

面向共享目标的科学大数据治理模型研究

彭洁

(中国科学技术信息研究所, 北京 100038)

摘要: 科学大数据治理是提升资源利用效率、提升科研透明度的重要保障。在解析科学大数据共享目标的基础上, 构建面向共享目标的科学大数据治理模型, 提出涵盖原则、范围、实施与评估3个维度的治理实施路径, 为开展科学大数据治理提供参考。

关键词: 科学数据; 数据治理; 数据服务; 科学数据共享; 治理模型

DOI: 10.3772/j.issn.1674-1544.2022.04.001

CSTR: 15994.14.issn.1674.1544.2022.04.001

中图分类号: G350

文献标识码: A

Research on Scientific Big Data Governance Model for Sharing Objectives

PENG Jie

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: Scientific big data governance is an important guarantee to improve scientific data resource utilization efficiency and scientific research transparency. On the basis of analyzing scientific big data sharing goals, this paper constructs the sharing goal oriented scientific big data governance model, and puts forward the governance implementation path covering the three dimensions of principle, scope, implementation and evaluation, which can provide a reference for the development of scientific big data governance.

Keywords: scientific data, data governance, data services, scientific data sharing, governance model

0 引言

科学数据是科技创新的重要战略资源, 主要包括在自然科学、工程技术科学等领域, 通过基础研究、应用研究、试验开发等产生的数据, 以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据^[1]。科学数据既是支撑科学研究的重要基础, 也是科学研究的重要产物和成果。为了避免重复的数据采集和管理, 提升数据使用效率, 增强科研过程透明性, 科学数据共享已成为各国政府和科学界的共识, 是科学研究和科研管理过程

中必须解决的问题。大数据时代, 科学大数据因有规模性 (Volume)、多样性 (Varity)、高速性 (Velocity)、价值性 (Value) 特征, 在数据利用过程中科学数据不断从非结构化向半结构化、结构化转变, 单一化向融合化、关联化转变, 数据共享更加复杂, 因而引发了对科学大数据的治理需求。

我国政府一直高度重视科学数据共享, 在加强数据资源整合、促进科学数据开放、推动科学数据服务方面取得了重要进展, 促进了多部门、多机构科学数据资源集聚, 在基础科学、资源环境、农业、人口与健康、工程与技术等多个重点

作者简介: 彭洁 (1965—), 女, 中国科学技术信息研究所研究员, 研究方向为信息资源管理。

收稿时间: 2022年6月2日。

领域建成了科学数据共享门户。特别是,2018年《科学数据管理办法》颁布以来,认定建设了20个国家科学数据中心,数据科学管理和有效共享的理念深入人心。但是,由于存在数据多头生产、分散管理、开放不足等问题,仍然没有从根本上解决共享问题,大量科学数据资源仍然没有实现有效集成和充分利用,支撑科学数据管理和共享的技术手段仍然有限。主要表现在以下几个方面:①可共享的基础数据资源建设不足。科学数据机构资源分散,服务较封闭,孤岛现象仍然存在,一些科学数据没有统一揭示、发现和互联互通,缺乏足够的技术与能力实现资源融合与服务集成。②共享意愿不强。政策和程序会对数据共享产生积极作用,但现有的政策在很大程度上影响了研究人员对数据共享的意愿,科学数据产权不清晰、隐私保护、安全考虑以及相关主体害怕科研竞争优势丧失导致不愿意共享数据资源,数据共享需要投入人力和时间等成本、缺乏补偿共享成本投入的激励措施也导致共享意愿不强^[2]。③共享的质量和效率不高。一方面,科学数据的内容不准确、质量不高,导致使用效果较差;另一方面,新技术应用不充分,导致科学数据的组织和揭示程度不充分,科学数据可获取性不高。不同领域的数据共享的差距较为明显,地理、生物、天文等大学科数据共享做得比其他学科好^[3]。④基于共享的服务创新不足。科学数据中心建设在很大程度上满足了若干重点专业领域的获取和使用需求,各机构的机构知识库在本机构人员发表的文献存档和共享方面有了一定的进展,但还存在共享服务针对的数据使用场景不清、服务对象和业务不够精准等问题。

鉴于此,本文引入数据治理理念,拟从体制机制的角度解决科学数据共享意愿、效率及科学数据质量问题,提升科学大数据共享水平,促进数据资产的价值增值。

1 科学大数据治理的研究现状

随着信息技术的发展,数据量日益增加,数据治理成为一个信息研究的领域。Watson等^[4]最

早提出了“数据仓库治理”的概念,开启了数据治理研究的领域。许多学者对数据治理的内涵进行了深入探讨。Wende^[5]提出数据治理就是要明确使用数据行为的决策权和责任框架,并定义数据使用角色,将责任决策区域分配给这些角色^[6],认为数据治理作为实施数据责任的通用方法,适合所有数据和组织的需求。数据治理协会(DGI)认为数据治理是对信息相关流程的决策权利和责任制度,按照商定的模型执行,描述谁可以采取什么行动^[7]。可以说,有效的数据治理已被认为是获取数据使用价值的关键^[8]。数据治理需要巩固和结合整个公司、机构的数据^[9],其重要性也得到了从业人员的认可,认为其将成为企业信息管理的新趋势^[10]。

在此基础上,数据治理已成为许多重要会议的重要主题,国际商业机器公司(IBM)数据治理委员、国际数据管理协会(DAMA)、DGI、COBIT 5等也开始重视数据治理,提出了数据治理定义。如《DAMA数据管理知识体系指南》认为数据治理是对数据资产管理行使权力和控制的活动集合^[11]。2004年,IBM制定了包括四大领域11个要素的数据治理框架和方法,指导数据治理工作的开展。2014年6月,在悉尼召开的ISO/IEC 7TC1/SC40(IT治理和TT服务管理分技术委员会)第一次全会上,中国代表团首度提出“数据治理”的概念。2014年11月,在荷兰召开的SC40/WG1(IT治理工作组)第二次工作组会议上,中国代表分析了包括DAMA、DGI、IBM、高德纳咨询公司(Gartner)等组织在内的主流的数据治理方法论、模型,提出了《数据治理白皮书》的框架设想^[12]。

近年来,国内外学者的研究逐渐增多。在国外,Lai Kuan Cheong等^[10]对数据治理的必要性方面进行了深入分析;圣加伦大学信息管理研究所的Kristin Wende进行了数据治理模型方面的研究,他的同事Kristin Weber等^[6]则提出数据治理的权变方法,Mohanapriya等^[13]提出了包括数据抽取、内容分析、数据处理计算、安全分发等内容的大数据治理实现路径。在此基础上,学者们

强调数据治理在各个领域的应用,包括将数据治理纳入智慧城市建设范畴^[14],将数据治理运用到健康领域等^[15]。在国内,学者对大数据治理的模式^[16]、方法^[17]、标准化^[18]、安全^[19]、技术^[20]、数据质量^[21]、隐私保护^[22]等问题进行了相关研究,在大数据治理应用方面开展了地学^[23]、人口健康^[24]、食品科学^[25]、科技^[26]等领域大数据共享与服务的讨论。

可以说,数据治理理论和实践的发展为科学大数据共享治理形成了重要的方法论指导。科学大数据的治理和管理活动须围绕共享的战略目标,进行人员组织、制度安排、技术支持,并明确在特定的时空范围内其组织实现科学大数据的合理开放、高效获取和满足用户使用的需求。面向开放共享讨论科学大数据治理,其治理的需求和达到的目标来自于共享,而治理的策略和实施发源于数据的所有者和管理者。因此,本文将以科学大数据资源建设和运行管理机构/组织为出发点,讨论在科学大数据共享过程中各个主体如何通过治理的体制、机制达成共享目标,保证科学大数据可见、可得和可用^[27]。

2 科学大数据共享目标解析

科学大数据共享指参与共享的主体达到共有、共用、共承受、共分担的状态。科学大数据共享从本质上来说是一种利益交换行为,参与利益交换的各个行为主体在数据资源稀缺和追求效益最大化的驱动下,实现数据的流动,共享的最终目的是最大限度地有效满足用户的需求,提升资源的利用效率^[28]。科学大数据共享的最终目的是数据需求方能够获得其满意的数据,共享的目标可以分解为科学大数据合理开放、科学大数据高效获取、科学大数据满足使用。

(1) 合理开放。合理开放是指科学大数据的拥有者在特定条件下开放其拥有的数据资源,对于公共财政资助产生的科学数据开放,需要满足潘顿原则,即这些数据应该能够通过互联网允许任何人以任何目的免费获取,包括下载、复制、分析、重新处理等;对于非公共财政资助产生的

科学数据开放,则可以通过协议或合同的形式面向特定的科研用户开放。开放的条件包括付出经费、进行等价交换、提供其他回报等。对于涉及隐私和安全的数据,如需共享,则必须进行脱敏处理,然后按照规定的渠道和方式进行共享。

(2) 高效获取。高效获取是指科学大数据开放的方式要保证数据需求方能够以便捷、经济的方式发现、获取、追踪数据,而对于数据提供方也要能够以便捷、经济的方式输出数据。对于受众单一的科学大数据获取,则采用数据提供方和数据接收方共同理解的标准、格式、状态表示,这种表示方式不一定被其他人所知晓。但对于具有公共性质的科学大数据资源,则需要采用通用的、标准化的数据描述方式、格式等,且渠道更加大众化。在当前特别需要通过互联网以门户、群组、论坛等方式揭示,且该渠道能为大多数的潜在用户所知晓。

(3) 满足使用。满足使用是指科学大数据能够被最终用户使用,且能在使用过程中产生价值,或者对其科研过程或其他业务产生正向影响。这就要求科学大数据质量较高,不仅要求形式上完整、一致,数据来源可信,内容准确,能够被用户理解,而且要求能够以更小的努力融入科研过程和科研管理过程。更为重要的是,科学大数据的生产部门,在获取和形成科学数据集的时候就要考虑最终潜在用户的需求,尽可能地以用户熟悉、了解的方法记录、传输和解析数据,从而使用户在使用数据时减少数据失真。

3 科学大数据治理目标的设立

面向共享目标的科学大数据治理的最终目标是实现科学大数据的价值,科学大数据的价值一般表现为两个方面:一是从科学大数据的生产者角度,二是从科学大数据的用户角度。对于科学大数据生产者,主要是使他们生产的数据有价值,数据实现再生性应用,超出最初的使用意图,产生新应用、拓展新领域或新方法;对于科学大数据的使用者,其价值主要体现在通过使用共享的数据,增加对现有数据、业务的理解并灵

活应用。

科学大数据的价值体现在科学大数据应用上，在科研领域，不仅需要科学大数据本身，更需要与该数据存在各种关联的数据，以实现对这些数据全方位、精准化分析。如将科学数据与科技文献数据关联、将科学数据与专利数据关联、将科学数据与仪器数据关联等。科学数据的使用（Data Uses）方式包括数据验证、数据聚合、数据挖掘、数据再利用等^[29]。科学数据从采集获取、管理、加工至分析、应用，整个链条中的每一个环节对于科学数据的价值实现和价值增值都是非常重要的。但是，在大数据时代，只有在充分共享这些科学数据的基础上，从纷繁复杂的数据关联中发现潜在的利用价值，才能真正实现其大数据的价值（图1）。

要实现科学大数据的价值和价值增值，就要使科学大数据共享的收益大于成本，至少要达到科学大数据共享的成本和收益均衡。科学大数据共享的成本主要包括：①数据生产和建设成本，包括数据采集、整理、保存、维护等方面的成本；②共享协调成本，科学大数据共享需要根据科研活动和管理活动的需求，对原来各自为政、互不往来的条块模式进行变革，原有的组织机制和 workflows 都可能发生变化，甚至一些部门的利益会受到影响，需要开展大量的协调工作；③数

据安全管理成本，科学数据共享会带来安全方面的问题，特别是大数据的共享会使原有共享体系复杂化，还会增加隐私、安全等被侵害的机会，因此需要加强数据安全；④其他成本，科学大数据共享的收益和效率主要表现为提升科研影响力与持久性、提高科学研究过程透明度、加快科学研究进程、提高科学数据资源使用效率、数据利益相关者获得学术认可等^[30]。在大数据时代，科学数据共享的收益会更加巨大，因此数据价值不一定依赖于质量、大小、规模、支持、稀有程度或费用，数据的价值会随使用而增加。

科学大数据治理要适应大数据时代科研合作带来的数据共享需求，特别要适应当前E-Science环境下科研组织方式的变革。当前，科学组织模式发生了巨大变革，从“个人英雄式”转变为“团队协作式”，建立在合作、协作基础上的科学研究活动，不仅需要创建庞大的基础数据资源，而且需要关注共享和重用已有数据。因此，在科研合作的背景下，解决数据共享问题需要包括研究机构、数据存储和管理机构等合作解决数据更广泛发现获取以及再利用的问题。主要表现在：第一，适应科研过程合作机构和人员规模不断扩大的数据共享需求；第二，适应科研过程跨学科、跨地区、跨国界合作带来的数据共享需求；第三，适应科研过程中产学研合作带来的

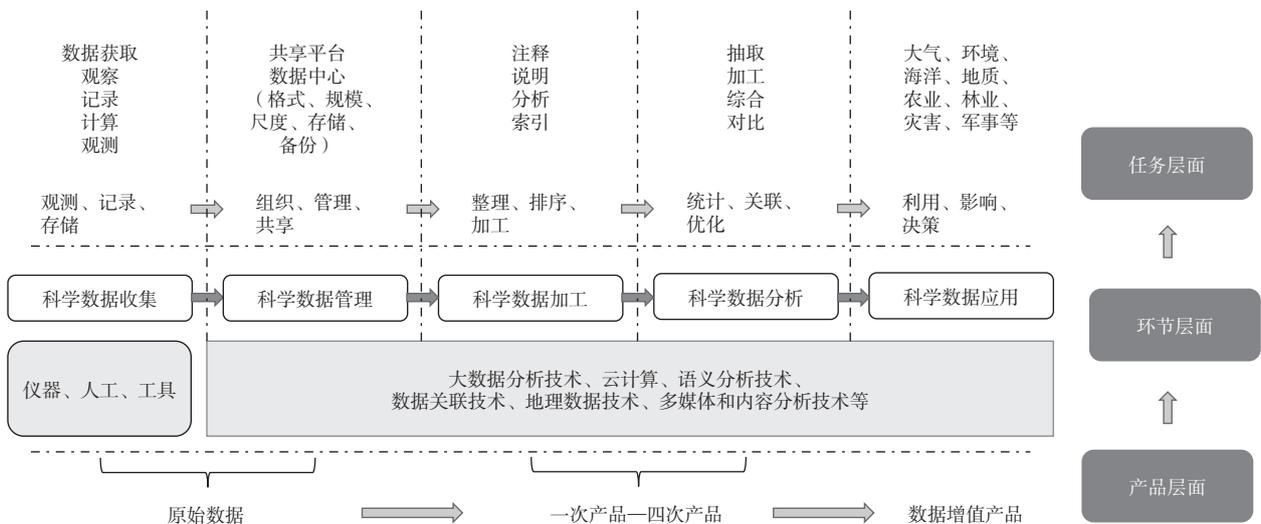


图1 科学大数据的价值实现和价值增值

科学数据向产业界转移带来的数据共享需求等。

科学大数据治理还要适应大数据时代不同学科领域特征带来的数据共享需求。在不同学科领域,自然科学和社会科学有不同的科研活动,不同科研活动产生不同的数据特征,数据类型、特征、规模、组织方式等各不相同,数据共享需求、意愿、方式等也不相同,在共享中出现的问题也不太一致。因此,治理的主体、规模、活动、决策域内容等也会有差异。如在社会科学领域的科学数据治理中,数据的隐私问题是重要的决策域;在自然科学领域的科学数据治理中,涉及地理、人种、资源等类型数据的安全问题是重要的决策域。

科学大数据治理要求是指科学大数据治理所遵循的、首要的、基本的指导性法则。对于科学大数据治理的要求,Khatri等^[31]指出有效数据治理原则是建立与业务的联系。如组织决定规范业务过程意味着应该由数据资产所有者通过描述数据的业务使用,树立数据是一个企业的资产的概念,同时数据治理原则也带来共享和重用数据的机会。数据治理原则将数据问题定义为组织政策、控制标准和程序的主题。为了管理这一资产,McGlinchey^[32]推荐应该决定大数据分析的目标,强制在整个组织的数据使用和集成,定义所有人员的数据管理和责任。在此基础上,提出了确保数据治理政策的责任矩阵(Responsibility Matrix)。数据原则应该清晰描述关于数据管理的政策和程序是如何制定的。Alhajj等^[33]指出,这些原则加速了企业组织结构的形成,从而决定了数据治理政策和程序,修复了通过数据治理程序的失败所产生的风险,沟通了数据治理决策,并提供了组织内数据管理政策形成的透明度,授权财务支出用于大数据的投资,包括技术和人员配置。

科学大数据治理要求对科学大数据治理实践起着指导作用,只有将要求融入实践过程中,才能实现科学大数据治理的战略和目标。科学大数据治理的基本要求是保证能够通过共享实现成本降低、效率提升、科研或服务可实现。科学大数

据治理与组织业务是依赖于敏捷(Agility)、信任(Trust)、智能(Intelligence)、透明(Transparency)4个支柱支持的最常见的业务。

在科学大数据治理过程中,特别需要关注以下几个问题。

第一,突出关注科学大数据“大”的特征。在大数据时代科学数据规模不断增加、类型和结构日益复杂、数据生产和处理速度提升、数据价值不断提升的背景下,科学数据的特征相比所谓的“小数据”时代发生了变化,同时相应的业务过程、应用场景、知识体系、学科体系等也都发生了变化。针对这样的变化,如何实现治理?如何去应对这样的变化?

第二,突出关注面向科学大数据“共享”目标。主要体现在:如何有效地传输、存储、保存、组织、发现、获取、集成、融合、分析、挖掘和可视化这些科学大数据;如何有效地建设支撑科学大数据全生命周期管理、支撑数据多源融合、跨学科应用和转移;如何在保障数据对象隐私和国家安全的前提下,最大程度地促进科学大数据的合理开放、高效获取和满足各种场景应用需求;如何培养符合科学大数据管理需要的各类人才,丰富其技能支持。

第三,突出关注保障科学大数据治理以“用”为主的价值导向。通过治理可以提升共享服务水平,支撑科学大数据的应用和服务。大数据时代,能够利用一些原来不能利用的方法如数据挖掘、数据融合、用户行为画像等新的方法进行各类服务和产品的设计、推广与应用,这是大数据环境下需要实现的服务,而在小数据时代无法实现,这种服务在大数据时代更加重要。科学大数据治理的原则、内容和实施都应该有助于实现这些服务,而这些服务大多是建立在数据共享的基础上才可能实现的。因此,通过实施科学大数据治理,应该能够有助于促进数据共享,进而促进数据服务创新的实现。

4 科学大数据治理模型的构建

面向共享目标的科学大数据治理模型主要

包括3部分(图2):一是业务层面,有关在科学研究、科技管理等活动中出现的科学大数据共享的业务内容,是科学大数据治理的目标,也是治理活动的需求的起源;二是技术层面,在共享目标指导下有关数据治理的对象和落脚点,所有治理活动的指导、监督、评估活动都是针对这些决策域的状态,即治理过程中的技术绩效指标(T-KPI),每一个决策域都可以单独开展治理活动,对其进行评估和修正,实现关键绩效指标的目标;三是原则层面,是共享的实施和行动,这些实施和行动都是在相应的共享原则指导下开展的,治理原则又以实现高效共享为目标。

面向共享目标的科学大数据治理模型3层之间的关系为:在共享原则和数据治理原则指导下,为了实现科学大数据高效共享,需要解决影响数据共享的意愿、质量和效率因素。只有提升数据共享意愿,提高数据共享效率和质量,才能更好地达成数据共享目标。这些共享影响因素的状态变化来源于数据治理决策域各个管理活动和状态的影响。因此,数据共享影响因素和数据治理决策域存在映射关系。而实施数据治理是基于决策域的指导、评估和监督活动,通过不断比

对、测评技术关键绩效指标(T-KPI)和业务关键绩效指标(B-KPI),校正数据治理的结果和数据共享效果间的关联关系,使之能够保证数据治理产生的结果可以正向改变共享的影响因素。

面向共享目标的科学大数据治理的基本逻辑是从科学大数据共享的目标出发,围绕科学大数据的合理开放、高效获取和合理利用,识别达到科学大数据共享的意愿、质量和效率表征的关键业务绩效指标(B-KPI),发现影响业务关键绩效指标的各种技术、政策、方法等影响因素。在科学大数据治理整体原则的指导下,以科学大数据架构、科学大数据生命周期、科学大数据质量、科学大数据安全与合规、隐私与伦理、科学大数据共享服务等关键决策域为抓手,明确达成科学大数据共享目标的关键技术绩效指标(T-KPI),包括科学大数据的获取和传输效率、数据质量、应用安全等,以这些技术关键绩效指标的正向变化促进共享为导向,在科学大数据共享的组织架构和权责体系的约束下,开展科学大数据共享治理,通过指导、监督和评估等治理活动,保证在技术关键绩效指标的指导下,治理活动能够调节和干预科学大数据共享的影响因素,正向

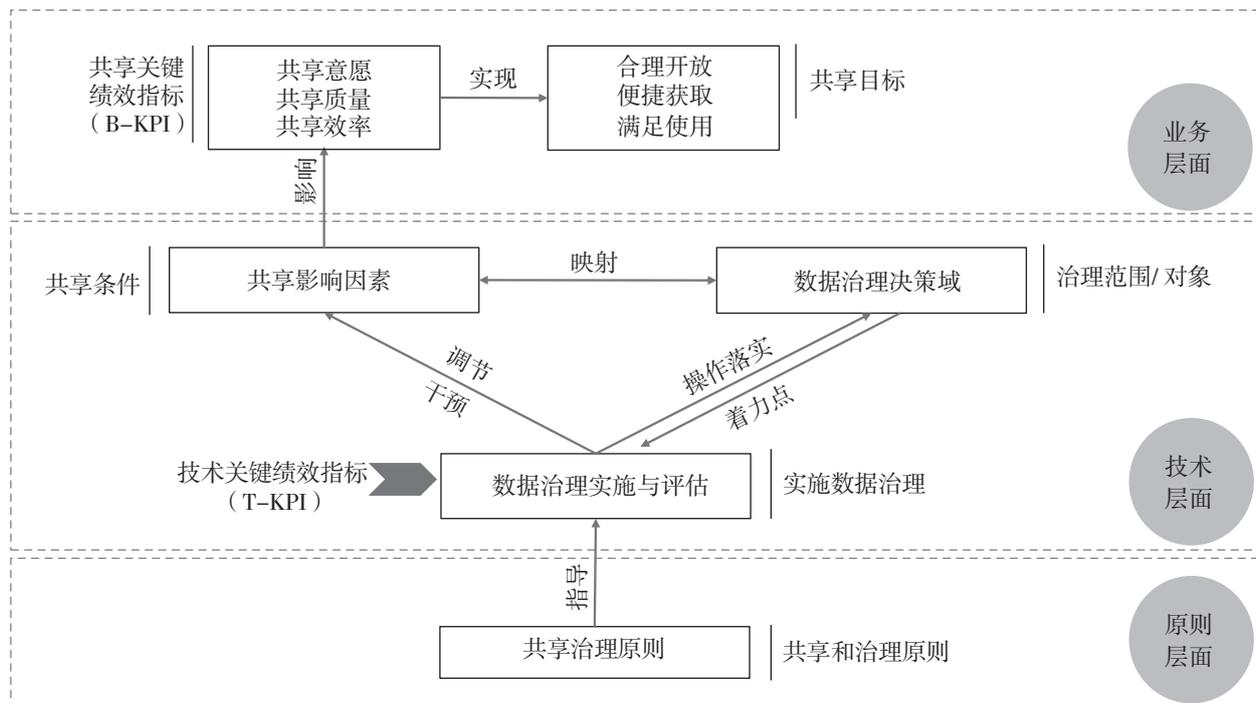


图2 面向共享目标的科学大数据治理的三层模型

影响科学大数据共享达成业务关键绩效指标。

在面向共享目标的科学大数据治理过程中，技术关键绩效指标（T-KPI）与业务关键绩效指标（B-KPI）存在着多对多的作用传导机制，每一个技术关键绩效指标（T-KPI）的提升都会带来所有业务关键绩效指标（B-KPI）的变化和提升，从而更好地促进实现共享。每一个业务关键绩效指标（B-KPI）又是由多个技术关键绩效指标（T-KPI）决定的。在不同学科领域、不同的共享模式选择下，不同的技术关键绩效指标对业务关键绩效指标作用的力度不同。如从共享模式角度看，在提供科学大数据的主体较少的共享模式下，需要重点解决从数据提供者到需求者的数据传递速度问题，突出共享的及时性。在提供数据的主体较多的共享模式下，重点要解决对科学大数据的产权界定和科学大数据共享中各利益相关者利益分配的问题，突出共享的合规性。从共享数据的类型角度看，对于自然科学领域的大数据共享，要重点针对科学大数据质量这一关键决策域来解决共享过程中数据获取、数据组织和数据揭示等问题，突出共享数据的准确性、一致性；对于社会科学领域的大数据共享，重点要解决的是数据在共享过程中的溯源问题及共享过程中对数据描述主体（特别是涉及人员）隐私和伦理等问题。总之，面向共享的科学大数据治理实施，核心是通过组织、制度、流程的建设和执行得以落实，数据治理执行效果就是这些组织、制度、流程落实的效果。

5 科学大数据治理的技术框架

根据数据治理的内涵分析，综合考虑已有数据治理框架模型的侧重点，以张绍华等^[34]、包冬梅等^[35]、张明英等^[12]提及的立体模型为基础，考虑科学大数据共享的基本要求和特征，在对上述模型修正的基础上提出科学大数据技术框架。如对于科学大数据来说，其知识产权情况复杂、难确权，因此在决策域中将科学大数据的权属纳入监管范围等。最终将面向共享目标的科学大数据治理模型三层框架模型中的技术层面展现为一个

立体三维结构的科学大数据治理技术框架模型，如图 3 所示。

科学大数据治理技术框架模型包括 3 个维度，即原则、范围、实施与评估。

（1）原则维度。原则维度给出了科学大数据治理工作所遵循的、首要的和基本的指导性法则，即治理战略与共享战略一致、隐私和安全风险可控、治理行动合规、共享和服务绩效提升，如图 3 所示的顶面。

（2）范围维度（决策域）。范围维度（决策域）描述了科学大数据治理的关键域，即科学大数据治理决策层应该在哪些关键领域内做出决策，如图 3 所示的正面。科学大数据治理的目标是实现科学大数据共享，围绕共享目标，治理要实现科学大数据的合理开放和有限开放，实现科学大数据的便捷、高效获取，实现科学大数据的精准共享服务，满足科研人员和管理部门的需求。在这个过程中，需要对实现这些目标的各个决策域开展工作，并进行评估、监督和指导，以保证这些流程和行动合规。该维度共包含 5 个关键域：科学大数据架构、科学大数据生命周期、科学大数据质量、科学大数据安全隐私与合规和科学大数据共享服务。这 5 个关键域就是科学大数据治理的关键决策领域（Decision Domain），面向共享开展科学大数据治理，就是要针对科学大数据治理中的关键决策域，实现科学大数据的合理开放和有限开放、便捷高效获取及精准服务。为了实现科学大数据合理开放，需要保证科学大数据共享中的安全、隐私和合规；为了实现科学大数据便捷高效获取，需要有合理、先进、实用、灵活的科学大数据基础设施和架构；为了实现科学大数据的共享和应用服务，必须提升科学大数据的质量，并不断进行共享服务创新。

（3）实施与评估维度。实施与评估维度描述了科学大数据治理实施和评估过程中需要重点关注的关键内容，如图 3 所示的侧面。该维度共包含 4 个部分：促成因素、治理实施过程、治理成熟度评估和审计。组织可根据原则维度中的 4 个指导原则，对范围维度中的关键决策域，按照实

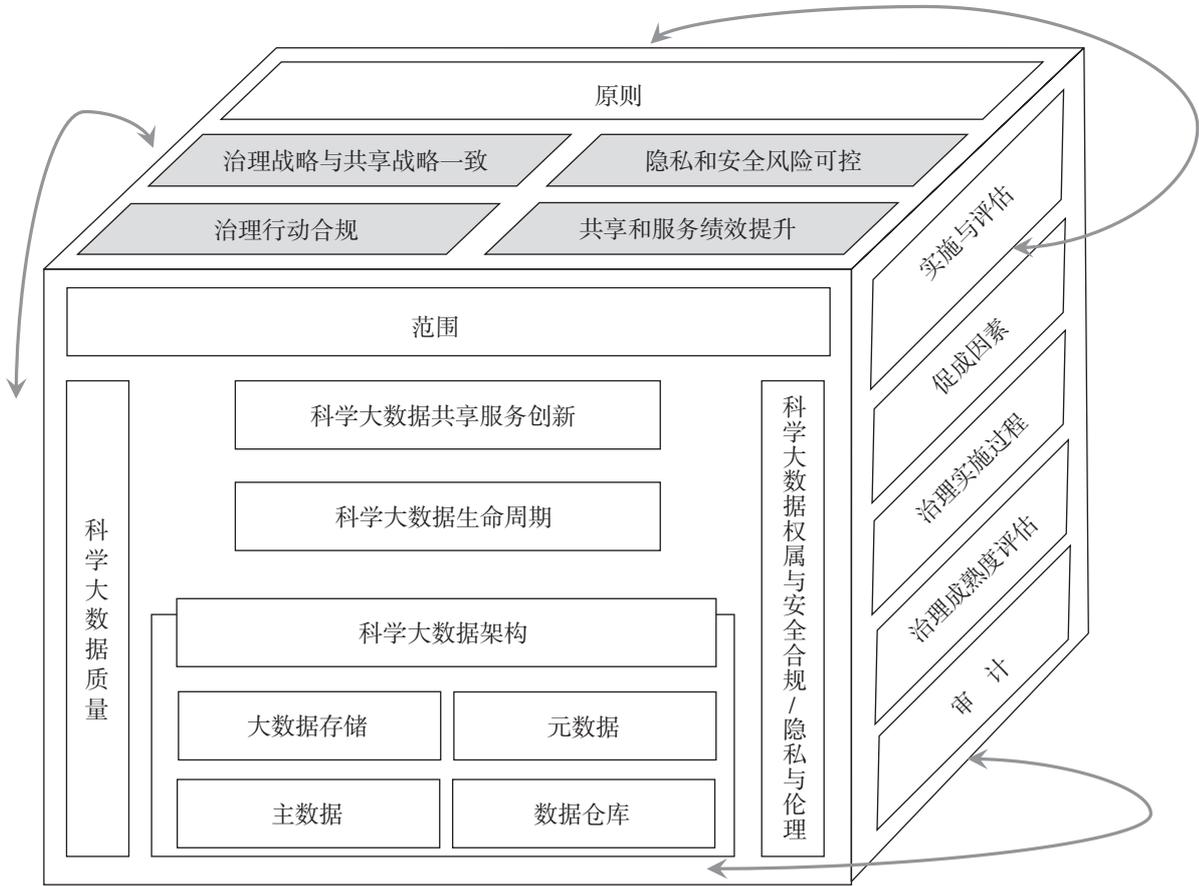


图3 科学大数据治理技术框架模型

施与评估维度中的方法论，持续稳步推进科学大数据治理工作。

科学大数据治理技术框架模型的3个维度之间相互关联，形成了数据治理的技术体系。科学大数据治理的原则决定了治理范围和实施流程、评估指标，科学大数据治理决策域是治理的具体对象，是治理实施的依托点和落脚点。将图3所示的数据治理技术框架三维模型放入科学大数据共享的环境中，构成面向共享目标的科学大数据治理技术的三层框架模型，如图4所示。从图4可以看出立体模型与三层模型的关系：在技术层面的治理范围/对策是立体模型的正面，在技术层面的实施数据治理是立体模型的侧面，而在原则层面的共享和治理原则是立体模型的顶层。

数据监管的责任体系和问责制度是数据治理的核心之一，数据标准是数据治理的重要工作内容，数据质量也是数据治理的一个重要课题。数据治理与数据生命周期管理过程各个环节紧密相

关，并与系统运行维护管理过程紧密相关。总之，面向共享目标的科学大数据治理需要依据共享战略和业务要求，准确定义实现高效共享的业务关键绩效指标，并映射到具体的数据治理技术绩效指标上，再依据科学大数据相关标准、制度和政策，通过对决策域实施指导、评估、监督来落实相关的治理活动，保证能够实现科学大数据的一致、完整、安全、准确、可用、高效，达到调节和干预科学大数据共享中的共享意愿、共享质量与共享效率影响因素的效果，促成科学大数据合理开放、高效获取和合理利用。

面向共享目标开展科学大数据治理，可以重点解决在开展科学数据共享过程中对共享业务绩效关键指标定位不准、具体治理活动行动目标不明、治理的组织保障不足、角色职责定位不清晰等相关问题。通过实施治理活动，有效保障科学数据共享不断朝着合理开放、高效获取、满足使用的目标前进。

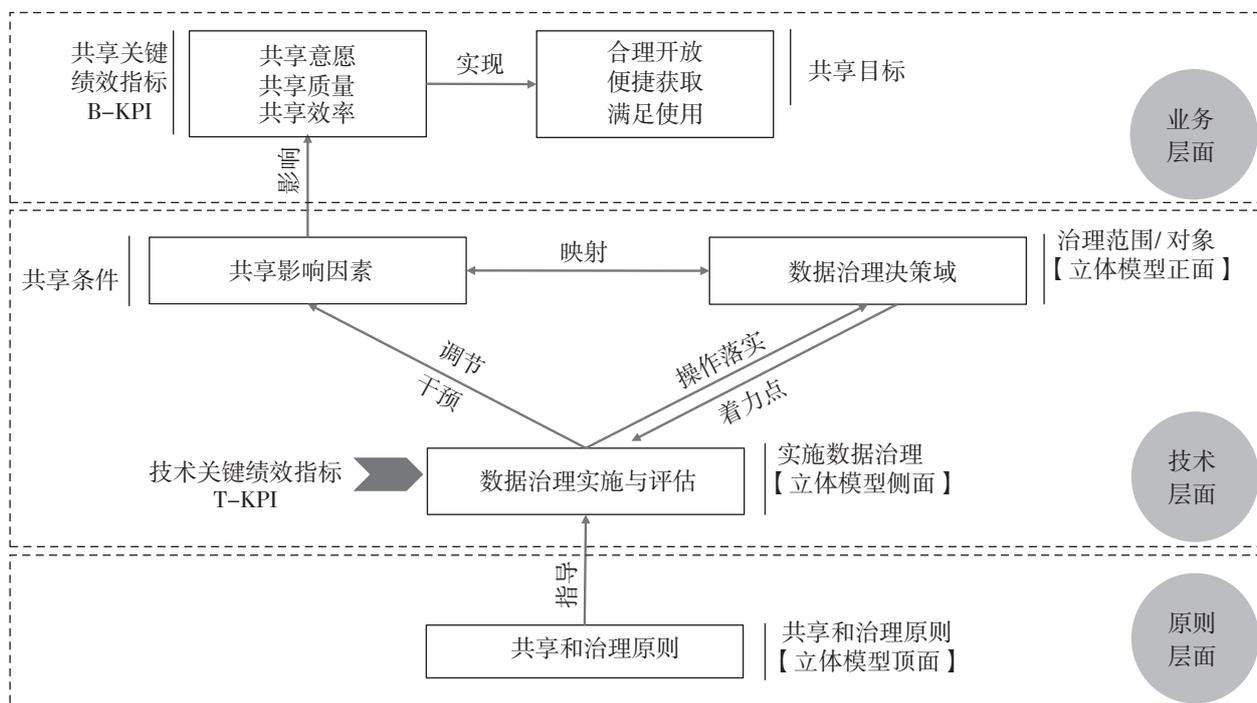


图4 面向共享目标的科学大数据治理技术框架模型与三层模型的关系

6 结语

本文提出的科学大数据治理模型可用于指导、评估和监督科学大数据的管理活动，调节和干预科学大数据共享的影响因素，进而促进科学大数据共享，为开展科学大数据治理提供参考。然而，科学大数据共享治理的实施过程，就是进行成熟度评估后，针对发现需要进行技术关键绩效指标（T-KPI）调整和提升的各个决策域，运用各种监管手段进行调节和干预的过程。特别是，围绕科学大数据共享活动，在定义业务问题、执行成熟度评估的基础上，定义数据治理路线图及定义治理相关绩效指标的度量值，开展数据治理相关活动。这些都是后续需要深入探讨的问题。

参考文献

- [1] 科学数据管理办法[EB/OL].[2022/01/20]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
- [2] 万莉,程慧平.基于元分析结构方程模型的科学数据共享行为因素研究[J].情报理论与实践,2021,44(7):125-130,158.
- [3] 屈宝强,彭洁,刘蔚,等.科学数据共享及其发展趋势[M]//情报学进展.北京:航空工业出版社,2020:381-420.
- [4] WATSON H J, FULLER C, ARIYACHANDRA T. Data warehouse governance: best practices at blue cross and blue shield of north carolina[J]. Decision support systems, 2004, 38(3): 435-450.
- [5] WENDE K. A model for data governance organising accountabilities for data quality management[C]//Proceedings of Australasian Conference on Information Systems(ACIS 2007). Toowoomba, Australia:ACIS, 2007.
- [6] WEBER K, OTTO B. One size does not fit all: a contingency approach to data governance[J]. Journal of data & information quality, 2009(1): 1-27.
- [7] THOMAS G. The DGI data governance framework[EB/OL]. [2017-03-26]. http://www.datagovernance.com/wp-content/uploads/2014/11/dgi_framework.pdf.
- [8] COHEN R. BI strategy: what's in a name? data governance roles, responsibilities and results factors[EB/OL]. [2017-02-16].<http://www.information-management.com/news/columns/-1057220-1.html>.
- [9] GAO J, KORONIOS A, SELLE S. Towards a process

- view on critical success factors in big data analytics projects[EB/OL].[2017-05-10].<http://pdfs.semanticscholar.org/2d17/7fc156d0fedca15ef8a8f066d5f54a515e60.pdf>.
- [10] CHEONG L K, CHANG V. The need for data governance: a case study[C]//Proceedings of the 18th Australasian Conference on Information Systems. Toowoomba, Australia: ACIS, 2007: 999-1008.
- [11] 李善斌. 银行业数据治理分析[J]. 金融科技时代, 2017(4): 36-38.
- [12] 张明英, 潘蓉. 数据治理白皮书国际标准研究报告要点解读[J]. 信息技术与标准化, 2015(6): 54-57.
- [13] MOHANAPRIYA C, BHARATHI K M, ARAVINTH S S, et al. A trusted data governance model for big data analytics[J]. International journal for innovative research in science & technology, 2014(7): 307-309.
- [14] KRASSIMIRA P, JAMES E, CHRISTOPHER M, et al. Data governance in the sustainable smart city[J]. Informatics, 2017, 4(4): 1-19.
- [15] ALOFAYSAN S, ALHAQBANI B, ALSEGHAYYIR R, et al. The significance of data governance in healthcare: a case study in a tertiary care hospital[C]//Proceedings of the International Conference on Health Informatics. Angers, France: ESEO, 2014: 178-187.
- [16] 李振, 鲍宗豪. 云治理: 大数据时代社会治理的新模式[J]. 天津社会科学, 2015(3): 62-67.
- [17] 徐建忠, 张亮, 李娇娇. 数据智能分类技术在数据治理中的应用研究[J]. 信息安全与通信保密, 2016(6): 88-90.
- [18] 阚鑫禹. 数据治理标准化发展现状与启示[J]. 信息技术与政策, 2022(2): 2-7.
- [19] 李晓楠, 宋阳. 国家安全视域下数据出境审查规则研究[J]. 情报杂志, 2021, 40(10): 72-84.
- [20] 盛小平, 郭道胜. 科学数据开放共享中的数据安全治理研究[J]. 图书情报工作, 2020, 64(22): 25-36.
- [21] 盛小平, 田婧, 向桂林. 科学数据开放共享中的数据安全治理研究[J]. 图书情报工作, 2020, 64(22): 11-21, 24.
- [22] 孟小峰, 张啸剑. 大数据隐私管理[J]. 计算机研究与发展, 2015, 52(2): 265-281.
- [23] 文蝶, 黄浩, 史宇坤, 等. 地球科学大数据的管理与共享: 以英国地质调查局为例[J]. 高校地质学报, 2021, 27(1): 45-57.
- [24] 钟明, 钱庆, 吴思竹. 基于场景化的人口健康科学大数据安全治理体系构建[J]. 中华医学图书情报杂志, 2019, 28(9): 6-12.
- [25] 崔晓晖, 李伟, 顾诚淳. 食品科学大数据与人工智能技术[J]. 中国食品学报, 2021, 21(2): 1-8.
- [26] 王良熙. 支撑新型智库的科技大数据治理体系研究及应用[J]. 科技成果管理与研究, 2020(11): 27-31.
- [27] 彭洁, 赵伟, 屈宝强. 基于用户视角的科技资源开放共享评价理论模型研究[J]. 中国科技资源导刊, 2013(2): 1-5.
- [28] 彭洁, 白晨, 赵伟. 资源共享的本质问题研究[J]. 中国科技资源导刊, 2013(1): 2-7.
- [29] GARY M, 杨冠灿, 芦昆. 科学数据管理: 保障数据质量, 促进iSchools新科学研究[J]. 图书情报知识, 2013(4): 4-9.
- [30] ANDREW W. How much is enough: metadata for preserving digital data[J]. Journal of library metadata, 2010, 10(2/3): 205-217.
- [31] KHATRI V, BROWN C V. Designing data governance[J]. Communications of the Acm, 2010, 53(1): 148-152.
- [32] MCGLINCHEY J. Putting the right processes in place to benefit from big data[J]. Computer weekly, 2013, 14.
- [33] ALHAJJ R, RIDLEY M, RIFAIE M. Data governance strategy: a key issue in building enterprise data warehouse[C]. The Eleventh International Conference on Information Integration & Web-based Applications & Services(LIWAS), New York, USA: Association for Computing Machinery, 2009: 587-591.
- [34] 张绍华, 潘蓉, 宗宇伟. 大数据治理与服务[M]. 上海: 上海科学技术出版社, 2016: 38.
- [35] 包冬梅, 范颖捷, 李鸣. 高校图书馆数据治理及其框架[J]. 图书情报工作, 2015(18): 134-141.