

面向全域空间的分布式数据治理范式：挑战、框架和应用

章敏, 高嵩, 贾晓丰

北京市大数据中心, 北京 101117

摘要

在以超大城市为代表的复杂系统下, 数据的大规模跨域流通具有数源分散、数据异质、场景交叉、规则多变等显著特征, 给多源数据的协同治理带来系统性挑战。针对该难题, 首先从多视角阐述了数据治理的范式演进, 分析了各代际范式的理念、技术和应用特征; 其次总结了全域空间下数据治理面临的技术挑战和科学问题, 创新性提出以数据共生为特征的数据治理 5.0 范式及其技术框架; 最后以智能问数系统为例, 介绍了数据治理 5.0 范式的典型应用。

关键词

全域空间; 分布式数据治理; 数据治理范式; 跨域流通; 供需匹配

中图分类号: TP315

文献标志码: A

doi:10.11959/j.issn.2096-0271.2026037

Distributed data governance paradigm for the entire spatial scope : challenges, framework and applications

Zhang Min, Gao Song, Jia Xiaofeng

Beijing Big Data Centre, Beijing 101117, China

Abstract

In complex systems represented by megacities, large-scale cross-domain data circulation exhibits distinct characteristics such as dispersed data sources, heterogeneous data, cross-scenario application, and dynamic rules, which pose systematic challenges to the collaborative governance of multi-source data. To address these challenges, this paper firstly elaborates on the evolution of data governance paradigms from multiple perspectives, analyzing the conceptual foundations, technological approaches, and application characteristics of each stage. It subsequently summarizes the technical challenges and scientific questions encountered in the entire spatial scope, innovatively proposing the Data Governance 5.0 paradigm characterized by data symbiosis along with its corresponding technical framework. Finally, the application of this proposed paradigm is demonstrated through a case study of an Intelligent Data Operating System.

Key words

entire spatial scope, distributed data governance, data governance paradigm, cross-domain circulation, supply-demand matching

0 引言

数据作为关键生产要素，正加快推动新质生产力的形成和价值释放。近年来，我国围绕数据要素进行了一系列重大战略布局，明确提出“构建城市数据资源体系”，建立“多方参与的协同治理”机制，并在《关于构建数据基础制度更好发挥数据要素作用的意见》中明确提出“数据来源可确认、使用范围可界定、流通过程可追溯、安全风险可防范”的具体要求。2025年，《深化智慧城市发展推进全域数字化转型行动计划》明确要求构建“城市智能中枢”、打造“城市大模型”和“城市智能体应用”。

数据治理最初源于组织内部的实践^[1]，近年来其研究范围逐步延伸和拓展，研究对象从单组织内的数据延伸至政府数据，再到公共数据、企业数据和个人数据^[2]；尺度上从微观、中观拓展至宏观；内容上从聚焦数据质量等基础治理延伸至数据权属、流通和资产化等体系性治理^[3]。国内外学者围绕组织内部的数据治理模型^[4]和工业、金融、医疗等特定领域的治理框架^[5-8]开展了相关研究，数据集成^[9]、数据挖掘^[10]、数据融合^[11]、语义融合^[12]等技术备受关注，以数据中台、湖仓一体^[13]、数据编织^[14]、国际数据空间^[15]（international data spaces, IDS）为代表的技术架构在数据治理中得到了广泛应用。其中，数据中台和湖仓一体主要聚焦于集中化的数据治理，通过分层设计实现原始数据、标准化治理数据和业务应用数据的解耦，但与数源端的实时联动性差；数据编织采用逻辑集中模式，利用统一连接和语

义映射等技术对数据进行虚拟化整合，但在跨域数据的语义对齐和多态场景的协同服务等方面存在不足；以IDS为代表的数字空间提供了跨组织、跨领域的数据共享机制，但主要聚焦于数据共享的合规框架和信任体系构建，在复杂系统数据治理上存在局限性。目前，数据治理在从商业智能到人工智能（artificial intelligence, AI）发展过程中的重要性日益凸显^[16]，但现有研究大多聚焦于特定领域或管理层面的跨域问题^[17]，对多维度跨域协同的研究仍存在诸多理论和技术挑战。

与组织级、领域级的数据治理相比，以城市为代表的复杂巨系统具有数据多源、关系多样、场景多变等显著特征^[18]，其数据治理是一个复杂性、多维度的问题，主要挑战如下：（1）数据多源，在组织级或领域级数据治理中，数据多源但规模有限，业务复杂但同质性强，城市复杂系统没有单一领域的场景，从而带来数源本身不确定或数源之间不一致等问题；（2）关系多样，城市复杂系统中数据流通规模呈指数级增长，数据的语义异质复杂^[19]，大规模数据跨域流通^[20]带来了多源数据和多态场景之间的语义错位问题；（3）场景多变，城市级复杂场景的成长性和开放性带来需求的高频变化，同时多场景耦合使需求呈现出动态复杂性，使得精准捕捉、预测和匹配等应用需求更加困难。由于数字世界和物理世界映射空间的动态复杂性、快速成长性、对特定场景的“零容错”，现有方法论和技术体系难以满足城市级复杂系统的数据治理需求^[21]。

基于此，本文提出了基于联邦机制的分布式数据治理范式，进一步给出了技术框架和典型应用，为以超大城市为代表的复杂系统跨域数据治理提供了可行路径。

1 数据治理范式演进

1.1 数据治理 1.0 到 3.0 范式:从数据分散到数据集中

从面临的问题出发，复杂系统的数据治理历经3个范式（如图1所示），各阶段范式都有其核心治理理念和代表性技术。

数据治理 1.0 范式的主要特征为“分散数据-分散管理”。面向微观尺度单系统的系统级治理，其核心是以数据读写为标志的“建岛”，治理主体以B端（机构）为主，治理对象以数据为主，数据来源为“单业务”，代表性技术是管理信息系统（management information system, MIS）。

数据治理 2.0 范式的主要特征为“分散数据-统一调度”。面向微观尺度多系统的部门级治理，其核心是以数据交换为标志的“建桥”，治理主体以B端（机构）为主，治理对象以数据为主，数据来源为“多业务”，代表性技术是“目录+交换”。

数据治理 3.0 范式的主要特征为“集中数据-集中治理”。面向介观尺度多部门的领域级治理，其核心是以数据汇聚为标志

的“灭岛”，治理主体以B端（机构）为主，治理对象以数据为主，数据来源为“类场景”，代表性技术是“云计算+大数据”。

1.2 数据治理 3.0 到 4.0 范式:从数据自治到数据共治

由于基于单点的数据开发难以发挥大数据的优势，数据治理已从局部的单主体、单组织、单领域治理走向多领域治理。数据治理 3.0 范式的集中管理不仅带来了安全和管理责任的转移，还进一步加剧了数源间的异质性和语义冲突。因此，在以超大城市为代表的复杂系统下，微观上“管不过来”和宏观上“建立秩序”的矛盾驱动了数据治理 4.0 范式（如图2所示）的形成，其核心思想是从数据自治到数据共治。

数据治理 4.0 范式通过链码机制实现分布式治理。链码机制通过统一标准、唯一标识的数字码和城市码分别统筹数字世界的所有数据和物理世界的所有实体，并通过区块链^[22-24]将二者链接在一起，形成数字世界和物理世界之间的动态匹配体系，实现城市复杂系统下多领域数据的逻辑管控。

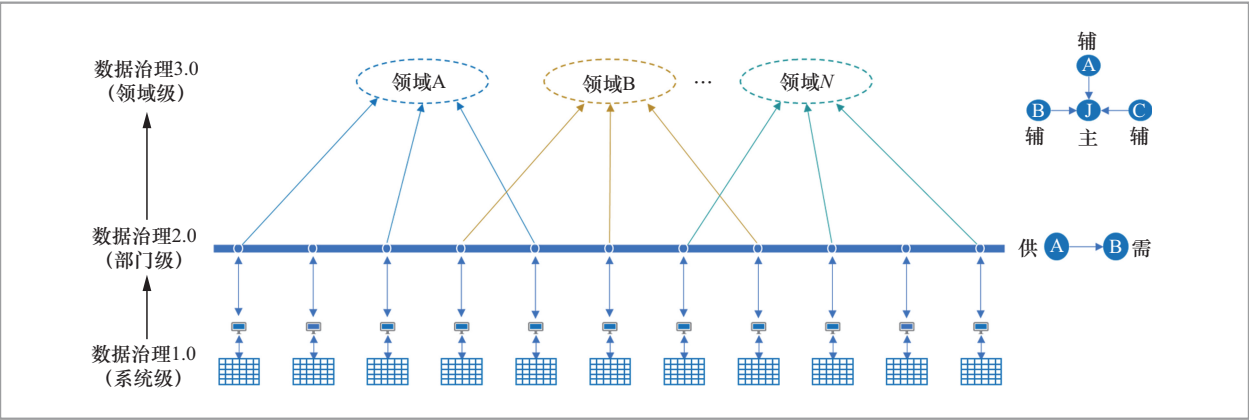


图1 数据治理1.0到3.0范式

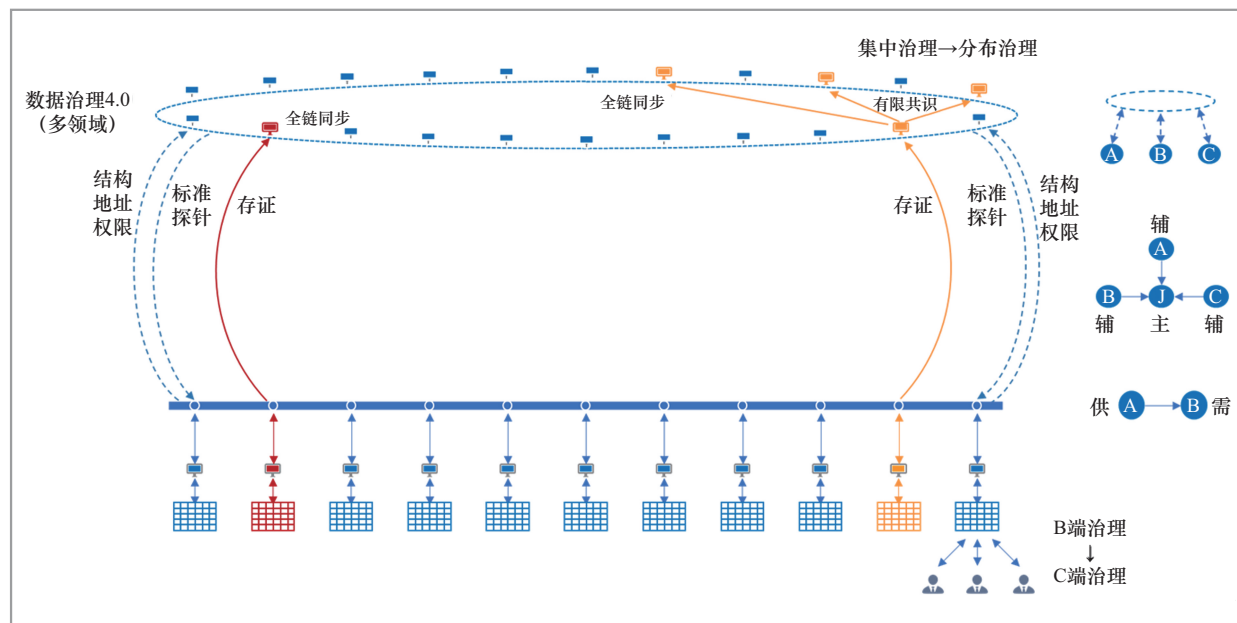


图2 数据治理4.0范式

数据治理4.0范式的主要特征为“分布数据-分布治理”。面向宏观尺度多领域的城市级治理，其核心是以数据链通为标志的“去孤存岛”，治理主体延伸为“B端（机构）+C端（用户）”，治理对象的重心从数据转变为关系，数据来源为“多场景”，代表性技术是“区块链+隐私保护”。

数据治理4.0范式的技术理念如下：一是触一发而动全身，从“分布式数据治理”到“数据分布式治理”；二是没有统一就没有分布，数据结构、授权、标准由链上控制，通过建立权限规则确保链上主体都能治理数据但不能随意治理；三是统得越少、管得越多，基于轻量级数据探针在边缘端开展计算，将数据治理下沉到终端。

数据治理4.0范式的应用理念如下：一是把每个机构都变成一个“大组织”，每个入口都是唯一入口，从而解决横向问题；二是把每个个体都变成一个“治理员”，每次服务都是一次治理，从而解决纵向问题；三是把每个需求方都变成一个“供给方”，

每次使用都是一次反馈，从而解决闭环问题。

2 数据治理5.0范式

2.1 技术挑战和科学问题

随着多场景耦合程度的持续加深，数据要素流通呈现“强跨域”特征。以超大城市为代表的复杂社会系统构成了一个“全域空间”，包括物理空间的管理域、数字空间的数据域和社会空间的应用域，如图3所示。其中，物理空间以跨主体、跨组织、跨区域、跨层级的管理跨域为主要特征，数字空间以跨语义、跨模态、跨系统、跨网络的数据跨域为主要特征，社会空间以跨场景、跨领域、跨尺度、跨信任的应用跨域为主要特征。跨域数据治理需要系统性解决三域内部及相互之间的问题，这大大提高了数据治理的复杂度。现有研

究主要聚焦于某一个特定域，如 Nezami 等^[25]聚焦跨组织之间数据共享流通的管理跨域，Radford 等^[26]关注多模态语义对齐相关的数据跨域，Froelicher 等^[27]侧重跨信任的应用跨域，但缺少对三域之间跨域问题的研究和实践。随着数据量的爆发式增长和数据开发利用的多元化拓展，复杂系统面临着越来越多的跨域协同治理需求。例如，在超大城市复杂系统下，交通治理场景不仅是交通领域的问题，还涉及气象（天气）、电商（快递）、医疗（医院）、教育（学校）等多个领域，且有些问题在表象上属于交通领域，但其根源在其他领域，其治理需要融合不同组织跨系统的库表、文本、图片、视频等多模态数据，使得数据及其相互关系的复杂度显著增加。

数据流通的跨域特征给数据治理带来了一系列新的问题和挑战：数源的分散性问题带来协同控制和优化的挑战；数据的异质性问题带来语义表示和融合的挑战；场景的交叉性问题带来关系发现和传递的挑战；规则的多变性问题带来匹配映射和生成的挑战；隐私的敏感性问题带来隐私保护演化和博弈的挑战。全域空间下数据治理面临的技术挑战可以从供给侧、供需间和需求侧总结为3个科学问题：（1）供给侧，针对数据多源带来的“岛化”数据可利用问题，研究如何在全域数据空间中兼顾“岛”的独立和“孤”的联动，构建多元孤岛的协同治理体系；（2）供需间，针对关系多样带来的“跨域”行为可管控问题，研究如何设计有效的分布式管控机制，实现异质空间下目标、规则与信任的对齐；（3）需求侧，针对场景多变带来的“多态”场景可适配问题，研究如何利用有效的决策共识方法，实现开放场景中供需匹配的自动化与智能化。

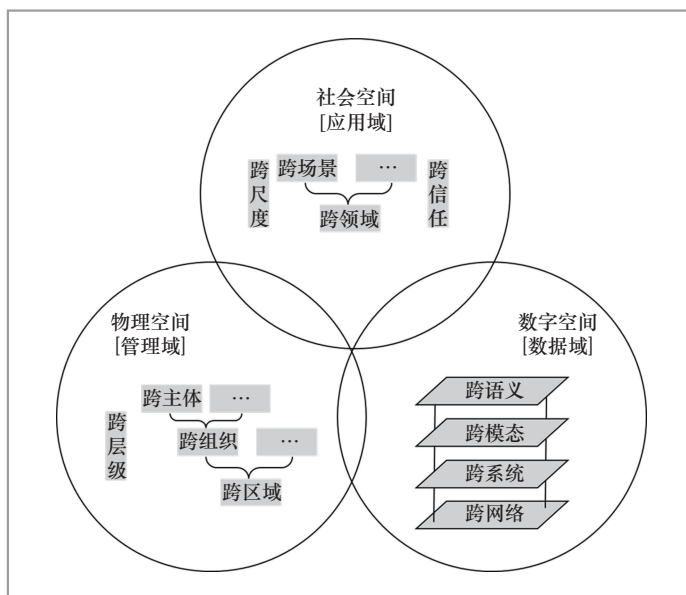


图3 数据流通的“三域十二跨”特征

2.2 数据治理5.0范式的提出：从数据共治到数据共生

针对上述科学问题和技术挑战，基于联邦理念和智能化能力，提出数据治理的5.0范式（如图4所示），其核心思想是从数据共治到数据共生。

首先，将每一个数源侧的联邦节点（如前置机等）替换为一组数据智能体（Agent），数据流转从基于前置规则的数据推送方式转为基于理解和推理能力的数据自组织方式，实现数字空间多源数据的自动识别与调度，打破数据模态与语义的壁垒，融合物理空间下多主体、多领域的跨域数据，通过强化学习优化数据与场景交互中的策略匹配，实现分布式智能协同计算。

其次，将数据目录由人工编制转为智能生成，结合应用场景和业务语义的业务语言以及信息系统中存储的机器语言，基于大模型自动生成可描述数据和场景基本信息及其复杂关系的数据目录，并将不同

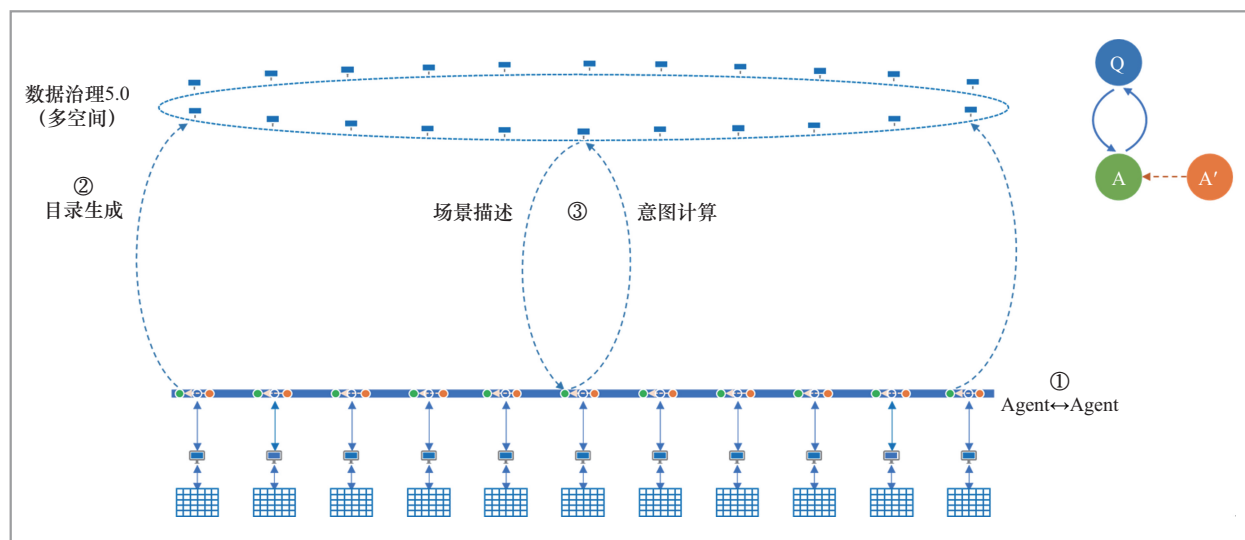


图4 数据治理5.0范式

领域的目录整合到一个统一的语义空间中，实现跨域数据的统一语义表示。

第三，通过意图计算重新定义数据与场景之间的交互界面，将社会空间的场景描述作为直接输入，结合智能体自主学习，挖掘场景的深层需求，建立领域知识库和动态知识图谱与场景反馈之间的闭环迭代机制，实现从数据自动化到知识自动化、再到知识智能化的智能转变。

数据治理5.0范式的主要特征为“分布场景-分权治理”。面向跨尺度的多空间（物理空间-数字空间-社会空间）治理，其核心是以数据智治为标志的“联孤强岛”，治理主体延伸为“B端（组织）+C端（用户）+A端（AI）”，治理对象的重心从关系转变为语义，数据来源从真实场景延伸至虚拟生成，代表性技术是“联邦+大模型”。

数据治理5.0范式的技术理念如下：一是由“M2M”（机器-机器）转变为“A2A”（Agent-Agent），由“数据进-数据出”转变为“场景进-场景出”；二是数据的价值在于“岛”，联邦的目的是“强

岛”；三是通过语义表示、关系发现、场景生成等技术组合的演化机制，实现数据治理技术体系的可进化。

数据治理5.0范式的应用理念如下：一是每个系统（供给方）都由一组Agent管理，以解决“铁路警察”问题，实现物理空间到数字空间的映射；二是每个场景（需求方）都由一组Agent服务，以解决“千人千面”问题，实现数字空间到物理空间的映射；三是每次响应（供需）都在Agent之间对话，以解决“鸡同鸭讲”问题，实现物理空间和数字空间之间的供需匹配。

数据治理1.0范式到数据治理5.0范式的演进如表1所示。

数据治理在1.0范式中完成了电子化，各代际范式的演变经历了从“孤岛林立”（1.0独立→2.0分散）到“分久必合”（2.0分散→3.0集中），再到“合久必分”（3.0集中→4.0分布），再到“分即是合”（4.0分布→5.0分权）的过程。其间，数据治理的自动化和智能化从个体智能向群体智能、再向联邦智能演变。

表1 数据治理各范式的主要特点

	尺度	范围	特征	理念	主体/对象	数据来源	技术代表
数据治理5.0	跨尺度	物理-数字-社会 (多空间)	分权	“联孤强岛” (数据智治)	B端+C端+A端 (语义)	场景+生成	联邦+大模型
数据治理4.0	宏观	城市级 (多领域)	分布	“去孤存岛” (数据链通)	B端+C端 (关系)	多场景	区块链+隐私保护
数据治理3.0	介观	领域级 (多部门)	集中	“灭岛” (数据汇聚)	B端 (数据)	类场景	云计算+大数据
数据治理2.0	微观	部门级 (多系统)	分散	“建桥” (数据交换)	B端 (数据)	多业务	目录+交换
数据治理1.0	微观	系统级 (单系统)	独立	“建岛” (数据读写)	B端 (数据)	单业务	管理信息系统

2.3 数据治理5.0 范式的技术框架

随着数据流通的规模越来越大，传统的数据治理无法有效满足多场景耦合的应用需求。为了在保护数据安全和隐私的同时打破数据孤岛，实现智能供需匹配，最大化发挥数据价值，基于联邦的理念提出一种面向跨域数据治理的分布式技术框架——DCS 框架（“D”表示 data，“C”表示 control、cognition 和 computing，“S”表示 services、security），利用人工智能、区块链、隐私保护等技术，实现多源数据的智能感知和协同治理。如图 5 所示，该框架由 6 部分组成：自下而上分别为联邦数据（federated data）、联邦控制（federated control）、联邦认知（federated cognition）、联邦服务（federated services），实现从分布式数据到分布式服务的全链条智能化；左右两侧分别为联邦计算（federated computing）和联邦安全（federated security），其贯穿数据治理的全流程，可根据不同场景的动态变化实现“一边计算、一边保护”，通过演化博弈机制实现可用性和安全性的动态平衡。纵向各层间存在 3 个层次的“生

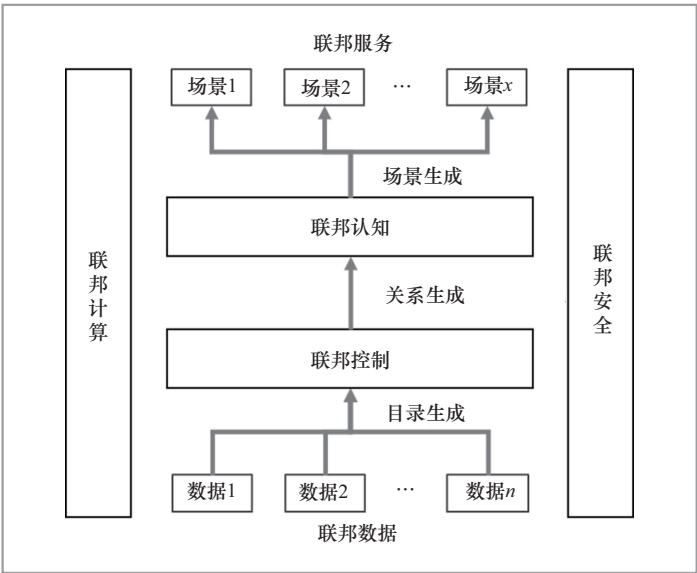


图5 面向分布式治理的DCS框架

成”能力，即从数源层到控制层实现目录生成（catalog generation）、从控制层到认知层实现关系生成（connection generation）、从认知层到应用层实现场景生成（在多场景耦合上体现为城市整体空间的生成，即 city generation），3 个层次的“生成”分别对应数据工程、意图工程、场景工程。不同于数据治理 3.0 范式的集

中治理和数据治理 4.0 范式的分布式治理，DCS 框架以“数据”的分布式协同为基础，进一步解决“智能”的分布式协同问题，实现从“数据共治”到“数据共生”，为全域空间下大规模数据流通和治理的感知、连接、控制、计算和保护提供了一种有效的技术路径。

(1) 联邦数据，其核心是将个体的数据源“邦联化”(confederationization)，以形成全域空间下“联邦化”(federalization)的整体。联邦数据是可被计算的、相互独立的数据单元，分布上具有超大规模、分散、异质的特点。联邦数据的载体有不同的尺度，可以是数据库中的一张表或一个文件，也可以是一个数据库、数据仓库或数据湖，抑或是更大尺度上的一个信息系统或一个可信数据空间。

(2) 联邦控制，解决跨域数据的统一管控调度问题。以多模态数据感知、数据探针和数据路由为核心技术，实现对分布式数据源的智能探测和识别，基于特征识别进行目录生成与校正，并通过数据路由合约实现对多源异构数据的自动寻址映射与跨链调度。

(3) 联邦认知，解决多维语义的表示、融合和关系推理问题。以语义表示、关系发现为核心技术，通过去中心化的目标分解与合成策略，构建分布式自组织的语义协同机制，实现跨域协同主体的目标对齐、跨域协同行为的规则对齐和跨域协同结果的信任对齐；通过时空向量模型和关系迁移学习，实现原子级大尺度弱关系的自动生成与时空映射下的强关系收敛。

(4) 联邦服务，解决多态场景的协同响应和自适应匹配问题。以场景工程、多视角学习和智能决策为关键技术，通过场景解构的多视角表示，构建分布式去中心化服务体系，形成多场景耦合下的动态决

策共识。

(5) 联邦计算，解决跨域数据的分布式协同计算问题。通过对联邦数据不同节点的分布式模型训练与推理，利用安全高效的模型参数更新与聚合机制，整合各节点数据特征，融合模型聚合与协同优化等关键方法生成高质量的全局模型，实现联邦数据的跨域高效协同计算。

(6) 联邦安全，解决不同信任域下隐私数据的协同保护问题。通过智能合约与协议机制将各方权责规则化，确保数据使用的透明性和合规性，并通过多方共识的协作规则，确保数据操作和计算结果的可验证性和不可篡改性，为联邦数据交互与协同计算提供隐私保护计算能力。

3 数据治理 5.0 范式的典型应用

3.1 场景描述:问题提出

不管是组织内部的管理和业务应用，还是组织间的数据共享、开放、交易等数据流通行行为，本质上都是数据供给侧和需求侧的匹配问题，其核心体现为“问数”。其中，“问”代表需求侧，来自物理世界对场景的描述；“数”代表供给侧，来自数字世界对数据的描述。“问数”场景的问题描述与智能化分级如图 6 所示。

从需求侧来看，“问数”场景可归纳为 3 类：一是服务类场景，问题描述具体，但存在表述不准确、内容冗余等挑战；二是管理类场景，问题描述与业务相关，但存在语境不一致、一数多源等挑战；三是决策类场景，问题描述准确，但存在信息不对称、领域交叉等挑战。

物理世界的 3 类场景需求可以映射为对数字世界的数据元、数据、信息和知识的 4 层问题。其中，问“有什么”对应数

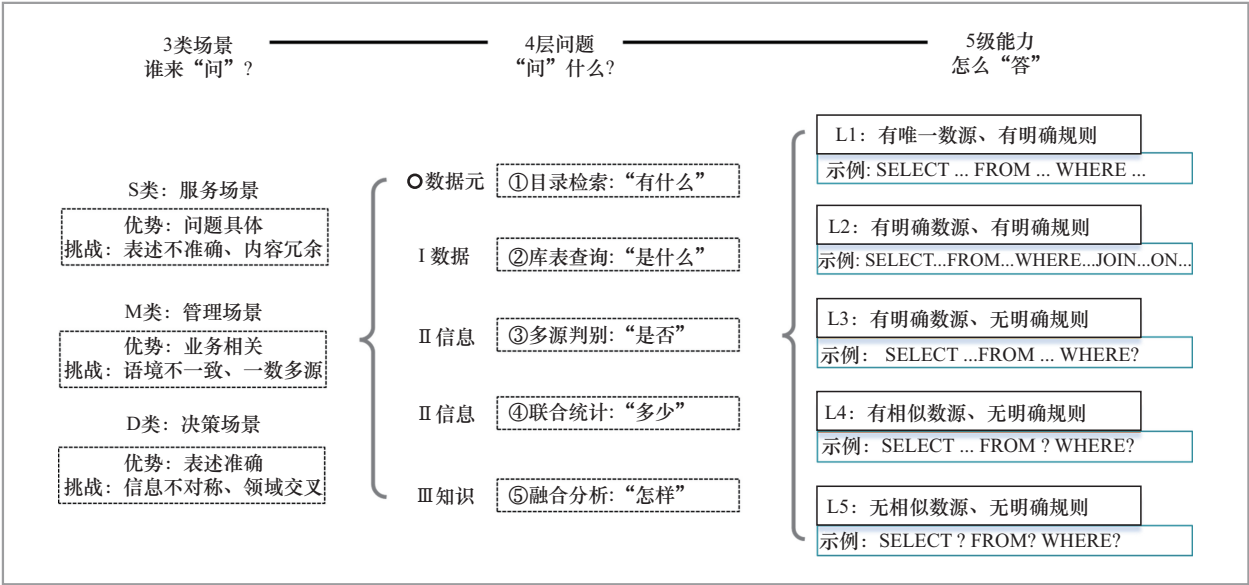


图6 问数场景的问题描述与智能化分级

据元层面的目录检索问题，问“是什么”对应数据层面的库表查询问题，问“是否”对应信息层面的多源判别问题，问“多少”对应信息层面的联合统计问题，问“怎样”对应知识层面的融合分析问题。

“问数”场景的智能响应程度，从数源和规则两个基本要素的确定上划分为L1到L5等级，其智能化水平逐级递增。其中，L1级具备规则执行能力，在有唯一数源、有明确规则时，可基于业务规则描述解析数据需求，实现场景需求到数据处理的标准化映射，以SQL查询语言为例（下同），能自动生成结构完整的SQL查询语句；L2级具备简单推理能力，在有明确数源、有明确规则时，可基于预设逻辑解析多源数据关联需求，实现多源数据协同场景下的复杂检索与逻辑关联，如自动生成结构完整且支持关联操作的SQL查询语句；L3级具备复杂推理能力，在有明确数源、无明确规则时，可理解并识别模糊业务需求，从非显性规则中推理数据处理逻辑，实现查询条件的智能化构建与语义映射，如自

动推理SQL查询中WHERE子句的条件逻辑；L4级具备认知学习能力，在有相似数源、无明确规则时，可通过学习挖掘相似数源的数据模式和关联关系，从模糊业务需求中自动提取数源映射规则，实现数据结构的智能化构建与多源数据语义映射，如自动推理SQL查询中FROM子句的多表关联逻辑与WHERE子句的条件语义；L5级具备自主智能能力，在无相似数源、无明确规则时，可自主解析全域空间下无模式关联的数源特征，从自然语言需求中提取完整数据逻辑链条，实现数据模型的智能化构建与全链路语义映射，最终完成从用户意图到知识的智能转化，如自动推理SQL查询中SELECT子句的字段映射关系、FROM子句的跨表关联逻辑及WHERE子句的动态条件语义生成。

3.2 技术实现：智能问数系统

基于数据治理5.0范式的理念和技术构建智能问数系统，其技术架构如图7所示。

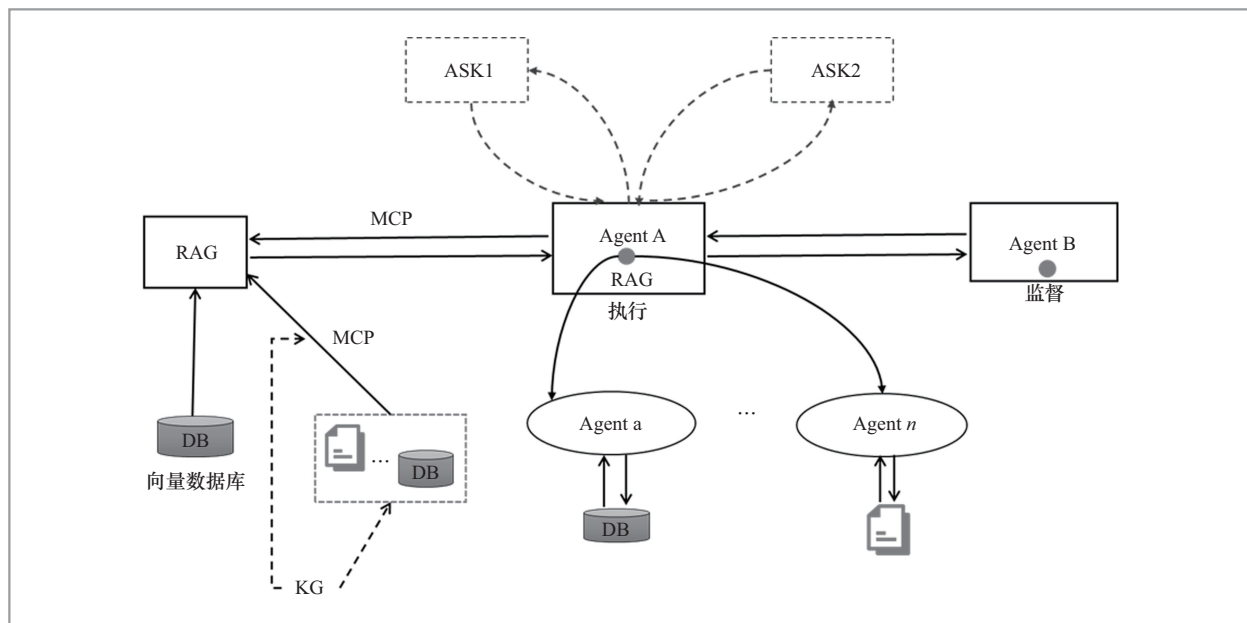


图7 智能问数系统的技术架构

智能问数系统通过一组 Agent 的相互协作实现 Text-to-SQL 的自动生成，屏蔽数据理解、检索、推理等复杂过程，解决跨域数据与多态场景之间的供需匹配问题，使场景方能快速匹配对应的数据服务。

其中，Agent A 负责理解和执行，Agent B 通过动态博弈机制对 Agent A 的决策过程、执行进度、访问权限、数据生成等进行合规性审查与监督，确保规则被正确执行。Agent A 接收到问题请求后，将问题拆解为多个可执行的操作单元（子 Agent），实现对相关数据的寻源定位、规则生成和多源融合，将场景描述转化为可执行的查询语言。在问题理解和规则生成过程中，智能问数系统通过模型上下文协议（model context protocol, MCP）连接多个检索增强生成（retrieval-augmented generation, RAG）模块，并按需连接不同领域治理后的知识图谱（knowledge graph, KG），增强复杂语义关系和知识检索的精准性。

3.3 应用实例：时空问数

时空作为数据和场景的基础属性，是城市运行、社会治理等传统领域及自动驾驶、低空经济等新领域的核心要素。智能问数系统在时空场景下的应用具有代表性，以时空问数为例进行阐述。

若想获取某市昨日降雨量，可利用 L1 级能力从跨域数据中快速匹配到气象部门的降雨数据，自动生成 SQL 查询语句，精准检索并返回特定时间和区域的具体降雨数值。若想获取某地积水相关数据，可利用 L2 级能力跨域匹配到水务部门的水文监测、气象部门的实时雨量、交通部门的道路积水监测等多源异构数据，进而根据场景描述开展多源数据协同分析。若想获取防汛相关数据，因防汛涉及多个部门且场景具有差异化和强动态性特征，暂无统一的业务规则 and 标准，可利用 L3 级能力推理关联逻辑并自动映射出雨情、水情、次生灾害、应急设施等相关信息。若想获取某

水库泄洪影响人数，因该问题缺乏标准化业务规则且无明确对应的数据来源，可利用L4级能力学习水库及潜在泄洪区域的人口、房屋、车辆、机构分布等相似数源和非结构化规则场景，解析数据处理和关联逻辑，自动实现多源数据的定位和关联统计。更进一步，若想获取某市某水库泄洪影响哪些人，因泄洪影响与时空维度高度耦合且人员动态流动，无相似数源且无明确规则，可利用L5级能力理解用户意图，融合分析高精度地址、居住人口、实时水量、多点位视频等多源数据，进行自主学习和模型构建，实现对特定群体的精准识别。

4 结束语

现有数据治理的理论和应用研究大多聚焦于特定领域或管理层面的跨域问题，本文针对“物理-数字-社会”全域空间下数据治理面临的岛化数据利用、跨域行为管控、多态场景适配等问题，创新性地提出了基于联邦机制的数据治理5.0范式，阐述了从数据共治到数据共生的核心理念和技术框架，并介绍了以智能问数为代表的典型场景应用。数据治理5.0范式可广泛应用于数据流通和开发利用的各类形态，如数据共享的高效供需匹配、数据开放的即席合规生成、授权运营的全链路穿透监管等。随着智能技术的快速迭代，数据治理在目的（why）、对象（what）、主体（who）、客体（whom）、环节（where）和方法（how）上经历深刻变革，本文的研究需要在数据与场景的多轮交互和智能问数系统的跨场景验证上进行持续优化。未来研究将重点关注以下几个方面：一是在范式层面，多模态大模型和智能体的深

度融合驱动数据治理从人工设计向自主智能演化；二是在技术层面，端到端、视觉-语言-动作（vision-language-action，VLA）和世界模型等技术加速离身智能向具身智能发展，通过联邦机制实现“数据-模型-智能体”跨域联邦，其意义将进一步凸显；三是在应用层面，意图工程将密切连接供给侧的数据工程和需求侧的场景工程，改变传统的数据整合和表达方式，以“极简+即席”的智能供需匹配重新定义数据与应用间的界面。上述研究对推动数据治理理论、技术和应用的发展具有前瞻性意义。

参考文献：

[1] Abraham R, Schneider J, Vom Brocke J. Data governance: a conceptual framework, structured review, and research agenda[J]. International Journal of Information Management, 2019, 49: 424-438.

[2] Lnenicka M, Nikiforova A, Luterek M, et al. Understanding the development of public data ecosystems: from a conceptual model to a six-generation model of the evolution of public data ecosystems[J]. Telematics and Informatics, 2024, 94: 102190.

[3] 孙金焯, 郭树行. 面向权属治理的分布式数据空间架构模型研究[J]. 大数据, 2025, 11(2): 140-151.

Sun J Y, Guo S X. Research on distributed data spatial architecture model for ownership governance[J]. Big Data Research, 2025, 11(2): 140-151.

[4] Biagi V, Russo A. Data model design to support data-driven IT governance implementation[J]. Technologies, 2022, 10(5): 106.

[5] Zorrilla M, Yebenes J. A reference framework for the implementation of

- data governance systems for industry 4.0[J]. *Computer Standards & Interfaces*, 2022, 81: 103595.
- [6] Kalaboukas K, Kiritsis D, Arampatzis G. Governance framework for autonomous and cognitive digital twins in agile supply chains[J]. *Computers in Industry*, 2023, 146: 103857.
- [7] Karkošková S. Data governance model to enhance data quality in financial institutions[J]. *Information Systems Management*, 2023, 40(1): 90–110.
- [8] Torabi F, Squires E, Orton C, et al. A common framework for health data governance standards[J]. *Nature Medicine*, 2024, 30(1): 26–29.
- [9] Althati C, Tomar M, Shanmugam L. Enhancing data integration and management: the role of AI and machine learning in modern data platforms[J]. *Journal of Artificial Intelligence General Science*, 2024, 2(1): 220–232.
- [10] 程学旗, 刘盛华, 张儒清. 大数据分析处理技术新体系的思考[J]. *中国科学院院刊*, 2022, 37(1): 60–67.
Cheng X Q, Liu S H, Zhang R Q. Thinking on new system for big data technology[J]. *Bulletin of the Chinese Academy of Sciences*, 2022, 37(1): 60–67.
- [11] Pawłowski M, Wróblewska A, Sysko-Romańczuk S. Effective techniques for multimodal data fusion: a comparative analysis[J]. *Sensors*, 2023, 23(5): 2381.
- [12] Masmoudi M, Ben Abdallah Ben Lamine S, Karray M H, et al. Semantic data integration and querying: a survey and challenges[J]. *ACM Computing Surveys*, 2024, 56(8): 1–35.
- [13] Appalapuram V S R. The lakehouse paradigm: converging data lakes and warehouses for integrated enterprise analytics[J]. *Journal of Computer Science and Technology Studies*, 2025, 7(4): 641–648.
- [14] Macías A, Muñoz D, Navarro E, et al. Data fabric and digital twins: an integrated approach for data fusion design and evaluation of pervasive systems[J]. *Information Fusion*, 2024, 103: 102139.
- [15] Otto B, Jarke M. Designing a multi-sided data platform: findings from the International Data Spaces case[J]. *Electronic Markets*, 2019, 29(4): 561–580.
- [16] 梅宏, 杜小勇, 金海, 等. 大数据技术前瞻[J]. *大数据*, 2023, 9(1): 1–20.
Mei H, Du X Y, Jin H, et al. Big data technologies forward-looking[J]. *Big Data Research*, 2023, 9(1): 1–20.
- [17] Micheli M, Ponti M, Craglia M, et al. Emerging models of data governance in the age of datafication[J]. *Big Data & Society*, 2020, 7(2): 2053951720948087.
- [18] 贾晓丰, 高嵩, 周琰, 等. 一种面向超大城市治理的数据高效跨域流通技术框架[J]. *数据与计算发展前沿*, 2023, 5(5): 35–45.
Jia X F, Gao S, Zhou Y, et al. A technical framework of efficient cross-domain data circulation for mega-city governance[J]. *Frontiers of Data & Computing*, 2023, 5(5): 35–45.
- [19] Tu J H, Fan J, Tang N, et al. Unicorn: a unified multi-tasking model for supporting matching tasks in data integration[C]//*Proceedings of the ACM on Management of Data*. New York: ACM, 2023: 1–26.
- [20] 杜小勇, 李彤, 卢卫, 等. 跨域数据管理[J]. *计算机科学*, 2024, 51(1): 4–12.
Du X Y, Li T, Lu W, et al. Cross-domain data management[J]. *Computer Science*, 2024, 51(1): 4–12.
- [21] 贾晓丰, 高嵩, 江茜, 等. 面向城市复杂系统的社会计算关键技术研究[J]. *智能科学与技术学报*, 2021, 3(2): 228–233.
Jia X F, Gao S, Jiang X, et al. Research on key technologies of social computing for urban complex system[J]. *Chinese Journal of Intelligent Science and Technology*, 2021, 3(2): 228–233.
- [22] Nguyen T L, Nguyen L, Hoang T, et al. Blockchain-empowered trustworthy data sharing: fundamentals, applications, and

- challenges[J]. ACM Computing Surveys, 2025, 57(8): 1–36.
- [23] Pincheira M, Donini E, Vecchio M, et al. A decentralized architecture for trusted dataset sharing using smart contracts and distributed storage[J]. Sensors, 2022, 22(23): 9118.
- [24] Alsamhi S H, Myrzashova R, Hawbani A, et al. Federated learning meets blockchain in decentralized data sharing: healthcare use case[J]. IEEE Internet of Things Journal, 2024, 11(11): 19602–19615.
- [25] Nezami M R, de Bruijne M L C, Hertogh M J C M, et al. Collaboration and data sharing in inter-organizational infrastructure construction projects[J]. Sustainability, 2022, 14(24): 16835.
- [26] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]// Proceedings of the 38th International Conference on Machine Learning. Cambridge: PMLR, 2021: 8748–8763.
- [27] Froelicher D, Troncoso-Pastoriza J R, Raisaro J L, et al. Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption[J]. Nature Communications, 2021, 12: 5910.

作者简介



章敏（1992–），女，中国人民大学信息学院博士生，北京市大数据中心高级工程师，主要研究方向为数据治理与数据智能。



高嵩（1986–），女，北京市大数据中心正高级工程师，主要研究方向为数据治理、分布式系统。



贾晓丰（1985–），男，博士，教授级高级工程师，北京市大数据中心数据治理部部长，主要研究方向为复杂系统数据治理、时空智能、意图计算等。

录用日期: 2025-12-15

通信作者: 贾晓丰, jiaxf@jxj.beijing.gov.cn

基金项目: 国家自然科学基金项目(No.62436010); 国家重点研发计划项目(No.2021YFB2104800)

Foundation Items: The National Natural Science Foundation of China (No.62436010), The National Key Research and Development Program of China (No.2021YFB2104800)