DOI：10.13998/j.cnki.issn1002-1248.26-0032

# 档案数据智能范式意蕴解析与图景建构

杨　鹏[1,2]，乔菁菁[1]

（1. 上海大学 文化遗产与信息管理学院，上海 200444；2. 上海大学 档案事业发展研究中心，上海 200444）

**摘　要：**[目的/意义]数智时代，大数据与人工智能技术深度融合，驱动档案资源形态从传统模拟态向数字态、数据态跃迁，档案管理对象变革必然引发学科范式的立体式重构，档案数据智能范式建构成为档案学科发展重大议题。[方法/过程]基于科学范式理论，结合档案学科范式基础，通过文献梳理、模型建构与理论融合，探索揭示档案学科范式从档案文件史料范式到档案信息资源范式，进而迈向档案数据智能范式的演化脉络与历史逻辑。[结果/结论]档案数据智能范式以档案数智化为内核，以数据要素赋能与人工智能赋能为引擎，以智慧档案为大脑，以算力中枢与算法模型为驱动，以高质量档案数据集为支撑，形成涵盖数智文化、数智技术、数智人才、数智管理于一体的档案数据智能理论知识框架与实践发展体系。通过解析档案数据智能范式的理论意蕴，以哲学基础与实践发展为依据，从本体论、认识论、方法论、技术论与价值论等维度，系统建构档案数据智能范式图景，为档案事业数智化转型提供理论借鉴与实践指引。

**关键词：**数智时代；档案数据；智能范式；数智赋能；图景建构；数据要素

**中图分类号：**G270　　**文献标识码：**A　　**文章编号：**1002-1248（2026）10-0039-11

**引用本文：**杨鹏,乔菁菁. 档案数据智能范式意蕴解析与图景建构[J]. 农业图书情报学报,2026,38(10):39-49.

## 0　引　言

当前，大数据、人工智能等新一代信息技术正以前所未有的广度、力度与深度重塑着人类生产方式与国家治理模式，进而引发社会全局性、系统性深刻变革。“数据为王、智慧引领”的数智时代已悄然来临，据国际数据公司（IDC）推测，2025年全球每年产生的数据将达到175ZB，每天大约会增加491EB的数据[1]，数据作为重要战略资源、信息资产和社会资本，已成为数智时代最核心的生产要素和最活跃的要素质态[2]。在面临数据驱动、智能赋能的数智时代浪潮下，作为记录社会活动、承载集体记忆的档案，其资源形态、管理对象与价值实现路径同样面临着根本性挑战与历史性变革。档案资源正从传统的“实体文件”与“信息资源”向海量多元的“档案数据”转型，档案学科固有的基础理念、管理框架与实践方法在应对数智洪流冲击时，正迎来一场内外双重革命，驱动档案学科范式重构。基于中国知网文献库，以“档案学范式”为主题进行检索，检索时间为2026年3月4日，共检索到文献169篇，通过分析发现，相关学者主要从档案观念转换[3]、记忆范式基础[4]、管理范式变迁[5]等角度探索档案学科范式。与此同时，已有学者关注到数智赋能对档案学科发展与档案事业创新带来的深刻影响[6]，档案数据智能范式已成为关乎档案学学科

收稿日期：2026-01-21

基金项目：国家社会科学基金青年项目“档案数据价值链运行机理与实现路径研究”（25CTQ049）

作者简介：杨鹏（1996- ），男，博士，讲师，研究方向为档案数据治理、档案数智化。乔菁菁（2001- ），男，硕士研究生，研究方向为档案数据治理

DOI：10.13998/j.cnki.issn1002-1248.26-0032

未来发展与档案事业兴衰的战略性议题。为此，立足数智时代环境，结合笔者前期研究基础[7]，深入解析与全面探究档案数据智能范式理论内涵，系统建构档案数据智能范式发展图景，推动档案学科范式数智化演进。

## 1　档案学科范式的数智变迁

“范式”是科学哲学家托马斯·库恩科学观中的核心概念，其代表着一个由特定共同体的成员所共有的信念、价值、技术等构成的整体[8]。范式是一门学科科学共同体共识性的、具有元科学意义的概念、理论、工具和方法论的集合，理论是其主要表现形式[9]。认知功能和规范功能是范式的基本功能，它们具体揭示了在某一特定历史时期内学科的重点关注议题，以及学术共同体在学科研究与工作实践中所共同遵循的一些基本理论、观点、方法和原则。

### 1.1　档案学科范式的逻辑起点

档案学的逻辑起点主要围绕对档案本质及其形成规律的探索而展开。这一起点将档案的本质及其形成规律的探索作为学科研究的核心对象，关注其产生、运动和发展的基本规律[10]。对逻辑起点的坚守与阐释则是范式发展的内在连续性保证。传统上，档案被视为一种静态的、具象化的历史记录与可信凭证，其形成规律遵循着文件生命周期的立体化运动发散模式。这一逻辑起点催生了以实体管理为核心、以保管利用为导向的档案文件史料范式，并延伸至以信息管理为核心、以藏用并重为导向的档案信息资源范式，二者理论与方法体系均建立在对物理载体及其内容信息的有效控制之上。

随着数智时代的到来，档案资源形态与管理对象从物理实体、数字信息跃迁为数据单元，文本、图像、音频、视频、三维多媒体等各类结构化、半结构化、非结构化的数据成为档案资源的主要存在形式与运行形态。相应地，可被机器识别理解、语义关联、计算分析的细颗粒度比特字节成为档案管理基本范畴，档案价值以凭证情报价值为根基，拓展至档案数据要素价值，如何挖掘释放档案数据要素价值潜能成为档案事业高质量发展的时代命题。人工智能技术的迅猛发展，为破解档案数据要素价值挖掘开发难题提供新路径、新手段、新方法，赋予档案数据要素智能化新内涵，统合档案的数据属性、智能属性和价值属性，全面推动档案学科范式变革，形成基于数据、赋以智能、创造价值的档案数据智能范式新图景。总体上看，档案学科范式无论如何变迁，其逻辑起点始终是档案管理实践的发展，围绕档案进行的实体物理管理、数字信息管理与数据智能管理，构成档案学科范式的实践根基。

### 1.2　档案学科范式的演进脉络

借鉴科学哲学家托马斯·库恩的范式理论来审视档案学科范式演进脉络，可以清晰地归纳出从档案文件史料范式到档案信息资源范式，再到档案数据智能范式三大阶段的不断跃迁[7]。这种变革进阶过程以档案资源形态“模拟态”“数字态”“数据态”为根基，进而导致档案工作资源对象、业务核心与技术形态的深刻变化，具体演进脉络如图1所示。

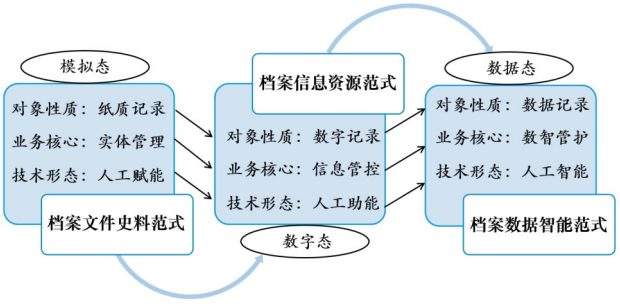



图1　档案学科范式演进脉络图

Fig.1　Evolutionary timeline of disciplinary paradigms in archival science

第一阶段——档案文件史料范式。这一范式自档案学科形成伊始，一直延续至20世纪中期。文件管理与史料整理是档案学兴起的两大源流，档案文件史料范式是中国档案学的传统经典范式。在此范式下，档案被视为社会活动的原始记录，档案工作业务活动主要围绕着实体模拟态纸质档案展开手工管理，以人工赋能文件档案的接收整理、编目保管等。档案作为历史的真实记录，以档为史、据档而研，能够系统发挥档案的情报价值、凭证价值。这一阶段，档案机构职能侧重于文件保管与史料整理，社会公众利用需求尚

未被广泛认知与全面激活，但这一范式毫无疑问为档案学科确立了历史根基。从历史的视角去探寻档案的文件属性、证据功能与整理方式，是档案文件史料范式的基本研究取向。

第二阶段——档案信息资源范式。随着20世纪中后期计算机、互联网等信息技术的发展与社会信息资源意识的崛起，档案资源对象、业务核心与工作手段也发生了深刻变化，推动档案学科范式转向档案信息资源范式。档案不再仅仅局限为物理实体文件，更被认为是一种重要的基础信息资源，各类数字态信息被纳入档案管理范畴。信息技术开始深度融入档案管理流程，电子文件管理系统被引入档案管理工作，人工与机器相辅相成共同提升档案管理自动化水平。

第三阶段——档案数据智能范式。当前大数据、人工智能等数智技术集群爆发，推动社会迎来智能奇点。档案资源对象发生质变，各类文本、图像、音频、视频等多结构化并存的海量数据记录成为档案信息存在形式。这催生了档案学科范式经历第二次深刻变革——迈向“档案数据智能范式”。档案数据智能范式中的“档案数据智能”意味着档案不仅是历史文献、信息资源，更是可被人工智能机器智能感知识别、读取理解、深度挖掘、计算分析、关联交互的细粒度数据单元。档案数据智能范式将成为数智时代档案学科演化方向，推动档案学科突破传统边界，与数据科学交叉融合，与人工智能技术深度链接，构建以数据为基础、智能为引擎、要素价值创造为目标的新生态。

## 1.3　档案学科范式的运动规律

范式作为常规科学赖以运作的理论基础和实践范式，是动态发展变化的[11]。档案学科范式的发展历程，是一部在外部环境与内部体系双重变革下，不断演进与更迭的历史。档案学研究中存在着人文、管理和技术三大研究取向，在新的社会历史条件下呈现出新发展方向[12]。此外，档案价值运动规律是档案学科发展中的重要理论根基。数智时代，档案价值形态发生变化，呈现出新的价值实现规律。“研究数字时代档案价值形态与价值实现规律的嬗变，对于夯实档案学学科自信、推进档案事业高质量发展、建构中国自主的档案学知识体系具有重要时代意义。”[13]根据档案学研究三大发展取向，结合档案价值实现规律，提炼归纳档案学科范式演进蕴含的技术、管理、人文与价值4方面的交织融合。

在技术迭代上，技术始终是驱动档案学科范式变革的核心动能与根本引擎。从造纸术、印刷术到计算机、互联网，再到当前的大数据、人工智能，每次技术突破都带来了新的资源对象、工具方法和业务环境，推动档案工作从实体保管向信息管理，再向数智管理范式演进。档案数据智能范式以数据智能为核心，综合运用虚拟现实技术、数据挖掘分析技术、高性能计算技术、人工智能技术、知识图谱技术，实现知识创新与智慧创造。在管理深化上，基于档案管理对象的模拟态、数字态、数据态变迁，档案管理对象颗粒度从粗颗粒的纸质文件到中颗粒的数字图像再到细颗粒的数据单元，管理对象的细粒化导致档案管理思维、管理制度、管理模式、管理形态、管理层级从粗放走向细化，实现语义降维与内容聚合，以智能机器为核心的精深化、关联化、标准化档案数据管理正成为档案管理新业态。在人文发展上，档案工作本质上是以人为本的事业，档案馆是提供公共文化服务的文化机构。因此，档案学科范式演化带有特定的人本色彩与文化烙印。无论是技术的发展还是管理的升级，都推动了文化理念的融合、文化场景的丰富与文化习惯的形成，促进档案学科范式从封闭走向开放、从独立走向交叉、从隔阂走向对话，彰显档案学科人文关怀、道德底蕴与伦理关切。在价值拓展上，技术革新、管理再造、人文创新的落脚点最终归结于档案价值的拓展。档案价值是档案学科范式的内生牵引力，指引着范式发展方向。在技术赋能加持下，档案学科价值追求也随之拓展：从保障文件史料证据真实，到提升档案信息资源传播效能，再到激活档案数据要素潜能。通过海量档案数据清洗加工、聚合关联、挖掘计算，推动档案数据开放共享、确权登记、估值入表、流通交易、质押融资，促进档案数据资产化资本化价值循环复用与乘数倍增。如2025年，苏州中国丝绸档案馆完成“近现代中国苏州丝绸文献遗产数据集”登记，构建“档案数字化建设-知识产权确权-数据资产登记-数据资产授权运营”管理模式，为深挖文献遗产数据价值开辟新路径[14]。

DOI：10.13998/j.cnki.issn1002-1248.26-0032

### 1.4 档案数据智能范式的出场背景

档案数据智能范式是数智时代背景下多重驱动力共同作用的产物，其出场逻辑是内外部环境的双向奔赴与统合互构。

从外部背景看，大数据、人工智能等数智技术的迅猛发展，使人类身处第四次工业革命浪潮，推动档案学与数据科学、计算机科学、信息技术等学科领域深度交叉共创，这种跨学科的碰撞与融合，为档案数据智能范式的生成提供方法论支撑。尤其是科学研究经历了以自然实验为主的第一范式实验科学、以理论假设推理为主的第二范式理论科学、以计算模拟为主的第三范式计算科学之后，目前已进阶到第四范式时代，即数据密集型科学范式，数据驱动的研究方法成为人文社科领域的宽口径通用范式，加速档案学新思维、新技术、新取向的渗透，开拓出档案数据治理、档案数据要素、计算档案学、智慧档案馆等新研究场域，为档案数据智能范式提供了丰富的理论源泉与实践土壤。此外，国家数智战略政策的引领与社会信息生态的变革，进一步推动档案学科融入数字中国、智慧社会、数据要素×、人工智能+、文化数字化等战略格局中，这些外部环境造就了档案数据智能范式形成的必要条件。

从内部背景看，技术环境变迁带来档案管理理念、档案资源形态、档案业务模式的全方位革新，传统档案管理流程在当今数据洪流面前已难以为继。文本、图像、音频、视频、三维模型等海量多媒体数据的涌现，要求档案工作必须具备处理相应数据量的能力，如何将具有长久保存价值的数据资源进行科学有效归档、安全可靠保存与深度开发利用，已成为数智时代档案学科无法回避的必答之问。此外，大语言模型等生成式人工智能技术的成熟应用，更是为档案数据内容深度理解、知识提炼开辟了全新路径。数智赋能不再仅仅是辅助工具，而已演变为重塑档案管理理念、流程与目标的建构性力量。传统档案学范式在应对数智环境时已显现出解释力与执行力的双重危机，旧有的理论框架与方法论体系，在面对数据态、智能化、场景化的档案管理新实践时，常常显得力不从心。这场刀刃向内的革命倒逼档案学科与时俱进、开拓创新，学科内在发展的危机与自觉催生了以数据智能为核心的科学新范式。

## 2 档案数据智能范式意蕴解析

范式转换是“旧有抽象思维方式的局限性带来的必然结果”[15]。档案数据智能范式的出场，是学科审视自身危机、寻求新的增长点而进行的自觉求变。在揭示档案数据智能范式产生背景后，有必要进一步解析档案数据智能范式理论内核，为图景构筑奠定基础。

### 2.1 档案数据智能范式内涵结构

档案数据智能范式的形成建构并非仅仅是技术工具的简单引入，本质上是一场触及学科核心的深刻革命[7]。这场变革发端于底层思维认知，贯穿于档案价值创造全过程，最终将重塑档案理论知识体系与档案事业发展格局。档案数据智能范式以档案数智化为内核，以数据要素赋能与人工智能驱动为引擎，以智慧档案为大脑（集档案资源智能管理、深度开发与精准服务于一体的现代化档案工作模式[16]），以算力中枢与算法模型为动能，以高质量档案数据集为支撑（塑造可信任、可流通、可共享的优质算据与高质语料，借助人工智能计算开发，激活档案数据资产价值），涵盖数智文化、数智技术、数智人才、数智管理四大版块的档案数据智能理论知识框架与实践发展体系。档案数据智能范式结构模型如图2所示。

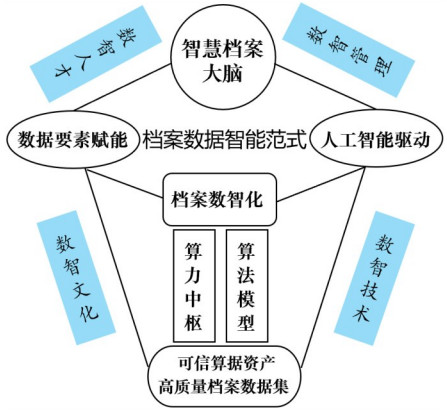



图2 档案数据智能范式结构示意图

Fig.2 Schematic diagram of the intelligent archival data paradigms

范式是一个由信念、理论、标准、方法、技术等构成的有机整体。档案学范式是档案学共同体成员所共同认可的价值取向、思维方式、理论主张、方式方法等内容成分的集合。档案数据智能范式进阶本质是从实体保管到数据驱动的认知论转向。在传统视野中，档案工作对象是物理载体，其性质源于对原始记录性的证据维护。数智时代，档案数据成为档案资源新对象，也带来档案收管存用全流程的系统变化，包括数据收集的实时性、数据管理的精准性、数据存储的拓展性、数据利用的高效性等，由此带来档案工作内容的革新，如档案数据整合集成、清洗修复、流通交易、计算开发与资产增值等。档案数据智能范式将档案视为关联化、可计算的数据场域。在此视角下，档案价值不仅在于其承载的文字信息，更在于其能否被转化为结构化的数据单元，并与其他数据节点归集聚合、语义互联。档案数据深层价值是一种待挖掘释放的“潜能”，其激活高度依赖于算法与模型的双轮驱动。这意味着，档案工作核心议题将围绕数智时代档案数据开发利用，驱动档案工作者从历史文献保管员转型为档案数据资产运营者与智能算法工程师，这一认知转向也驱动了档案价值重心与核心要素的战略变革。

## 2.2　档案数据智能范式核心对象

档案数据智能范式是建立在必要基础之上，即确保该科学范式得以成立的根基。学科对象是学科理论发展的起点，学科对象的变化必然引发学科范式的系统性变革。尽管技术环境剧变，但档案数据智能范式的建构不能脱离档案学自身逻辑起点，需要从档案这一核心对象根本属性出发。基于档案学科理论特质与实践坐标，档案数据智能范式构建基础是档案数据这一档案资源新对象，以确保档案数据证据价值、维护历史记忆真实完整为导向。

档案数据作为具有保存价值的各种数据形态记录，是可被机器深度读取的数据颗粒单元，能够作为算法模型的底层燃料在算力网络中自由流转。当前，档案数据来源广泛、类型多样、格式复杂，涵盖各类结构化、半结构化、非结构化数据资源，不再仅仅是历史的记忆载体，更是海量的数据要素集合，需要通过采集捕获、分类分级、规范处理和长期保存，确保来源可靠、程序规范、要素合规。把握档案数据的根本属性有助于准确认知档案数据智能范式的建构逻辑，档案数据的本质属性融合了档案的“原始记录性”与数据的“计算可用性”，在重新诠释档案数据原始记录性的同时，也需要将档案数据系统地视为一种高质量、高价值的数据资产，拓展档案数据功能认知。借助自然语言处理、数据挖掘、机器学习、智能算法等，围绕档案数据生成、采集、处理、存储、利用全生命周期进行科学管控，将档案管理的方式从信息组织升维至数据治理与智慧赋能，以深化、巩固和拓展档案数据要素价值，使其从静态的信息记录转变为可计算、可关联、可推理、可交互的数据语料与可信资产，进而支撑经济发展、民生服务与科学决策。如衢州市“医档宝盒——基于参保人员‘一人一档’多维医保守护模式”获2025年“数据要素×”大赛全国总决赛医疗保障赛道一等奖，基于超过6 800万份标准化的“一人一档”全生命周期健康档案数据，构建“111N”医保数据治理与服务新模式，衍生出全民参保、病贫共济等“*N*个”数智应用场景，推进参保动员效率提升50%，成功帮扶数十万潜在因病致贫人口，并有效缩短商保理赔核保时间40%[17]。

## 2.3　档案数据智能范式主要向度

基于托马斯·库恩的范式理论，范式是理论层面的认知取向、实践层面的行为活动与科学层面的共同体构成的基本观点、基本方式、基本规范，形成体系化、通用化、专业化理论模型与实践框架。结合数智时代档案学发展态势，档案数据智能范式可从价值信念、实践范例和科学共同体3个核心维度进行解析。

基于思维信念向度，档案数据智能范式实现了从实体保管到数智赋能的理念转变，体现出档案价值不止于守护历史凭证，更是激活档案数据作为生产要素和智慧源泉的潜能，从而驱动社会治理、文化创新与知识服务。具体而言是双向赋能的理论机理：一方面是由外向内赋能，即积极拥抱大数据、人工智能等新兴技术浪潮与国家战略，以政策、数据和技术为导向重塑档案学科研方法与档案治理体系；另一方面是由内向外赋能，即依据档案数据的独特权威性，探索档案数据开发利用新理论、新模式，开辟档案数据治

DOI：10.13998/j.cnki.issn1002-1248.26-0032

理、档案数据要素、智慧档案数据等学科新的学术增长点，以战略思维服务宏观决策、以人本思维促进社会公平、以先行思维提供知识服务、以协作思维构建共治生态，从而反向赋能数字中国建设与新质生产力创新发展。纵观2020—2024年中国信息资源管理学界十大学术热点，从“数字记忆建构的档案开发与服务”“图情档与数字人文融合”“世界变局之中的信息安全与数据治理”，到“国家治理体系下的档案治理”“国家数据基础制度体系建设”“国家文化数字化战略与公共文化服务”，再到“通用人工智能时代的信息资源管理学”“图书情报档案机构的数智化转型与创新”“档案数据要素与档案价值鉴定”，一条从数字化到数据化、数智化的关注焦点迁移脉络正逐渐清晰浮现。

基于实践行为向度，档案数据智能范式的落脚点是档案事业数字化、数据化转型与智能化、智慧化升级。档案数智化重构必然投射于业务流程与治理模式的系统性重塑中。面对档案数据的指数级、规模化增长，需要打破传统业务流程，运用数据挖掘、机器学习、计算机视觉、自然语言处理、社会网络分析等技术，构建动态循环、精准智慧的实践体系，打造智慧档案创新实验室、档案数据智能计算实验室，形成智能驱动、算力驱动的档案数据治理闭环。治理逻辑从事后被动响应，转向事前预测与主动干预，治理结构也从封闭科层制，迈向开放联通、多元协同、共生共创的网络生态，拓展档案数据治理应用场景。如重庆市档案馆聚焦婚姻档案数据关键信息采集不全、查全率和查准率不高等问题，利用大语言模型等技术，积极推进婚姻档案数据治理，实现婚姻档案精准识别、高效检索，打破资源数据孤岛，赋能婚姻登记“全国通办”，提升了档案工作公共服务效能[18]。

基于科学共同体向度，档案数据智能范式的形成与巩固依赖于一个身份认同与技能重塑的新型共同体，这个共同体不再局限于传统档案学家、历史学家、档案馆员，而是扩展并吸纳数据科学家、人工智能专家、首席数据官、数据管理员、数据服务商、数智运维师等新角色，成为一个跨学科、跨行业、跨部门的档案数据智能范式共同体。该共同体拥有一套新型技能共识与职业规范：学术研究者需具备理论创新能力与实践应用能力；实践工作者需掌握档案数据治理、智慧档案开发、数据资产流通等新兴职业技能，完成从档案管理员向档案数智化管理师转变。这需要共同体成员秉持技术与人文相统一的职业操守，既善于运用数智技能，又坚守档案工作守护历史真实、传承优秀文明的本质责任，捍卫档案馆作为数字记忆宫殿的社会形象。如上海大学等5所高校，以及上海市档案局馆等14个实践部门共同开展档案学专业政产学研协同育人机制创新，主动适应档案数智化步伐，引入大数据、人工智能等新一代信息技术，在教学内容和课程体系中持续加大信息技术与数智管理力度，促进专业建设与人工智能深度融合，培养档案新质人才[19]。

## 2.4 档案数据智能范式表现形态

档案学科范式通常以档案学共同体进行的学术研究与实践操守为外在表现形式[20]。步入数智时代，档案数据智能范式的表征既包含形成的理论新知，更包含档案数据智能管理的运行空间与价值实现过程。依据档案学科范式演进的客观规律，档案数据智能范式的具体表现形态呈现出学术争鸣与典范涌现两大特征。

在学术争鸣上，近年来围绕档案数据化、智能化、数智化，学界开展多项学术活动，包括长三角数智档案一体化研讨会、以“AI引领数智未来”为主题的第三届智慧档案科技博览会、以“迈向‘十五五’的档案数智转型”为主题的第十五届中国电子文件管理论坛暨2025档案数据产教融合大会、以“数智赋能档案工作创新”为主题的首届粤港澳大湾区档案管理研讨会等。此外，国家社科基金、相关学术期刊等聚焦档案数智化主题，设立了诸多科研项目、发表了大量学术论文，这些学术研究对档案数智化理论内涵进行了系统探究，推进了档案数据智能科学共同体的形成与发展，筑牢档案数据智能范式的理论根基。

在典范涌现方面，在国家档案局、各省市档案局馆对档案数智化进行系统布局的基础上，各地积极探索档案数据赋能经济建设、公共服务、文化创新、产业转型等方面的应用场景，推动档案数据作为关键生产要素融入数字中国建设，涌现出首都智慧档案联合创新实验室、浙江省档案数据共享中心、南京市档案馆云锦档案数据授权商业开发、青岛市智慧档案馆建设、苏州中国丝绸档案馆近现代中国苏州丝绸文献遗

产数据登记确权、嵊州个人全生命周期综合智治、嘉兴原生数据档案化治理等典型案例，形成与理论热潮相呼应的实践样板。

## 3　档案数据智能范式图景建构

当前，人类社会已不可避免地迈入数智时代，这不仅是一场划时代的技术革命，也是一次全方位的技术革命，档案数据智能范式也将成为档案学科发展新范式。档案数据智能范式图景建构是一项系统性工程，需要从哲学基础到实践路径的全面革新，“从哲学研究3个基本层次，系统探索档案数据本体论、认识论和方法论，构筑档案数据理论知识体系，指导档案数据管理实践。”[21]以哲学研究本体论、认识论、方法论、价值论等为理论基础，结合人工智能技术等实践应用，形成本体论、认识论、方法论、技术论和价值论“五论一体”的档案数据智能范式图景建构依据，为档案事业数智化发展提供理论支撑和实践指引，如图3所示。5个维度相互关联、相互协调，共同构筑档案数据智能范式完整图景，从而推动实现档案数据理论内涵深化与实践管理深耕目标，激活档案数据要素价值，助力档案数据要素赋能新质生产力。




图3　档案数据智能范式五维图景

Fig.3　Five-dimensional framework for intelligent archival data paradigms

### 3.1　本体论：夯实档案数据基座源能

数据资源体系是档案数据智能范式的母体基础，是档案学科发展生存的水之源、木之本、塔之基、炊之米。这一体系以规模庞大、质量优异、特色鲜明的档案数据仓储为核心。当前，档案数据资源建设面临着数量规模不足与质量参差不齐的双重困境。根据国家档案局统计，截至2024年底，全国各级综合档案馆馆藏电子档案2 812.7TB，馆藏档案数字化成果30 084.5TB，总计约32PB[22]。与此同时，2024年，全国年度数据生产总量已达41.06ZB，同比增长25%；全国数据存储总量为2.09ZB，同比增长20.81%[23]。可见，档案数据总体规模无论是与国家数据总体规模还是各类机构数据存储规模相比，都存在几何级数的数量鸿沟。一言以蔽之，档案部门的数据量远不及各类机构的档案数据生产量和其他部门保存的档案数据量[24]。此外，在纳入档案数据资源范畴视野中，数据失真失读、数据污染异构、数据孤岛壁垒、数据冗余残缺、数据碎片割裂等质量问题层出不穷，成为掣肘档案数据互联互通、共建共享的瓶颈。

档案数据作为喂养训练人工智能大模型的优质语料与高质量数据，需要满足数量规模足够大、质量禀赋足够优、特色鲜明足够亮、类型结构足够多等基本特点。贯彻“大档案观”，对政务、金融、文旅、医疗、交通等各行业各领域各形态档案数据应收尽收、应归尽归、应管尽管，加快传统档案资源数据化进程，持续缩小档案数据“多点、多体、多域、多源”生态与档案部门有限数据收管能力之间的“剪刀差”。通过深度标注、检测去重、清洗修复、语义增强等系统化档案数据质量控制，确保档案数据来源可靠、程序规范、要素合规、安全可用，构建高质量档案数据集。2025年，中国将建成高质量数据集总量超500PB，其中直接或间接地都与高质量档案数据集建设相关。在国家数据局发布的首批高质量数据集典型案例中，“高质量电子疾病档案数据集”“公共信用档案高质量数据集”等档案数据集入选，而内容涉及档案数据质量控制的则更多[25]。

### 3.2　认识论：塑造多元协同共建动能

从认识论上而言，数智时代档案学科范式需要多方主体协同共创，即突破传统档案管理中的单一主体模式，构建开放包容、均等普惠、共建共享的多元主体协同新格局。在主体结构上，形成以档案行业为主

DOI：10.13998/j.cnki.issn1002-1248.26-0032

导，政府部门、数据管理部门、数据交易机构、信息技术公司、科研院所和社会公众共同参与的多元协作体系与创新生态，打破以往档案界相对封闭、保守、孤立地解决问题的惯性模式。档案行业作为资源持有、规则制定的重要主体，是范式建构的主导力量，发挥着资源统筹、制度设计、学术研究、实践开拓、技术应用、平台搭建、服务优化等功能，其他主体则发挥辅助决策、科技支撑、市场消费、业务协同、志愿参与、监督评价等作用，共同推动档案事业可持续发展。多元协同共创模式，将有效激发社会各方参与档案事业创新发展的积极性和创造力，实现档案学科体系系统性重塑。

数智时代，档案数据智能范式多元主体协同共建本质上是打造科教产教共同体，对接数字中国、人工智能+、数据要素×等国家重大战略，适应新质生产力发展需求，通过资源共享、能力交融、优势互补、互利共赢，打破馆际壁垒、部门围墙与数字鸿沟，加速档案数智化基础理论创新、关键技术突破、高端人才培养与应用场景拓展，推动档案数据与公共数据、政务数据、企业数据、民生数据等集成汇聚，融入大数据海洋，促进档案数据跨区域、跨行业、跨部门、跨业务、跨系统互联互通，从数据孤岛走向知识群岛，形成能够理解感知档案数据内容、自动建立知识图谱、链接语义关联网络的智能认知空间，支撑更广泛的政府治理、经济发展与社会运行。2023年，为应对档案数智化转型过程中面临的技术更新、人才短缺、协作不足等挑战，中国人民大学信息资源管理学院倡议成立档案数据产教融合发展平台，通过汇集高校的科研能力、企业的技术创新能力、档案部门的实践经验，构建合作共享、互利共赢的生态系统，促进档案数据教育、行业、科技的一体化融合，目前成员单位总数已达51家[26]。

### 3.3 方法论：增强学科交叉创新效能

学科交叉融合与理论互渗是当代科学发展的必然趋势，也是科研突破与学术创新的动力机制。档案学科范式的运动过程与演变规律表明，档案学知识体系具有明显的开放性、融合性、拓展性，只有不断吸纳新知识、新方法、新实践，新旧范式之间才能形成批判性对话，推动学科在继承中创新，在辩证中发展。因此，档案数据智能范式的形成建构是多学科深度交融的产物，通过文理交叉、文工交叉、文文交叉，借鉴数据科学、计算机科学、人工智能、社会学、法学、生态学等多学科领域的理论方法，推动档案学科范式转向以数据智能为引擎、更具前瞻性的现代科学范式，建设适应档案数据智能范式的学术规范、理论法则、基础设施和实践工具，打造“大数据（算料）+大平台（算力）+大模型（算法）”深度融合的新场景，形成定性与定量结合、实证与解释互构、思辨与理性交织的科学方法体系，对海量档案数据进行深度分析，增强档案学研究整体性、原创性、学理性，提升档案工作预测性、时效性、精准性[7]。

档案数据智能范式的学科创新主要体现在3个维度：首先是研究视角的交叉融合，以档案学专业知识为根基，吸纳借鉴数据科学的整体视角、计算机科学的计算视角、人工智能的主动视角、经济学的生产视角、社会学的治理视角、法学的规范视角、生态学的平衡视角等，推动学科深度对话，不断丰富档案学科研究视野和思维论域[27]。其次是研究手段的创新集成，在坚守档案学传统文献分析、历史研究考证、哲学抽象思辨、实证案例验证等研究方法基础上，引入数据挖掘、机器学习、自然语言处理、社会网络分析、文本计算、虚拟现实、数字孪生、文化组学、大语言模型推理等新型研究方法，推动档案学研究从小数据样本分析向大数据全景洞察转变。最后是研究领域的开拓延展，通过学科跨界融合，开辟计算档案学、档案数据治理、档案数据要素、智慧档案馆、档案数字记忆等学术增长点，拓展档案学研究空间与边界，增强学科社会影响力与专业认同感。如计算档案学强调将“计算思维”融入档案学研究中，重视档案数据与人工智能技术的深度融合，体现了大数据时代档案学研究范式的转变[28]。

### 3.4 技术论：推动人工智能场景赋能

技术是档案数据智能范式兴起发展的核心引擎。当前，以生成式人工智能为代表的新一代智能技术，已成为赋能档案事业高质量发展的关键动能。以人工智能技术为核心驱动力，加强算力算法、数据中台、

5G网络、边缘计算、具身智能等数智基础设施建设，系统性构建档案数据科技工程体系，打造广覆盖、强牵引的应用场景。引入自然语言处理技术，实现档案文本深度理解、自动标引、内容汇聚和智能处理；借助计算机视觉技术，推动档案图像智能识别与开放审核；通过知识图谱技术，从海量异构的档案数据中抽取实体、关系和事件，建构跨领域、跨时空的档案数据语义关联网络，支持智能检索和知识发现；构建档案领域垂直大模型，促进档案数据智能分类、标注、问答、推送、对话；依托数字孪生技术，开展沉浸式、交互式档案展览体验。如广东省档案馆联合中国人民大学档案学院依托DeepSeek大模型的语义理解、模态分析、逻辑推理等能力，实现对档案目录项与原文多维度、语义化、逐件逐页的智能审核，通过可解释、可理解、可追溯的人机协作，推动档案审核从“经验驱动”向“数智驱动”转变[29]。

技术是一把双刃剑，技术赋能的背后也隐藏着信任赤字与伦理危机。安全是档案工作的底线和红线。在拥抱人工智能的同时，需要坚守人的主体性，防范黑客攻击、网络病毒、算法黑箱、AI幻觉、操作失误、隐私泄露、版权侵害、数字鸿沟等安全风险，筑牢档案数智化伦理基石，塑造以人为本、人机共治、安全可靠的技术应用生态，避免被技术异化。2024年，全球最大的互联网档案馆遭受一系列分布式阻断服务攻击，导致其服务下线。此次黑客攻击发生时，互联网档案馆正准备披露其此前遭遇的一次网站入侵，共导致3 100万多条记录遭泄露[30]。

### 3.5　价值论：激活档案数据要素潜能

档案数据要素价值释放是档案数据智能范式建构的落脚点，超越传统信息查阅与传播价值，转向数据深度挖掘与深层洞见价值。但在思维保守、制度约束、技术瓶颈与安全风险等多重因素掣肘下，大量档案数据资源仍处于冷藏、深闺、休眠、沉睡的困境，开放范围不足、挖掘程度不深、产品创新不够，档案数据资产化资本化价值未能释放[31]。根据《中国统计年鉴》，2020—2024年5年间，国家综合档案馆档案开放率（开放档案/馆藏档案）分别为15.89%、16.77%、17.91%、17.45%和16.90%；档案利用率（利用档案/馆藏档案）分别为2.25%、2.30%、1.95%、2.24%和5.32%[32]。可见，近80%的档案数据长期处于深闺沉睡状态，近90%的档案数据无法被有效共享利用，档案数据要素价值潜能亟待激活。

数智时代，档案价值向档案数据要素价值拓展延伸，需要建立健全档案数据确权登记、估值定价、流通交易、收益分配、利益保护标准规范，推出档案数据汇聚畅流归集工程、市场配置改革工程、开放共享示范工程、服务创新品牌工程，探索公共档案数据授权运营与企业档案数据资产入表新模式[33]，构建全国一体化档案数据市场平台与可信档案数据空间，培育壮大档案数据产业链、产业体、产业园、产业群、产业网，营销档案数据创意产品、文化精品与知识优品，全方位实现档案数据要素价值。如2025年，上海玉龙光碧文化投资有限公司对2PB影像大数据进行管理开发，推出“影像档案数据资产管理系统”“影像档案可信数据空间管理平台”数据产品，在上海数据交易所挂牌上市，为行业发展提供高质量影像数据产品服务，不断提升企业市场竞争力[34]。

## 4　结　语

数据浪潮奔涌，智能时代已至。在数智赋能深刻催化下，档案管理对象正经历从实体、信息到数据的阶段性跃迁，业务模式向数据化、智能化、智慧化方向加速演进，时空场域亦在数实融合中不断延展。这一系列变革不仅重塑着档案工作实践形态，更对档案学学科定位、价值理念、方法论域、生态空间产生体系化重构，驱动档案学科范式加速向档案数据智能范式转型进阶。本研究立足数智时代背景，探索档案数据智能范式的演变规律、内涵意蕴、模型结构等理论基础，从本体论、认识论、方法论、技术论和价值论五大层面系统缔构档案数据智能范式新图景。展望未来，面对浪潮滚滚的数智洪流，档案学须坚持理论深化与实践深耕并重、技术应用与伦理防范并举、数据理性与人文价值并存，在人、机、物三元智慧融合中，塑造更具时代气息、历史厚度与科技魅力的档案数据智能范式，推动档案事业数字化数据化转型与智能化智慧化升级。

DOI：10.13998/j.cnki.issn1002-1248.26-0032

## 参考文献：


[1] 王益民 . 数据论[M]. 北京: 中共中央党校出版社, 2021: 37.
Wang Yimin. The theory of data[M]. Beijing:Central party school press of the communist party of China, 2021: 37.

[2] 杨鹏, 金波 . 档案数智化的理论内涵与实践路径[J]. 图书情报工作, 2025, 69(11): 90-100.
Yang Peng, Jin Bo. Theoretical connotation and practical path of digital intelligence in archives[J]. Library and Information Service, 2025, 69(11): 90-100.

[3] 特里·库克, 李音 . 四个范式: 欧洲档案学的观念和战略的变化——1840 年以来西方档案观念与战略的变化[J]. 档案学研究, 2011(3): 81-87.
Terry C, Li Yin. Four paradigms: Changing ideas and strategies about archives in the western world, 1840 to 2011, and beyond[J]. Archives Science Study, 2011(3): 81-87.

[4] 徐拥军 . 档案记忆观的理论基础[J]. 档案学研究, 2017(6): 4-12.
Xu Yongjun. The theoretical basis of archival memory viewpoint[J]. Archives Science Study, 2017(6): 4-12.

[5] 韩立波 . 档案管理范式的转型与社会变迁[J]. 中外企业家, 2020(3): 136.
Han Libo. Transformation of archives management paradigm and social changes[J]. Chinese and Foreign Entrepreneurs, 2020(3): 136.

[6] 章燕华 . 以数智化驱动引领档案事业现代化的发展进程与实施路径[J]. 档案学通讯, 2023(6): 4-13.
Zhang Yanhua. The trends and paths of "digitization and intelligence" to drive the archival modernization[J]. Archives Science Bulletin, 2023(6): 4-13.

[7] 金波, 杨鹏 . “数智赋能”驱动档案学学科范式变革[J]. 中国图书馆学报, 2024, 50(4): 85-99.
Jin Bo, Yang Peng. "Data intelligence empowerment" drives the discipline paradigm change of archival science[J]. Journal of Library Science in China, 2024, 50(4): 85-99.

[8] 托马斯·库恩 . 科学革命的结构[M]. 2 版 . 金吾伦, 胡新和, 译 . 北京: 北京大学出版社, 2012: 147.
Thomas S K. The structure of scientific revolutions[M]. 2nd ed. Beijing: Peking University Press, 2012: 147.

[9] 孙大东 . 基于范式论批判的中国档案学发展研究[M]. 北京: 科学出版社, 2017: 49.
Sun Dadong. Research on the development of archives science in China based on paradigm criticism[M]. Beijing: Science Press, 2017: 49.

[10] 胡鸿杰 . 中国档案学的理念与模式[M]. 北京: 中国人民大学出版社, 2005: 11-12.
Hu Hongjie. The idea and mode of China's archives science[M]. Beijing: China Renmin University Press, 2005: 11-12.

[11] 赵瑞雪 . 走向第四范式——数据密集型科学研究[M]. 北京: 科学出版社, 2024: 4.
Zhao Ruixue. Towards the fourth paradigm - Data-intensive scientific research[M]. Beijing: Science Press, 2024: 4.

[12] 丁华东, 张燕 . 论档案学的三大研究取向及其当代发展[J]. 档案学通讯, 2019(6): 4-10.
Ding Huadong, Zhang Yan. Three research orientations of archival science and their contemporary development[J]. Archives Science Bulletin, 2019(6): 4-10.

[13] 张斌, 张旭 . 数字时代档案价值的新思考[J]. 档案学通讯, 2025(3): 4-10.
Zhang Bin, Zhang Xu. New thinking on the value of archives in the digital age[J]. Archives Science Bulletin, 2025(3): 4-10.

[14] 谢震香, 顾晓琦 . 苏州丝绸档案率先完成数据登记[N]. 中国档案报, 2025-06-30(2).

[15] 胡红生 . 论社会科学发展的内在机制——从主体思维生产的角度看 [J]. 华中师范大学学报 ( 人文社会科学版), 2002, 41(4): 35-40.
Hu Hongsheng. On the internal mechanism of the development of social science - Research from the angle of the production of thought[J]. Journal of Central China Normal University (Humanities and Social Sciences), 2002, 41(4): 35-40.

[16] 梁尔真 . 智慧档案管理概论[M]. 北京: 中国原子能出版社, 2025: 18.
Liang Erzhen. Introduction to intelligent archival management[M]. Beijing: China Atomic Energy Press, 2025: 18.

[17] 浙里办 . 数创高地书写“浙”里新答卷 浙江 13 个项目获 2025“数据要素×”大赛奖项[EB/OL]. (2025-11-26)[2026-03-04]. https://mp.weixin.qq.com/s/G--cuxPy7NPE611gZGIwKQ.

[18] 王君正 . 重庆运用大语言模型赋能婚姻档案数据治理[N]. 中国档案报, 2025-10-09(1).

[19] 金波, 杨鹏 . 将“以学生为中心”的理念转化为具体行动[N]. 中国档案报, 2025-09-15(4).

[20] 陈祖芬 . 档案学范式的历史演进及未来发展[M]. 上海:上海世界图书出版公司,2010:24.
CHEN Z F. The historical evolution and future development of archival science paradigms[M]. Shanghai: World publishing corporation, 2010:24.

[21] 杨鹏, 金波 . 档案数据论: 大数据时代档案学新论域[J]. 档案学通讯, 2025(1): 38-45.
Yang Peng, Jin Bo. Archival data theory: A new field of archival science in the era of big data[J]. Archives Science Bulletin, 2025(1): 38-45.

[22] 国家档案局 . 2024 年度全国档案主管部门和档案馆基本情况摘要 ( 二)[EB/OL]. (2025-09-10)[2026-01-20]. https://www.saac.gov.cn/daj/zhdt/202509/6cac5e5684874efcb454f860a4b94927.shtml.

[23]《全国数据资源调查报告(2024 年)》正式发布[EB/OL]. (2025-04-30)[2026-01-20]. https://baijiahao.baidu.com/s id=1830879584753578428&wfr=spider&for=pc.

[24] 金波, 杨鹏 . “数智”赋能档案治理现代化: 话语转向、范式变革与路径构筑[J]. 档案学研究, 2022(2): 4-11.
Jin Bo, Yang Peng. "Data intelligence" empowers modernization of archives governance: Discourse turn, paradigm change and path construction[J]. Archives Science Study, 2022(2): 4-11.

[25] 国家数据局 . 高质量数据集典型案例名单发布[EB/OL]. (2025-09-12) [2026-01-20]. https://mp. weixin. qq. com/s/KheAEt17MljWJsWR3B583Q.

[26] 马林青 . 档案数据产教融合发展平台的实践与经验[J]. 北京档案, 2024(11): 53-55.
Ma Linqing. Practice and experience of development platform for integration of production and education of archival data[J]. Beijing Archives, 2024(11): 53-55.

[27] 金波, 杨鹏, 刘娟娟 . 档案数据要素价值内涵要义与生成机理[J]. 档案学通讯, 2024(6): 4-12.
Jin Bo, Yang Peng, Liu Juanjuan. The connotation and mechanism of archival data elements value[J]. Archives Science Bulletin, 2024(6): 4-12.

[28] 刘越男, 杨建梁, 何思源, 等 . 计算档案学: 档案学科的新发展[J]. 图书情报知识, 2021, 38(3): 4-13.
Liu Yuenan, Yang Jianliang, He Siyuan, et al. Computational archival science: The new development of archival science[J]. Document, Information & Knowledge, 2021, 38(3): 4-13.

[29] 段博 . 广东推进档案开放审核数智化发展[N]. 中国档案报, 2025-10-13(1).

[30] 牟沁妍 . 互联网档案馆因黑客攻击而陷入瘫痪[N]. 中国档案报, 2024-12-12(3).

[31] 金波, 杨鹏 . 大数据时代档案数据治理研究[J]. 档案学研究, 2020(4): 29-37.
Jin Bo, Yang Peng. Research on archival data governance in big data era[J]. Archives Science Study, 2020(4): 29-37.

[32] 国家统计局 . 中国统计年鉴(2025)国家综合档案馆基本情况[EB/OL]. [2026-01-21]. https://www.stats.gov.cn/sj/ndsj/2025/indexch.htm.

[33] 金波,杨鹏 . 大数据时代档案数据治理研究[M]. 北京:科学出版社,2025:16.
JIN B, YANG P. Research on archival data governance in big data era[M].Beijing: Science Press, 2025:16.

[34] 上海数据交易所 . 玉龙光碧挂牌影像档案数据产品, 激活影像数据价值[EB/OL]. (2025-01-10)[2026-01-21]. https://mp.weixin.qq.com/s/72ccKMvC-Ue2nWs6IOCazw.

# Perspective and Graph Construction of the Archival Data Intelligence Paradigm

YANG Peng[1,2], QIAO Jingjing[1]
(1. School of Cultural Heritage and Information Management, Shanghai University, Shanghai 200444; 2. Shanghai University Archives Development Research Center, Shanghai 200444)

**Abstract:** [Purpose/Significance] In the contemporary era, which is characterized by digital intelligence, the profound and unprecedented integration of big data analytics and artificial intelligence technologies is fundamentally transforming archival resources. This paradigm-shifting transition transforms traditional, static analog archival records into highly dynamic, interconnected, data-centric digital records. The fundamental transformation of these primary archival management objects triggers an inevitable, multi-dimensional, holistic reconstruction of the prevailing disciplinary paradigm. Consequently, the rigorous conceptualization and structural construction of the "archival data intelligence paradigm" has definitively emerged as a paramount and indispensable research issue for the sustainable development and continuous modernization of archival science. This issue demands meticulous academic scrutiny and strategic foresight. [Method/Process] Grounded firmly in the foundational scientific paradigm theory and meticulously combined with the pre-existing theoretical frameworks of the archival science paradigm, this comprehensive study methodically explores and elucidates the evolutionary lineage and underlying historical logic of the discipline. Through rigorous systematic literature reviews, advanced conceptual model construction, and cross-disciplinary theoretical integration, this research systematically traces the epistemological and methodological trajectory of the archival science paradigm. It provides a critical analysis of the transformative journey from the paradigm of traditional archival documents and historical materials, to the transitional paradigm of archival information resources, and finally to the cutting-edge, technology-driven paradigm of archival data intelligence. [Results/Conclusions] The newly articulated archival data intelligence paradigm intrinsically positions archival digital intelligence as its fundamental operational core. It is powerfully propelled by the dual interactive engines of data element empowerment and artificial intelligence augmentation. Within this sophisticated architectural framework, the "smart archive" functions as the central cognitive brain, meticulously driven by robust computing power hubs and advanced algorithmic machine-learning models, while being reliably supported by the continuous curation of high-quality, machine-readable archival datasets. This synergistic integration fundamentally culminates in a comprehensive theoretical knowledge framework and a robust practical development system, seamlessly synthesizing digital intelligence culture, innovative technologies, specialized talent cultivation, and adaptive administrative management strategies. Furthermore, by rigorously deconstructing the complex theoretical implications of this novel paradigm and firmly grounding it in both profound philosophical foundations and empirical practical development, this study systematically delineates a holistic structural landscape of the archival data intelligence paradigm. This conceptual landscape is mapped out meticulously across five critical dimensions: ontology (redefining the nature of archival data), epistemology (reconceptualizing how archival knowledge is validated), methodology (innovating analytical approaches), technology (implementing advanced AI architectures), and axiology (reevaluating the inherent societal value of archives). Ultimately, this comprehensive theoretical construct provides invaluable conceptual references and strategic, practical guidance. It actively facilitates the successful digital intelligence transformation and future-proofing of contemporary archival institutions and the broader archival profession worldwide.

**Keywords:** digital intelligence era; archival data; intelligent paradigm; digital intelligence empowerment; graph construction; data element