开放政府数据生态系统中的数据利用: 瓶颈与组态*

汤志伟^{1,2} 罗意³

(1. 电子科技大学公共管理学院,成都 611731; 2. 电子科技大学(深圳)高等研究院,深圳 518110; 3. 电子科技大学经济与管理学院,成都 611731)

摘要: 当前中国开放政府数据利用水平普遍较低,揭示高开放政府数据利用水平生成的必要条件和组态条件对释放数据价值、促进数字经济发展有重要意义。基于创新生态系统理论构建开放政府数据生态系统分析框架,并运用NCA和QCA混合方法,基于中国27个省域的数据检验高开放政府数据利用水平生成的必要条件和组态条件。研究发现,开放者、开放平台和技术环境是生成高开放政府数据利用水平的必要条件,其分别在74.70%、77.00%和45.00%的水平上构成瓶颈。高开放政府数据利用水平有供给主导型、供需联动型和综合驱动型3种生成类型。要提高开放政府数据利用水平,各地政府既要补好"漏水桶",又要打好"组合拳",并注重发挥需求侧的拉动作用。

关键词: 开放政府数据; 开放政府数据利用; 必要条件分析; 定性比较分析; 生态系统理论中图分类号: D63 DOI: 10.3772/j.issn.1673-2286.2024.06.002

引文格式: 汤志伟, 罗意. 开放政府数据生态系统中的数据利用: 瓶颈与组态[J]. 数字图书馆论坛, 2024, 20(6): 12-22.

数字经济时代,数据已经被正式纳入生产要素范畴,在赋能经济高质量发展、促进国家治理模式转变和推动社会生活方式变迁等方面发挥着日益重要的作用。在中国,各级政府部门掌握了80%以上的信息数据资源"1,这些数据资源蕴含着巨大的政治、经济与社会价值。《"十四五"数字经济发展规划》指出,要提升公共数据开放水平,鼓励社会力量开发利用公共数据资源,释放数据红利^[2]。在数字经济发展和数字中国建设背景下,提高开放政府数据(Open Government Data,OGD)利用水平,释放数据价值,是政府面临的关键任务之一。中国从2012年开始践行政府数据开放,截至2023年8月已经建成226个OGD平台,在开放数量、开放质量、开放领域等方面都有了长足进步。然而,2023

年《中国地方公共数据开放利用报告(省域)》显示,中国各省域OGD利用水平普遍较低,利用水平亟待提升^[3]。这表明,尽管OGD在中国蓬勃发展且潜力巨大,但OGD利用水平却不容乐观,数据无法充分释放蕴含的巨大价值,也无法满足数字经济、数字治理、社会创新等领域的需求。

OGD涉及多边利益,其利用过程也往往涉及众多系统要素,数据的供需不匹配、公众参与不足、社会合作不充分、互动反馈机制不健全、法规政策不完善等问题都可能导致OGD利用水平低下。那么,到底哪些因素在影响OGD利用?这些因素之间应如何组合以生成高OGD利用水平?哪些因素会构成OGD利用的"瓶颈"?这些问题随之产生并亟待回答。OGD生态系统能较好

收稿日期: 2023-12-19

*本研究得到国家社会科学基金重点项目"人工智能商业化应用的社会影响与治理体系研究"(编号:22AZD135)、国家自然科学基金专项项目"公共数据授权运营机制与方案设计"(编号:72241423)资助。

刻画数据利用过程面临的复杂环境和互动关系,成为OGD及OGD利用研究的新兴视角。该视角下,OGD利用过程发生在一个复杂的OGD生态系统中,不仅需要政府发布和管理高质量数据,也需要充分调动其他生态系统主体的参与互动,并与外部环境相适应。因此,本研究从OGD生态系统视角入手,探究高OGD利用水平生成的必要条件和组态条件,尝试回答上述问题,以期为促进数据利用和领域内研究提供新的思路。

1 文献综述

当前,学术界对OGD利用的概念本身尚无过多专门性探讨,也暂未形成相对公认的概念。从既有研究可以发现,在过程上,OGD利用涉及数据在数据管理机构、公众、数据开发团体等多元主体之间的流动;在结果上,OGD利用能够创造数据产品和数据服务,并最终释放数据蕴含的价值^[4-5]。结合上述内容,本研究将OGD利用界定为以OGD为基本要素,通过数据管理机构、公众、数据开发团体等多元主体的互动反馈,创造数据产品、提供数据服务、生成新的数据,并最终实现数据价值增值的循环可持续过程。该定义下,OGD利用以数据为基本要素,以数据流动为关键进程,以数据价值增值为最终目标,以循环可持续为主要特征。

OGD研究早期主要从政府、数据、用户和外部环境等方面探讨了OGD利用的影响因素。从政府视角出发的研究多关注政府在OGD利用中的主导作用^[6-8];从数据角度出发的研究强调数据本身的特性^[9-10];从用户角度出发的研究则更加关注OGD用户的体验、互动参与和使用障碍等^[11-13]。也有一些研究讨论了府际竞争、文化环境、政策法规等外部因素对OGD利用的影响^[14-15]。随着研究和实践的推进,OGD生态系统相关研究逐步涌现,旨在从生态系统的视角审视各复杂因素对OGD利用的影响,试图找到OGD利用水平不理想的原因^[16]。有关研究可大致分为要素视角和流程视角两种类型。

要素视角下的相关研究主要从OGD生态系统中的要素结构与特征出发阐释OGD利用,关注OGD生态系统的构成要素。范丽莉等[17]测度了中国省级政府的OGD生态系统建设状况和差异,指出OGD生态系统建设对OGD利用有积极意义,良好的OGD生态系统应当考虑环境、主体、客体和价值目标。门理想等[5]认为,包

括数据利用在内的OGD建设成效有赖于各系统要素基于数据链的互动,并受到开放者行政权力、开发者开发能力、消费者采纳能力、外部环境等要素特征的影响,其强调了OGD生态系统中数据池的作用。Erna等^[18]认为OGD生态系统是一种复杂系统,系统中的使用者、开放平台、受益者等要素相互作用,从而直接或间接为OGD利用作出贡献,实现单一主体难以产生的效果。Van Schalkwyk等^[19]则强调了OGD生态系统中的中介机构对OGD利用的重要促进作用,指出中介机构增加了数据的可访问性、实用性、可持续性和流动性,并有望使OGD利用民主化。

流程视角下的研究主要从数据流动的不同环节出 发阐释OGD利用,关注数据流动过程和生命周期。张 晓娟等[20]分析了生态系统理论对OGD实践的适用性, 指出OGD的开放、利用和价值释放需要在可持续和动 态的OGD生态系统中实现,数据是系统中各利益相关 者生产、传递、利用和消费的对象。赵需要等[21]构建了 一个OGD生态链结构模型,指出OGD生态链通过数 据资源化、数据资产化和数据资本化环节,实现数据利 用和数据价值释放。侯晓丽等^[22]进一步指出, OGD生 态链上各节点以数据流转为核心,按照价值激活、价值 汇聚和价值扩散的顺序实现数据利用和价值释放,各 节点的良性互动是推进数据利用和价值释放的有效路 径。Dawes等[23]和Runeson等[24]发现,建设OGD生态系 统是一种能有效结合现实条件、消除OGD利用障碍并 实现数据预期价值的方法,一般会涉及发布数据、批准 和取得数据、加工数据、理解和反馈数据、利用数据等 内容。Zuiderwijk等[25]结合数据生命周期,归纳了OGD 生态系统的关键要素,强调要从系统视角完善OGD生 态系统中各相互依赖的环节,包括释放和发布数据、获 取数据及相关许可、加工并可视化数据、探讨数据并提 供反馈,以顺利实现OGD利用。

OGD生态系统相关研究的兴起是对早期研究的一种反思,表明学术界已然意识到要推进OGD利用,不能仅关注单向度的数据供给、数据本身或者用户使用,更应考虑复杂的OGD生态系统,从而改善OGD利用过程,充分释放OGD的价值^[26]。遗憾的是,OGD生态系统作为OGD研究的新兴领域,尚在探索阶段,研究成果相对稀少且多为理论探讨,致使经验证据不足。既有的实证研究较少从必要性和充分性双重逻辑出发审视研究问题,且对要素之间的互动关注不够,无法回答"影响因素的必要性和必要程度如何"和"不同影响

因素应如何组合"的问题。然而,要在复杂OGD生态系统中尝试提升OGD利用水平,这些问题是无法回避的,这就为本研究创造了空间。

2 理论基础与分析框架

2.1 理论基础

作为一个新兴研究领域,OGD生态系统研究在发展过程中广泛借鉴了生态系统理论,如信息生态系统理论^[28]、而业生态系统理论^[29]等,形成了各有侧重的OGD生态系统模型。随着研究的推进,OGD生态系统的创新属性得到强调^[25]。马海群等^[16]指出OGD生态系统不仅是一个有技术与社会特征的信息生态系统,也是一个创新系统。孟雪等^[30]进一步论证了OGD生态系统的创新属性,指出OGD价值的实现是一个创新的过程。

创新生态系统理论最早可以追溯到1993年。根据 Moore^[31]的观点,创新生态系统被定义为一个囊括了诸 多要素的动态系统结构。随后,Marco等^[32]指出,创新 生态系统中不同的主体互相关联,某个主体的生态位变 化会使其他主体的生态位随之变化。Adner等^[33]进一步 强调了创新生态系统中主体之间的协同关系,指出创新 生态系统是一套协同机制,并注意到外部环境对系统 创新的影响。可以看到,多元主体共存互动,形成一个价值共创网络,并受到外部环境的影响,这是创新生态系统理论的题中之义。

OGD生态系统与创新生态系统具有类似的特征。一方面,创新生态系统能够刻画OGD生态系统中各主体之间的复杂互动关系。OGD生态系统与创新生态系统类似,存在着包括开放者、开发者、消费者、开放平台在内的众多承担不同功能的主体,不同主体之间相互影响和依赖。例如开发者和消费者有赖于开放者提供的高质量数据与服务,同时开放者也需要从开发者和消费者处获取数据和反馈。另一方面,一个完整的OGD生态系统也涉及相应的外部环境,这与创新生态系统的主要观点一脉相承。在OGD生态系统中,各主体的互动过程会受到技术环境、经济环境和社会环境等外部环境的影响。

以上分析表明,创新生态系统理论不仅能够刻画 OGD生态系统中的要素特征,还能较好地体现数据利 用和价值创造的过程。因此,本研究以创新生态系统理论为基础搭建OGD生态系统分析框架。

2.2 分析框架

目前,尽管学界对OGD生态系统的构成要素暂未 形成一致观点,但大多认为构成要素涉及行动主体与外 部环境^[17]。其中,行动主体主要包括开放者、开发者、消 费者等,外部环境包括经济环境、社会环境和技术环境 等。此外,以数据为链条的各要素的互动逻辑也被考虑 在内,主要包括各行动主体之间的数据流动、反馈和倡 议^[5]。在不同类型的OGD生态系统模型中,开放平台未 得到重视。从实践来看,OGD平台是最典型的OGD实现 模式^[34],不仅是各主体发挥其功能和作用的途径,也是 各个主体互动反馈的桥梁。从理论来看,OGD生态系统 强调主体之间的及时、可持续的反馈互动和便利的数据 获取,这有赖于承载了海量优质数据且建设优良的OGD 平台^[18]。因此,本研究亦将OGD平台纳入分析框架。

基于以上分析并结合创新生态系统理论的基本内容,设计OGD生态系统分析框架(见图1)。该分析框架主要由生态系统主体和外部环境构成,以开放者、开发者、消费者为生态系统主体,各主体之间以数据链相互连接,通过开放平台实现互动和反馈,实现数据利用和价值创造,并受到经济环境、社会环境和技术环境的影响。

开放者是指OGD管理机构,在中国主要指政府部门,负责数据采集、发布和管理。既有研究表明,开放者的组织水平、制定标准和政策法规等的能力对OGD利用水平有重要影响^[14]。数据开放者的开放能力越强,越能提供海量优质数据,从而为数据利用创造良好的基础。

开发者是指OGD的开发者,如企业、科研团体、行业团体等,是数据需求主体之一,利用OGD开发数据产品和服务,在推动OGD利用的同时也是开放者和消费者的桥梁。开发者的数据开发能力越强,越能挖掘OGD蕴含的价值,提供丰富优质的数据产品和服务,从而提升OGD利用水平^[35]。

消费者是指OGD产品和服务的终端使用者,主要包括民众、社会团体,也包括政府本身。消费者在从数据产品和服务中获益的同时,也是OGD的重要反馈来源和数据来源。消费者对数据产品与数据服务的消费

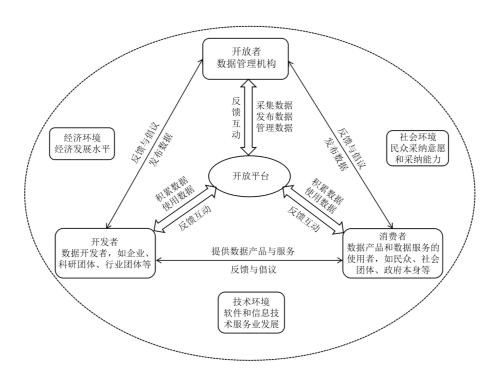


图1 OGD生态系统分析框架

能力影响着OGD能否最终实现增值利用[18]。

开放平台是指OGD门户网站,是数据载体和互动渠道,以在线平台为主,如中国各省级和地市级政府建设的OGD线上平台。开放平台是联系OGD生态系统中各主体的枢纽。建设良好的开放平台能够使行动主体便捷地获取和反馈数据,提升行动主体的使用体验,对OGD利用起着至关重要的作用^[36]。

同时,该系统还涉及技术环境、经济环境和社会环境。OGD伴随着数字社会和数字政府的发展而形成,而数字技术的产生和利用影响着数据采集、存储、传输、分析、安全保障等环节,对于提升数据利用水平至关重要。良好的经济环境有利于培育数据需求和数据开发者群体,也能为开放者和开放平台的发展提供更加充足的资金支持,有利于促进OGD利用^[37]。社会环境的内涵相对宽泛,OGD生态系统相关的研究多强调民众的信息采纳意愿和能力^[5,17],这反映了民众对数据产品和服务的接受能力和需求意愿是推进OGD利用的重要社会性因素。

在OGD生态系统中,数据集成于开放平台,并在各个生态系统主体之间流动,贯穿从开放到价值创造再到数据回流的全过程,形成连接整个生态系统的数据链条。具体地,数据最初由开放者通过开放平台发布和管理,然后流向开发者并被加工成对应的数据产品和数据

服务,接着流向消费者,进而在对应的场景中实现价值创造。伴随着OGD生态系统中各行动主体之间的反馈互动,各环节的流动过程也会创造和积累新的数据,新数据成为下一轮数据开放的对象,实现回流和循环。

3 OGD利用瓶颈与组态的实证分析

3.1 研究方法

必要条件分析法 (Necessary Condition Analysis, NCA) 是Dul^[38]提倡的一种必要条件检验方法,可以弥补传统回归分析不能发现必要条件的逻辑缺陷,且能够体现前因条件的必要程度,进行瓶颈水平分析。而定性比较分析法 (Qualitative Comparative Analysis, QCA) 是一种多案例比较分析方法,可发现各前因条件之间的联动关系,适合考察多要素联动影响的复杂情况^[39]。同时,相对于传统回归分析,QCA的要求更灵活,可处理不同数量规模的样本,且QCA算法基于布尔代数逻辑,无须考虑遗漏变量造成的偏差^[40]。

本研究需要同时探索前因条件对结果的必要性和充分性,检验必要条件和组态条件,因此NCA与QCA的混合方法能够较好适配本文的研究目的。首先,采用NCA检验必要条件,确定各前因条件是否构成高OGD

利用水平的必要条件及其必要程度,并进一步找出构成瓶颈的前因条件。其次,采用QCA检验组态条件,分析各前因条件之间的联动关系,找出可以实现高OGD利用水平的前因条件组合。

3.2 样本、变量与数据

3.2.1 样本与变量

研究样本为我国27个省级行政区(不含港澳台及直辖市),理由如下:一方面,省级政府作为省域最高行政主体统筹省域OGD工作,是OGD的代表性行动主体;另一方面,样本广泛分布于中国各地理区域,并呈现出不同的OGD利用水平,能够反映OGD生态系统的多样性和复杂性。

利用水平是结果变量,用2023年《中国地方公共数据开放利用报告(省域)》中的利用层指数衡量^[3]。该指数综合考虑了OGD利用成果的数量和质量,能够全面合理地衡量OGD利用水平^[41]。前因条件包括开放者、开发者、消费者、开放平台、技术环境、经济环境、社会环境7个变量。其中,开放者变量用《中国地方政府数据

开放报告(省域)》中的准备度指数衡量,该指数从法规政策、标准规范和组织推进3个方面综合评估,反映了数据管理机构在制度设定、重视程度和组织架构方面的情况,能够较为全面地体现开放者的数据开放能力^[42]。开发者变量用省域内信息传输、软件和信息技术服务从业人员比例衡量^[5],反映了对领域资源的利用能力^[43-44]。消费者变量用各省域人均信息消费规模衡量,反映了当地数据产品和服务消费能力^[17]。开放平台变量用《中国地方政府数据开放报告(省域)》中的平台层指数衡量,该指数从数据获取、使用体验、反馈机制等角度综合评估了数据开放平台的建设水平^[3]。

经济环境变量用人均GDP衡量^[38,45-46],体现了当地对OGD利用过程的经济支撑能力。社会环境变量用每10万人受高等教育人数衡量^[5]。受高等教育人数会影响公民对数据产品和服务的认知和判断^[47],进而影响其对数据产品和服务的采纳意愿和采纳能力^[48-49]。技术环境变量用软件业务收入衡量^[17],体现了当地软件和信息技术服务业对OGD利用过程的技术支撑能力。为避免因果倒置带来的影响,将所有前因条件的数据年份滞后于结果变量1年,变量设定与数据来源情况见表1。

| 变量类型 | 变量名称 | 变量描述 | 指标说明 | 数据来源 |
|------|---|-----------|--------------------------------------|------------------------|
| 结果变量 | 利用水平 | 数据利用水平 | 2023年《中国地方公共数据开放利用报告(省域)》 利用层指数得分 | 复旦大学数字与移动 治理实验室官方网站 |
| | 开放者 | 数据开放和管理能力 | 2022年《中国地方政府数据开放报告(省域)》 准备度指数得分 | 复旦大学数字与移动 治理实验室官方网站 |
| | 开发者 数据开发利用能力 2022年信息传输、软件和信息技术服务 从业人员比例 (%) | | 国家统计局官方网站 | |
| V N | 消费者 | 信息消费能力 | 2022年各省域人均信息消费规模(元/人) | 国家统计局官方网站 |
| 前因条件 | 开放平台 | 平台建设状况 | 2022年《中国地方政府数据开放报告(省域)》 平台层指数得分 | 复旦大学数字与移动 治理实验室官方网站 |
| | 技术环境 | 技术支撑能力 | 2022年软件业务收入(亿元) | 国家统计局官方网站 |
| | 经济环境 | 经济支撑能力 | 2022年人均GDP (元/人) | 国家统计局官方网站 |
| | 社会环境 | 信息采纳意愿与能力 | 2022年每10万人受高等教育人数 | 国家统计局官方网站 |

表1 变量设定与数据来源

3.2.2 数据校准

去掉数据奇异值后,以5%和95%分位数为"完全不隶属"和"完全隶属",以50%分位数为"交叉点",并结合实际情况酌情调整,确定校准点,赋予各前因条件集合隶属关系。校准点和描述性统计结果见表2。

3.3 高OGD利用水平的必要条件与瓶颈

首先运用NCA开展分析,以识别各变量是否能单独作为必要条件。NCA结果如表3所示,包括上限回归函数(Ceiling Regression, CR)和上限包络函数(Ceiling Envelopment, CE)两种算法的结果。若某个变量为必要条件,则效应量>0.1且P<0.01。表3数据显

| 变量 | 校准点 | | | 描述性统计结果 | | | |
|------|------------|-----------|-----------|-----------|-----------|-----------|------------|
| | 完全隶属 | 交叉点 | 完全不隶属 | 均值 | 标准差 | 最小值 | 最大值 |
| 利用水平 | 12.26 | 0.70 | 0.42 | 2.72 | 3.79 | 0.30 | 13.77 |
| 开放者 | 13.01 | 3.02 | 0.97 | 4.72 | 3.70 | 0.80 | 15.14 |
| 开发者 | 4.90 | 1.91 | 1.14 | 2.41 | 1.29 | 1.07 | 6.63 |
| 消费者 | 1 640.38 | 1 138.95 | 966.93 | 1 203.56 | 241.24 | 846.27 | 1 702.50 |
| 开放平台 | 15.34 | 5.89 | 0.00 | 6.01 | 5.61 | 0.00 | 16.10 |
| 技术环境 | 12 294.32 | 596.45 | 4.57 | 2 623.75 | 4 475.79 | 0.00 | 17 613.00 |
| 经济环境 | 122 239.20 | 69 019.00 | 51 187.40 | 75 027.52 | 23 704.53 | 44 646.00 | 143 466.00 |
| 社会环境 | 4 418.80 | 3 277.00 | 1 964.50 | 3 330.48 | 729.79 | 1 666.00 | 4 989.00 |

表2 校准点和描述性统计结果

表3 NCA结果

| 变量 | 算法 | 精确度/% | 上限区域 | 范围 | 效应量 | P |
|---------------|----|--------|-------|------|-------|-------|
| 开放者 | CR | 63.00 | 0.276 | 0.89 | 0.310 | 0.000 |
| | CE | 100.00 | 0.361 | 0.89 | 0.404 | 0.000 |
| 开发者 | CR | 88.90 | 0.094 | 0.91 | 0.103 | 0.248 |
| 丌及有 | CE | 100.00 | 0.032 | 0.91 | 0.035 | 0.513 |
| 消费者 | CR | 88.90 | 0.135 | 0.92 | 0.146 | 0.080 |
| | CE | 100.00 | 0.096 | 0.92 | 0.104 | 0.123 |
| 774.77. | CR | 48.10 | 0.278 | 0.89 | 0.312 | 0.000 |
| 开放平台 | CE | 100.00 | 0.380 | 0.89 | 0.425 | 0.000 |
| 技术环境 | CR | 96.30 | 0.165 | 0.90 | 0.183 | 0.001 |
| | CE | 100.00 | 0.207 | 0.90 | 0.229 | 0.000 |
| 经济环境 | CR | 96.30 | 0.023 | 0.94 | 0.024 | 0.778 |
| | CE | 100.00 | 0.038 | 0.94 | 0.041 | 0.718 |
| 社人IT 拉 | CR | 96.30 | 0.122 | 0.92 | 0.133 | 0.138 |
| 社会环境 | CE | 100.00 | 0.145 | 0.92 | 0.158 | 0.086 |

示,开放者、开放平台和技术环境是生成高OGD利用水平的必要条件,这表明在OGD生态系统中,能力强的开放者、优质的开放平台和良好的技术环境对于生成高OGD利用水平是不可或缺的。

瓶颈水平分析可体现结果达到某一水平时前因条件需要达到的程度,从而发现制约结果的瓶颈^[50]。瓶颈水平分析结果如表4所示,可见要达到80%的OGD利用水平,需要开放者、开放平台和技术环境分别达到74.70%、77.00%和45.00%的水平。这表明,要想实现一个较为理想的利用水平,须警惕开放能力不足以及平台建设和技术环境不佳造成的"木桶效应",即无论OGD生态系统中其他要素状况如何,若开放者、开放平台和技术环境存在短板,会使整体利用水平不高,整个系统犹如一个漏水之桶。

3.4 生成高OGD利用水平的组态

利用组态分析可以探究各前因条件是如何组合进而生成高OGD利用水平的。分析过程中的原始一致性、PRI一致性和案例频数阈值分别设置为0.80、0.80和1。考虑到结论的启示度和解的简洁性,本研究汇报了中间解,结果如表5所示。解中共包含3个组态,案例总体覆盖度达到0.44,总体一致性达到0.96,各组态中展示了核心条件、边缘条件和缺失条件的组合情况。

结合图1所示的分析框架将各组态划分为供给主导型(组态1)、供需联动型(组态2)和综合驱动型(组态3),以下展开分析组态的特征和对应典型案例。

(1)供给主导型。在该类型的组态中,核心条件主要分布在数据供给侧,包括开放者和开放平台。组态1是供给主导型的代表,该组态中开放者、开放平台作为核心条件出现,和技术环境共同生成高OGD利用水

表4 NCA瓶颈水平分析结果

单位:%

| 利用水平 | 开放者 | 开发者 | 消费者 | 开放平台 | 技术环境 | 经济环境 | 社会环境 |
|------|-------|-------|-------|-------|-------|-------|-------|
| 0 | NN |
| 10 | NN |
| 20 | NN |
| 30 | NN |
| 40 | NN |
| 50 | 7.10 | NN | NN | 4.60 | NN | NN | NN |
| 60 | 29.60 | NN | NN | 28.70 | 10.80 | NN | NN |
| 70 | 52.20 | NN | 11.70 | 52.90 | 27.90 | NN | 12.70 |
| 80 | 74.70 | 16.70 | 35.80 | 77.00 | 45.00 | NN | 32.80 |
| 90 | 97.20 | 49.50 | 59.90 | NA | 62.10 | NN | 53.00 |
| 100 | NA | 82.30 | 84.00 | NA | 79.20 | 94.40 | 73.10 |

注:采用CE计算, NN表示不必要, NA表示必要。

表5 生成高OGD利用水平的组态

| 变量 | 组态1 | 组态2 | 组态3 |
|-------|-----------|------|------|
| 开放者 | • | • | • |
| 开发者 | \otimes | • | • |
| 消费者 | | | • |
| 开放平台 | • | • | • |
| 技术环境 | • | • | • |
| 经济环境 | \otimes | • | |
| 社会环境 | \otimes | • | • |
| 一致性 | 0.97 | 0.95 | 0.94 |
| 原始覆盖度 | 0.19 | 0.36 | 0.31 |
| 唯一覆盖度 | 0.07 | 0.03 | 0.00 |
| 总体一致性 | | 0.96 | |
| 总体覆盖度 | | 0.44 | |

注: ●/•表示条件出现, 其中●表示核心条件, •表示边缘条件; ⊗表示条件缺失; 空白表示条件对结果不重要。

平,其他条件则缺失或者不重要。该类型组态对应的案例中,高OGD利用水平生成主要依赖良好的供给侧条件,包括开放实力较强的政府和建设水平较高的OGD平台,以及良好的技术环境,而数据开发者和数据产品的最终消费者发挥的作用相对有限,因此属于供给主导型的OGD生态系统,典型案例有广西和贵州。

以广西为例,其在2023年《中国地方公共数据开放利用报告(省域)》中的利用层指数得分为4.98,位居全国第三。在广西OGD生态系统中,OGD实践得到了政府长期重视,相应的平台建设也非常完善,呈现出供给主导的特点。广西OGD工作由广西壮族自治区大数据发展局统筹推进,该机构是直属于广西壮族自治区人民政府的正厅级机构,长期以来围绕《广西公共数

据开放管理办法》《广西壮族自治区大数据发展条例》 《广西加快数据要素市场化改革实施方案》等代表性 文件,聚焦公共数据资源的开放与授权运营,持续探索 数据利用创新方式,在引导数据利用、赋能经济社会发 展方面取得了长足进步。2023年《中国地方公共数据开 放利用报告(省域)》显示,广西公共数据开放长期保 持领先水平,在2020—2023年省域累计得分排名中, 4年累计得分为161.96,全国排名第七。广西于2020年 9月上线公共数据开放平台,截至2023年底,平台数据 总量已经超过70亿,包含数据目录近万个,接口调用量 近170万次,累计输出上百个数据应用成果。2022年广 西平台层指数得分为12.58,位居全国第五,紧跟浙江 等数字强省之后,优秀的平台建设水平为数据利用提供 了便捷的调用渠道和反馈互动机制。此外,2022年广西 软件业务收入达到819.65亿元,信息传输、软件和信息 技术服务业发展处于全国中上游水平,在供给侧优势的 基础上为OGD利用提供了外部环境助力。

(2)供需联动型。在该类型的组态中,核心条件分布在数据供给侧和需求侧两端,包括开放者、开放平台、开发者。组态2是供需联动型的代表,该组态中开放者和开放平台作为供给侧的核心条件出现,开发者作为需求侧的核心条件出现,结合良好的技术、经济和社会外部环境,共同生成高OGD利用水平。该类型组态对应的案例不仅拥有开放实力较强的政府和建设水平较高的OGD平台,还拥有开发利用能力较强的开发者,使得需求侧拉动作用得到了一定体现,并得到了良好的外部环境支持。高OGD利用水平得益于供需两侧条件的共同作用,因此属于供需联动型的OGD生态系统,典型

案例有山东和安徽。

以山东为例, 其在2023年《中国地方公共数据开放 利用报告(省域)》中的利用层指数得分为13.20,位居 全国第二,呈现出领先的OGD利用水平。在山东OGD 生态系统中,不仅开放者能力和开放平台建设水平优 异,开发者也具有良好的开发能力,为OGD利用提供 了动力, OGD生态系统呈现出供需协同驱动的特点。 一方面, 山东在2022年《中国地方政府数据开放报告 (省域)》中的准备度指数得分为14.93,平台层指数得 分为15.64,均位居全国第二,表明山东OGD生态系统 在供给侧具备显著优势。山东OGD工作由直属于省政 府的山东省大数据局牵头, 其围绕《数字山东发展规 划(2018-2022年)》《山东省数字政府建设实施方案 (2019-2022年)》《2023年新增公共数据开放清单》 等政策文件,带领山东实现政府部门和主要公共事业 单位数据开放全覆盖,省域公共数据开放平台数量和 开放数据总量全国领先,实现地理信息、环境保护、交 通运输等重点领域数据深度开放,为OGD利用提供了 坚实的基础。另一方面,信息传输、软件和信息技术服 务从业人员数量众多,具有巨大的数据开发利用潜力, 这为OGD利用创造了需求侧优势。山东结合供需两侧 优势和良好的外部环境,超越一些沿海发达省份,实现 高OGD利用水平。

(3)综合驱动型。该类型的组态几乎包含了OGD 生态系统中的全部要素,行动主体和外部环境中的变量共同生成高OGD利用水平,组态3是综合驱动型的代表。该类型的案例中,政府的开放实力较强且能够提供高水平的OGD平台,为OGD利用提供了良好的供给侧保障。同时,需求侧的开发能力和消费能力也表现良好,为充分释放数据价值提供了开发能力支持和消费基础保证。此外,该类型的案例往往还拥有良好的社会技术外部环境。高OGD利用水平得益于多维度条件共同作用,因此属于综合驱动型的OGD生态系统,典型案例有广东和四川。

以广东为例,该省OGD利用水平在2023年《中国地方公共数据开放利用报告(省域)》中排名全国第七,呈现出良好的OGD利用水平。在广东OGD生态系统中,开放者实力和开放平台建设水平全国领先,且广东拥有领先的数据开发能力、信息消费能力和优越的社会技术环境支持,为OGD利用创造了综合驱动条件。首先,广东省政务服务和数据管理局作为省政府直属部门,先后颁布了《广东省促进大数据发展行动计划

(2016—2020年)》《广东省公共数据管理办法》《广东省首席数据官制度试点工作方案》等相关法规政策,明确了数据开放利用和共享的基本内容、程序和标准,为OGD利用提供了坚实可用的数据资源基础和组织保障。其开放数据平台集成了省内21个地级市和53个省级政府部门的数据,数据容量全国领先。其次,广东作为数字经济强省,信息传输、软件和信息技术服务从业人员比例居于全国之首,数据开发实力全国领先,信息消费规模也处于全国前列。最后,广东本身拥有优越的社会技术环境,为OGD利用提供了良好的外部支持。广东信息传输、软件和信息技术服务业的发展长期全国领先,2022年本地的受高等教育人数占比位列全国第一,且广东吸纳了大量来自各地的高素质人才,为数字技术的创新应用以及数据产品服务采纳提供了良好的社会技术基础。

3.5 稳健性检验

为了测试数据分析结果的稳健性,本研究参考既有研究^[51],采用上调一致性阈值的方法开展稳健性检验。具体地,将原始一致性阈值从0.80上调到0.85后,观察新的数据分析结果和上述分析结果的变化。结果显示,采用更高的原始一致性后,上述分析结果几乎没有变化。新的数据分析结果中的组态内容、组态数量和对应参数与表5所示的内容一致。因此,有理由认为本研究的数据分析结果具备一定稳健性。

4 结论与启示

本研究基于创新生态系统理论,构建了一个OGD 生态系统模型,将开放者、开发者、消费者视为生态系统主体,将经济环境、社会环境和技术环境作为外部环境,并考虑开放平台在连接系统要素和促进要素互动反馈中的作用,在该框架下分析影响OGD利用水平的瓶颈条件和组态条件。采用了NCA和QCA混合方法进行数据分析,结果显示开放者、开放平台和技术环境是生成高OGD利用水平的必要条件,并在相应水平上构成OGD利用的瓶颈。这表明了开放者、开放平台和技术环境对于提升OGD利用水平的关键性作用,也进一步回答了实现较高OGD利用水平需要上述变量达到何种水准。QCA组态分析结果表明,存在生成高OGD利用水 平的3种组态类型,分别是供给主导型、供需联动型和综合驱动型,不同组态类型体现出不同的变量分布特征。供给主导型的组态中,核心条件主要分布在数据供给侧,包括开放者和开放平台;供需联动型的组态中,核心条件主要分布在供给侧和需求侧两端,包括开放者、开放平台、开发者和消费者;综合驱动型的组态则涵盖了几乎全部的条件。本研究在OGD生态系统研究勃兴的背景之下,提供了来自中国的实证证据,一方面为OGD生态系统理论构建提供新的参考,另一方面也为相关实践提供启示。

一方面,要提高OGD利用水平,应打好"组合拳"。 应当从政府单向度开放数据转向生态系统多主体驱 动,并更加注重数据开发者和消费者的需求、反馈和 互动,打造使用者导向和循环式OGD生态系统。另一方 面,要提高OGD利用水平,须补好"漏水桶"。应警惕 开放者、开放平台和技术环境存在短板而造成的"木桶 效应"。政府作为OGD的主导者,应当注重法规制度体 系化搭建和必要的组织建设,不断优化自身数据开放 综合能力,为数据利用链条铸就良好开端,并为OGD 生态系统营造良好外部技术条件。特别地,开放平台作 为OGD生态系统中桥接多方主体的枢纽,在促进数据 利用的过程中发挥着至关重要的作用。瓶颈水平分析 结果表明,要达到一个比较理想的OGD利用水平,对平 台建设的要求是不低的。从实践来看,一些地区的OGD 平台建设在数据发布、互动反馈、权益保护和使用体验 等方面尚有巨大进步空间,应当加强OGD平台建设, 打造透明高效、权益保护得当、供需适配的开放平台, 尽快补齐短板。此外,应注重发挥OGD生态系统中需 求侧的拉动作用。组态分析结果表明,当前中国的OGD 生态系统中,需求侧在促进OGD利用中发挥的作用尚 且有限,这无疑不利于OGD利用水平进一步提升。因 此,一方面应该减少数据开发障碍,提供开放接口和调 用指南,必要时设置开发者专区,设置常态化和多样化 的联合创新机制,培育更多信息化人才,最大限度释放 开发者潜能。另一方面,应随着数字经济的发展,创造 更多更丰富的开放数据应用场景,提升公民数字素养 和数据意识,吸引更多消费者参与OGD利用。

本研究的局限性在于,受QCA算法逻辑的有限多样性^[52]和数据可得性等客观条件限制,当下的变量设置和赋值与复杂现实难免有异。此外,随着未来OGD生态系统理论继续发展直至形成相对独立的范畴,当前分析框架可得到进一步优化。

参考文献

- [1] 北京日报. 李克强: 信息数据"深藏闺中"是极大浪费[EB/OL]. [2024-02-05]. https://www.gov.cn/xinwen/2016-05/13/content_5073036.htm? url_type=39&object_type=webpage&pos=1.
- [2] 国家发展和改革委员会规划司. "十四五"数字经济发展规划[EB/OL]. [2024-02-05]. https://www.ndrc.gov.cn/fggz/fzzlgh/gjjzxgh/202203/t20220325_1320207_ext.html.
- [3] 复旦大学数字与移动治理实验室. 开放树林指数[EB/OL]. [2024-02-05]. http://ifopendata.fudan.edu.cn.
- [4] 郑磊, 刘新萍. 我国公共数据开放利用的现状、体系与能力建设研究[J]. 经济纵横, 2024(1): 86-92.
- [5] 门理想,王丛虎. 中国地方政府数据开放建设成效的影响因素探究:基于生态系统理论框架[J]. 现代情报,2021,41(2):152-161.
- [6] 冉从敬. 公共部门信息再利用制度研究[J]. 中国图书馆学报, 2008, 34(6):103-108.
- [7] ZUIDERWIJK A, JANSSEN M, DWIVEDI Y K. Acceptance and use predictors of open data technologies: drawing upon the unified theory of acceptance and use of technology[J]. Government Information Quarterly, 2015, 32 (4): 429-440.
- [8] ZELETI F A, OJO A, CURRY E. Exploring the economic value of open government data[J]. Government Information Quarterly, 2016, 33 (3): 535-551.
- [9] VETRÒ A, CANOVA L, TORCHIANO M, et al. Open data quality measurement framework: definition and application to open government data[J]. Government Information Quarterly, 2016, 33 (2): 325-337.
- [10] BARCELLOS R, VITERBO J, MIRANDA L, et al. Transparency in practice: using visualization to enhance the interpretability of open data[C]//The 18th Annual International Conference on Digital Government Research. New York: ACM Press, 2017: 139-148.
- [11] ZUIDERWIJK A, JANSSEN M, CHOENNI S, et al. Sociotechnical impediments of open data[J]. Electronic Journal of e-Government, 2012, 10 (2): 156-172.
- [12] WIRTZ B W, WEYERER J C, RÖSCH M. Open government and citizen participation: an empirical analysis of citizen expectancy towards open government data[J]. International Review of Administrative Sciences, 2019, 85 (3): 566-586.
- [13] WANG H J. Adoption of open government data: perspectives

- of user innovators[J]. Information Research, 2020, 25 (1): 849-849
- [14] 张晓宇,周文泓,王萍,等. 国外促进政府开放数据社会化利用的实践全景及其启示:基于十国政府开放数据平台的调查[J]. 现代情报,2023,43(1):100-109.
- [15] 汤志伟,王研. TOE框架下政府数据开放平台利用水平的组态 分析[J]. 情报杂志, 2020, 39(6): 187-195.
- [16] 马海群,王本刚. 开放数据生态系统:一个系统性文献综述[J]. 图书馆杂志, 2022, 41 (11): 8-18, 39.
- [17] 范丽莉,姚倩钰,石蕾. 我国省级政府数据开放水平测度及评价:基于生态系统视角[J]. 数字图书馆论坛,2023,19(2):18-26.
- [18] ERNA R, JUSTIEN D, ALBERT M. Studying complex systems through design interventions probing open government data ecosystems in the Netherlands[J]. Public Management Review, 2023, 25 (1): 129-149.
- [19] VAN SCHALKWYK F, WILLMERS M, MCNAUGHTON M. Viscous open data: the roles of intermediaries in an open data ecosystem[J]. Information Technology for Development, 2016, 22 (sup1): 68-83.
- [20] 张晓娟, 莫富传, 王意. 政府数据开放生态系统的理论、要素与模型探究[J]. 情报理论与实践, 2022, 45 (12): 42-49.
- [21] 赵需要,郭义钊,姬祥飞,等.政府开放数据生态链上数据要素价值分析及评估模型构建:基于"数据势能"的方法[J].情报理论与实践,2022,45(12):50-59.
- [22] 侯晓丽,赵需要,周庆山,等.政府开放数据生态链价值共创机 理与节点优化策略研究[J].情报理论与实践,2024,47(5): 67-77.
- [23] DAWES S S, VIDIASOVA L, PARKHIMOVICH O. Planning and designing open government data programs: an ecosystem approach[J]. Government Information Quarterly, 2016, 33 (1): 15-27.
- [24] RUNESON P, OLSSON T, LINAKER J. Open data ecosystems: an empirical investigation into an emerging industry collaboration concept[J]. Journal of Systems and Software, 2021, 182 (12): 111088.
- [25] ZUIDERWIJK A, JANSSEN M, DAVIS C. Innovation with open data: essential elements of open data ecosystems[J]. Information Polity, 2014, 19 (1/2): 17-33.
- [26] VAN LOENEN B, ZUIDERWIJK A, VANCAUWENBERGHE
 G, et al. Towards value-creating and sustainable open data
 ecosystems: a comparative case study and a research agenda[J].

- JeDEM-eJournal of eDemocracy and Open Government, 2021, 13 (2): 1-27.
- [27] 袁红,王焘.政府开放数据生态系统可持续发展实现路径的系统动力学分析[J].图书情报工作,2021,65(17):13-25.
- [28] HARRISON T M, PARDO T A, COOK M. Creating open government ecosystems: a research and development agenda[J]. Future Internet, 2012, 4 (4): 900-928.
- [29] JETZEK T. Analytics, innovation, and excellence-driven enterprise sustainability[M]. New York: Palgrave Macmillan, 2017: 137-168.
- [30] 孟雪, 郝文强. 面向数字经济发展的政府数据开放价值创造系统建构与运行机制研究:基于创新生态系统的理论分析[J]. 情报杂志, 2023, 42(2):134-141, 174.
- [31] MOORE J F. Predators and prey: a new ecology of competition[J]. Harvard Business Review, 1993, 71 (3): 75-86.
- [32] MARCO I, ROY L. Strategy as ecology[J]. Harvard Business Review, 2004, 82 (3): 68-78, 126.
- [33] ADNER R, KAPOOR R. Value creation in innovation ecosystems: how the structure of technological interdependence affects firm performance in new technology generations[J]. Strategic Management Journal, 2010, 31 (3): 306-333.
- [34] 王林川, 寿志勤, 吴慈生. 政府数据开放平台服务绩效评价指标体系研究: 基于公共价值视角[J]. 中国行政管理, 2022(1): 40-47.
- [35] 周文泓,张晓宇,田欣,等. 我国地方政府开放数据的社会利用 现状与促进对策[J]. 图书馆论坛, 2022, 42(4): 76-85.
- [36] 朱红灿,方心悦. 社会技术系统理论视角下政府数据开放平台 生态系统成熟度评估研究[J]. 电子政务, 2022 (9): 88-102.
- [37] 高富平. 数据经济的制度基础: 数据全面开放利用模式的构想[J]. 广东社会科学, 2019 (5): 5-16, 254.
- [38] DUL J. Necessary condition analysis (NCA) logic and methodology of "necessary but not sufficient" causality[J].

 Organizational Research Methods, 2016, 19 (1): 10-52.
- [39] 里豪克斯, 拉金. QCA设计原理与应用: 超越定性与定量研究的新方法[M]. 杜运周, 李永发, 译. 北京: 机械工业出版社, 2017: 2-15.
- [40] 杜运周,刘秋辰,程建青. 什么样的营商环境生态产生城市高创业活跃度?:基于制度组态的分析[J]. 管理世界,2020,36(9):141-155.
- [41] 温祖卿,郑磊. 地方政府开放数据的利用与产出研究[J]. 电子政务, 2019 (9): 23-31.
- [42] 上官莉娜,潘晨. 制度逻辑视角下省级政府数据开放绩效的影



- 响因素及生成机制:基于20省的模糊集定性比较分析[J].情报杂志,2021,40(12):150-157,176.
- [43] 符嵘. 参与主体视角下的政府数据开放研究: 以上海市为例[J]. 科技促进发展, 2017, 13(4): 237-241.
- [44] 汤志伟,郭雨晖,顾金周,等. 创新扩散视角下政府数据开放平台发展水平研究:基于全国18个地方政府的实证分析[J]. 图书馆理论与实践, 2018 (6):1-7.
- [45] 刘园园, 段尧清, 俞享. 基于超效率SBM-fsQCA的政府数据开放有效路径研究[J]. 情报科学, 2023, 41 (9): 146-154.
- [46] 张晓娟, 莫富传, 冯翠翠. 政府数据开放价值实现的机理: 基于系统动力学的分析[J]. 情报理论与实践, 2022, 45(5): 75-83.
- [47] 王立华, 苗婷. 农民对电子政务服务的采纳意愿及影响因素的 实证分析: 基于陕西省西安市农民的调查数据[J]. 当代经济科 学, 2012, 34(6): 109-114, 126.
- [48] HARRISON T M, SAYOGO D S. Transparency, participation,

- and accountability practices in open government: a comparative study[J]. Government Information Quarterly, 2014, 31 (4): 513-525.
- [49] 吴金鹏, 韩啸. 开放政府数据何以成功?: 生态关系理论视角的 跨国实证研究[J]. 图书馆论坛, 2020, 40(8): 87-95.
- [50] DING H E. What kinds of countries have better innovation performance? a country-level fsQCA and NCA study[J]. Journal of Innovation &Knowledge, 2022, 7 (4): 100215.
- [51] 郑方辉, 胡映佳, 黄蓝. 数字乡村治理绩效中价值理性与技术理性如何统一:以H县党员干部"直联群众"数字化为例[J]. 中国行政管理, 2022 (12): 49-56.
- [52] MISANGYI V F, ACHARYA A G. Substitutes or complements? a configurational examination of corporate governance mechanisms[J]. The Academy of Management Journal, 2014, 57 (6): 1681-1705.

作者简介

汤志伟, 男, 博士, 教授, 研究方向: 数字治理。 罗意, 女, 博士研究生, 通信作者, 研究方向: 数字治理, E-mail: luoyiuestc@163.com。

Data Utilization in Open Government Data Ecosystem: Bottlenecks and Configurations

TANG ZhiWei^{1,2} LUO Yi³

- (1. School of Public Administration, University of Electronic Science and Technology of China, Chengdu 611731, P. R. China;
- 2. Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518110, P. R. China;
- 3. School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 611731, P. R. China)

Abstract: Currently, the level of open government data (OGD) utilization in China remains generally low. Revealing the necessary conditions and configurations for achieving a high level of OGD utilization is crucial for unlocking data value and fostering digital economic development. Based on the theory of the innovation ecosystem, this paper constructs an analysis framework for the OGD ecosystem. It combines the necessary condition analysis and qualitative comparative analysis methods to examine the necessary conditions and configurations for achieving high level OGD utilization across 27 provincial regions in China. It is found that providers, OGD platforms, and technical environment are the necessary conditions for generating high level OGD utilization, forming bottlenecks at the levels of 74.70%, 77.00%, and 45.00%, respectively. Three types of configurations leading to high level OGD utilization are found: supply type, supply-demand type, and comprehensive type. To enhance OGD utilization, local governments should address the leaky bucket issue, adopt a combination blow, and focus on the driving effect of the demand side.

Keywords: Open Government Data; Open Government Data Utilization; Necessary Condition Analysis; Qualitative Comparative Analysis; Ecosystem Theory

(责任编辑: 王玮)