期模型 ^[27]	环形结构	数据采集(收集、筛选、质量控制、描述)、数据处理(处理、质量管理、分析)、数据保存(分类、质量管理、归档、发布)	该模型致力于解决数据6Vs的完整性挑战, 是一种可应用到任何领域和学科的抽象模型, 即可应用于任何不可知场景, 又可以灵活地适应特定场景下的数据管理需求和挑战
DCC 数据监管生命周期模型 ^[28]	环形层次结构	顺序操作: 数据的产生和收集、评估和筛选、接入和融合、长期保存和存储、获取使用和再利用, 数据转换; 不固定操作: 销毁、重新评估、迁移	该模型为数据的成功保存和存储提供了一个模型, 它包括不同的层: 全生命周期操作, 顺序操作和不固定操作。其中, 顺序操作提供数据生命周期的基础管理模型架构, 不固定操作为模型保证数据的质量和安全管理提供保障
DataONE 数据生命周期模型 ^[29]	环形结构	数据的规划、收集、质量控制、描述、长期保存、发现与获取、集成与分析	该模型形成的路径既可以是只包含某些阶段的线性路径, 也可以是包含全周期阶段的多重循环。此模型是专门为数据保存和再利用而开发, 对数据安全没有任何关注
DDI 组合生命周期模型 ^[30]	非线性迭代结构	研究设计与数据规划、数据收集、数据处理、数据存档、数据分配、数据发现、数据分析、数据重用	该模型以社会科学研究数据为研究对象, 从数据应用的视角总结, 包含数据处理阶段的一个二项选择路径与数据分析之后的反馈回路。但对数据质量和安全管理没有关注
政府开放数据生命周期模型 ^[31]	环型结构	政府数据创建与采集、组织与处理、存储与发布、发现与获取、增值与评价	该模型面向政府数据的开放实践, 共包括5个相互关联、连续迭代的阶段, 但缺乏数据质量和安全管理

*注: COSA-DLC: Comprehensive Scenario Agnostic Data LifeCycle, 全面的与场景无关数据生命周期模型; DCC: Digital Curation Centre, 英国数据管理中心; DataONE: Data Observation Network for Earth, 地球数据观测网; DDI: Data Documentation Initiative, 社会科学研究信息的国际标准

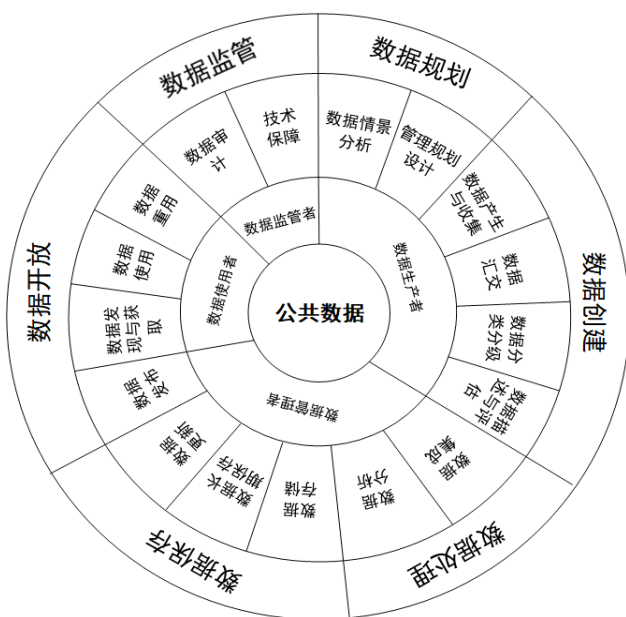


图2 公共数据全生命周期的循环层次模型

Fig.2 Cyclic hierarchical model for the whole life cycle of public data

的需求目录、使用规范、监管方案和人员素质要求等；
 ②管理规划设计是指在固定场景下，对公共数据在生命周期各阶段中参与任务的基础标准、管理标准、法律规定等关键问题的规范和设计。

(2) 数据创建。
 ①数据汇交：在基层政府和企业产生并收集公共数据之后，根据数据规划内容，在不同政府部门和机构之间进行公共数据共享，包含同级政府间的横向数据共享、上下级政府-政府和政府-企业/机构之间的纵向数据汇报，形成公共数据在数据管理者之间纵横交错的向上逐层汇交网络，以构建跨部门、跨行政区划、跨行业的协同治理体系。
 ②数据分类分级：公共数据数量大，种类繁多，数据敏感程度不一。以目标场景分析和原则为前提，将本部门公共数据进行先分类后分级的流程，以保证数据安全。
 ③数据描述与评估：根据公共数据的场景用途，采用适当的元数据描述标准对数据进行全面精准描述。在数据创建的最后阶段增设数据质量评估，对收集的基础公共数据执行质量标准控制，以有效降低数据处理和流通成本。

(3) 数据处理。数据处理是对收集到的数据进行简单清理和初步分析，使数据面向不同的场景标准化，

提高数据质量，为数据的场景应用和增值做准备。

①数据集成：把不同系统、结构、编码方案的数据集根据计算逻辑进行有效集中和转换，形成机器可读的同质性数据集，以支持数据迁移、数据挖掘、数据可视化等工作。
 ②数据分析：运用计量学、机器学习、数据建模等方法，从数据中挖掘出所需要的信息并尽可能实现可视化，提供给公共数据平台中的用户，使其获取信息。

(4) 数据保存。数据保存是对有价值的公共数据进行有效存储、并具有选择性地长期保存，进行数据更新，以保证数据安全和质量。
 ①数据存储：对不同类型的静态数据选择不同介质的保存方式，明确存储内容、位置、时间等，并通过存储介质维护与升级、访问权限设置、定期检查、数据备份等保证数据安全和质量。
 ②数据长期保存：公共数据的历史数据价值巨大，可用于历史验证和科学预测，更好地把握社会发展方向，提供更精准的公共服务。因此需要识别长期保存的公共数据，将其提交给专门的数据存档机构，减小存储成本。
 ③数据更新：公共数据处于持续产生与更新的状态，需要根据不同数据特征和功用通过定期社会调查、数据传感器等方法及时更新公共数据，保证数据时效性和可用性。

(5) 数据开放。经过公共数据前4个阶段的流通，公共数据具有数量大、质量高和标准相对统一的特点^[32]，具有非常大的潜在价值。为实现公共数据增殖和再利用，公共部门利用公共数据平台向用户开放可开放等级的、可机器读取的公共数据，“唤醒”公共数据价值。
 ①数据发布：根据相关政策和市场需求，将分级后可开放等级的公共数据分类在平台进行开放，并设置用户使用权限，保证数据的合法获取。
 ②数据发现与获取：用户根据自己的需求，在不同的平台中发现数据信息，并通过逐条下载、合并爬取等技术手段获取公共数据。其中数据获取平台包括各级公共数据开放网站、政务部门开放网站、新闻网站等。
 ③数据使用：用户对获得的数据通过数据挖掘、数据可视化等数据技术分析结果获得信息和知识，以支撑管理和服务决策，提供公共管理和公共服务。
 ④数据重用：

包括两种, 一是已保存的数据除了进行开放工作亦可用作公共部门内部数据库的构建升级和学术研究; 二是已授权或被获取的数据形成的公共产品可被其他用户有选择性地截取有用的数据和信息, 进行二次创作, 充分挖掘公共数据的价值。

(6) 数据监管。公共数据的流通和开放需要实现对公共数据全生命周期的监管。①数据审计: 需要通过各阶段数据质量评估模型实现, 包括开放前基础数据质量控制和开放后公共数据产品的质量监管。②技术保障: 可通过运用安全多方计算和区块链等技术开发现代化的监管沙箱等实现穿透式、全方位的监管, 确保数据流通的可溯源、可审计, 推动数据流转的规范化、制度化。

本文的公共数据全生命周期模型是一个较为全面的模型, 可根据不同场景灵活选择阶段运用, 包括了公共数据从规划使用到创建、收集、处理、保存等的基础数据管理阶段, 公共数据开放、形成公共产品、最后流通回数据产生的数据应用过程, 形成了循环结构。同时数据审计贯穿其中, 以保证数据质量和数据安全。在此大框架之下对每个阶段进行细分, 使每个阶段更加具体化和精准化。根据《浙江省公共数据管理条例》中“谁主管谁负责、谁提供谁负责”的公共数据提供原则和“谁使用谁负责、谁管理谁负责”的利用原则, 本文在公共数据全生命周期模型中嵌入公共数据生产者、管理者、使用者和监督者, 以明确其在公共数据循环的各个阶段中的责任主体, 是公共数据全生命周期管理的关键。

4 公共数据伦理准则框架

公共数据伦理是指在数据全生命周期中, 能够以符合道德准则的方式去规划、创建、处理、保存、开放和监管数据, 以保障公共数据的正常存在和运行, 实现公共利益最大化, 进而提高社会治理能力和公共服务水平。公共数据特征决定了公共数据伦理框架内容的准确性。因此, 本文以公共数据特征为基础的公共数据全生命周期模型为一、二级指标, 剖析公共数

据在各阶段可能会发生的伦理风险问题, 构建公共数据伦理框架, 以规制公共数据生命周期各阶段的数据参与者对公共数据的道德化操作和决策。

4.1 公共数据伦理参考准则

公共数据伦理准则, 是指在全生命周期的公共数据管理中, 公共数据参与主体在公共数据活动中进行伦理判断或伦理评价时应遵循的原则, 是公共数据伦理框架形成的综合性参考准则和保障性基础, 起着提纲挈领的作用。数据伦理是数字技术伦理的底层逻辑基础^[33], 而公共数据伦理属于数据伦理的一部分, 亦可称之为其底层逻辑基础。包括算法伦理^[34]、区块链伦理^[35]、人工智能伦理^[36]等已较为成熟的数字技术伦理^[37], 其准则主要侧重于宏观方面分析技术发展不足导致应用时产生的社会伦理问题, 因此以倡导算法精准和技术应用合乎道德为主要准则; 而包括教育数据伦理^[7]、交通数据伦理^[15]、医学数据伦理^[11]在内的公共数据伦理作为技术的底层逻辑, 也是各公共领域数据伦理的参考标准, 其准则更侧重于微观分析各阶段公共数据参与者进行数据决策时产生的误差导致技术问题, 因此以倡导数据参与者, 也就是人, 的数据决策合乎道德为主要准则 (图3)。

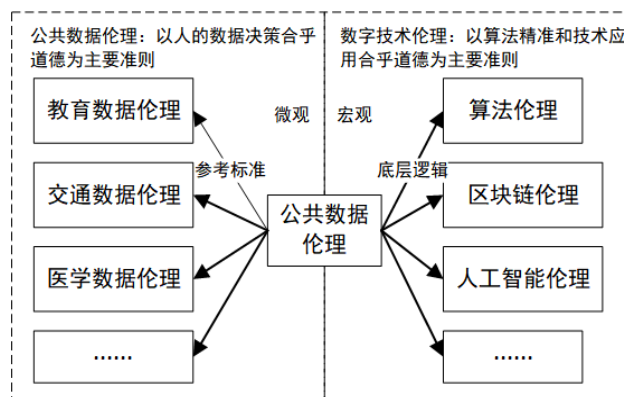


图3 公共数据伦理准则与数字技术伦理准则之间的区别

Fig.3 Difference between public data ethics guidelines and digital technology ethics guidelines

结合公共数据本身和应用时的特点, 并以公共数据参与者的数据决策合乎道德为前提, 总结出以下由深入浅、由理论到实践的公共数据伦理准则。

4.1.1 尊重：以人为本

数据伦理要遵循人本主义数据伦理观，公共数据亦如此，公共数据的本意是保障服务人民的权益而非奴役人民。人本主义伦理观主张数据自由和数据权利的主体是人而非数据，“以人为本”提倡以人的权利为本。人的权利具体包含了两部分，一是基本权利。二是数据权利。数据权利是人的基本权利在数据领域的具体体现，具体包括：隐私权、知情同意权、删除权、数据携带权等^[38]；与之相反的是数据主义伦理观，数据主义尊崇数据至上，一切都要数据化而无视人的自由意志，使人成为数据巨机器上无差别的螺丝钉^[39]。公共数据伦理在全生命周期中呼吁从数据主义回归到人本主义，维护人的尊严，尊重人的权利。

4.1.2 安全：保护隐私

公共数据的全生命周期流通必须尊重人的权利，其中包括尊重用户的个人隐私和个人信息。如数据的收集目的、范围等必须提前告知用户并征询用户的同意，同时秉持最少原则或必要原则，避免数据的重复采集^[38]。除此之外，公共数据伦理支持数据共享，但在公共数据的共享中应进行主动隐私保护设计，比如数据清洗、脱敏、脱密、格式转换和个人隐私泄露后的保护措施和问责机制等。反对任何窃取、篡改、泄露和其他非法收集利用个人数据的行为。

4.1.3 平等：公平透明

公共数据的获取、共享和开放等流通需要遵循公平透明原则，公平分配社会和社会内部的利益和平等机会，以公开透明化的规定保障社会公众平等拥有数据确权。如通过以完整、开放、可理解、易于访问和免费的方式发布各个项目的数据^[23]，以确保公共信息的透明度和公众平等享有对公共数据资源的使用权；明确并公开公共数据流通过程中各阶段的数据责任主体，以确保公众平等地对数据流通过程具有知情权和数据修改权；并在公共数据利用方面促进培训，让人们平等地获得公共数据管理和开放带来的收益权。尤其是对弱势地区和社会团体，避免将弱势人群置于更不利的地位。

4.2 构建公共数据伦理框架

公共数据伦理框架是从微观视角规范公共数据全生命周期各阶段数据处理操作细节，为数据处理工作画下一条红线，避免引发数据过度收集、数据获取鸿沟等社会伦理问题。同时，为了使公共数据开放在公民隐私权利保护、公共数据开放和公共数据应用之间达成适当平衡，本文借鉴公共数据全生命周期各阶段研究、公共数据伦理准则和各省公共数据条例，融合公共数据特征和公共数据全生命周期模型，构建了公共数据伦理框架，如表2所示。

公共数据伦理框架包括一级指标、二级指标、伦理风险、伦理细则和来源5部分。其中，一级指标采用公共数据全生命周期模型中的6大阶段，是公共数据全生命周期流通的主要方向和过程，提纲挈领；二级指标采用公共数据全生命周期模型中的17个子阶段，是对公共数据流通阶段的细化，支分节解，有助于厘定细化后各阶段的伦理风险；伦理风险旨在揭示公共数据二级指标的具体情境假设下可能会发生的伦理风险问题，缕析条分，为列举公共数据伦理细则提供预防目标；伦理细则是为预防公共数据流通各阶段发生的伦理风险对公共数据参与主体和责任主体提出的数据处理操作细节，剖烦析滞，为公共数据责任主体和参与主体规约公共数据伦理问题提供参考和契机；来源部分包括国内外关于公共数据的政策条例、数据伦理的成熟框架和公共数据全生命周期某阶段的伦理研究，是对部分伦理风险和伦理细则的出处进行标注，以增加公共数据伦理框架的可解释性和说服力。

5 结 语

公共数据流通的全生命周期管理是促进数据经济发展、释放数据红利的保障，但公共数据的蓬勃发展也有其科林格里奇困境，因此我们要预防潜在风险。中国地方政府对于公共数据的规范和事后法律责任的探索已如火如荼，但学者们对公共数据事前伦理责任的探究尚处于起步阶段。科技向善，伦理先行，公共

表2 公共数据伦理框架

Table 2 Framework of public data ethics

一级指标	二级指标	伦理风险	伦理细则	来源
数据规划	数据情景分析	数据歧视; 情景分析偏差导致数据过度收集	1.梳理公共数据流通的各类场景,明确各类场景下公共数据的适用边界、公共利益和数据参与者的需求; 2.保证数据的完整性,考虑完整性与成本的平衡	陆莉 ^[40] ;《浙江省公共数据条例》 ^[25]
	管理规划设计	数据过度收集;漠视主体权益;隐私泄露	1.对数据全生命周期中的各参与者进行责任界定; 2.根据各应用场景编制公共数据目录及其编制标准; 3.平台规划和测试时对测试数据进行数据水印等脱敏脱密操作	刘静羽 ^[41] ;《浙江省公共数据条例》 ^[25]
数据创建	数据产生与收集	过度收集造成侵犯公民隐私;知情权;选择权;数据来源伪造;数据篡改;数据滥用	1.数据产生时必须符合实施情况,不得伪造数据; 2.数据收集时被收集人必须知情,且禁止提供虚假数据; 3.数据收集遵循最小必要原则,对采集清单和采集人员进行透明化; 4.提供数据相关的情境信息,包括数据质量标准、收集方法、过程描述、知识产权等; 5.确定收集数据的来源、载体、格式等收集数据模板并确定其符合数据质量标准,增加数据可追溯性; 6.可以通过共享获取数据的,不得重复收集,不得强制要求个人采用多种方式重复验证或者特定方式验证	《浙江省公共数据条例》 ^[25] ;郑小军 ^[42] ;张誉元 ^[7] ;嘎拉森 ^[43]
	数据汇交	数据传输中断造成数据损毁、篡改、滥用	1.明确数据汇交的边界和条件,减少不加区分的数据归集; 2.预判数据集多源融合可能造成的风险,根据场景设计可将数据集定为更高级别开放或要进行二次脱敏	郑晓军 ^[42] ;《浙江省公共数据条例》 ^[25] ;周毅 ^[44]
	数据分类分级	分类标准不明显导致数据获取鸿沟	1.明确公共数据分类分级依据、目的、标准和安全应对问题; 2.预防因级别变更和动态调整的安全问题; 3.数据分类分级遵循“重大怀疑从无”原则	《江苏省公共数据管理办法》 ^[45] ;周毅 ^[44] ;陈祥玲 ^[46]
	数据描述与评估	数据失真;数据篡改	1.应当强调建立规范、清晰的描述文档,便于后期进行解码; 2.禁止数据描述过程的数据泄露; 3.评估中确定科学合理的共享开放属性(无条件、受限、禁止); 4.对可能产生涉密、敏感数据进行安全评估	《江苏省公共数据管理办法》 ^[45]
数据处理	数据集成	访问权限不受控导致数据非法获取;数据误用	1.强调数据变更信息的记录,避免数据误用; 2.禁止数据汇交后“意料之外的二次使用”	郑晓军 ^[42]
	数据分析	数据异化;第三方服务机构数据泄露;数字懒政	1.决策者应从数据分析和实际调查结合分析,而非只凭数据分析结果决策; 2.委托第三方协助分析时,不得擅自留存、使用、泄露或者向他人提供相关公共数据; 3.委托第三方协助分析结束后,及时监督第三方服务机构以不可逆方式删除相关数据	郑晓军 ^[42] ;《浙江省公共数据管理条例》 ^[25]
数据保存	数据存储	数据泄露、篡改;集中存储极易成为黑客攻击目标	1.降低对公共数据存储依托的云平台安全隐患,做好系统等级保护定级备案、测评整改; 2.对存储中的数据进行检索、更改等操作进行留痕,增强数据可追溯性; 3.根据加密等级禁止非授权访问; 4.及时进行到期鉴定,对过期数据进行完全销毁	嘎拉森 ^[43]
	数据长期保存	数据丢失;数据不可用	1.明确各类数据的数据存储期限; 2.对存储中的数据进行检索、更改等操作进行留痕,增强数据可追溯性	《浙江省公共数据条例》 ^[25]
	数据更新	数据更新不及时导致信息茧房、形象固化等	1.根据需求明确更新频率,及时更新已变更、失效数据; 2.根据时限、需求、技术等条件,及时更新公共数据的共享开放类型	《江苏省公共数据管理办法》 ^[45]

表2 (续)

Table 2 Continue

一级指标	二级指标	伦理风险	伦理细则	来源
数据共享与开放	数据发布	漠视主体权利；隐私泄露；信息不对称	<ol style="list-style-type: none"> 1.对敏感数据和隐私数据进行数据水印等脱敏脱密操作； 2.考虑用户如何快速简便地查阅并获得需要的公共数据； 3.优先开放与民生紧密相关、社会迫切需要、行业增值潜力显著和产业战略意义重大的公共数据，并听取组织、企业、专家和社会公众的意见； 4.通过门户网站、新闻媒体等途径向社会宣传推广公共数据开放相关信息，便于公民、法人和其他组织知晓 	英国《数据伦理框架》 ^[21] ；《浙江省公共数据条例》 ^[25] ；《江苏省公共数据管理办法》 ^[45]
	数据发现与获取	数据获取鸿沟；通过不合法渠道获取	<ol style="list-style-type: none"> 1.明确数据获取的目的； 2.数据获取者（包括共享和开放过程）在获取数据时必须明晰所承担的数据安全责任； 3.通过共享获得的数据，不得以任何形式提供给第三方，不得用于其他任何目的 	《浙江省公共数据条例》 ^[25]
	数据使用	数据非授权访问；数据窃取；数据泄露；数据异化	<ol style="list-style-type: none"> 1.政府正向引导、鼓励公民依法开发利用公共数据资源； 2.数据使用者依法获取并加工形成的公共产品和服务不得危害国家安全、公共利益和其他人的合法权益； 3.使用者形成公共数据产品和服务时，应当在成果中注明数据来源和机构； 4.运用大数据算法等技术挖掘时，结合算法结果和实际情况进行决策 	《江苏省公共数据管理办法》 ^[45]
	数据重用	数据衍生化导致权属分配争议；数据滥用；数据异化；数据过度挖掘	<ol style="list-style-type: none"> 1.避免因数据重用而对个人或群体产生不利影响； 2.数据重用者在形成成果时注明数据来源和机构，确保拥有使用该数据的权限； 3.运用大数据算法等技术挖掘时，结合算法结果和实际情况进行决策，防止数据异化； 4.界定个人数据和公共数据的权属分配问题 	英国《数据伦理框架》 ^[21] ；《江苏省公共数据管理办法》 ^[45] ；嘎拉森 ^[43]
数据监管	数据审计	数据审计信息壁垒；数据审计失信	<ol style="list-style-type: none"> 1.建立公共数据流通各环节中的审计规程，实时对数据（包括第三方平台）安全性进行评估，增强数据溯源能力； 2.保留审计记录，定期对警告事件进行查看和处理； 3.反复审视用户的需求和公共利益； 4.设置问题数据紧急处理员，明确问题数据的责任追究，对滥用权力或失信行为依法进行失信约束 	郑晓军 ^[42] ；《浙江省公共数据条例》 ^[25] ；陈凤霞 ^[47]
	技术保障	数据泄露；过度标签化；过度挖掘；数据依赖	<ol style="list-style-type: none"> 1.及时跟踪数据科学领域发展进展，特别要关注数据收集、存储、脱敏加密等的新方法、新技术，降低风险； 2.为数据安全事件应急预案提供技术支持，并定期预演，确保其可追查和追溯； 3.建立健全公共数据安全防护技术标准和规范，提高数据安全保障能力 	英国《数据伦理框架》 ^[21] ；《浙江省公共数据条例》 ^[25]

数据亦如此。为公共数据生产者、管理者、使用者和监督者等参与人员在公共数据流通前就树立数据安全和伦理意识，规约自己的行为，从而避免隐私泄露、数据过度收集和数字鸿沟等公共数据伦理问题的发生，本文结合国内公共数据管理政策对公共数据进行了全

面性的定义，总结了公共数据的特征，在此基础上探究了公共数据全生命周期模型。最后，以公共数据全生命周期模型的6个总阶段和17个子阶段为一、二级指标，结合公共数据伦理准则，分析每一阶段会发生的公共数据伦理风险，再针对这些风险对参与人员的

行为列举伦理细则, 形成公共数据伦理框架。此框架有助于公共数据参与人员规约自己的行为, 落实事前伦理责任, 并且在公共数据伦理风险发生之前进行预判、规避和化解。此框架的不足之处在于其应用性和实践性尚需检验, 这也是我们下一步研究工作的方向, 从而使框架更加精确实用, 投入使用。但仍可为涉及到公共管理和服务的各界学者提供公共数据伦理问题的参考和契机, 助力激发公共数据要素价值, 推进公共管理和服务数字化进程, 帮助实现社会公平、争议和可持续发展。

参考文献:

- [1] 刘金亚, 谌悦, 章成志, 等. 数据伦理政策的量化研究: 基于天文领域的管理方法[J]. 信息资源管理学报, 2021, 11(6): 63-75.
LIU J Y, CHEN Y, ZHANG C Z, et al. Quantitative research on data ethics policy: A data management approach based on astronomy[J]. Journal of information resources management, 2021, 11(6): 63-75.
- [2] CHEAH P Y. Institutional policy to boost data sharing[J]. Nature, 2019, 565(7739): 294-294.
- [3] 李思艺. 迈入技术与信任相融合的数字治理时代: 加拿大数字政府建设的启示[J]. 情报理论与实践, 2022, 45(1): 205-212.
LI S Y. Towards an age of digital governance integrated with technology and trust: Implications of digital government programs in Canada[J]. Information studies: Theory & application, 2022, 45(1): 205-212.
- [4] 凡景强, 邢思聪. 大数据伦理研究现状分析及未来展望[J/OL]. 情报杂志: 1-7 [2023-02-23]. <http://hfffg5fce84748f1d4cc2hunfcp-pvvbokq69qq.fgy.hlju.cwkeji.cn/kcms/detail/61.1167.G3.20230106.1529.022.html>.
FAN J Q, XING S C. Big data ethics research status analysis and future prospects [J/OL]. Journal of intelligence: 1-7 [2023-02-23]. <http://hfffg5fce84748f1d4cc2hunfcp-pvvbokq69qq.fgy.hlju.cwkeji.cn/kcms/detail/61.1167.G3.20230106.1529.022.html>.
- [5] 龙坤, 徐能武. 对美军推出人工智能伦理准则的剖析[J]. 情报杂志, 2022, 41(3): 1-8.
LONG K, XU N W. An analysis of the U.S. DoD's AI ethical principles[J]. Journal of intelligence, 2022, 41(3): 1-8.
- [6] MONTAGUE E, DAY T E, BARRY D, et al. The case for information fiduciaries: The implementation of a data ethics checklist at Seattle Children's Hospital[J]. Journal of the American medical informatics association: JAMIA, 2021, 28(3): 650-652.
- [7] 张誉元, 张海. 数据生命周期视阈下教育数据伦理问题及规约之径[J]. 中国电化教育, 2022(10): 118-125.
ZHANG Y Y, ZHANG H. Ethical issues and regulations of educational data from the perspective of data life cycle[J]. China educational technology, 2022(10): 118-125.
- [8] 韩磊, 胡广伟. 政府数据开放平台建设效率评估及其启示[J]. 数字图书馆论坛, 2018(9): 52-59.
HAN L, HU G W. Efficiency evaluation on open government data portals based on DEA[J]. Digital library forum, 2018(9): 52-59.
- [9] 刘佳静, 郑建明. 公共数据开放利用体系框架研究[J]. 现代情报, 2022, 42(10): 90-98.
LIU J J, ZHENG J M. Research on the framework of public data open utilization system[J]. Journal of modern information, 2022, 42(10): 90-98.
- [10] 童楠楠, 杨铭鑫, 莫心瑶, 等. 数据财政: 新时期推动公共数据授权运营利益分配的模式框架[J]. 电子政务, 2023(1): 23-35.
TONG N N, YANG M X, MO X Y, et al. Data finance: A model framework for promoting benefit distribution of public data authorization operation in the new era[J]. E-Government, 2023(1): 23-35.
- [11] 关键. 医学科学数据共享与使用的伦理要求和管理规范(一)前言[J]. 中国医学伦理学, 2020, 33(2): 143-146.
GUAN J. Ethical requirements and management standards for the sharing and re-use of scientific data in health care and medicine (I) preface[J]. Chinese medical ethics, 2020, 33(2): 143-146.
- [12] 刘金亚, 顾立平, 张潇月, 等. 开放科研数据环境下科研人员的伦理框架研究[J]. 情报理论与实践, 2021, 44(2): 83-89.
LIU J Y, GU L P, ZHANG X Y, et al. Framework of data ethics for researchers in open research data environment[J]. Information studies: Theory & application, 2021, 44(2): 83-89.
- [13] 刘霞, 徐博艺. 信息伦理对 G2C 电子政务系统用户接受行为的影响研究[J]. 情报杂志, 2010, 29(1): 22-26.
LIU X, XU B Y. Research on the influence of information ethics on user's acceptance behavior of G2C e-government system[J]. Journal of intelligence, 2010, 29(1): 22-26.

- [14] 王亚强. 论电子政务信息系统的负效应与伦理诉求[J]. 价值工程, 2010, 29(35): 136-137.
- WANG Y Q. Discussion on the negative effect of the e-government information system and ethical demands[J]. Value engineering, 2010, 29(35): 136-137.
- [15] 吴爱荣. 政府交通出行数据利用中的数据伦理问题研究[D]. 保定: 河北大学, 2021.
- WU A R. Research on data ethics in government traffic trip data utilization[D]. Baoding: Hebei University, 2021.
- [16] WANG Y. Information ethics in reuse of public sector information in China[J]. IEEE technology and society magazine, 2013, 32(3): 73-79.
- [17] VARLEY-WINTER O, SHAH H. The opportunities and ethics of big data: Practical priorities for a national council of data ethics[J]. Philosophical transactions series A, mathematical, physical, and engineering sciences, 2016, 374(2083): 20160116.
- [18] DREW C. Data science ethics in government [J]. Philosophical transactions Series A, mathematical, physical, and engineering sciences, 2016, 374(2083): 20160119.
- [19] JOSEPH R C. Data breaches: Public sector perspectives[J]. IT professional, 2018, 20(4): 57-64.
- [20] DREW C. Design for data ethics: Using service design approaches to operationalize ethical principles on four projects[J]. Philosophical transactions Series A, mathematical, physical, and engineering sciences, 2018, 376(2128): 20170353.
- [21] Central digital and data office. Data ethics framework[EB/OL]. [2020-09-16]. <https://www.gov.uk/government/publications/data-ethics-framework>.
- [22] ROSEN D L, BUCHBINDER M, JUENGST E, et al. Public health research, practice, and ethics for justice-involved persons in the big data era[J]. American journal of public health, 2020, 110(S1): S37-S38.
- [23] MOLLDREM S, SMITH A K J, MCCLELLAND A. Predictive analytics in HIV surveillance require new approaches to data ethics, rights, and regulation in public health[J]. Critical public health, 2023, 33(3): 275-281.
- [24] 沈斌, 黎江虹. 论公共数据的类型化规制及其立法落实[J]. 武汉大学学报(哲学社会科学版), 2023, 76(1): 67-77.
- SHEN B, LI J H. Typed regulation of public data and its legislative implementation[J]. Wuhan university journal (philosophy & social science), 2023, 76(1): 67-77.
- [25] 浙江省公共数据条例[N]. 浙江日报, 2022-02-09(008). DOI:10.38328/n.cnki.nzjrb.2022.000468.
- [26] 中华人民共和国数据安全法[N]. 人民日报, 2021-06-19(007). DOI: 10.28655/n.cnki.nrmrb.2021.006406.
- [27] SINAEEPOURFARD A, GARCIA J, MASIP-BRUIN X, et al. A comprehensive scenario agnostic data life cycle model for an efficient data complexity management[C]// 2016 IEEE 12th International Conference on e-Science (e-Science). Piscataway, New Jersey: IEEE, 2017: 276-281.
- [28] RHEE H L. A new lifecycle model enabling optimal digital curation[J]. Journal of librarianship and information science, 2022. <https://doi.org/10.1177/09610006221125956>.
- [29] DataONE[EB/OL]. [2012-10-28]. <http://www.dataone.org/>.
- [30] BALL A. Review of data management lifecycle models[EB/OL]. [2019-05-25]. <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf>.
- [31] 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[J]. 情报理论与实践, 2018, 41(2): 7-13.
- HUANG R H, LAI T. Barriers of open government data in China from the perspective of data life cycle[J]. Information studies: Theory & application, 2018, 41(2): 7-13.
- [32] 马颜昕. 公共数据授权运营的类型构建与制度展开[J]. 中外法学, 2023, 35(2): 328-345.
- MA Y X. Classification and system development of authorized operation of public data[J]. Peking university law journal, 2023, 35(2): 328-345.
- [33] FLORIDI L, TADDEO M. What is data ethics? [J]. Philosophical transactions Series A, mathematical, physical, and engineering sciences, 2016, 374(2083): 20160360.
- [34] 陈昌凤, 吕宇翔. 算法伦理研究: 视角、框架和原则[J]. 内蒙古社会科学, 2022, 43(3): 163-170, 213.
- CHEN C F, LV Y X. Research of the algorithmic ethics: Perspectives, frameworks and principles[J]. Inner Mongolia social sciences, 2022, 43(3): 163-170, 213.

- [35] 范毅强. 区块链技术的风险与伦理规约的路径研究[J]. 自然辩证法研究, 2019, 35(2): 53-57.
FAN Y Q. Research on the risk and the path of ethical stipulation on blockchain technology[J]. Studies in dialectics of nature, 2019, 35(2): 53-57.
- [36] 李升. 论人工智能伦理准则的细化与完善——基于 50 部人工智能伦理准则的梳理与解读[D]. 杭州: 浙江大学, 2020.
LI S. On the refinement and perfection of the ethical codes of artificial intelligence[D]. Hangzhou: Zhejiang University, 2020.
- [37] 肖红军, 阳镇. 数字科技伦理监管: 美国进展与中国借鉴[J]. 财经问题研究, 2023(6): 73-86.
XIAO H J, YANG Z. Ethical regulation of digital technology: Progress in the United States and reference from China[J]. Research on financial and economic issues, 2023(6): 73-86.
- [38] 李伦, 黄关. 数据主义与人本主义数据伦理[J]. 伦理学研究, 2019(2): 102-107.
LI L, HUANG G. Data doctrine and humanistic data ethics[J]. Studies in ethics, 2019(2): 102-107.
- [39] 李伦. “楚门效应”: 数据巨机器的“意识形态”——数据主义与基于权利的数据伦理[J]. 探索与争鸣, 2018(5): 29-31.
LI L. "Truman effect": "Ideology" of data giant machine - Data doctrine and data ethics based on rights[J]. Exploration and free views, 2018(5): 29-31.
- [40] 陆莉, 沙勇忠, 徐雪峰. 基于生命周期的公共安全数据管理模型研究[J]. 图书与情报, 2019(4): 13-21.
LU L, SHA Y Z, XU X F. Research about public safety data management model based on the life cycle[J]. Library & information, 2019(4): 13-21.
- [41] 刘静羽, 贾毓洁, 黄金霞, 等. 开放科学创新生态构建中的数据处理伦理准则框架研究[J]. 农业图书情报学报, 2022, 34(9): 29-43.
LIU J Y, JIA Y J, HUANG J X, et al. Ethics principle framework of data handling for open scientific innovation ecology[J]. Journal of library and information science in agriculture, 2022, 34(9): 29-43.
- [42] 郑晓军. 反思公共数据归集[J]. 华东政法大学学报, 2023, 26(2): 53-67.
ZHENG X J. Rethinking public data aggregation[J]. ECUPL journal, 2023, 26(2): 53-67.
- [43] 嘎拉森, 徐拥军. 公共数据开放视域下的个人信息保护风险及其治理逻辑[J]. 图书与情报, 2022(6): 83-90.
GA L S, XU Y J. The risk and governance logic of personal information protection in the view of public data open[J]. Library & information, 2022(6): 83-90.
- [44] 周毅, 徐梦. 公共数据分类分级标准建设探析[J]. 情报探索, 2023(3): 24-31.
ZHOU Y, XU M. Exploring the construction of public data classification and grading standards[J]. Information research, 2023(3): 24-31.
- [45] 江苏省公共数据管理办法[J]. 江苏省人民政府公报, 2021(20): 5-18.
The administrative measures for public data in Jiangsu Province[J]. Gazette of the people's government of Jiangsu province, 2021(20): 5-18.
- [46] 陈祥玲. 政府数据分类分级保护的逻辑、现实困境与实践路径[J]. 征信, 2023, 41(4): 36-44.
CHEN X L. Theoretical logic, realistic dilemma and practical path for government data classification and grading protection[J]. Credit reference, 2023, 41(4): 36-44.
- [47] 陈凤霞, 姜宾. 大数据技术赋能政府审计全覆盖: 动因、困境与路径[J]. 财会月刊, 2023, 44(7): 108-112.
CHEN F X, JIANG B. Big data technology empowers full coverage of government audit: Motivation, dilemma and approach[J]. Finance and accounting monthly, 2023, 44(7): 108-112.

A Framework of Ethics Guidelines on Public Data from a Whole Life Cycle Perspective

MA Haiqun^{1,2}, LI Jinling¹, YU Tongtong¹, ZHANG Tao^{1*}

(1. School of Information Management, Heilongjiang University, Harbin 150080; 2. Research Center of Information Resources Management, Heilongjiang University, Harbin 150080)

Abstract: [Purpose/Significance] As a basic national strategic resource, public data plays an important role in promoting high-quality economic development, enhancing social governance, and improving government service levels. The construction of an ethics framework helps participants in public data circulation such as public data producers, managers, users, and supervisors to establish an awareness of data security and ethics and regulate their behavior before public data circulation, avoiding public data ethics issues such as privacy breaches, excessive data collection, data abuse, and digital gap. [Method/Process] Based on previous studies and existing public data policies, this study summarizes the definition of public data according to legal logic and the principle of maximization and describes the differences between public data and government data in terms of data subject, data openness, and data profitability as its characteristics. Then, combined with the typical data life cycle model, the cyclic hierarchy model of the public data life cycle was constructed. In addition, by comparing the principles of public data ethics and the principles of digital technology ethics, the focus of the principles of public data ethics was analyzed, and the principles of public data ethics were summarized from the perspectives of respect, security, and equality. Finally, taking the principles of public data ethics as the criterion and the six general stages and 17 sub-stages of the public data lifecycle model as the first and second-level indicators, the ethical risks of public data at each stage were analyzed. By referring to the research at all stages of the public data lifecycle, academic achievements of domestic and international data ethics principles, provincial public data regulations, and ethical rules were listed for participants to manage their behaviors given these risks at each stage and to build a framework of public data ethics. [Results/Conclusions] This framework helps public data participants to regulate their actions, and implement prior ethical responsibilities to anticipate, avoid, and mitigate public data ethical risks before they occur throughout the life cycle. The shortcoming of this framework is that its applicability and practicability need to be tested, which is also the direction of our next research work, so as to make the framework more accurate and practical and put it into use. Additionally, it can provide references and opportunities for scholars involved in public management and services on public data ethics to stimulate the value of public data elements, promote the digitization process of public management and services, and help achieve social equity, solve disputes, and gain sustainable development.

Keywords: public data; data ethics; whole life cycle model; guideline framework