

政府优先开放高价值数据集多维解析*

闫志开^{1,2}

(1. 江苏师范大学法学院, 徐州 221116; 2. 东南大学法学院, 南京 211189)

摘要: 在对开放数据运动探索与反思基础上, 包括我国在内的不少国家提出优先开放高价值数据集。高价值数据集自产生以来经历了横向扩张与纵向深化的发展过程, 并呈现出规制依据从政策向法律递进、优先开放策略取决于各国国情、指涉范围呈动态扩张趋势等演进特点。从本体论角度解析, 可将高价值数据集作为复合型概念, 定义为经政府和用户等利益攸关方共同参与识别、采用非专有开放共享格式、在经济社会等领域具有较高再利用价值的数据集合。从认识论角度解析, 政府、用户等利益攸关方需要通过参与多轮对话来识别高价值数据集, 对不同主体的价值分歧予以调适。从方法论角度解析, 政府开放高价值数据集在开放内容、开放形式及开放程序方面均须符合特有要求。

关键词: 高价值数据集; 公共数据; 数据开放; 数据价值; 数据集

中图分类号: G2; G35 **DOI:** 10.3772/j.issn.1673-2286.2023.04.004

引文格式: 闫志开. 政府优先开放高价值数据集多维解析[J]. 数字图书馆论坛, 2023 (4) : 29-38.

开放数据运动提倡以机器可读、开放许可的形式在线、批量、免费提供数据^[1]。它源自对开放政府数据的现实需求, 并采取了开放源码定义的技术标准。由于开放数据具有多重积极意义, 其自产生便迅即传播开来。不过, 开放数据在实践中也遇到了一些新的问题。①开放数据的范围受限。各国政府作为开放数据运动的主导者, 往往更为关注增加开放数据的数量而非改进数据的质量^[2]; 发展中国家数据供给与需求之间的不匹配现象更为明显, 开放的数据常是最容易被开放的而非最需要的^[3]。②开放数据的互操作性不佳。开放数据在来源和形式上常是异构的, 阻碍了有意义的检索、分析和整合, 限制了数据价值发挥^[4]。③开放数据在提高政府透明度和改进问责制方面未达到预期目标。实践表明, 政府开放数据本身不足以推动开放数据的透明治理, 其改善公共事务治理的效果引发质疑^[5]。为了确保开放数据能够真正创造价值, 政府面临的紧迫任务是识别并开放对公众具有高价值的数据^[2]。因此, 在对开放数据运动持续探索与深入反思基础上, 不少国家

提出优先开放“高价值数据集”的计划, 并与公共部门信息再利用等政策有机结合, 取得了较好成果。

2022年1月, 国务院办公厅印发《要素市场化配置综合改革试点总体方案》, 提出“优先推进企业登记监管、卫生健康、交通运输、气象等高价值数据集向社会开放”^[6]。这为我国高价值数据集开放提供了动力与契机。但是, 一年多来, 这项工作进展缓慢。其直接原因在于, 高价值数据集开放的顶层设计较为抽象, 相关工作部署操作性不强。根本原因则在于, 我国相关主体对高价值数据集这一新的概念认知不够。从本体论角度而言, 什么是高价值数据集? 从认识论角度而言, 如何识别高价值数据集? 从方法论角度而言, 如何实施高价值数据集的开放工作? 在这些基础问题尚未厘清的情况下, 政府优先开放高价值数据集将可能面临困难。鉴于此, 本研究在梳理国外开放高价值数据集实践历程的基础上, 逐层分析高价值数据集的基本概念、识别标准、实施规则等, 为我国政府和用户等相关主体共同推进高价值数据集开放实践提供参考。

收稿日期: 2023-02-19

*本研究得到江苏师范大学教师科研支持项目“大数据时代网络信息犯罪问题比较研究”(编号: 19XFRS007)资助。

1 政府开放高价值数据集的历史进程

从理论上讲,开放的数据越多越好。但鉴于政府机构人力资源有限,应优先考虑开放高价值数据集。这一认识也是在开放数据运动的发展过程中,针对信息技术的发展与经济社会的需求而逐渐演进的。

1.1 高价值数据集的发展阶段

根据对国外相关文献的梳理,以相关的典型文件出台为标志,可将高价值数据集的发展历程大致分为3个阶段。

第一阶段:概念提出。信息技术发展与信息自由法律共同促使美国开放政府数据。2009年1月,时任美国总统奥巴马在上任伊始即发布《关于透明度和开放政府的备忘录》(*Memorandum on Transparency and Open Government*)^[7],要求相关部门制定开放政府指令的具体建议。根据随后发布的《开放政府指令》(*Open Government Directive*)^[8],各政府机构应在45天内确定并以开放格式在线发布至少3个高价值数据集(*high-value datasets*);每个开放政府网页还应包含供公众对已发布信息的质量进行反馈和评估的指引规定,以确定哪些信息应优先发布,且各机构应定期对公众意见作出回应。该指令下达之后,美国政府的多个机构在网站上发布了近300个高价值数据集。但此后不久,政策发生调整,特别是2012年《数字政府战略》(*Digital Government Strategy*)^[9]的发布意味着政府关注重心转向数据的促进创业功能。“透明度”仅在2009年被政策提及,不再是2012年后的主要政策目标,从而导致美国开放高价值数据集计划没有沿着预定思路发展。至今,美国许多州虽有开放数据门户,但没有相应政策来推动数据有效使用。而且由于操作时间较短,实践不够完善,这些决策部署也遭到了质疑,典型观点有:“这些数据是对政府来说高价值,还是对美国人民来说高价值?”^[10]“这些门户网站往往营养不良、不太实用,对用户反应也不太积极”^[11]。

第二阶段:横向扩展。开放政府数据是促进信息相关领域经济发展的重要资源,开放数据形式不同导致各国之间互操作性不足、共享困难和数据重复现象,也影响数据的跨境流动。针对这些新的问题,一些西方国家认为有必要采取集体行动以破除障碍。2013年八国集团《开放数据宪章》(*G8 Open Data Charter*)^[12]提出一

套原则以确保开放数据的效果,其中“质量和数量”原则认为需要相互协商以确定优先发布或改进哪些数据,“所有人都可使用”原则强调尽可能以开放格式发布数据。为贯彻这些原则,该宪章提出高价值数据集发布行动计划,不仅提出“默认公开”和“保质保量”两大原则,还确立高价值数据集发布的“三步走”战略:首先,由各成员国共同提供有关国家统计、国家地图、国家选举和国家预算的关键数据集,并细化数据集的颗粒度、提高可访问性;其次,增加有关国家关键职能的数据集供应;最后,根据各成员国的国家框架发布其余类别下的数据集。与美国相比,八国集团的这一计划在操作规则方面更为具体可行,因此发布后即在各成员国得到较好实施。

第三阶段:纵向深化。公共部门信息是欧洲最大的单一信息来源,其溢出价值甚至比行业本身价值还要大^[13]。欧盟《公共部门信息再利用指令(修订版)》

(*Amendment of Directive 2003/98/EC on the Re-use of Public Sector Information*)^[14]在“数据集标准”一节中提出,应为开放数据倡议设定目标、优先事项和范围,重点放在能够识别高价值数据集的标准以及反映各利益攸关方的观点和需求上。然而,由于欧盟成员国在数据集领域具有不同的价值立场、法律体系及现实需求,该指令规则考虑不周、设计不合理,导致该指令在欧盟成员国实施效果参差不齐。2014年欧盟委员会发布了《关于推荐标准许可证、数据集和文件再利用收费的指南》(*Guidelines on Recommended Standard Licences, Datasets and Charging for the Reuse of Documents*)^[15],强调对快速发布若干高价值数据集的必要性达成的普遍共识,并为以后在指令中定义高价值数据集奠定了基础。欧盟《开放数据和公共部门信息再利用指令》[*Directive (EU) 2019/1024 on Open Data and the Reuse of Public-Sector Information*]^[16]对高价值数据集作出新的规定,并精心构建多方互动的运作机制以强力推进。它将原来的《公共部门信息再利用指令(修订版)》和对其的修正合并为一个法案,并明确该指令不适用于受第三方版权或相关权利保护的内容,且必须遵守欧盟法律框架和国际义务。

1.2 高价值数据集的演进特点

对上述高价值数据集的发展过程加以归纳分析,可发现高价值数据集的演进呈现以下特点。

第一，高价值数据集的规制依据从政策向法律递进。第一阶段，在美国政府提出高价值数据集概念时，这一概念还只是开放政府数据政策中新增加的一个普通要素，其优先开放地位尚不牢固，政府备忘录还强调该政策不会改变现有法律对数据开放利益攸关方的权利义务配置。第二阶段，八国集团认识到高价值数据集的多重价值，又将其确立为解决数据开放问题的优先策略，不仅使这一概念的应用范围迅速扩张，也使其操作规范更加完备，实现了从政策依据到法律依据的渗透。正如八国集团《开放数据宪章》所要求的，各成员国应在国家政治和法律框架内开展工作。第三阶段，欧盟基于实践反思再次确立高价值数据集的优先地位，并发布多个法律文件以不断调整规制框架，表明其意识到高价值数据集涉及政府权力与公民权利分配等深层问题，因此将确认高价值数据集这一敏感问题上升到授权法案的高度。

第二，高价值数据集的优先开放策略取决于各国国情。尽管多个国家已经吸收采纳高价值数据集的概念，但不同国家的优先开放策略并不一致。如，美国开放高价值数据集策略比较自由，欧盟则强调统一部署。这不仅源于不同国家对网络数据的规制传统，还与各国数据经济发展情况相关。正如《欧洲数据战略》（*A European Strategy for Data*）^[17]指出，美国少数大型科技公司掌握着世界上大部分的数据，这可能会降低数据驱动型企业在欧盟涌现、成长和创新的动力，欧盟研究制订开放高价值数据集法律时也要考虑照顾欧洲中小企业的特殊需求。八国集团《开放数据宪章》也要求各成员国根据国家行动计划规定的最佳技术做法和时间表予以实施。

第三，高价值数据集的指涉范围有动态扩张的趋势。虽然高价值数据集产生于开放政府数据运动浪潮，但这并不意味着它必须是政府数据。近年来，高价值数据集的边界有从政府数据集向外扩张的趋势，如：英国的Open Banking和Open Active等倡议开放由私营部门持有的关键数据集^[18]；美国谷歌云平台免费提供部分高价值数据集，其中USA name数据集就包含了1879年以来的美国所有社会保障名称申请^[19]。可见，高价值数据集的范围可以是本地的、本国的或全球的，其来源可以是政府部门也可以是企业。这种动向既对现有的高价值数据集规制框架带来挑战，也蕴含着高价值数据集的发展趋势。

2 高价值数据集的本体论解析

概念是人类认识世界的重要工具，也是构造理论的基本要素。虽然部分国家和国际组织在政策文件中多次提出“高价值数据集”，但由于缺乏相应的理论研究，该概念内涵不清、外延不明，影响了实践进展，因此有必要从概念建构角度对高价值数据集加以剖析。

2.1 高价值数据集概念建构的关键问题

随着数据开放运动的深入发展，数据的类型日益复杂、属性愈发丰富，原有的数据概念已经不能满足多样的实践需求。从开放数据实践效果评估来看，将数据的潜在影响与实际影响自动联系起来会导致对开放数据效果的认识误区^[20]。因此，本研究从以下3个角度对高价值数据集概念予以阐释。

2.1.1 高价值数据集概念的复合构词形态

高价值数据集来源于“价值”（value）与“数据”（data）2个自由词素和“高”（high）与“集”（set）2个粘附词素。概念间的组合会导致突显特征的出现^[21]，高价值数据集概念的突显特征就源于其复合性。从形式上看，政府开放的数据已经转向数据集；从实质上看，政府关注的重心已经从数据的价值转向高价值。政府收集的各类数据来源不同、形式有异、分布零散，主要呈现异质性，但也有潜在的关联，有待整合和优化以释放数据价值。因此，从数据到数据集的形式转化会产生大小不等的增量价值，与“高价值”建立因果关系。由此，高价值数据集概念的各要素之间形成了互相依附、相互影响的关系，该复合构词概念具有合理性基础。

2.1.2 高价值数据集概念的实用主义色彩

高价值数据集概念是伴随信息技术的推广应用而在开放数据实践中被主观建构起来的，其目的是回应与解释开放数据面临的新问题，破解原有话语解释力不足带来的实践和理论困境。它既有数据集集合化这一形式要求，也蕴含着对数据价值的认定要求，因此更具客观与主观相互验证的可操作性。

2.1.3 高价值数据集概念的价值规范意涵

高价值数据集概念虽然从表面上看是对同类事物的描述,但其实具有明显的价值倾向。在信息技术发展初期,人们一度认为网络或数据的概念是价值无涉或价值中立的,然而,随着信息技术的更新换代,技术对整个个人类行为、意识和社会结构产生了日益明显而复杂的影响。高价值数据集虽然是包含多种元素的组合词汇,但其核心在于“高价值”,而价值又取决于数据主体的主观需求、特殊情感和偏好差异。从价值方面对数据加以分类和筛选,正是高价值数据集概念的要义所在,也是该概念落地实践的难点所在。

2.2 高价值数据集的内涵

内涵是概念反映的思维对象所特有的属性。高价值数据集是一种数据集,它与其他数据集的区别在于“高价值”。对数据集的“高价值”,可从内容与形式2个维度进行分析。

2.2.1 从内容维度分析,“高价值”涵盖数据开放领域的多种价值

随着大数据、云计算等新兴技术的推广应用,数据的价值体现在多个领域。政府数据开放的最初目标是发挥数据的政治价值,如美国提出开放政府数据的价值为透明、参与和合作^[8],但在推进过程中数据集的多种价值逐渐得到实务界和理论界的认识。一方面,多国政府和国际组织的官方政策文件体现了实务界对开放数据的价值追求:八国集团侧重于数据的再利用潜力和对目标受众等的贡献程度^[12];欧盟强调高价值数据集应具有产生重大社会、经济或环境效益的潜力,以及与其他数据集结合的潜力等^[22];联合国也认为开放数据能够在促进私营部门发展、创造就业机会、促进经济增长及社会治理方面发挥多种价值^[23]。另一方面,学界普遍认为数据开放在政治、经济、社会、技术等方面产生价值。这些价值之间并非总是正相关关系,有时也会有竞争或冲突,比如开放数据可以在某种方式上促进政治平等,同时又会导致其他不平等^[24]。

尽管各界对于开放数据集价值的认识不断丰富,但忽略了数据集开放过程如何保障安全的问题。信息技术诞生以来,网络上的技术黑客、政治上的国家博

弈、经济上的间谍行为等都对数据安全构成了持续的威胁^[25]。数据开放是政府数据开放共享、再利用的前提。政府数据开放不仅要在政治、经济、社会等方面达成积极价值,也应当通过技术预防个人隐私、商业秘密或国家秘密泄露,预防数据“规模致敏”,以实现数据的安全价值。与传统数据相比,新型数据面临多源大量数据聚合和分析可能带来的安全风险^[26]。简言之,开放高价值数据集的价值并非单一的政治价值,而是多个相关领域的复合价值,这些价值在政府开放数据语境中有着优先位阶顺序:在保障数据安全价值的前提下,基本实现数据的政治价值,进而最大限度地发挥数据在经济、社会等领域的其他价值。

2.2.2 从形式维度分析,“高价值”源自从数据到数据集的转化

在大数据时代,信息技术的发展催生海量的数据。然而,数据量并不能直接转化为数据价值。原始的、零散的数据来源丰富、数量庞大、种类繁多,这些数据的不断积累导致越来越高的“信息熵”,影响了数据潜在价值的实现,为此需要采用数据集的形式对杂乱无章的数据加以整合。

数据集是数据的有序集合,美国《开放数据政策备忘录》[Open Data Policy (M-13-13)]^[27]将其定义为以表格或非表格形式呈现的数据集合,我国《信息与文献 参考文献著录规则》^[28]将其界定为一种由数据组成的集合。数据集最常见的格式是电子表格或CSV格式,但有时数据集也可能是包含多个数据表和相关数据的文件夹。

数据集是帮助人们将所有粒度数据转化为易于理解、直观且有用的决策信息的有力工具。一般而言,高价值数据集是呈介于单个的原始数据与统计数据之间的形态:既有原始数据的真实性与丰富性,从而保证数据的质量,又有统计数据的可操作性与可机读性,以便公众深入分析;既符合开放数据的基本格式要求,又有特定的数据集格式要求。

国外研究指出,数据价值与整合度呈抛物线关系,20%的整合度可以达到80%的效用价值^[29]。通过数据集对零散数据进行控制和筛选,使之从无序的数据转为有意义的数据集,是“信息熵”减少的过程,也是数据价值提升的过程。简言之,数据集能够实现高价值的依据在于,数据集是数据聚合的成果,也是数据价值的

载体。从数据到数据集的形式整合转化融入政府部门及其工作人员的“新的活劳动”，成为开放数据实现高价值的必要条件，使数据实现了从内在价值到现实价值的跨越。

2.3 高价值数据集的外延

外延是指具有概念反映的本质属性的一切对象。高价值数据集的外延即高价值数据集的分类。欧美国家对高价值数据集的分类已有多年探索，对我国研究者更好地理解高价值数据集具有启发意义。

2.3.1 世界范围内对高价值数据集的分类探索

当前，世界范围内对高价值数据集的分类主要有3种模式。①美国对于高价值数据集的分类没有统一规划和硬性要求，总体由各部门或各州自行决定是否发布及如何发布高价值数据集。如，在美国联邦政府门户网站上可发现338 032个数据集，依“话题”分为农业、气象、能源、地方政府、健康等7类^[30]。在美国各州政府官网上发布的高价值数据集亦各有特色，如：得克萨斯州每月定期发布高价值数据集，主题聚焦于对现有囚犯的人口统计信息^[31]；明尼苏达州将数据集分为人口普查、教育、环境等9类^[32]。②八国集团于2013年签署《开放数据宪章》，提出高价值数据发布的行动计划，其中列出的高价值数据集分为公司、地球观测、教育、健康、统计等14类。③欧盟于2019年发布《开放数据和公共部门信息再利用指令》，其附件包括一份初步的高价值数据集列表，确认的高价值数据集分为地理空间数据、气象资料、统计等6类。

2.3.2 对高价值数据集类型的比较分析

从上述3种分类模式可见，高价值数据集的类型复杂多样，既有共同点，又有不同点。

(1) 共同点。①上述国家或国际组织均采用了类型化思维。类型化不是对现实的描述，但旨在为这种描述提供明确的表达手段^[33]。对于高价值数据集这样的综合性概念，需要借助典型形态的类型化思维来理解和把握。这些类型划分依据虽然是对原有开放数据所在领域划分的延续，但对解释新的概念具有较好的启发与说明作用。②上述国家或国际组织列出的高价

值数据集指向或涵盖的范围有较多共同之处，这体现了对于开放数据最优先事项的共识，如：气象资料被美国、八国集团和欧盟列入高价值数据集类型；地球观测、统计、公司、交通等数据则被八国集团和欧盟共同设定为高价值数据集类型。其原因在于，上述国家或组织具有一定程度的交叉重合性，如：美国既是开放高价值数据集的首倡者，又是八国集团的主导者；法、德等国同时是八国集团的成员和欧盟成员国，从而导致相关政策具有延续性。另外，上述国家或组织均为西方发达国家，因此在数字经济发展和数字开放领域具有相似的立场，在开放高价值数据集方面可互相借鉴。

(2) 不同点。①从类型数量上看，上述国家或组织界定的高价值数据集类型数量不同。八国集团对高价值数据集的分类较为具体，种类较多，覆盖范围相对广泛；美国和欧盟界定的类型数量相对较少，因为美国较早开放高价值数据集，因此分类较为粗放，而欧盟受到其特有的组织结构和价值定位影响，对开放高价值数据集非常谨慎，意在初步开放少数数据集，后续再逐步扩充。②从类型名称上看，相似领域的数据集名称并不一致。美国将能源单列一类，八国集团将环境与能源归入一类，而欧盟则将环境与地球观测归入一类。即使在名称相同的数据集类型下，涵盖的具体数据集也有不同，其原因在于各国或国际组织对高价值数据集的内涵理解不一，且类型命名也会受到本国或本组织的制度语境及主体偏好等因素影响。

综上，在考察高价值数据集的概念属性基础上，基于对其内涵与外延的系统剖析，可以将高价值数据集定义为经政府和用户等利益攸关方共同参与识别、采用非专有的开放共享格式、在经济社会等领域具有较高再利用价值的数据集合。

3 高价值数据集的认识论解析

价值属于关系范畴，表示客体属性与主体需要之间的效用关系。数据集的价值不是由创造它所需的资源量确定的，而是根据其场景和用户而变化的，主要取决于它对主体的有用程度。开放数据工作涉及多个利益攸关方，他们对于数据开放有不同的关注点，受到的价值损益也是不同的。在开放高价值数据集时，涉及的主体主要有作为数据供给方的政府和作为数据需求方的用户（主要包括企业和个人）。因此，需要基于主体视角对数据集的“高价值”展开分析。

3.1 对数据集“高价值”的认识分歧

数据开放运动是政府为了提高透明度而提出的重要举措。美国政府虽然在启动开放数据运动时即提出开放高价值数据集，但由于理论认识与配套制度不足，基本上均由政府部门及其工作人员来识别高价值数据集。因此，早期对数据集价值的识别不可避免地以政府价值为评判基准。然而，在“政府即平台”场景下，政府不再单纯以超然利益关系治理者的角色出现，而是以利益攸关方的主体身份登场，它同时也是最大的个人信息收集、处理、储存和利用者^[34]。在政府既是“运动员”又是“裁判员”的情况下，数据开放预期目标的实现具有不确定性。

针对政府数据开放实践中出现的效果偏差问题，多国相应调整政策导向，更加着眼于用户需求，这固然是对原有数据开放的价值认知的纠偏，但是政策落地亦面临多重困境。首先，用户包括各类人群，他们之间的数据需求与价值追求不尽相同。毕竟企业用户比个人用户具有更强的数据能力，不同企业之间的数据挖掘分析能力也有差异，而个人信息素养的差异也意味着政府数据会引发新的不平等。其次，实施高价值数据集开放政策的操作主体是政府而非用户，政府是否有能力准确理解用户的数据需求值得考量。最后，数据开放的实践也与预期目标有差距。常见的情形是，公共部门缺乏对其责任和数据集开放价值的认识，同时企业和个人用户对其权利和数据集的可用性知之甚少。

3.2 对数据集价值认识分歧的调适

无论是政府部门还是用户个体，对开放数据的价值评判都有其独特的利益考量和认知思维。根据伯林的价值多元论，人类的目标是多样的：一方面，不同主体追求的价值之间存在差异，而且相互之间往往处于永久的敌对状态；另一方面，不同的价值观并无高低优劣之分，因此无法给出确定的等级排序^[35]。鉴于在数据开放中政府与用户均有自己的价值偏好，高价值数据集的价值基准应当是超越二者价值观局限的公共价值，这也与政府数据作为公共物品的定位契合。然而，即使获得实质合法性，管理者在决策时还经常面临巨大的不确定性，即应由谁来决定开放数据价值的高低？公共物品所满足需求的不可分割性和承载利益的公共性决定了其供需必须依赖集体行动，而任何集体行动本质

上都是人与人之间的合作^[36]，为此，有必要构建以交往行为为基础、以公共价值为导向的高价值数据集识别机制。

在既有实践中，鉴于不同主体对于数据集的价值认识不同，美国、八国集团和欧盟均确立了相对固定的机制，吸纳多主体共同参与，以使最终确定的高价值数据集尽可能实现公共价值。

(1) 美国将对优先开放高价值数据集的识别机制纳入《开放政府数据法》(*Open Government Data Act*)^[37]的规制框架。主要考虑以下因素：①由政府部门确定数据集的成熟度，且数据集应有足够高的质量；②各机构使用学习议程(learning agenda)来确定本机构关键问题及所需数据，以明确优先顺序；③政府评估和平衡利益攸关方的需求；④确保数据集开放合法合规；⑤根据利益攸关方的反馈，定期评估和处理公众对数据价值等方面的认知，以改进数据集开放计划。

(2) 八国集团对高价值数据集的认定遵循《开放数据宪章》。其主要环节有：①政府部门进行初步判断，主要结合各自国家的行动计划，判断这些领域的数据是否可以促进政治民主或可以鼓励数据的创新再利用；②各成员国的政府工作人员就政府提出的初步方案咨询技术专家后确立“最佳实践”；③八国集团结合各成员国的实践，在考虑新兴技术解决方案并总结实际操作经验后对初步认定的高价值数据集进行修订。

(3) 欧盟对于高价值数据集类型的识别和增减工作都按照法定程序进行。①《开放数据和公共部门信息再利用指令》对“高价值数据集”作出初步分类，并在附件中列出清单。该清单主要借鉴全球开放数据索引以及八国集团《开放数据宪章》的技术附件标准。②欧盟委员会组织专家磋商。为了确保平等参与，欧洲议会和理事会与成员国的专家同时收到所有文件，并可参加专家组的会议，共同处理数据集识别工作。③根据《开放数据和公共部门信息再利用指令》，欧盟委员会通过专门立法，在原有的高价值数据集的主题类别列表中添加更多种类。

综合上述美国、八国集团和欧盟处理高价值数据集认识分歧的调适机制，可见一般先由政府部门根据政府职能关联度、数据集成熟度等因素初步制定高价值数据集清单，接着由技术专家根据数据集的合法性、合理性等因素提出修改建议，再由用户根据使用需求发表意见，最后由政府部门基于数据集的再利用价值对高价值数据集清单予以调整，这是一个多主体参与

识别高价值数据集的过程。当然，由于数据集和用户需求均处于变动之中，某一轮高价值数据集识别过程的结束并非彻底终结，而是下一轮高价值数据集识别的起点，从而形成基于实践的多轮对话循环，使开放高价值数据集不断完善。

4 高价值数据集的方法论解析

高价值数据集受制于一套单独的规则^[38]。政府优先开放高价值数据集是对传统开放数据运动的反思与改进，它界定了一种新兴技术的社会关系，也形成了一个具体领域的理论框架，为解决实践问题提供了新的思路。由于高价值数据集不同于数据的概念特质，它在开放内容、开放形式及开放程序等方面均须符合特有要求。

4.1 开放内容规则要求

开放高价值数据集有着多种潜在的积极价值，这些价值会对政府部门及用户形成强烈的吸引力，然而也容易出现潜在的安全风险。为了保障数据发挥积极效果，应当从一开始就参考“设计隐私”理念，将个人隐私、商业秘密和国家安全等因素纳入数据开放的流程和系统，确保数据安全提前嵌入数据开放的组织体系。数据去识别化技术是平衡隐私和透明度目标的重要工具，以“一刀切”的方式将涉及个人或企业的所有信息匿名化当然能最大限度地降低数据安全风险，但这也也会影响高价值数据集潜在价值的发挥。在现代社会，对个人信息的保护不能限制信息的前端传播，而应立足规制信息的中端使用^[39]，这就要求对开放高价值数据集的内容范围有更高水平的设计。

具体来说，开放高价值数据集的内容禁止范围划定了高价值数据集开放的边界，直接关系到数据安全，也决定了数据价值的实现程度。从正面看，与社会和经济重要利益相关的数据集在原则上都应开放；从反面看，除了未被识别为高价值的数据集外，凡是违反法律规定的数据集也不宜开放。这种共识也是数据开放“默认公开，例外封闭”原则在高价值数据集领域的体现。①基于“默认公开”原则，至少需要确定潜在数据集内容的初步评估标准。可以通过设定相关问题，根据回答内容来判断，如相关数据是否有助于改善普通人的生活、提高各利益攸关方的关注度、增加公众对政府运作的了解

等。②针对“例外封闭”的情况，需要确立相关规则，将符合封闭标准的潜在数据集排除。开放数据机构在创作新数据集之后与发布之前的时间段里，应依据相关的信息技术安全政策、隐私政策和管理政策审核数据集，对于基本符合发布条件的也应核验是否已采取去识别化措施。

4.2 开放形式规则要求

在开放数据集之前，数据资源碎片化地散落在不同来源之中，数据往往存在形态多样、质量参差、价值稀疏等问题。研究表明，开放数据访问困难以及政府未能定期更新阻碍了组织和民众及时获取公共部门开放数据^[40]。尽管用户意识到开放数据的有用性，但其倾向于避免使用复杂的产品或服务^[41]。形式方面的因素导致数据内容与数据价值之间产生了差距。从数据价值链来说，原始数据的价值较低，数据集不仅在形式上整合了异质性的数据，还创造了不同信息之间的诸种关系，从而显著增加了数据的价值。然而，数据集的推广应用容易遮蔽非结构化的原始数据，过于整合的数据集格式会显著增加数据集开放成本，也可能消减数据利用的多种可能性。

为了高效且公平地促进高价值数据集开放，对于形式应当有基本的统一要求，但也不宜设定过高的门槛。结合数据再利用的发展需要，应以非专有、机器可读的格式提供数据。早在2010年，Berners-Lee^[42]就为关联开放数据提出了五星级标准，该标准已被用于对数据集的排序。高价值数据集至少应以三星级标准发布^[43]，才能做到以非专有的开放格式促进数据再利用。

4.3 开放程序规则要求

就数据开放的利益攸关方而言，对于高价值数据集的认知因涉及主观判断而有差异，但也存在部分共同点。公共政策中平衡式的价值冲突解决方案难以奏效，因为政策制定者很难计算每一个价值的重要性^[44]，应当诉诸程序机制以确认公共价值的合理路径。公共价值的确定与实现不是公共部门的单向责任，而是基于对话和参与的双向互动过程^[45]。使各种不同意见充分讨论和表达是找到隐含在所有人心中的共同价值倾向的最佳手段。在这个实践意义上，价值分歧与争议本身就是价值建构过程的一部分^[46]。

高价值数据集开放涉及多个环节,每一环节都涉及不同的利益攸关方,因此应当分阶段明确操作程序规范。①初步认定环节。首先,制定数据集提案,其来源可以是相关数据开放机构或某个公民或企业。其次,识别高价值数据集。为减少失误,可安排2个主体分别对数据集进行初审。最后,确定该数据集的优先级,这需要在具有最高价值(必要性)和最充分(可行性)的数据集之间做出平衡。②审核发布环节。首先,应当以可接受的格式编制数据集;其次,要审查相关的数据集是否合法;最后,由具体人员将数据集发布到公共网站。虽然每个国家或组织对开放数据集的处理方式不完全相同,但目前很多国家已发布开放数据工具包(open data toolkit)以提供简要指南。③反馈维护环节。数据开放机构应按照数据集清单记录的频率更新数据集,以保持数据的最新状态。此外,用户反馈也是数据更新的重要依据。对于开放政府数据,用户可以通过网站的在线评论或填写反馈表的形式提出建议,机构收到反馈意见后应当及时回复,并根据需要对数据集予以更新。

5 结语

政府优先开放高价值数据集的主要目标是确保具有最高社会经济潜力的公共数据能够在受到最小法律和技术限制的情况下免费再利用^[47]。然而,由于政府资源紧张现象持续存在、民众数据需求不断变化,开放高价值数据集将是一个长期任务。本研究基于对国外高价值数据集开放历程的梳理,提炼实践中有待澄清的普遍问题,从本体论、认识论、方法论3个维度对高价值数据集作了较为深入的系统分析。首先,从本体论的角度界定高价值数据集的概念,意在为推进高价值数据集开放提供逻辑基点;其次,从认识论角度分析如何识别高价值数据集,意在指出相关各方主体的价值分歧及其调适机制;最后,从方法论角度探讨如何推进高价值数据集工作,意在明确相关的内容规则、形式规则和程序规则,为当下我国开放高价值数据集实践提供具体可行的参考意见。

值得注意的是,高价值数据集是开放数据的升级,呈现出更为多元、复杂的形态。政府是数据开放的主导者,用户也是数据开放生态系统中的利益攸关方。与传统的开放数据相比,高价值数据集这一概念的重要意义正在于将用户的价值吸纳进来,从而要求用户参与高

价值数据集的识别过程。易言之,高价值数据集的确定是开放主体(政府)与其服务对象(用户)等利益攸关方共同参与、多轮对话的结果。本研究虽然基于政府数据开放的视角开展,但对用户参与高价值数据集的识别也具有参考意义。

政府优先开放高价值数据集是一项新的系统工程,涉及各利益攸关方的多种复杂行为。本研究虽然从理论层面论证了政府和用户等主体参与高价值数据集开放的必要性与可行性,但仍是一种整体性的构想,尚待进一步实践检验,在具体实施环节更是有待深入的研究。比如,在开放高价值数据集的启动环节,需要由政府部门工作人员识别数据集价值高低并初步确立开放内容清单,其中如何把控自由裁量权、如何合理确定识别标准需要专门研究。再如,开放高价值数据集是不断调整优化的动态过程,只有各利益攸关方真正参与对话才能促进数据集开放的良性循环,因此,如何优化数据开放平台的功能设置,建立各利益攸关方的激励相容机制,尽力实现数据集公共价值的最大化,也是将来值得关注和研讨的重要话题。

参考文献

- [1] World Wide Web Foundation. What is open data & why is it important? [EB/OL]. [2022-11-12]. <https://opendatabarometer.org/open-data/>.
- [2] BARBARA U. Open government data: towards empirical analysis of open government data initiatives [EB/OL]. [2022-11-12]. <https://www.oecd-ilibrary.org/docserver/5k46bj4f03s7-en.pdf?expires=1643110165&id=id&accname=guest&checksum=02B497F98E279F8481DC37A95B9FAA97>.
- [3] W3f. Exploring the emerging impacts of open data in developing countries [EB/OL]. [2022-11-12]. <http://www.opendataresearch.org/emerging-impacts>.
- [4] 陈水湘. 国外开放政府数据研究进展 [J]. 图书馆, 2016 (12): 86-92.
- [5] SIMON C. A brief history of open data [EB/OL]. [2022-11-12]. <http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data>.
- [6] 国务院办公厅. 要素市场化配置综合改革试点总体方案 [EB/OL]. [2022-11-12]. http://www.gov.cn/zhengce/content/2022-01/06/content_5666681.htm.
- [7] OBAMA B. Memorandum on transparency and open

- government [EB/OL]. [2022-11-12]. <https://www.presidency.ucsb.edu/documents/memorandum-transparency-and-open-government>.
- [8] Orszag P R. Open government directive [EB/OL]. [2022-11-12]. <https://obamawhitehouse.archives.gov/open/documents/open-government-directive>.
- [9] Whitehouse. Digital government strategy [EB/OL]. [2022-11-12]. <https://digital.gov/resources/2012-digital-government-strategy/#:~:text=The%202012%20Digital%20Government%20Strategy%20mission%20drives%20agencies%2C,government%20that%20works%20better%20for%20the%20American%20people>.
- [10] HOOVER J N. High value government data made public [EB/OL]. [2022-11-12]. <https://www.informationweek.com/government/-high-value-government-data-made-public>.
- [11] STEFAAN G V, ANDREW Y, ANDREW J Z, et al. The emergence of a third wave of open data [EB/OL]. <https://opendatapolicylab.org/images/odpl/third-wave-of-opendata.pdf>.
- [12] Open Data Charter and Technical Annex [EB/OL]. [2022-11-12]. <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>.
- [13] Digital Agenda Assembly. Report from the workshop “Open Data and Re-use of Public Sector Information” [EB/OL]. [2022-11-12]. http://ec.europa.eu/information_society/events/cf/daa11/item-display.cfm?id=5963,#daa11psi.
- [14] European Commission. Directive 2013/37-Amendment of Directive 2003/98/EC on the Re-use of Public Sector Information [EB/OL]. [2022-11-12]. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013L0037>.
- [15] European Commission. Guidelines on Recommended Standard Licences, Datasets and Charging for the Reuse of Documents [EB/OL]. [2022-11-12]. [https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014XC0724\(01\)](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014XC0724(01)).
- [16] European Commission. Directive (EU) 2019/1024 on Open Data and the Reuse of Public-Sector Information [EB/OL]. [2022-11-12]. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L1024>.
- [17] European Commission. A European strategy for data [EB/OL]. [2022-11-12]. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN>.
- [18] Open Data Institute. Open Data Barometer: now more than ever, we must strengthen open data initiatives across the world [EB/OL]. [2022-11-12]. <https://theodi.org/article/open-data-barometer-now-more-than-ever-we-must-strengthen-open-data-initiatives-across-the-world/>.
- [19] VIK P. 21 places to find free datasets for data science projects [EB/OL]. [2022-11-12]. <https://www.dataquest.io/blog/free-datasets-for-projects/>.
- [20] JANSSEN M, CHARALABIDIS Y, ZUIDERWIJK A. Benefits, adoption barriers and myths of open data and open government [J]. Information Systems Management, 2012, 29 (4): 258-268.
- [21] PRINZ J, CLARK A. Putting concepts to work: some thoughts for the twentyfirst century [J]. Mind and Language, 2004, 19 (1): 57-69.
- [22] DANNY L. What data counts in Europe? [EB/OL]. [2022-11-12]. <https://blog.okfn.org/2019/01/16/what-data-counts-in-europe-towards-a-public-debate-on-europes-high-value-data-and-the-psi-directive/>.
- [23] GURIN J L, PETROV O V, MANLEY L. Open data for sustainable development [EB/OL]. [2022-11-12]. <https://openknowledge.worldbank.org/server/api/core/bitstreams/82626bbb-5b1e-58b1-bc46-8bf197435c45/content>.
- [24] 布鲁斯·宾伯. 信息与美国民主: 技术在政治权力演化中的作用 [M]. 刘钢, 译. 北京: 科学出版社, 2011: 28.
- [25] 闫志开. 人工智能时代司法效率的理论逻辑与实践模式 [J]. 湖南社会科学, 2022 (2): 81-89.
- [26] 刘金瑞. 数据安全范式革新及其立法展开 [J]. 环球法律评论, 2021, 43 (1): 5-21.
- [27] OMB. Open Data Policy (M-13-13) [EB/OL]. [2022-11-12]. <https://digital.gov/resources/open-data-policy-m-13-13/>.
- [28] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 信息与文献 参考文献著录规则 [EB/OL]. [2022-11-12]. <https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?heno=7FA63E9BBA56E60471AEDAEBDE44B14C>.
- [29] GOODHUE D L, WYBO M D, KIRSCH L J. The impact of data integration on the costs and benefits of information systems [J]. MIS Quarterly, 1992, 16 (3): 293-311.
- [30] U. S. General Services Administration. Technology Transformation Service. Topics [EB/OL]. [2022-11-12]. https://catalog.data.gov/dataset?metadata_type=geospatial&vocab_category_all_limit=0&groups_limit=0.
- [31] State of Texas. High value dataset July 2021 [EB/OL]. [2022-11-12]. <https://data.texas.gov/dataset/High-Value-Dataset>

- July-2021/ymcy-45cn.
- [32] State of Minnesota. Minnesota open data [EB/OL]. [2022-11-12]. <https://www.state.mn.us/opendata/data.html>.
- [33] 马克斯·韦伯. 社会科学方法论 [M]. 杨富斌, 译. 北京: 华夏出版社, 1999: 185-186.
- [34] 张新宝. 从隐私到个人信息: 利益再衡量的理论与制度安排 [J]. 中国法学, 2015 (3): 38-59.
- [35] 以赛亚·伯林. 自由论 [M]. 胡传胜, 译. 南京: 译林出版社, 2003: 245.
- [36] 张晋武, 齐守印. 公共物品概念定义的缺陷及其重新建构 [J]. 财政研究, 2016 (8): 2-13.
- [37] Congress. Open, Public, Electronic and Necessary (OPEN) Government Data Act [EB/OL]. [2022-11-12]. <https://www.congress.gov/bill/115th-congress/house-bill/1770>.
- [38] European Commission. Open data [EB/OL]. [2022-11-12]. <https://digital-strategy.ec.europa.eu/en>.
- [39] 刘艳红. 公共空间运用大规模监控的法理逻辑及限度: 基于个人信息有序共享之视角 [J]. 法学论坛, 2020, 35 (2): 5-16.
- [40] KASSEN M. A promising phenomenon of open data: a case study of the Chicago open data project [J]. Government Information Quarterly, 2013, 30 (4): 508-513.
- [41] DAVIS F D. Perceived usefulness, perceived ease of use, and user acceptance of information technology [J]. MIS Quarterly, 1989, 13 (3): 319-340.
- [42] BERNERS-LEE T. Linked Data [EB/OL]. [2022-11-12]. <https://www.w3.org/DesignIssues/LinkedData.html>.
- [43] LOOZEN N. Best practice: dataset criteria [EB/OL]. [2023-03-01]. <https://www.w3.org/2013/share-psi/bp/dc/>.
- [44] THACHER D, REIN M. Managing value conflict in public policy [J]. Governance, 2004, 17 (4): 457-486.
- [45] 常莉, 胡晨寒. 公共价值与公共服务: 逻辑内化与现实偏离 [J]. 行政论坛, 2020, 27 (5): 39-45.
- [46] 郑少翀. 走出价值多元主义的困境 [J]. 福建论坛 (社科教育版), 2007 (4): 65-69.
- [47] MANCINO D. High-value datasets -an overview through visualisation [EB/OL]. [2022-11-12]. <https://data.europa.eu/en/publications/datastories/high-value-datasets-overview-through-visualisation>.

作者简介

闫志开, 男, 1981年生, 博士, 讲师, 研究方向: 数据治理, E-mail: yanzhikai@126.com。

Multidimensional Analysis of Government Prioritizing Opening of High Value Datasets

YAN ZhiKai^{1,2}

(1. Law School, Jiangsu Normal University, Xuzhou 221116, P. R. China; 2. Law School, Southeast University, Nanjing 211189, P. R. China)

Abstract: On the basis of exploration and reflection on the open data movement, many countries, including China, have proposed prioritizing the opening of high-value datasets. Since their inception, high-value datasets have undergone a development process of horizontal expansion and vertical deepening, and have shown evolutionary characteristics such as regulatory basis progressing from policy to law, priority opening strategy depending on national conditions, and dynamic expansion trend of reference scope. From the ontological perspective, a high-value dataset can be defined as a composite concept that is jointly identified by stakeholders such as government and users, adopts a non-proprietary open sharing format, and has high reuse value in economic and social fields. From the epistemology perspective, identifying high-value datasets requires the government, users, and other stakeholders to adjust the value differences of different subjects by participating in multiple rounds of dialogue. From the methodological perspective, opening high-value datasets by the government must meet unique requirements in terms of open content, open form, and open procedures.

Keywords: High Value Dataset; Public Data; Data Openness; Data Value; Dataset

(责任编辑: 王玮)