

基于耗散结构理论的政府数据治理研究*

付学敬, 陈 珍, 章君玉

(上海市市场监督管理局信息应用研究中心, 上海 200032)

摘 要: 针对政府数据利用中数据的可查找、可理解、可使用等典型问题, 为有效缓解高质量数据供给不平衡不充分的矛盾, 综合分析政府数据治理的理论成果与目标任务, 进一步研究政府数据治理的基础理论和相应的实践路径, 从系统科学角度探究基于耗散结构理论的政府数据治理的内在逻辑机理, 提出基于元数据管理和一体化建模的政府数据内生治理机制建设的技术方法, 以期从基础层面强化数据标准、提升数据质量、促进数据利用, 提升政府数据治理效能。

关键词: 政府数据治理; 耗散结构; 元数据管理; 一体化建模

中图分类号: N949; TP391

文献标识码: A

DOI: 10.19358/j.issn.2097-1788.2024.02.010

引用格式: 付学敬, 陈珍, 章君玉. 基于耗散结构理论的政府数据治理研究 [J]. 网络安全与数据治理, 2024, 43(2): 64-72.

Research on government data governance based on dissipative structure theory

Fu Xuejing, Chen Zhen, Zhang Junyu

(Information Application Research Center of Shanghai Municipal Administration for Market Regulation, Shanghai 200032, China)

Abstract: To address typical issues such as the availability, comprehension and usability of data in government data utilization, and to effectively alleviate the imbalance and inadequacy of high-quality data supply, this paper conducted a comprehensive analysis of theoretical findings and goals related to government data governance, and further investigated the theoretical foundation of government data governance and corresponding practical paths, explored the inherent logical mechanism of government data governance based on dissipative structure theory from the perspective of systems science, and proposed technical methods for building an endogenous governance mechanism for government data based on metadata management and integrated modeling. This aims to strengthen data standards, improve data quality, and facilitate data utilization at the fundamental level, thereby enhancing the effectiveness of government data governance.

Key words: government data governance; dissipative structure; metadata management; integrated modeling

0 引言

数据作为新型生产要素, 是数字化、网络化、智能化的基础, 深刻改变着生产方式、生活方式和社会治理方式^[1]。政府数据作为一种重要的公共数据, 在数字政府建设和治理数字化转型中具有重要的基础支撑作用, 有助于实现政府决策科学化、公共服务高效化、社会治理精准化, 从而进一步推进国家治理体系和治理能力现代化。

数字化时代, 经济社会发展对高质量数据存在广泛需求, 而高质量数据的供给也存在不平衡不充分的问题,

数据要素市场建设仍处于前期推进阶段, 数据要素价值未能充分发挥。提高数据质量和规范性, 推进政府数据开放共享, 是加快培育数据要素市场的重要基础工作^[2]。加强数据治理, 依法依规促进数据高效共享和有序开发利用, 充分释放数据要素价值, 是构建开放共享的数据资源体系的重要任务^[3]。政府数据治理是应对政务信息化过程中信息孤岛、数据壁垒等问题的解决方案和促进政务信息系统整合共享目标实现的有效保障^[4]。在数字中国建设背景下, 政府数据治理是国家治理体系和治理能力现代化的重要内容^[5]。

本文针对政府数据利用中的突出矛盾和典型问题, 分析数据治理中的目标与任务, 进一步研究政府数据治

* 基金项目: 2023 年度上海市市场监督管理局科技项目 (2023-56)

理的理论基础和相应的实践路径,从理论上探究政府数据治理的逻辑机理并从实践上提出相应的实施策略与技术方法,旨在从基础层面强化数据标准、提升数据质量、促进数据利用,提升政府数据治理效能。

1 政府数据治理的相关研究

1.1 数据治理的概念与数据治理成熟度模型

基于不同的业务目标,数据治理在国际标准和国家标准中存在多种定义。

国际标准组织在相关文件^[6-11]中将数据治理定义为组织为实现数据资产价值而对数据管理活动开展的评价、指导和监控的战略协同活动,包含数据相关政策的制定与实施、确立数据所有权、明确数据管理责任等宏观层面的统筹规划^[12]。国际数据管理协会(DAMA)将数据治理定义为对数据资产管理行使权力和控制的活动集合,而国际数据治理研究所(DGI)将数据治理定义为通过一系列根据共识模型执行的信息相关的过程来实现决策权和职责分工的系统^[13-17]。

我国的国家标准从不同角度将数据治理定义为“数据资源及其应用过程中相关管控活动、绩效和风险管理的集合”“对数据进行处置、格式化和规范化的过程”“基于数据生命周期,进行数据全面质量管理、资产管理、风险管理等统筹与协调管控的过程”^[18]。

综上,数据治理可以概述为组织为实现信息资产有效使用,对数据从采集、归集、整合到分析、利用进行管理、评估、指导和监督的过程。数据治理是获得高质量数据不可或缺的核心管理和控制规程,通过数据治理可以提升数据的可用性和可靠性。

数据治理要取得好的成效,需要高的数据治理成熟度。数据治理成熟度是指一个组织进行数据治理所达到的阶段,核心衡量指标包括数据质量、数据访问管理和数据素养。为了达到高的数据治理成熟度,相关组织需要引入数据治理成熟度模型以度量数据治理活动,并通过该方法学习如何有效地管理数据从而确保数据的高质量。目前具有代表性的数据治理成熟度模型主要有IBM公司的5级成熟度模型、Gartner公司的6级成熟度模型以及我国的国家标准《数据管理能力成熟度评估模型》(DCMM)等^[16]。DCMM定义了8个核心能力域及28个能力项445条标准,用于建立和评价组织的数据管理能力。数据治理是一种自上而下的系统性工作,负责指导、监督和评价数据管理;数据管理则侧重于执行数据治理形成的决策,是数据治理在执行层的具体细化^[19]。DCMM聚焦数据管理能力并更加注重数据标准的作用,是面向行业应用的数据治理成熟度模型。

1.2 政府数据治理体系与治理能力

1.2.1 政府数据治理体系

数据治理体系是数据治理的承载系统,广义上由组织体系、制度体系、运行体系、评价体系和保障体系构成,是一个涉及战略目标、组织架构、数据标准、管理规范、技术工具的综合体。

安小米等^[20-23]认为数据治理体系是一个具有社会建构和技术建构双重建构属性的复杂体系,并提出由大数据治理主体、治理客体、治理活动、治理工具、治理目标等要素组成的政府大数据治理体系和基于戴明环(PD-CA,计划-执行-检查-处理循环)的综合集成框架,以及融合宏观、中观与微观三个层次体系构成要素的政府大数据治理实现路径,包括通过业务驱动对元数据和主数据进行管理以保证数据的完整性、一致性和准确性。尧淦等^[24]研究构建了由组织结构、规则标准、技术支持三个关键域组成的政府大数据治理体系,从顶层设计、过程管控、基础保障三个方面把握政府大数据治理的框架和内容,从而满足数据的完整性、一致性、合规性要求。宋懿等^[25]依照“治理挑战识别-治理要素提炼”思路构建了由“战略、保障和工具”三个维度及若干要素构成的“面向共享和开放的政府数据治理体系”分析框架。蒋敏娟^[26]从狭义层面提出了政府数据治理的三个主要内容模块:数据标准及质量管理、数据共享与开放、数据安全与风险,构建基于治理主体、治理流程、治理层次、治理机制的数据治理行动框架。胡媛等^[27]对美国政府数据治理进行了结构性分析并得出以下结论:(1)构成要素日臻全面,包含微观层面上对数据质量和数据安全隐私的要求,中观层面对运用生命周期理论维系全要素过程管理的要求和宏观层面的协同管理规划要求;(2)过程管理更高标准,主要体现为风险管理和全生命周期良性循环;(3)技术体系构建更加完备,突出表现为新增元数据管理要求,并将其视为数据治理的基础。黄静等^[28]按照信息生命周期管理理论构建了政府数据治理框架,更加强调数据的标准化、真实性及准确性,并重点关注元数据管理、数据集成、数据安全等关键环节。左美云等^[17]在综合研究国内外数据治理框架的基础上,提出我国跨部门政府数据治理框架,强调政府数据治理的特殊性决定了统一数据标准、保障数据质量的重要性。

综上,政府数据治理体系是一个具有多属性、多要素、多视域、多过程、多层次的复杂体系,包括治理目标、组织架构、制度流程、标准规范、信息架构、技术工具等,更加重视基础层面的数据标准与质量以服务数据共享和开放,元数据管理在其中发挥越来越重要的

作用。

1.2.2 政府数据治理能力

政府数据治理能力可以划分为数据采集能力、数据分析能力和数据应用能力三个部分^[29]。从根本上说,数据治理的目标就是提高数据质量、促进数据利用,因此数据应用能力在政府数据治理能力中具有重要地位。安小米等^[18,30]从政府数据依法治理、源头治理、精准治理和长效治理国家战略需求出发,提出由政府数据依法治理与数据可用能力、政府数据源头治理与数据有用能力、政府数据精准治理与数据易用能力、政府数据长效治理与数据善用能力构成的多维政府数据治理与利用能力框架。杨智勇等^[31]从“对数据的治理”和“依数据去治理”两个角度出发,通过政府数据质量治理、政府数据仓库建设、数据共享平台搭建提高治数效能,通过“赋智”“赋能”“赋权”实现数治能级的提升。

面向数据应用,政府数据的有效流通是数据治理能力的重要体现,它内含了数据治理中最基本的标准、质量、安全、服务等要素。因此,数据模型与数据指标的构建和管理能力是数据治理的两项重要的基础能力。

1.3 本文工作

学术界关于政府数据治理已有较多研究,并从不同的视角形成了一系列的理论成果,这些研究成果对于政府数据治理具有重要指导意义。与此同时,政府数据治理在理论和实践上仍有进一步研究和探索的空间:一是进一步丰富和完善基础理论体系,从而给予实践更坚实的理论支撑;二是在基础理论框架下,进一步突出实践路径的起始原点与实施重点。本文分析政府数据治理的目标与任务,在基础理论层面,从系统科学的角度进一步探究数据治理的内在逻辑机理,旨在更好地指导政府数据治理实践;在应用实践层面,基于现有政府数据治理体系与治理能力框架,围绕数据治理基础层面的数据标准规范,研究在系统科学理论指导下的更具实践意义的政府数据内生治理机制。

2 政府数据治理的目标与任务

政府数据治理具有系统要素的多元性、多层次的协同性、数据边界的开放性、过程的动态性、矛盾的复杂性五重系统特性,提升政府数据治理水平,需要提高政府数据治理的整合能力、协同能力、开放能力、预测能力及解决复杂问题的应对能力^[32]。

2.1 政府数据治理面临的主要问题

随着政务信息化建设的不断推进,政府机构积累和沉淀了海量、多源、异构的公共数据,同时也不同程度地产生如下问题:

(1) 数据孤岛:各部门信息化系统各自独立建设,形成了事实上的系统烟囱和数据孤岛,如果不对这些业务数据进行整合融合一体化管理,数据的关联分析和深度挖掘将很难展开。

(2) 数据冗余:因为缺乏统一的数据规划,造成各业务部门重复采集数据,同时数据的含义和规格也不尽相同,不仅造成存储浪费,而且对数据的完整性和一致性造成损害。

(3) 质量低下:政务信息化系统是政府数据产生的源头,建设部门对数据质量的责任意识以及开发团队对数据质量的管控能力不足,造成数据质量低下,利用价值降低。

(4) 使用困难:数据的各自定义、各自存储、各自管理造成数据标准不一、数据质量低下等问题,对业务人员理解和利用数据造成了巨大的障碍。

这些问题的解决有赖于数据治理工作的实施;数据治理已成为大数据时代数字政府建设的核心任务,也是政府数据价值实现的重要前提^[25]。

2.2 政府数据治理的基本目标

数据资产是数据治理的对象,它决定了数据治理的范围和边界。数据治理的目标旨在构建一套完善的数据管理机制,对指导、规划、执行、监督等数据管理活动进行规范和激励,具有保障数据质量、防控数据风险、提升数据价值等多元价值诉求^[33]。

从政府数据治理的角度看,曹惠民^[5]提出政府数据治理目标应包括数据和信息等核心资源能够被安全、有效地利用和整合,公众能够享受到政府数据治理改革技术所带来的高品质公共服务。实践上,政府数据治理的基本目标可以概括为两个:一个是“可供给”,一个是“可获得”。“可供给”主要解决数据规范和数据质量问题,实现数据更广泛更充分的共享和开放利用,更好地满足各类业务场景的数据应用需求。“可获得”主要解决数据安全性和访问便利性问题,让各类用户能够自主地、便捷地找到所需要的数据。

2.3 政府数据治理的主要任务

作为数据治理的一个特定领域,政府数据治理具有一般数据治理的特征,同时由于政府数据具有更为突出的共享与开放属性,因而政府数据治理的任务又突出体现在以下三个方面:

(1) 提升数据质量:数据质量不仅是业务工作价值的体现,也能够反向赋能业务工作,高质量的数据底座是支撑数字化转型的重要基础。数据规范是数据质量的基础依托,需要从数据采集、数据标准、模型建设、指标体系等方面建立完整的数据规范,通过设定标准或规

则,实现异常数据的及时发现和依规处置。

(2) 保障数据安全:只有在保障数据安全的前提下才能更好地开发数据价值,作为最基本的治理要求,数据安全治理贯穿于数据的全链条全生命周期,包括分类分级定义以及相应的配套措施,如数据加密、数据脱敏、访问控制、数据审计、数据监控等。

(3) 优化数据服务:政府数据治理的最终目标是数据利用,包括利用各种技术方法进行数据挖掘以提升数据价值和实现辅助决策。提升数据质量和保障数据安全,从根本上说也是为了促进数据被更好地利用,优化数据服务是提供优质数据供给不可或缺的手段。

3 耗散结构理论与政府数据治理

数据治理是一项复杂的系统工程,因此系统科学的相关理论对于指导数据治理实践具有内在的逻辑统一性。

3.1 耗散结构理论

耗散结构理论认为:一个远离平衡态的开放系统,通过不断与外界进行物质、能量与信息的交换,在系统参量变化达到某一特定阈值时,系统在某些非线性动力学过程作用下,通过涨落变化实现非平衡相变,由原来的混乱无序状态转变为一种时间、空间或功能上新的有序状态,这种有序状态需要不断地与外界进行物质、能量和信息交换才能维持,这种远离平衡态的、达到有序状态的稳定结构被称为“耗散结构”^[34-37]。这里的平衡态是指系统内部运动过程中推动力无限小的状态,在这种状态下,系统由于发展动力丧失而表现为静止或无序状态^[34]。熵是系统无序性的表征,根据热力学第二定律,自然条件下系统的熵值总是趋于不断增大,即系统总是向着混乱的方向发展。当熵值达到最大时,系统的无序程度也最高,此时系统达到了平衡态^[38]。根据耗散结构理论,在开放体系中,影响系统结构和稳定性的系统熵模型为:

$$dS = d_i S + d_e S \quad (1)$$

其中, $d_i S$ 是内部熵,为系统内部在自然运动过程中以不可逆的方式产生的熵的变化,适用热力学第二定律,因此内部熵恒为正值,也称为熵产。 $d_e S$ 是外部熵,是系统与外部环境进行物质、能量交换引起的熵的变化,为非自然运动过程,不适用热力学第二定律,因此外部熵可为正值、负值或零,也称为熵流。 dS 是系统的总熵,是内部熵与外部熵共同作用的结果,如果系统的总熵趋于增大,系统趋于混乱;如果系统的总熵趋于减小,则系统趋于有序,因此系统的总熵反映系统质的变化,也称为熵变。

耗散结构理论是研究系统耗散结构的性质及其形成、稳定和演化规律的科学,从有序和无序相互转化的机制

和条件问题揭示自组织系统的机制和规律,同时耗散结构特征可将不同的风险因素用熵进行统一的表达^[39]。通过耗散结构理论可以分析系统的耗散结构特征演化规律,也可以通过构建系统熵模型和相应的指标体系,通过计算关联熵和运行熵对系统的耗散结构特性进行实证分析或有序性评价^[37]。熵和耗散结构理论已成为研究系统由无序状态变为有序状态的重要理论基础^[36],该理论可以回答开放系统如何从无序走向有序的问题,对物理系统、化学系统、生物系统、生命系统、社会系统和思维系统的研究具有普遍指导意义^[34]。

3.2 政府数据系统的耗散结构特征

维系政府数据的采集、归集、整合、分析、利用等工作开展的机构人员、制度规程、软硬件设施以及各类公共数据,共同构成了政府数据系统。

3.2.1 开放性

政府数据系统作为政府部门履行公共管理与公共服务职能的重要支撑载体,具有数据的采集、归集、管理、加工、共享和开放等功能,其投入要素包括人力资源、数据资源、技术工具、系统设备等,产出为经过治理的规范的数据和经过优化的便捷服务。政府数据系统的职能定位,决定了其必然与外部环境进行常态化的高频次的物质、能量、信息的交换,天然具有开放性特征。开放性为数据系统提供运行能量和发展动力,同时运行良好的数据系统也必然以数据为载体向外界输出更高级别的信息和能量。

3.2.2 非线性作用

在数据治理过程中,外部环境的政策法规、标准规范等会通过管理体制、运行机制的调整以及资源投入、要素配置的变化等对数据系统施加结构性或功能性影响。对数据系统而言,人的主观能动性在数据治理活动中始终具有突出的作用,深刻影响着数据治理活动开展和数据治理工作效能;技术工具作为人的能力延伸手段,在数据治理活动中则发挥效能倍增器的作用,这些都会促使各子系统或者系统要素之间发生协同、相干效应,具有典型的非线性作用。

3.2.3 远离平衡态

数据系统的管理体制和运行机制与政府的公共管理与服务职能紧密相关,体现了政府在履行公共责任、实现公共目标、向社会提供公共物品和服务过程中所体现出来的功能与效用,因此对数据系统具有至关重要的影响力。同时,系统各要素作用机制不同,运行势能也不同,其中人员队伍、操作规程、技术工具发挥着关键性作用。在上述内外部因素的共同作用下,数据系统具备远离平衡状态的条件。

3.2.4 存在涨落

数据系统作为复杂开放系统,系统各要素相互关联、相互制约、相互促进。一方面数据系统与外部环境进行物质、能量、信息的交换,数据交互、技术更新、监管要求、外部反馈等都会影响系统内部运行;另一方面,管理人员、运维人员、技术方法、平台工具、操作规程等各要素或各子系统之间相互作用,使得数据系统在运行过程中存在诸多动态指标的涨落现象,体现了系统在不断运动和变化情况下的宏观状态。

3.3 基于耗散结构的政府数据内生治理机制

数据治理的过程是数据系统对抗熵增的过程。数据系统是典型的管理系统,任佩瑜^[40]提出并证明了管理系统内部产生的管理总熵并不恒大于零的结论,其管理系统熵增公式为:

$$dMS = d_i S + d_{in} S + d_e S \quad (2)$$

其中, dMS 为管理系统的管理总熵,概念内涵同式(1)的 dS ; $d_i S$ 、 $d_e S$ 的概念内涵同式(1); $d_{in} S$ 为系统内部人工干预或管理活动所产生的熵增,由于这是非自然条件下的运动过程,因此不遵循热力学第二定律,数值上可以是正数、负数或零,具有不确定性。

自然条件下,系统内部要素或子系统之间通过相互、协同作用而自发产生特定结构或功能的过程叫“自组织”^[34]。基于耗散结构理论与熵增原理,要使数据治理具有持久的生命力,从而产生可持续的治理效果,必

须对数据的生产、加工进行规范化、秩序化管理和控制,通过系统自组织功能建设有效减少内部熵的增加,减少数据冲突和混乱,增强数据治理的内生动力,从而形成数据内生治理机制。

实践上,政府数据治理一般通过责任部门联合专业机构开展实施,采取自上而下的治理路径,是一种外生型数据治理。单纯的外生型数据治理模式会造成数据治理与数据生产的割裂与脱节,治理成本高昂,治理效果有限。而内生型政府数据治理基于数据系统的自身运行,以数据生产和数据加工作为过程控制的关键环节,采取自下而上的治理路径,强调数据生产与数据治理的整合、融合、一体化,将数据生产和数据加工与数据治理逻辑、治理机制融为一体。

基于耗散结构理论,政府数据治理需要从技术方法和管理机制上构建数据系统的负熵流和内部管理负熵,特别是要聚焦数据系统关键要素和数据处理关键环节,从标准规范层面夯实数据治理基础,通过构建内生型数据治理的自组织功能实现“生产即治理”^[41]的目标,避免“先生产后治理”的弊端。在这种数据治理形态下,外生型数据治理为数据系统提供负熵流,内生型数据治理作为外生型数据治理负熵流的作用“靶点”,为数据系统提供管理负熵,从而形成双向治理路径的协同运行机制和动力保障机制。政府数据治理的耗散结构概念模型如图1所示。

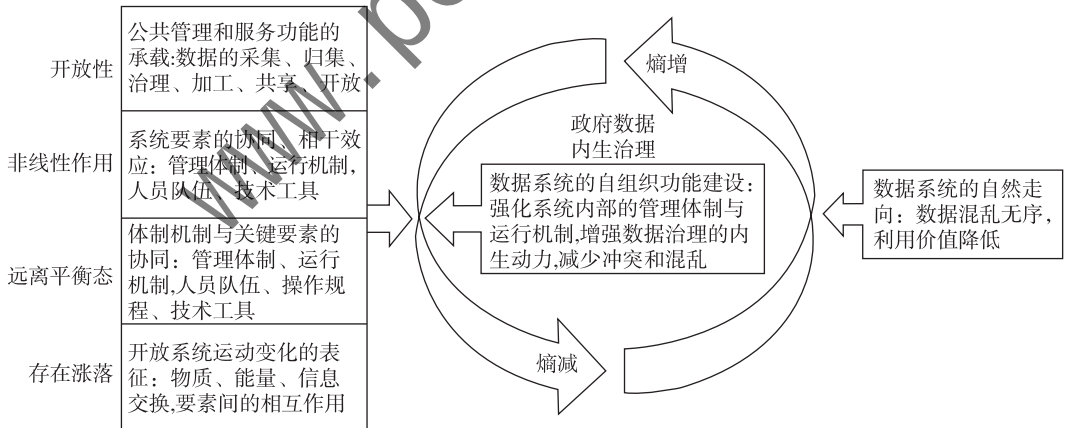


图1 政府数据治理的耗散结构概念模型

4 政府数据内生治理的逻辑机理与技术实践

政府部门在履行公共管理和公共服务职能的过程中采集和产生了大量数据,这些公共数据资源对于促进经济社会高质量发展具有重要价值。数据治理是充分激发数据价值的一项基础性、系统性工作,同时也是一项艰巨、繁杂和具有挑战性的工作,政府数据治理在理论和实践上都需要进一步的研究和探索。

数据治理涉及组织体系、标准体系、流程体系、技术体系和评价体系五方面的工作领域,包括数据标准、数据质量、主数据、元数据、数据安全等多个方面内容^[42]。实践上,在数据系统建设过程中,实质性过程管控的缺乏导致不能对模型设计活动进行有效规范,信息架构设计与信息系统实现不能有效贯通,数据需求、模型设计、物理实现相互脱节,不仅造成基础层面数据资

源的缺失,同时也造成数据的冗余、交叉、不一致等数据质量问题,影响数据资产价值。

基于耗散结构理论,为有效实现政府数据治理目标,可以采取以下措施:通过应用层面广泛的数据共享与开放利用,进一步拓展数据系统的开放性;通过基础层面强化元数据管理和一体化建模增强数据序化势能,使数据系统远离平衡态;通过组织层面加大人员和技术要素投入,使数据系统保持非线性作用;通过管理层面加强运维监测和监督检查,保持数据系统的涨落机制。通过上述治理活动,使数据系统因应外部环境变化和内部功能需求而形成新的耗散结构,并具备维持这种稳定有序结构的能力。其中,元数据管理和一体化建模是具有核心地位的基础性工作,可有效增强数据系统的自组织功能,推进内生型数据治理,形成内外协同效应,提升政府数据治理效能。

4.1 元数据管理

数据共享和开放是政府数据的突出属性和鲜明特征。可用性是数据共享和开放中影响数据价值实现的首要因素。元数据可以从基础层面夯实数据标准规范建设,增加数据系统的序化势能而使其远离平衡态,为数据共享和开放提供有力支撑。利用元数据,数据供给方可以支持数据的分类分级和共享开放,数据需求方可以理解数据的业务含义和技术规格。

元数据是对信息资源的结构化描述,可以帮助用户发现、识别、确认、记录数据,具有管理、检验和保障数据等功能^[43]。元数据通常分为三类:业务元数据,用于记录数据的业务属性,对数据的业务含义、业务规则进行描述;技术元数据,用于记录数据的技术属性,包括物理模型的表与字段、ETL规则、集成关系等;管理元数据,用于明确数据所有者、数据安全等级、共享开放属性等内容。以元数据为基础,在数据处理过程中进行完整性、有效性、一致性、唯一性、正确性、准确性等数据质量管理,能够充分有效地夯实数据价值基础^[44]。

元数据是数据共享利用的基础,广泛存在的数据可查找、可理解、可使用问题,底层原因在于业务元数据与技术元数据未能有效统一,导致用户无法对数据进行快速搜索、理解和获取。通过具有约束力的元数据管理办法和操作规程,连通数据的业务属性与技术属性,形成的高质量元数据可以支撑数据地图等可视化工具对数据目录、数据标准、数据模型、数据分布等信息架构进行集中展示,使数据的来源、数量、分布、标准、流向、关联关系得以清晰呈现,让用户能够高效地查找数据和正确地理解和使用数据。

作为数据治理的基石,元数据管理^[45]贯穿数据生产、

汇聚、加工、利用的全链条,主要功能包括:承载业务规则落地,增强数据业务属性,从源头夯实数据质量基础;发现沉淀的数据,提高数据曝光和利用率,有效激活数据价值;支撑数据指标和业务报表的高效构建;助力敏捷开发,快速响应需求等。

在关键的元数据生产环节,对于增量数据场景,逻辑建模时需要逐层建立业务对象、逻辑实体、属性等的对应关系,并将实体的属性映射为数据标准,物理建模时需要将物理表、数据视图与逻辑实体一一对应,数据表的字段与逻辑实体的属性一一对应,从而实现业务元数据与技术元数据的统一;对于存量数据场景,可以利用元数据采集工具,按照统一的元模型从数据源中获取元数据。元数据管理概念模型如图2所示。

4.2 一体化建模

政府数据突出的共享与开放属性决定了标准规范的特殊重要性。虽然行业主管部门会制定数据规范加强数据的标准化建设,但政务信息化建设任务的外源性和实施主体的多元化,为数据标准规范的落地实施增加了难度和风险。在依托政务信息化系统进行的数据采集、流转、加工等操作中,需要更加注重实质性过程控制,强化自组织功能,从而增大管理负熵,使数据系统远离平衡态。

一体化建模^[45-46]旨在维护信息架构与数据资产的一致性,实现数据定义、模型设计、数据生产的内在统一,确保数据规范得到有效执行并为数据利用提供元数据的基础支撑,避免因数据标准未能得到有效执行而造成数据系统混乱。在一体化建模中,模型管理是核心,元数据贯穿整个流程。

数据标准管理是一体化建模的基础和前提。在政务信息化建设中,通常可以利用数据规范的基础数据元构建数据标准池,在数据建模过程中约束实体属性只能从数据标准池中选择,从源头加强数据标准落地。

对系统开发涉及的数据生产活动,开展源头数据治理,对数据设计过程进行实质性管控,即首先对业务指标进行数据的结构化定义完成逻辑模型设计,形成业务元数据;然后基于逻辑模型设计,在业务元数据的规范和约束下完成物理模型设计,建立物理数据表与数据资源目录的映射关系、物理字段与数据指标的映射关系,形成技术元数据。通过上述过程,确保数据需求、模型设计、物理实现的统一,在形成高质量模型资产的同时产生高质量的元数据,达到“设计即开发,生产即治理”的一体化效果。

对数据仓库建设等数据生产活动,开展链路数据治理,对落地数据进行标准核验,即定期自动采集生产系

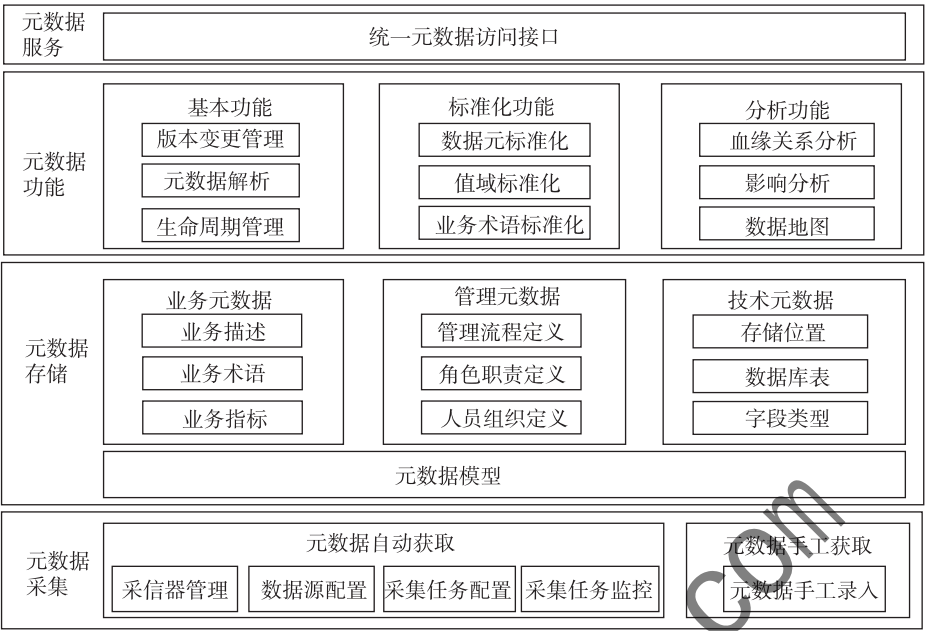


图2 元数据管理概念模型

统的技术元数据并进行模型验证，依操作规程对数据的物理实现与模型设计不一致的情况进行核查处理；随模型发布同步完成元数据、数据资产目录及数据标准的更新，确保数据模型的实际落地，实现对数据的全生命周期管理。

一体化建模框架如图3所示。

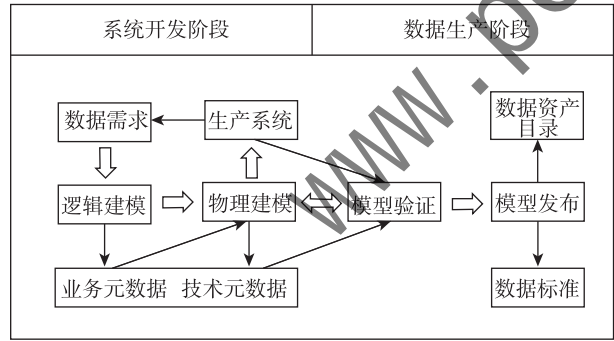


图3 一体化建模框架

通过一体化建模的规范机制，数据系统进一步完善了保障数据完整性、准确性和一致性的自组织功能。

4.3 内生型治理

基于耗散结构理论，政府数据治理一方面需要通过持续的数据共享和开放，保持数据系统的开放性，并借由要素投入、外部反馈和监管要求，对数据系统持续注入负熵流；另一方面，需要通过数据系统自身的有序化管理和自组织功能建设，持续产生管理负熵，实现内生型的数据治理，确保数据系统稳定、有序运行。

实现政府数据内生型治理，技术上需要建设相应的平台工具系统，对数据生产与数据加工等关键环节加强过程控制，通过以元数据为基础支撑的一体化建模，实现数据模型的一体化设计以及数据模型与元数据的融合管理，形成完善、有效的数据架构；管理上需要建立责任监督机制，完善数据治理的管理体制与运行机制，对数据质量进行实时监测和定期评估，加强数据质量管控；制度上需要制定业务流程和操作规程，对数据治理活动加以规范，保障制度的有效执行和机制的有效运转。通过技术、管理、制度三位一体的协同作用，在保持数据系统的开放性和远离平衡态的同时，强化非线性作用和涨落机制，使数据系统始终在耗散结构下稳定运行。

5 结论

数据治理是激发数据要素价值的重要手段，政府数据治理在数字政府建设和治理数字化转型中具有重要的基础支撑作用。数据治理是一项系统工程，需要从数据战略、组织架构、制度流程、技术系统等多个层面开展工作。传统的“先生产后治理”模式，不仅数据治理成本高昂，而且数据质量也难以得到有效保障。从系统科学的角度看，数据治理的过程就是避免数据混乱、对抗系统熵增的过程。基于耗散结构理论，建立数据内生治理机制，实现“生产即治理”的数据系统运行态势，对于提升数据治理效能具有重要意义。基于元数据驱动的一体化建模，通过有效的过程管控规范数据生产，实现数据定义、模型设计、物理实现的内在统一以及元数据和数据模型管理的有效融合，有助于形成稳定有序的数据治理体系。

据系统,实现基于耗散结构理论的政府数据内生治理机制。

参考文献

- [1] 新华社. 中共中央国务院关于构建数据基础制度更好发挥数据要素作用的意见 [EB/OL]. [2023-07-30]. https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm.
- [2] 新华社. 中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见 [EB/OL]. [2023-07-30]. https://www.gov.cn/zhengce/2020-04/09/content_5500622.htm.
- [3] 国务院关于加强数字政府建设的指导意见 [EB/OL]. [2023-07-30]. https://www.gov.cn/zhengce/content/2022-06/23/content_5697299.htm.
- [4] 马广惠, 安小米. 政府大数据共享交换情境下的大数据治理路径研究 [J]. 情报资料工作, 2019, 40 (3): 62-70.
- [5] 曹惠民. 数字中国建设背景下的政府数据治理改革 [J]. 理论探索, 2023 (3): 83-89.
- [6] ISO. Asset management-overview, principles and terminology (ISO55000-2014) [S]. Geneva: ISO Copyright Office, 2014.
- [7] ISO, IEC. Information technology: governance of IT for the organization (ISO/IEC 38500-2015) [S]. Geneva: ISO Copyright Office, 2015.
- [8] ISO. Health informatics-identification of medicinal products-core principles for maintenance of identifiers and terms (ISO/TR 14872-2019) [S]. Geneva: ISO Copyright Office, 2019.
- [9] ITU-T. Technical specification D2.1 data processing and management framework for IoT and smart cities and communities; ITU-T FG-DPM technical specification D2.1: 2019 [S]. Geneva: Telecommunication Standardization Sector of ITU, 2019.
- [10] ISO. Data quality-part 2: vocabulary (ISO8000-2: 2020) [S]. Geneva: ISO Copyright Office, 2020.
- [11] ISO, IEC. Information technology-big data reference architecture-part 3: reference architecture (ISO/IEC 20547-3: 2020) [S]. Geneva: ISO Copyright Office, 2020.
- [12] 安小米, 许济沧, 王丽丽, 等. 国际标准中的数据治理: 概念、视角及其标准化协同路径 [J]. 中国图书馆学报, 2021, 47 (5): 59-79.
- [13] DAMA International. The DAMA guide to the data management body of knowledge [M]. New York: Technics Publications, 2009.
- [14] 数据管理协会. DAMA 数据管理知识体系指南 [M]. 北京: 机械工业出版社, 2020.
- [15] The DGI data governance framework [S]. Orlando: The Data Governance Institute, 2006.
- [16] 肖益, 董晶. 数字孪生城市多视域数据治理框架研究 [J]. 信息技术与标准化, 2023 (4): 30-35.
- [17] 左美云, 王配配. 数据共享视角下跨部门政府数据治理框架构建 [J]. 图书情报工作, 2020, 64 (2): 116-123.
- [18] 安小米, 许济沧, 黄婕, 等. 政府数据治理与利用能力研究: 现状、问题与建议 [J]. 图书情报知识, 2021, 38 (5): 21-33.
- [19] 王雷, 程主. 大数据时代的政府数据治理研究综述 [J]. 信息技术与标准化, 2023 (6): 71-76, 94.
- [20] 安小米, 王丽丽. 大数据治理体系构建方法论框架研究 [J]. 图书情报工作, 2019, 63 (24): 43-51.
- [21] 安小米, 宋懿, 郭明军, 等. 政府大数据治理规则体系构建研究构想 [J]. 图书情报工作, 2018, 62 (9): 14-20.
- [22] 安小米, 白献阳, 洪学海. 政府大数据治理体系构成要素研究——基于贵州省的案例研究 [J]. 电子政务, 2019 (2): 2-16.
- [23] 安小米, 郭明军, 洪学海, 等. 政府大数据治理体系的框架及其实现的有效路径 [J]. 大数据, 2019, 5 (3): 3-12.
- [24] 尧淦, 夏志杰. 政府大数据治理体系下的实践研究——基于上海、北京、深圳的比较分析 [J]. 情报资料工作, 2020, 41 (1): 94-101.
- [25] 宋懿, 黄婕. 欧盟政府数据治理: 体系建设及其经验启示 [J]. 现代情报, 2023, 43 (6): 139-148.
- [26] 蒋敏娟. 机构改革背景下的政府数据治理: 逻辑理路与行动框架 [J]. 学海, 2023 (3): 33-41.
- [27] 胡媛, 黄思慧. 美国政府数据治理战略与行动研究分析及启示 [J]. 情报杂志, 2023, 42 (10): 48-55.
- [28] 黄静, 周锐. 基于信息生命周期管理理论的政府数据治理框架构建研究 [J]. 电子政务, 2019 (9): 85-95.
- [29] 张松波, 李辉, 孙博文. 政府数据治理能力与治理效果关系研究: 一项来自 2456 个微观数据的实证 [J]. 信息技术与管理应用, 2023, 2 (2): 18-28.
- [30] 安小米, 王丽丽, 许济沧, 等. 我国政府数据治理与利用能力框架构建研究 [J]. 图书情报知识, 2021, 38 (5): 34-47.
- [31] 杨智勇, 谢雨欣. 数据如水: 以治水思维析政府数据治理 [J/OL]. 情报科学: 1-17 [2023-07-30]. <https://kns.cnki.net/kcms2/detail/22.1264.G2.20230614.1725.012.html>.
- [32] 苏玉娟. 政府数据治理的五重系统特性探讨 [J]. 理论探索, 2016 (2): 71-75.
- [33] 谭章禄, 王美君. 智能化煤矿数据治理概念模型及技术架构研究 [J]. 矿业科学学报, 2023, 8 (2): 242-255.
- [34] 单明, 刘彦青, 马荣江. 我国农村自然村落的未来发展——基于耗散结构理论的分析 [J]. 中国工程科学, 2021, 23 (1): 141-148.
- [35] 木合塔尔·艾买提, 肖鹏南, 徐涛. 基于耗散结构理论的粮食生产系统发展状态分析 [J]. 江苏农业学报, 2021, 37 (4): 1043-1050.

- [36] 王成,周明茗,李颖颖,等. 基于耗散结构系统熵模型的乡村生产空间系统有序性研究 [J]. 地理研究, 2019, 38 (3): 619-631.
- [37] 王振宇,贾永飞,李金萍,等. 基于耗散结构理论的科技人才政策系统有序性评价: 以山东省为例 [J]. 科技管理研究, 2022, 42 (2): 36-45.
- [38] 付焱焱,蒋兵,李振宇. 耗散结构理论视域下山东省水资源系统演化机制研究 [J]. 中国农村水利水电, 2022 (5): 178-183.
- [39] 靳泽鹏,赵江平,刘茜. 基于耗散结构理论的煤矿安全风险演化机理 [J]. 陕西煤炭, 2022, 41 (3): 160-163, 178.
- [40] 任佩瑜. 从自然系统到管理系统——管理系统的熵、耗散结构、信息与复杂性 [M]. 北京: 科学出版社, 2018.
- [41] 虎兴龙. 腾讯欧拉数据治理平台思考与实践 [EB/OL]. [2023-08-30]. <https://zhuanlan.zhihu.com/p/579654173>.
- [42] 李题印,宣成,郁建兴,等. 数智赋能时代企业数据治理能力模型研究 [J]. 情报科学, 2022, 40 (11): 20-25.
- [43] 刘智锋,王继民,李倩. 元数据质量评价研究综述 [J]. 情报理论与实践, 2022, 45 (7): 42-48.
- [44] 周艳会,曾荣仁. 基于元数据的数据质量管理研究 [J]. 信息技术与信息化, 2020 (7): 26-29.
- [45] 华为公司数据管理部. 华为数据之道 [M]. 北京: 机械工业出版社, 2020.
- [46] 王鹏,新兴,晓飞. 数据治理一体化实践之体系化建模 [EB/OL]. [2023-08-30]. <https://tech.meituan.com/2022/02/24/systematic-modeling-of-data-development-and-governance-integration-practice.html>.

(收稿日期: 2023-10-11)

作者简介:

付学敬 (1971-), 通信作者, 男, 硕士, 高级工程师, 主要研究方向: 数据治理、数据挖掘、深度学习。E-mail: ascend_ai@163.com。

陈珍 (1989-), 女, 硕士, 工程师, 主要研究方向: 数据挖掘、深度学习。

章君玉 (1984-), 女, 硕士, 工程师, 主要研究方向: 数据治理、数据挖掘。



版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.com