

# 数智时代人文社科数据资源开发利用的跃迁：从观念到行动

孙建军 李 阳

**摘 要** 近年来数据资源开发利用的热潮已然在人文社科领域扩散开来,人文社科的数据资源观正在从传统的有限样本资料、“小数据”向大数据、智慧数据转变。针对当前人文社科数据资源开发利用仍然存在的“保护”“深闺”“孤岛”“同质”“展示”现象,本文构建了以主体层、内容层、保障层为核心的人文社科数据资源开发利用路径框架:主体层强调顶层规划主体、建库主体、数据用户之间的协同合作;内容层凸显逻辑维、资源维、方法维、验证维的组合路径;保障层涉及政策法规、资金项目、人才资源、文化氛围等保障措施。面向未来,人文社科数据资源开发利用需要积极打造人文社科数据资源“治理术”,在基本理念上需要从“资源基础论”向“治理能力论”转变,在具体行动上可从数据基础设施体系搭建、多学科数据资源共生共长机制构建、数据资源产权与安全有效保护、基于数据资源的服务链建构等方面着力推进。图 1。参考文献 15。

**关键词** 数智时代 人文社科 数据资源 专题数据库 数字人文 社会计算

**分类号** G203 G250

## The Leap of the Data Resources Development and Utilization in Humanities and Social Sciences in Data Intelligence Era: From Concept to Action

SUN Jianjun, LI Yang

### ABSTRACT

In recent years, data resources have been described as a new source of knowledge discovery and academic innovation in humanities and social sciences and the upsurge of data resources development and utilization has already spread in those fields. From the perspective of development, the fields of humanities and social sciences have actually experienced a process of continuous understanding and construction of data resources. The view of data resources is changing from traditional limited sample data and “small data” to big data and smart data.

At present, the phenomenons of “protection”, “seclusion”, “isolation”, “homogeny” and “show” still exist in the data resources development and utilization in humanities and social sciences. To this end, the authors think that the data resources development and utilization in humanities and social sciences need to meet the national strategic needs and assume the functional positioning of memory, production, traceability, prediction and dissemination under the perspective of humanities and social sciences in the new era. For better guidance, this paper constructs the development and utilization path framework of humanities and social sciences data resources from the theoretical perspective, which takes the subject layer, the content layer and the guarantee layer as the core. The subject layer emphasizes the cooperation among the top planning

通信作者:孙建军, Email: sjj@nju.edu.cn, ORCID: 0000-0003-2144-8253 (Correspondence should be addressed to SUN Jianjun, Email: sjj@nju.edu.cn, ORCID: 0000-0003-2144-8253)

subjects, database building subjects and data users. The content layer highlights the combination path of logic dimension, resource dimension, method dimension and verification dimension. The guarantee layer involves policies and regulations, capital projects, human resources, cultural atmosphere and other safeguard measures.

Facing the future, the data resources development and utilization in humanities and social sciences need to actively build the governmentality of humanities and social sciences data resources, and change the concept from the “resource-based theory” to the “governance capabilities theory”. Concrete actions can be promoted from the data infrastructure system construction, multi-disciplinary data resources co-growth mechanism construction, data resources property rights and security and effective protection, data resource-based service chain construction and other aspects.

The exploration and discovery of this paper is expected to provide better guidance for the construction, management and service of data resources in humanities and social sciences, and improve the data discourse construction in humanities and social sciences. 1 fig. 15 refs.

### KEY WORDS

Data intelligence era. Humanities and social sciences. Data resources. Thematic database. Digital humanity. Social computing.

## 0 引言

一直以来,资源被认为是组织竞争力和竞争优势的来源,组织的资源禀赋和资源集成能力决定其战略发展方向和高度,由此也催生了资源基础观、动态资源观、知识基础观等相关经典理论。信息资源是资源观理论架构中的重要组成部分,尤其是信息社会到来后,信息资源被视为继物质、能源之后的第三大战略资源,其重要性与日俱增。从政策层面看,早在2004年,国家就出台了《关于加强信息资源开发利用工作的若干意见》,以促进不同领域、区域、行业等信息资源开发利用工作的协调发展。到了大数据时代,数据资源一度上升到生产要素的地位,各行各业都投身到数字空间、数据空间建设之中,国家“十四五”规划也专门提出要“推动数据资源开发利用”。从资源观的演变路径可以看出,人们对资源的关注经历了一个从实体的物质类资源向虚拟的信息类资源转变的过程,而信息类资源又经历了从信息资源到数据资源的转变过程,这也恰恰符合当前的数智赋能环境。而结合信息链理论,信息类资源观的这种转变正是从下游视角揭示事物发展本质和机遇的关键。新型的数据资源观本质上就是对数据资源以及基于数据资源的相关基础理论、方法工具、应用服务等变革的观念和行动体系,而建构新环境下的数据资源观对于各行各业发展具有战略意义。

数据资源在不同领域和方向正在被不同程度地激发,但受限于基础、认知、文化等因素,数据资源建设与管理显示出典型差异。从科学研究角度看,数据资源正在被视为“珍宝”,尤其是第四范式的兴起已然催动了新兴数据资源图景的浮现。从自然科学与人文社会科学的角度看,一直以来,自然科学领域是数据资源使用的主要阵地,在数据追逐之路上也积累了较为丰富的经验,而人文社科领域由于长期受归纳、演绎和推理式研究范式的影响,对数据价值的认知相对较晚。在大数据、开放数据、智能技术革新等背景下,人文社科领域的理论、实证研究等逐渐受到数据洪流的影响,数据驱动逐渐成为人文社科“追捧”的新兴范式,数据资源则被描述为人文社科知识发现与学术创新的新兴来源,在

此背景下,人文社科领域开启了很多旧场景新问题以及新场景新问题的探索。数据资源对于人文社科领域的影响体现在多个方面,包括更容易且以较低的成本实现知识多维度刻画、降低田野研究等研究方式中的干扰性与不可重复性、增加研究与应用的“科学性”、通过相关性研究可支持人文社科的因果知识发现等。当然,也有部分学者对这种数据视域下的人文社科研究与应用产生了迟疑、质疑、否定等态度,包括数据对人的意涵的剥离、基于数据的知识发现创新度不凸显、情感和乐趣的牺牲(如“远读”)等。尽管如此,作为时代潮流中的核心因子,数据之于人文社科的奥义仍然受到绝大多数人的认可、支持和推进,数据作为资源、资产这一认知也正在被不断建构、发展和呈现。国家推动的首批教育部哲学社会科学实验室建设,其中一个重要目标就是推进人文社科领域海量数据资源建设及相关研究与服务工作。

有鉴于上述分析,本文将研究问题聚焦于人文社科领域数据资源的开发利用,试图揭示人文社科的向“数”之路,并针对人文社科数据资源开发利用现状,探索数智时代人文社科数据资源的开发利用路径,以期更好地指导人文社科领域的的数据资源建设、管理与服务,提升人文社科数据话语建设。

## 1 人文社科领域的的数据资源观转变

### 1.1 人文社科“数据取向”的缘起:从定性主导到量化崛起

人文社科是以人的活动和社会现象为研究对象的知识体系,目的在于认识世界、解释世界,进而改造世界。由于人文社科本身的属性特征和定位,长时间以来很多人文社科研究主要是通过思辨、解读、想象、洞察、推理、演绎、批判等方式来解剖问题和发现知识,对数据的依赖性并不高。与自然科学相比,人文社科对数据的使用相对滞后,且总体普及性不高。在这种定性、理论占据主导的环境下,人文社科的“数据取向”较为薄弱,对社会、文化等现象和过程的理解,往往多依赖于学者的直觉和体悟。究其原因,一方面源于人文社科本身的学科特质和固有文化,即在传统认知范畴中人文社科研究问题的解决往往需要经过持续性的“雕刻”“深挖”,而不是基于数据的单一化展现,由此导致一些学者对数据分析抱有迟疑甚至拒绝的态度;另一方面源于目前人文社科学者相对薄弱的的数据意识和数据能力,现实中有较多人文社科学者难以运用数据分析来剖析研究问题,进而习惯于用传统手段来建构研究路径。

随着科学研究的量化与规范化发展,人文社科领域也逐渐与“数据”打起交道,以数据为基础的研究呈现显著的增长趋势,量化思维逐渐渗透到人文社科的方方面面。如在经济学领域,一项研究指出,顶级经济学期刊的“理论”论文比例从1963年的50.7%下降到了2011年的19.1%,经济学已无可辩驳地成为了实证学科<sup>[1]</sup>。在社会学领域,研究人员指出,定量研究论文在1995年比例为29.3%,2012年以后稳定在35%以上,2016年该比例攀升至41.9%<sup>[2]</sup>。另外一项针对《中国社会科学》期刊中科学数据使用情况的调查表明,人文社科实证数据使用主要集中在经济学、社会学、管理学等学科,其中社会学比例达到67%,经济学达到73%<sup>[3]</sup>。此外,诸如新闻的量化转身<sup>[4]</sup>、数字人文研究中数字数据运用<sup>[5]</sup>等无不揭示了人文社科领域对数据使用的逐渐重视。可以看出,人文社科领域很多方向都经历了从“定性主导”到“量化崛起”的转变,数据之于人文社科的重要性逐渐凸显,其背后本质上是数据资源观的形成与普及。

## 1.2 人文社科数据资源观:从“小数据”到“大与智”

如前所述,人文社科领域实际上经历了一个对数据资源不断认识和建构的过程。早期,人文社科相关研究范式不完全依托于数据资源,在此“小数据”逻辑下,多重庞杂的学术范式缺乏清晰公认的程序,相关学术知识的产出也缺乏共同的评估标准<sup>[6]</sup>。实际上,人文社科领域一直充斥着各种数据资源,只是在具体研究中多采用有限样本资料、小规模抽样数据、局部数据或零散数据,而非海量和全量数据。尽管在“量化崛起”阶段也有一些研究涉猎“大数据”范畴,但总体并不多见。总体来看,这种“小数据”资源观在一定时期为人文社科的繁荣发展奠定了坚实的基础,在一些社会现象和发展规律的相关剖析和因果推理上起到了关键作用,但也存在“管中窥豹”的不足。随着以大数据、人工智能等信息技术为代表的数智化环境的到来,人文社科“小数据”资源观难以适应很多感知数据、大规模集成数据的新场景新问题研究需求,也不能契合大人文社科融合的研究与发展需要,在此背景下,“小数据”资源观亟待新的建构和突破。

从外在环境看,在信息化、网络化、数据化的时代演进中,过去封闭、被埋藏、“被保护”的人文社科资料信息逐渐被数据化处理,相关数据资源的整合集成度与开放共享度明显提升,数字人文等典型领域更是赋予了数据资源的全过程记录和立体化开放应用。由于数据环境的变化和新技术的发展,人文社科研究对数据资源的需求已经不仅仅拘泥于传统的图书、期刊、报纸等(数字化)文献资料以及汇编数据、调查抽样数据等,政策文本、社会化媒体数据、舆情数据、人际网络数据等新兴数据源,也成为人文社科研究的重要数据基础,人文社科研究的数据来源更加广泛、多源与多向,逐步走向了“大数据化”<sup>[7]</sup>。受此影响,学理研究越来越依赖于大数据资源的支撑,以通过大规模数据分析与挖掘,实现对人类行为与社会现象的精准描述、计算和认知。在此背景下,人文社科领域逐渐建构起大数据资源观。例如,很多高等院校、学术组织、图书馆等纷纷不遗余力地开展人文社科数据资源的管理、组织、服务与共享等工作<sup>[8]</sup>,力求为人文社科纵深化研究的开展提供数据资源保障。

不仅如此,为了满足集成化、关联化与个性化的知识需求,学术界提出了一个新兴理念——智慧数据——数据并非越多越好,智慧化才是大数据的核心。与大数据资源观所强调的全面、开放、协同、共享等理念相比,智慧数据则强调从大数据的“大”聚焦到知识层面的数据精华。智慧数据被认为是通过对任何规模的可信的、情境化的、相关切题的、可认知的、可预测的和可消费的数据的使用,来获得重大的见解和洞察力,揭示规律,给出结论和对策<sup>[9]</sup>。可以看出,智慧数据被视为大数据的新一轮革命,更加关注大数据“价值”(value)的实现方式和支持模式。在数字人文领域,智慧数据理念已经被广泛应用于相关学理研究与项目实践,在专注于原始资料 and 大规模数据库支持的图档博领域,智慧数据的理论与应用一直在不断深入和推进。

总之,当前人文社科领域以“大与智”为典型特征的数据资源观正在悄然形成和建构,大数据资源观拓展了人文社科研究的问题域和场景域,智慧数据资源观则为人文社科的智慧化发展注入了新的活力,促使人文社科研究不断孕育智慧基因,走向更大的“智慧”。

## 2 人文社科领域数据资源开发利用的现象

### 2.1 “保护”现象

目前来看,数据资源对于人文社科领域而言仍然是一种“稀缺”资源。实践中,持有数据资源的

研究机构、行业组织、数据商等,受规章、利益、文化等因素的综合影响,对数据资源的封闭保护倾向较严重。很多机构将数据资源看作是固有资源,从数据资源占有到数据资源开发往往都是“黑箱”过程,由此进一步拉远了人文社科学者与数据资源之间的距离。对很多人文社科研究者而言,受限于平台、资金、技术条件等方面的因素,支持其研究的数据资源的可靠性、质量与可持续供给难以得到有效保障。

## 2.2 “深闺”现象

在长期的积淀下,人文社科领域不乏一些高质量、原创、特色的专题数据库等数据资源,但很多数据资源却存在典型的“深闺”现象——相关受众的知晓度以及对外的影响力偏低,优质数据资源的潜能没有得到充分激发,相关数据服务平台应有的功能和应用价值没有得到充分挖掘。从根源上看,这可能与人文社科领域在数据驱动范式上的融入壁垒,以及相关数据资源宣传推广力度的不足等因素相关。从人文社科数据主权以及人文社科话语建设角度看,人文社科数据资源“酒香也怕巷子深”现象实际上带来了极大的资源浪费。

## 2.3 “孤岛”现象

数据孤岛在各行各业都广泛存在,人文社科领域也不例外,尤其是人文社科传统“单枪匹马”式的研究路径,使得这种数据资源的“孤岛”现象更为特殊。在人文社科领域,无论是存量数据资源,还是新建的各类增量数据资源,都存在一定程度上的局部化、碎片化问题,人文社科自有的数据资源建设呈现出多重范式体系。尽管一些机构也在摸索和推进人文社科数据资源的开放与共享,但由于数据共享方式、利益保护机制、知识产权、风险规避、激励机制、约束机制等因素,相关数据资源的协同共享机制存在不足。目前实践中主要是一些专题领域、区域联盟内的数据资源协同开发,或者是个别自建专题数据库平台的开放与共享,但其实现路径多是传统的核心机构主导模式,数据资源的价值辐射空间有限。

## 2.4 “同质”现象

在量化崛起以及数据驱动影响下,重视对数据资源的使用已然成为现代人文社科研究的新时尚,数据资源开发利用甚至被认为是人文社科的新兴救援力量,诸如数字人文研究等也因此获得大量资金、项目以及相关系列性的支持。在此背景下,各类人文社科数据平台、数字人文项目数据库、综合性开放数据服务平台等如雨后春笋般不断涌现,数据类研究成果也相应地成为人文社科研究的新业态。但由于“孤岛”现象以及“一拥而上”倾向的存在,相关领域也衍生出了一些相似或雷同的数据平台或研究成果,人文社科数据资源开发利用呈现出一定程度的同质化现象。一些人文社科数据平台存在“为建库而建库”的问题,数据资源的特色化不明显,对数据资源现状的调研以及用户数据需求的剖析也不够,甚至存在已有同类数据资源,却要“另起炉灶”的现象,做了较多无谓的重复性工作。

## 2.5 “展示”现象

数智时代的到来使得很多传统人文社科研究方法逐步让位于代码、脚本语言等,然而人文社科数据资源的开发利用不是单纯地将数据资源进行“展示”<sup>[10]</sup>,或者通过技术手段进行自动化“展现”,而是要在数据资源的基础上进一步深度挖掘,为决策主体、科研用户乃至公众提供更好的服务,包括但

不限于管理支持、文化记忆、公众科普等。人文社科研究需要有智识、有思想的探索,数据和信息不等于知识,将数据资源汇聚到一个平台仅仅是提供了一个“壳”,还必须要有的数据科学的“质”和“料”的融入,人文社科数据资源才能更好地得以开发利用。一些人文社科学术组织、人文社科智库热衷于“专题知识库建设”“指数研究”等,并致力于将相关数据资源的开发利用作为彰显机构特色的“标签”或“重点工程”。但目前来看,一些数据资源机构停留在数据资源的“多”和“大”的层面,其所提供的数据资源往往存在粗糙化、过时化等问题,难以发挥实际性的数据开放价值和学术引导价值,这种只注重“面子效应”忽视“内容质量”、只注重“短期效益”忽视“长期发展”的行为损害了人文社科数据资源开发利用的整体生态。

### 3 数智赋能的人文社科数据资源开发利用路径框架

#### 3.1 面向人文社科的数据资源开发利用基本定位

数据资源开发利用是时代语境下的新命题,是对过去“信息资源开发利用”的一种传承和升级。针对上文所提及的人文社科数据资源开发利用的几种典型现象,有必要明确并进一步提升人文社科数据资源开发利用的战略定位,以支持新环境下的人文社科知识创新。在新环境下,人文社科数据资源开发利用要面向国家战略需求,运用先进、适用的信息技术和数据技术,通过对人文社科领域数据资源的获取、汇聚、分析、处理、共享、保存等,构建基于数据资源的人文社科知识生产线,进而推出适用于人文社科场景的成果、产品与服务。

在整个社会与文化的大进程体系下,人文社科数据资源开发利用具有重要的战略意义。尤其在数智赋能环境下,人文社科数据资源开发利用被赋予了更多的功能内涵,具体表现在记忆、生产、溯源、预测、传播等方面。在记忆方面,基于海量人文社科数据资源的开发利用,可以打造出面向国家、城市、乡村等的数字记忆工程,进而推动相关文化遗产的保护与传承。在生产方面,基于多源异构数据资源的整合与分析,开拓数据驱动的知识发现模式,可以有效提升人文社科知识的生产能力,推动各类数字文化产品的创意开发。在溯源方面,通过对数据的全生命周期追溯与管理以及数据资源关联而产生的上下游关系,可以探寻事物根本、寻求历史根源、发掘关键信号和证据等。在预测方面,基于大规模数据的运算、仿真和推理,可以实现态势感知、风险评估和事件预警,为组织战略决策、风险管控等提供支持。在传播方面,基于数字技术打造的各类文化产品、节目与服务,在形式与内容方面有更具创意的表达和呈现,能够助推文化传播、增强文化认同。

#### 3.2 人文社科数据资源开发利用的路径框架

目前来看,现有的与数据资源开发利用相关的政策法规体系、资源共享平台等主要为自然科学领域提供规范和服务,而适用于人文社科领域特点的数据资源开发利用路径缺少系统性、针对性的认知和建构。为此,本文结合人文社科学科属性及现状,以人文社科数据资源独特的生命周期为参考,构建包含主体层、内容层、保障层在内的人文社科数据资源开发利用路径框架,如图1所示。其中,主体层承担着“发起”的功能,内容层承担着“实施”的功能,保障层承担着“维护”的功能。

图1所示的路径框架既可以作为整体性的人文社科数据资源开发利用战略实施参考,也可作为微观层面领域数据资源开发利用的路径参考。

(1) 主体层。主体层是人文社科数据资源开发利用的驱动力量,涉及顶层规划主体、建库主体、

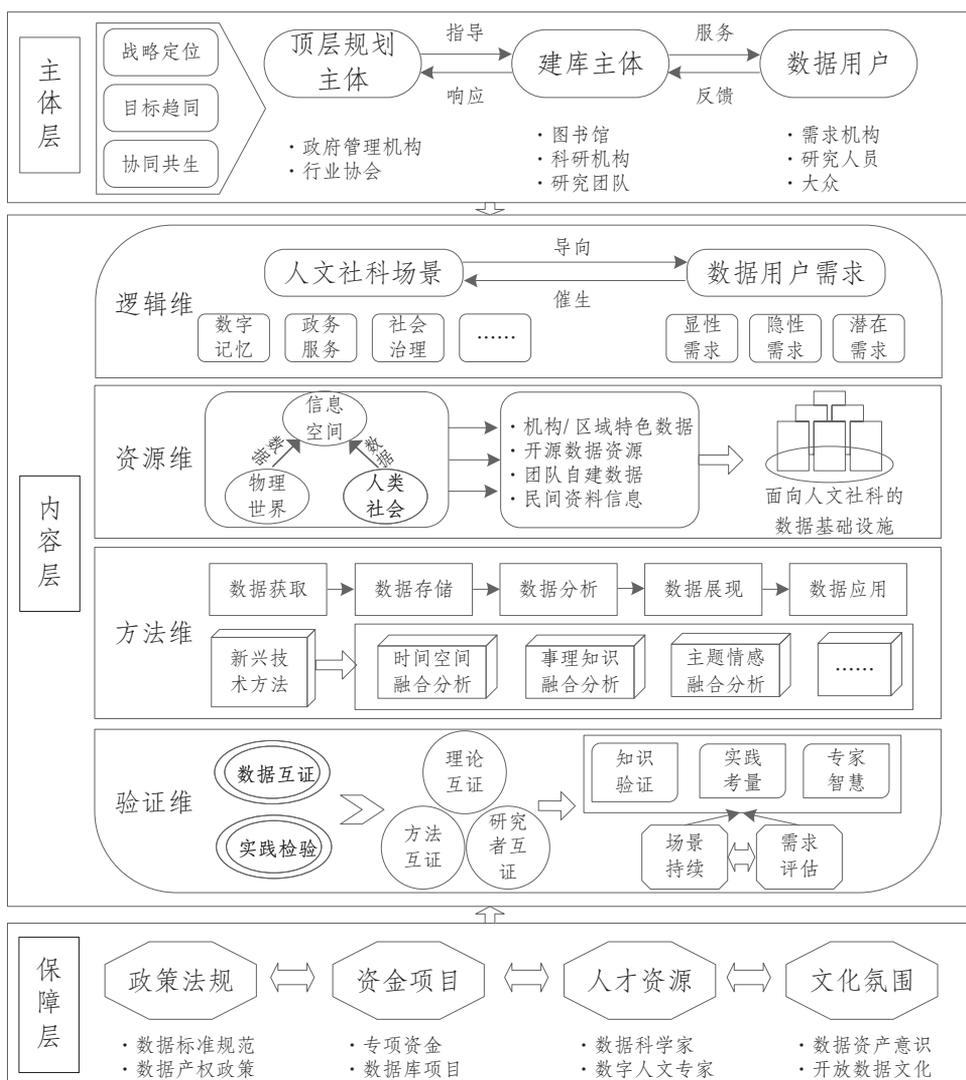


图1 人文社科数据资源开发利用路径框架

数据用户等。顶层规划主体是指统筹国家、区域、行业等人文社科数据资源的机构或组织,起着战略规划指导的作用;建库主体是具体负责特色数据资源或专题数据库项目“建、管、服”的实施者,是人文社科数据资源开发利用的核心,涉及学术型图书馆、公共图书馆、科研机构、研究团队等;数据用户是人文社科数据资源开发利用的“享用者”,既涉及业务层面的相关文化机构,也涉及科研层面的人文学者,还涉及文化学习层面的普通公众。

(2)内容层。内容层将人文社科数据资源开发利用具体化,包括逻辑维、资源维、方法维、验证维四个方面。

逻辑维是人文社科数据资源开发利用的起点。很多数据驱动研究与应用是先有数据再做分析,最后用结果反馈现实,这种逻辑脱离了现实场景与需求。随着场景驱动的兴起,以场景和场景需求为基点的数据资源开发范式能够从更为全局、更切实际的角度支持解决具体问题、满足现实需求。在数智赋能环境下,很多人文社科领域的场景体系得以开拓,如数据资源与数字记忆项目、数字政府与政务服务、智慧数据与社会治理等,通过对接业务需求、科研需求、用户需求,充分挖掘多元主体的显性、

隐性以及潜在需求,可以支持人文社科领域知识生产与创新。

资源维是人文社科数据资源开发利用的源泉,没有数据资源就如同“无米之炊”。在数智赋能环境下,人文社科领域的的数据资源观正在朝着“大与智”的总体逻辑发展。从“大”的角度看,基于三元世界大数据的资源建构方式能够从更广义视角支持人文社科研究与应用,即通过信息空间、物理空间、人类社会三者联动与映射搭建面向领域的数字资源空间,进而实现数据资源的集成、分析与服务;而从“智”的角度看,需要切中数据体系中的关键节点,将那些具有语义描述或在整体系统环境中能够得到解释或得以行动的数据资源挖掘并逐级提炼出来,以进一步凸显数据增值环节<sup>[11,12]</sup>,实现智慧服务。目前来看,以三元世界大数据与智慧数据为逻辑,充分融合人文社科领域相关机构/区域特色数据资源、开源数据资源、团队自建数据资源、民间资料信息资源等,推动跨学科、众包等数据集聚模式创新,在此基础上建构具有示范意义的人文社科数据基础设施,才能够打造人文社科领域的“数据长城”。

方法维是人文社科数据资源深度开发利用的核心。传统人文社科研究多采用系统性综述、案例研究、实证研究、扎根理论等研究方法,而在数智赋能环境下,诸如文化组学、社会感知计算、复杂网络分析、机器学习与深度学习、知识图谱、虚拟现实等新的研究方法和技术工具的不断涌现,使得过去不能研究或难以研究的人文社科问题得以“释放”。在此基础上,打造围绕人文社科数据资源生命周期的新型技术方法体系尤为重要,这种技术方法体系需要以人文思想为指导和灵魂,只有在技术与人文智慧耦合的基础上设计出凸显人文核心概念的技术方法体系,才能实现本体论和社会性层面的人文知识建构<sup>[13]</sup>。目前来看,一些新兴分析视角值得引起人文社科领域的重点关注。如时间空间融合分析视角通过时空数据分析实现“时间叙事”与“空间叙事”的融合,进而完成人文知识、城市记忆等的“深度绘图”和“历史轨迹探寻”;事理知识融合分析视角通过构建具体场景事件的知识图谱和事理图谱,可以从静态和动态联动的角度实现领域知识挖掘以及事件发展脉络梳理;主题情感融合分析视角通过事件特定主题内容与相关情感倾向的结合,可以进一步增强知识发现的用户表达。以上视角还可以通过集成和叠加来实现人文社科具体场景的分析、服务与应用。

验证维是保障人文社科数据资源产品与服务科学性并实现应用转化的关键。随着循证社会科学的兴起以及学术界对“可重复性危机”的日益关注,越来越多的人文社科研究开始注重与实践的对接,这就从反方向对人文社科“数据服务”的可靠性、可用性等提出了更高要求。可以从数据互证与实践检验两个细分维度去综合考量:一方面,为了避免人文社科数据资源本身及相关数据分析的不完整、不可靠、不精确等导致的连带性影响,有必要建构以数据互证为核心的多元互证体系,如理论互证、方法互证、研究者互证,通过这些互证体系可以避免研究结果的冲突和不一致,进一步提高数据产品与服务的品质;另一方面,还需要通过实践检验来进一步判断数据分析结果的可靠性,包括结合专家智慧、机构发展态势、用户意见等来进一步优化数据服务,为此,面向场景的周期性跟踪与用户满意度测评就显得尤为重要。可以说,验证维的运用对于人文社科数据资源开发利用是一种升华,是人文社科研究实现人文性与科学性融合的关键。

(3)保障层。保障层是人文社科数据资源开发利用的外在环境保障,涉及政策法规、资金项目、人才资源、文化氛围等方面。政策法规是指与人文社科领域数据资源标准化、开放性、安全性等问题相关的政策、制度、规章、指南等,它是保障数据资源规范化管理的关键。目前与科学数据资源相关的政策法规主要是从综合大类以及自然科学的角度建构的,围绕人文社科特点的数据资源管理政策法规比较缺乏,有待专门性推进和完善。资金项目是从“硬件投入”视角保障人文社科数据资源开发利

用工作,包括专项资金设置、对各类专题数据库的资助和支持等。人才资源也尤为重要,由于人文社科领域的数据技术基础相对薄弱,因此,培育大量的数据科学家、数字人文专家、社会计算专家等,是人文社科数据资源开发利用不断升级的智力保障。文化氛围则是“软效应”,通过培育积极的人文社科数据文化,包括数据资产管理意识、数据驱动认可文化、数据开放共享文化等,才能营造出多元主体广泛参与和开发利用人文社科数据资源的良好氛围。

## 4 面向未来的人文社科数据资源“治理术”

### 4.1 理念升级:从“资源基础论”向“治理能力论”演进

数据资源是人文社科数据驱动研究与应用的基础,但这种基础不应仅仅停留在资源建设层面。尽管有积累、有积淀、有特色的资源是机构或科研团队把握优势并脱颖而出的基础,但是面向未来环境,数据源更加多样化、开放化,技术更新使得数据开发成本更加低廉,单纯的“资源基础论”不足以应对未来人文社科数据资源开发利用的各种挑战。为此,本文认为,人文社科数据资源开发利用亟需在理念上升级,其认知体系要从“资源基础论”向“治理能力论”转变,即不再将单一的数据资源建设作为开发利用的关键任务和指向,而是要从全局治理的角度推进数据资源的开发利用,不断建构数据资源生命周期应用生态。从本质上看,“资源基础论”强调的是静态化的“资源—绩效”逻辑,这种逻辑往往局限于内部管控,忽视了外部环境以及长远创新活力。与之相比,“治理能力论”凸显的是动态化的“能力—生态”逻辑,即不只局限于数据资源保障体系建设,而是囊括了数据资源建设、管理与服务的全链条活动。

总体来看,人文社科数据资源开发利用的“治理能力论”是国家治理体系与治理能力现代化的重要组成部分,唯有强化治理理念和能力观,才能不断建构人文社科数据话语,推动人文社会科学繁荣发展。

### 4.2 面向未来的人文社科数据资源治理关键行动

#### 4.2.1 打造立体化层次化的人文社科数据基础设施体系

人文社科数据基础设施建设对于人文社科研究范式转型以及实践活动创新都具有重要支撑作用。以往很多优秀的人文社科研究成果案例都得益于数据基础设施的搭建,如斯坦福大学开发的启蒙运动知识界通信地图系统<sup>[14]</sup>。人文社科数据基础设施是面向人文社科特点建构的以数据为核心的各种单元的组合,包括数据资源、软件、服务(交流)平台、网络空间等。在数智时代,数字人文、社会计算的发展推动全新的人文社科数据基础设施建设,如复旦大学社会科学数据平台、北京大学开放研究数据平台、南京大学人文社科大数据平台等。但总体来看,目前人文社科数据基础设施体系仍然不够完善,在数量、质量方面都有待提升。随着更加开放化、协同化环境的到来,亟需从国家、行业、机构、个体出发,协同不同主体的数据力量,打造立体化层次化的公益性人文社科数据基础设施体系,不断推进典型领域和方向的人文社科实验室建设,以在哲学社会科学自主知识体系建设中下好“先手棋”。

需要特别指出的是,构建动态化的数据资源共享目录应是整个数据基础设施体系建设的重要内容。人文社科数据资源共享目录是指在人文社科相关科研探索、产业实践等过程中所构建出的各类(专题)数据资源索引目录。我国人文社科领域在长期发展实践中,尤其是在数智赋能的牵引下已经

建设和开发出较多优质、特色的数据库、知识库,规范、高效地共享这些数据资源和“拳头”数据库产品,可以融合不同优势学科力量 and 不同学科特点,集智攻关重要方向和关键领域。为此,可以通过构建动态化的人文社科数据资源共享目录,帮助学界、业界快速了解、获悉、导览以及合理使用具体领域的现有(专题)数据资源,以此驱动人文社科领域的交叉融合创新。当前需要按照人文社科领域的数据库资源主题、类型、关联领域等,在全国高校、科研院所范围内对数据库资源尤其是专题数据库资源进行摸底和统计调查,整理人文社科数据资源共享目录清单。目录清单内容应涉及数据库资源主体、名称、项目信息、数据项信息、共享提供方式、数据提供方式、数据更新周期等,以此为基础形成完善的目录体系。除此之外,还要重视目录清单的评估与考核工作,强化人文社科数据资源共享目录的动态过程管理,才能更好地发挥共享目录的价值。

#### 4.2.2 构建多学科数据资源共生共长机制

在人文社科整体量化发展和一路向“数”的过程中,不同学科的数据基础、数据取向、数据能力等实际上存在一定的差异。一般认为,相比人文领域而言,社科领域对数据库资源的使用要更多一些,但随着数字人文的发展,人文领域对于数字数据库资源的使用开始呈急速上升趋势。另外,在长期的数据库探索中,不同学科常常是“各自为战”,其所依赖和建构的数据库资源体系多局限于学科内部领域,这使得很多数据库类研究思维受限,研究质量和影响力也就被间接“打折”。随着跨学科交叉领域研究的兴起,数据库互补、知识互补、技术互补、场景互补等需求增加,人文社科不同学科之间以及与自然科学学科之间的借鉴交流显得愈加重要。在这样的背景下,构建多学科数据库资源共生共长机制迫在眉睫。

多学科数据库资源共生共长机制构建需要重点关注以下三个方面内容。首先,各个学科需要形成数据库资源价值共创的基本认知,弱化学科壁垒,以跨学科融通思维来共同推进人文社科数据库资源开发利用。其次,需要充分发挥不同学科在数据库技术、领域知识等方面的特长,通过构建数据库资源联合体、数据库资源联盟等推动大规模协同跨界合作,基于研究互补需要、团队协作需要、项目攻关需要等积极开展数据库交流和对话。例如,可以以重大人文社科项目、重大人文社科知识工程等形式集成和聚合典型领域或方向的优质数据库资源,开展更多跨学科数智赋能交叉创新研究与应用,由此能够产生更具品质和特色的人文社科“大成果”。最后,需要一些学科发挥在数据库资源开发利用中的“中介”功能,即发挥桥梁和纽带作用。图书情报与档案管理(现为“信息资源管理”)学科长期开展信息资源建设、数据库分析、信息需求和数字学术需求等相关领域的研究,理应在人文社科数据库资源治理行动中承担起重要责任,通过传承数据库信息层面的“基因密码”,打造新时代数字空间“专精特新”型品质。近年来,图书情报与档案管理学科积极顺应数智赋能环境,投身区域、行业、机构的数据库资源平台建设和数字人文项目数据库建设,主动承担人文社科数据库资源主流阵地建设的任务。其他与“数据库”关联的学科也应积极发挥各自优势和特长,共同推进多学科数据库资源共生共长机制建设。

#### 4.2.3 重视人文社科视域下数据库资源产权与安全保护

国家“十四五”规划提出要建立数据库资源产权、交易流通、跨境传输和安全保护等基础制度和标准规范。人文社科领域数据库资源开发利用除了关注学界和业界普遍提及的基本标准与规范问题外,也需要特别重视对数据库资源的产权、安全等的保护,这一重要方向在人文实践中往往会被忽视。

从数据库资源产权的角度看,由于受传统研究范式的影响,人文社科领域过去并不太关注数据库资源产权问题,而且一些领域或方向的数据库资源产权本身也难以界定,如公民科学项目中的数据库产权问题、网络数据库资源产权问题、机构合作过程中的数据库产权问题、数据库交易市场中的数据库产权问题。尤其随着知识付费、文化消费、“数据库超市”等新业态的到来和深入,一些新兴数据库场景不断呼吁数据库资

源产权问题的规范化。2022年5月,国家颁布《关于推进实施国家文化数字化战略的意见》,提出“强化文化数据要素市场交易监管”“完善文化资源数据和文化数字内容的产权保护措施”等。因此,必须完善人文社科领域数据资源产权政策法规,结合人文社科领域特点明晰各方数据行为和权利边界,进一步提高人文社科数据资源利用效率。

从数据资源安全的角度看,人文社科领域实际上存在很多与意识形态、国家政治、历史文化、公民个体相关的档案、资料、数据等,这些数据资源涉及舆论思想、国家安全、个人隐私等敏感性、导向性、保密性的信息,因此不能随意流通、加工或跨境传输<sup>[15]</sup>。在实践中,《中华人民共和国数据安全法》提出“国家建立数据分类分级保护制度”“加强对重要数据的保护”等,国家互联网信息办公室颁布的《数据出境安全评估办法》针对数据出境活动提出数据处理者向境外提供重要数据时应申报安全评估的要求。目前来看,人文社科领域缺乏对这种“关键数据资源”<sup>①</sup>的专门定义和政策规范。因此,需要对这些“关键数据资源”实施专项管控,根据人文社科领域的特点及数据资源的重要程度、数据泄露后的危害程度等,建立分类分级保护标准与政策、审批审读制度、安全风险评估机制等,确保人文社科领域数据资源、数据服务产品和传播内容等可管可控。

#### 4.2.4 推动人文社科领域建构基于数据资源的服务链

数据资源开发利用最终需要落脚到“用”的层面,为此,需要积极探索人文社科导向下的数据要素市场培育,引导数据主体以应用场景为导向,将数据服务应用与场景紧密结合起来,并进一步带动数据资源的整合、共享与协同,发挥数据资源在人文社科应用场景中的基础资源作用、文化记忆作用、创新引擎作用等,进一步推动具有人文社科特征的、基于数据资源的服务链的建构。

服务链建构需要重点关注以下三个方面。第一,积极开拓新场景。在数智时代,人文社科场景呈现不断丰富化、多元化、复杂化等趋势,需要以国家数字文化、数字经济、数字政府等战略需求为指引,不断开拓新兴服务场景,挖掘数据资源在国家战略需求下的价值潜能。第二,注重全链条式搭建。针对传统人文社科数据服务链单一、断层等不足,需要建构出涵盖数据生产、数据加工、数据运营、数据交易、数据消费、数据体验等全流程的服务链。由于不同行业或方向的运作特点和需求存在差异,可以以重点领域为突破点,如古籍数字化、文化数字化、应急大数据、健康信息化等,通过充分调动领域内各方主体参与数据资源增值开发,构建“数据资源+技术方法+应用场景”的应用生态。第三,创新服务供给模式。聚焦文化与科技融合、科技与社会治理融合等理念,不断拓展数据产品与服务的深度和广度,建构精准化、可定制、全时段的服务供给方式,开拓互联网+、虚拟现实等新兴服务供给模式,进而发挥服务链的辐射带动作用 and 集群效应。

## 5 小结

数据资源是新时代人文社科创新发展的“新能源”,数据资源开发利用是实现人文社会科学繁荣发展的关键路径。本文试图勾勒和描述数智时代人文社科数据资源开发利用的跃迁之路,尽管在总

<sup>①</sup> 对于“关键数据资源”目前并未有明确统一的定义,如文中提及的《中华人民共和国数据安全法》提到了“重要数据”这一概念,并指出“关系国家安全、国民经济命脉、重要民生、重大公共利益等数据属于国家核心数据”;《数据出境安全评估办法》涉及“重要数据”和“个人信息”概念;另外,《信息安全技术 术语》(GB/T 25069—2022)将数据划分为“重要数据”“核心数据”“一般数据”等。本文认为,“关键数据资源”可理解为给国家、组织或个体带来竞争优势或者安全风险的一类特殊数据资源,以上分类逻辑可供具体行业领域作为“关键数据资源”界定的参考。

体上论证了人文社科领域数据资源观的转变,但也不得不承认当前人文社科的“数据文化”仍然处于起步阶段,数据驱动研究范式仍然在推崇和质疑中辩证前进。针对当前人文社科数据资源开发利用的几种现象,从宏观与微观相结合的角度提出了数智时代人文社科数据资源开发利用的路径框架,希望能够指导不同领域、不同方向的人文社科数据资源建设、管理与服务。最后,着眼于未来,提出了对人文社科数据资源治理理念和关键行动的若干思考,包括从“资源基础论”到“治理能力论”的进阶,以及在基础设施、学科参与、安全保护、服务链等方面的行动措施,这些也是当前人文社科数据资源开发利用亟需解决或易被忽视的痛点和难点。

总体来看,人文社科数据资源开发利用是一个复杂的系统工程,一方面需要重视数智赋能效用优势,以“数”为“力”,不断拓展人文社科的应用场景和服务空间,另一方面需要坚持“技术理性”与“人文价值”的统一,将数据资源的“刚性”与人文实践的“柔性”结合起来,推动有温度的数据驱动研究与应用,由此才能更好地体现出人文社科的“智慧”。面向未来,我们期待更多的人文社科数据资源开发项目和工程,不断开拓新文科导向下的数据驱动研究、数据服务产品,从“数据自信”角度推动人文社会科学整体话语体系的建立。

**致谢:**本文系国家自然科学基金重大项目“人文社科专题数据库建设规范化管理研究”(项目编号:18ZDA326)的研究成果。

## 参考文献

- [ 1 ] THALER R H. Behavioral economics: past, present, future[J]. *The American Review*, 2016, 106(7): 1577-1600.
- [ 2 ] 陈云松, 贺光烨, 吴赛尔. 走出定量社会学双重危机[J]. *中国社会科学评价*, 2017(3): 15-27, 125. (CHEN Y S, HE G Y, WU S E. Overcoming the dual crisis of quantitative sociology[J]. *China Social Science Review*, 2017(3): 15-27, 125.)
- [ 3 ] 沈婷婷. 人文社科领域科学数据使用特征分析——基于《中国社会科学》样本论文的实证研究[J]. *大学图书馆学报*, 2015, 33(3): 101-107. (SHEN T T. The characteristic of research data application in humanities and social science: an empirical study of publications in *Social Science in China*[J]. *Journal of Academic Libraries*, 2015, 33(3): 101-107.)
- [ 4 ] 彭增军. 从人文到技术: 新闻的量化转身[J]. *新闻记者*, 2020(5): 36-40. (PENG Z J. From humanities to technology: the quantitative turn of journalism[J]. *Shanghai Journalism Review*, 2020(5): 36-40.)
- [ 5 ] 石静, 李阳. 数据和技术运用视角下我国数字人文“数字型”研究现状及启示[J]. *图书情报工作*, 2021, 65(21): 141-150. (SHI J, LI Y. Present situation and enlightenment of digital humanities “digital-type” research in China from the perspective of data and technology application[J]. *Library and Information Service*, 2021, 65(21): 141-150.)
- [ 6 ] 于文轩, 樊博. 公共管理学科的定量研究被滥用了吗?——与刘润泽、巩宜萱一文商榷[J]. *公共管理学报*, 2020, 17(1): 159-166, 176. (YU W X, FAN B. Is quantitative research methods abused in public management research?[J]. *Journal of Public Management*, 2020, 17(1): 159-166, 176.)
- [ 7 ] 李阳, 孙建军. 人文社科大数据研究的价值追寻[J]. *图书与情报*, 2019(1): 1-7. (LI Y, SUN J J. Value pursuit of humanities and social sciences big data research[J]. *Library and Information*, 2019(1): 1-7.)
- [ 8 ] 孟祥保, 钱鹏. 数据生命周期视角下人文社会科学数据特征研究[J]. *图书情报知识*, 2017(1): 76-88. (MENG X B, QIAN P. Characteristics of research data in humanities and social sciences from the perspective of data lifecycle[J]. *Documentation, Information & Knowledge*, 2017(1): 76-88.)

- [ 9 ] 曾蕾,王晓光,范炜. 图档博领域的智慧数据及其在数字人文研究中的角色[J]. 中国图书馆学报,2018,44(1):17-34. (ZENG M L,WANG X G,FAN W. Smart data from libraries,archives and museums and its role in the digital humanities researches[J]. Journal of Library Science in China,2018,44(1):17-34.)
- [ 10 ] BRENNAN T. The digital-humanities bust—after a decade of investment and hype,what has the field accomplished? Not much[EB/OL]. (2017-10-15) [2022-08-25]. <https://www.chronicle.com/article/The-Digital-Humanities-Bust/241424>.
- [ 11 ] 刘炜,嵇婷. “云瀚”与智慧图书馆:以开放创造未来[J]. 中国图书馆学报,2021,47(6):50-61. (LIU W,JI T. Yunhan and the smart library:adapting to the future with openness[J]. Journal of Library Science in China,2021,47(6):50-61.)
- [ 12 ] 张云中,刘嘉琳. 智慧数据研究综述:概念辨析、价值取向、关键技术与应用框架[J]. 图书情报工作,2021,65(10):141-150. (ZHANG Y Z,LIU J L. Smart data research:concept discrimination,value orientation,key technologies and application framework[J]. Library and Information Service,2021,65(10):141-150.)
- [ 13 ] 冯惠玲. 数字人文——在跨界中实现交融[N]. 中国社会科学报,2017-12-21(008). (FENG H L. Digital humanities:blending in cross border[N]. Chinese Social Sciences Today,2017-12-21(008).)
- [ 14 ] 刘炜,谢蓉,张磊,等. 面向人文研究的国家数据基础设施建设[J]. 中国图书馆学报,2016,42(5):29-39. (LIU W,XIE R,ZHANG L,et al. Towards a national data infrastructure for digital humanities[J]. Journal of Library Science in China,2016,42(5):29-39.)
- [ 15 ] 王晓光. 加强人文社科数据资源建设与管理[N]. 光明日报,2018-07-05(11). (WANG X G. Strengthening the construction and management of data resources in humanities and social sciences[N]. Guangming Daily,2018-07-05(11).)

孙建军 南京大学信息管理学院教授,博士生导师。江苏 南京 210023。

李 阳 南京大学信息管理学院副教授,硕士生导师。江苏 南京 210023。

(收稿日期:2022-10-04)