

# 国内外原生数字资源研究：回顾与展望<sup>\*</sup>

王勇臻 王贤文

(大连理工大学科学学与科技管理研究所暨 WISE 实验室, 大连 116024)

**摘要:** [目的/意义] 通过对国内外原生数字资源研究现状进行梳理与总结, 旨在为后续的探索实践提供一定的借鉴和参考。[方法/过程] 本文选取 2000 至 2022 年间有关原生数字资源的中英文文献作为调查对象, 综合运用文献计量方法和知识图谱工具, 扼要地回顾了国内外学界在该研究领域的发文情况和热点主题, 并予以对比分析。[结果/结论] 我国在原生数字资源研究领域的发文量与美国等国家相比尚存差距, 且发文期刊国际显示度相对较低, 还伴有发文内容同质化现象, 所探讨的主题较为单一。有鉴于此, 本文结合我国数字化转型的基本内涵对我国原生数字资源研究的未来走向进行了展望, 并为今后开展相关研究提出了四点方向性建议: 资源转化与利用、价值评估与定价、知识重组与创新以及权属界定与流通。

**关键词:** 原生数字资源 文献计量 知识图谱 可视化分析

**分类号:** G203

**DOI:** 10.31193/SSAP.J.ISSN.2096-6695.2023.02.05

## 0 引言

原生数字资源 (Born-digital Resources), 意指直接通过计算机等各类数字设备生成, 并在计算机系统及相关硬件系统中保存、管理和使用的数字信息资源, 既是信息技术进步的必然产物, 亦是数字社会赖以生存与发展的物质基础<sup>[1]</sup>。随着人类迈入数字化时代, 形形色色的原生数字资源正在源源不断地被创造出来, 潜移默化地重塑着现代日常生活的方方面面, 日益深刻地影响着当今世界经济格局及其演变。近年来, 我国数字经济欣欣向荣, 规模总量稳居全球第二, 仅次于美国, 已成为驱动我国经济发展的强劲引擎<sup>[2]</sup>。诚然, 丰裕的原生数字资源乃是我国参与国际数字经济竞争的核心力量。一项调查显示, 截至 2021 年 12 月, 我国网民规模为 10.51 亿, 互

<sup>\*</sup> 本文系辽宁省社会科学规划基金项目“5G 时代的网络原生数字资源的知识图谱构建研究”(项目编号: L21CTQ001) 的研究成果之一。

**[作者简介]** 王勇臻 (ORCID: 0000-0001-7306-1291), 男, 副教授, 硕士生导师, 研究方向为数据挖掘与信息计量, Email: yongzhenwang@dlut.edu.cn; 王贤文 (ORCID: 0000-0002-7236-9267), 男, 教授, 博士生导师, 研究方向为科学计量与科技管理, Email: xianwenwang@dlut.edu.cn (通讯作者)。

联网普及率接近 75%，俨然形成了全球最为庞大、生机勃勃的数字社会<sup>[3]</sup>。然而，值得注意的是，我国原生数字资源的经济价值还远未得到深度挖掘，其资产属性尚未得到充分体现。2022 年伊始，习近平总书记在《求是》杂志上发表了一篇题为“不断做强做优做大我国数字经济”的重要文章，其中指出“数据作为新型生产要素，对传统生产方式变革具有重大影响”，同时也提到“同世界数字经济大国、强国相比，我国数字经济大而不强、快而不优”<sup>[4]</sup>。有鉴于此，笔者认为，如何最大限度地发挥我国作为原生数字资源富地的天然优势，已是摆在我国政产学研用各界面前亟待解决的一个现实问题，直接关乎我国数字经济能否加速实现高质量发展。在此背景下，系统地梳理国内外原生数字资源研究的历史脉络显得尤为必要，不仅有助于更为细致地了解我国与其他国家在原生数字资源开发利用方面存在的共性和差异，还有助于更加准确地把握原生数字资源价值体系的演进逻辑及其未来走向。在此，本文以有关原生数字资源的中英文文献作为调查对象，从发文情况和热点主题两个角度切入，围绕国内外原生数字资源研究现状有针对性地展开评述，并予以对比分析，旨在探寻我国特色原生数字资源建设的潜在突破口，为我国科研人员开展后续的探索实践提供一定的借鉴与参考。

## 1 样本获取与预处理

为了保证样本文献的代表性，本文选取 Web of Science (WOS) 核心合集与中国知网 (CNKI) 核心期刊作为中英文文献来源，继而在研读经典论著和咨询权威专家的基础之上，分别拣选“原生数字”和“born-digital”充当线索词并在标题、关键词、摘要三个字段中进行精确检索，最后从返回的查询结果当中剔除诸如书评、征稿启事、会议通知之类的无效记录。截至 2022 年 6 月 3 日，笔者共计检索出 291 篇相关文献，最早可以追溯至 2000 年，其中有中文文献 49 篇 (CSSCI 收录 44 篇、EI 收录 1 篇，累积被引 516 次)，英文文献 242 篇 (SCI 收录 38 篇、SSCI 收录 84 篇、A&HCI 收录 40 篇、BKCI 收录 23 篇、CPCI 收录 7 篇、ESCI 收录 80 篇，累积被引 1657 次)。需要补充说明的是，早在“原生数字资源”这一提法出现之前，国内外专家学者就已针对其所表达的概念之外延，例如博客资源、网络论坛资源、网络百科资源，展开了一系列的调查研究。尽管如此，本文并不打算讨论这些具体的外延，主要原因如下。首先，虽然上述外延与“原生数字资源”在内涵上具有一定的相似性，但是彼此之间也存在相互区别的特质。例如，有的网络百科资源是由其他载体上的信息转化而来，并非天生即为数字形式。因此，如果将此类外延纳入检索范围之内，反而可能冲淡本文所关注的焦点。其次，“原生数字资源”所涵盖的外延难以详尽地罗列出来。不难想象，如果将全部潜在的外延统统都纳入检索范围之内，名目繁多的线索词势必会导致样本文献的数量大幅上升，进而显著地增加计量分析的复杂程度。

为了尽可能确切地抓住样本文献的主旨，本文采用人工检阅的方式对每篇文献的关键词集合进行了清洗，包括但不局限于以下两种情况。第一，归并同义词。例如，将“著作权”并入“版权”，将“archive”和“archives”单复数两种形式加以合并。第二，整合词义范围。以“international business”“international marketing”“internationalization strategy”之类的一组同族词为例，先取其概念交集，再将所得公共概念重新命名为“internationalization”。此外，由于样本文献

的数量相对有限, 且伴有关键词缺失现象 (缺失率约为 27.5%), 本文进一步利用每篇文献的标题与摘要两部分内容来扩充其关键词集合, 以期较全面地掌握有关原生数字资源的主题知识, 大致可以划分为如下三个步骤: ①将所有经清洗过的关键词添至一份动态列表  $L$  之中; ②根据  $L$  逐一匹配样本文献, 若某个关键词  $\alpha \in L$  能够成功地匹配某篇文献的标题或摘要, 则将关键词  $\alpha$  添至该文献的关键词集合之中; ③重新清洗每篇文献的关键词集合。如此循环往复, 直至  $L$  不再发生明显的变化。

## 2 国内外发文情况评述: 发表数量与学科分布的视角

通常来说, 一个研究领域的论文发表数量可以在很大程度上反映其活跃水平与发展状况<sup>[5]</sup>。图 1 展示了 2000 至 2022 年间国内外原生数字资源研究的历年发文量。综合来看, 该研究领域大致经历了三个发展阶段: 萌芽阶段 (2000 至 2008 年)、起步阶段 (2009 至 2015 年) 和成长阶段 (2016 年至今)。在萌芽阶段, 年发文量仅 1~4 篇。这一时期正值 Web 2.0 浪潮的迅猛兴起, 以博客、网络论坛、网络百科等为代表的社群软件相继问世, 网络环境下的数字内容创作从“精英时代”转向了“草根时代”<sup>[6]</sup>, 随之产生的诸如版权保护之类的社会议题逐渐引起了专家学者们的思考<sup>[7-8]</sup>。在起步阶段, 年发文量提高至 10~16 篇。在此期间, 随着智能手机的日益流行, 移动互联网在全球范围内急速地扩散开来。截至 2015 年底, 全球互联网普及率逼近 50%, 网络成为了真正意义上的社会信息基础设施<sup>[9]</sup>。与此同时, 关于如何合理高效地开发利用“网络原生数字资源”的探讨在国内外学界中掀起了一股研究热潮<sup>[10-11]</sup>。在成长阶段, 年发文量进一步攀升至 18~34 篇。2016 年, 李世石与 AlphaGo 的人机对弈标志着人工智能技术的发展进入了一个全新纪元。自此之后, 人工智能技术和互联网平台相互融合的趋势愈加明显, 智能物联网 (AIoT, 即人工智能 AI+ 物联网 IoT) 作为一种新型网络接入方案已跃然大众视野当中, 主张在传统互联网的基础之上实现人与人、人与物、物与物以及人与环境之间的广泛连接<sup>[12]</sup>。面对即将到来的超联结社会以及依附在这之上的海量信息, 如何防范潜藏的意识形态风险? 如何兼顾数据开放共

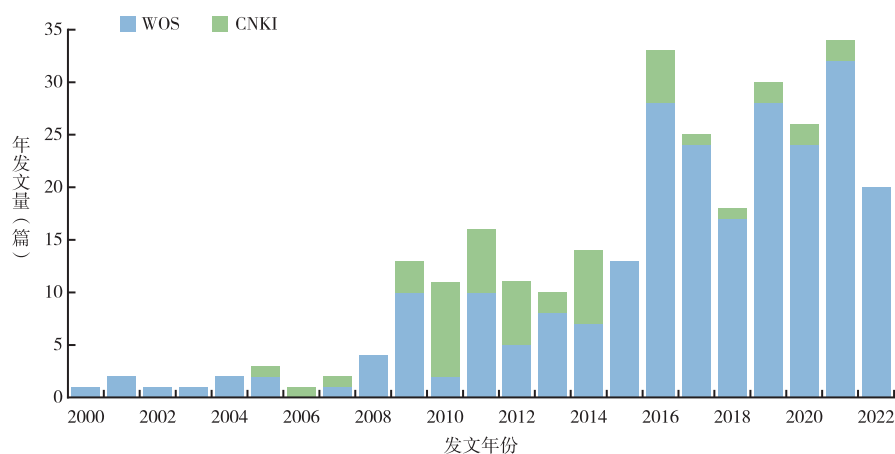


图 1 国内外原生数字资源研究的历年发文量

注: 2022 年的数据统计截止到 2022 年 6 月 3 日。

享和个人隐私保护？如此种种由全球数字化转型所带来的一连串严峻挑战，正在吸引越来越多的科研机构与科研人员投身其中<sup>[13-14]</sup>。除了以上所述三波阶段性进展，不难发现，在原生数字资源研究领域，中文文献的出现时间稍稍晚于英文文献，且主要集中于起步阶段，而在接下来的成长阶段中却相对稀罕。

表1 WOS数据库中在原生数字资源研究领域发文量排名前十的国家

排名	国家	发文量占比(%)	排名	国家	发文量占比(%)
1	美国	42.15	6	意大利	3.31
2	英国	15.29	7	中国	2.89
3	加拿大	6.20	7	法国	2.89
4	印度	4.55	7	爱尔兰	2.89
5	德国	4.13	7	澳大利亚	2.89

注：中国、法国、爱尔兰和澳大利亚，排名不分先后。

为了更加清楚地了解原生数字资源研究的国际化程度，本文相应地统计了世界各国在WOS数据库中的发文量，并列举了排名前十的国家，如表1所示，其占比合计高达70.25%（合著不予重复计数）。不难看出，美国目前在原生数字资源研究领域处于绝对领先的地位。与此同时，美英两国在该研究领域的发文量合计超过了英文相关文献总数的一半。相较之下，我国专家学者在国际刊物上发表关于原生数字资源的研究成果仅7篇，尚存差距。有鉴于此，笔者认为，现今阶段有关原生数字资源的中英文文献能够在一定程度上，分别视作国内外学界在该研究领域的知识积累或工作进展。在此基础之上，依据WOS和CNKI两大数据库的学科划分标准，即前者的“Category”字段和后者的“专题”字段，图2比较了国内外原生数字资源研究的学科分布情况。从整体上来看，图书情报学（WOS Category: Information Science & Library Science; CNKI专题：图书情报与数字图书馆）无疑是原生数字资源研究的中坚力量，英文相关文献中有32%属

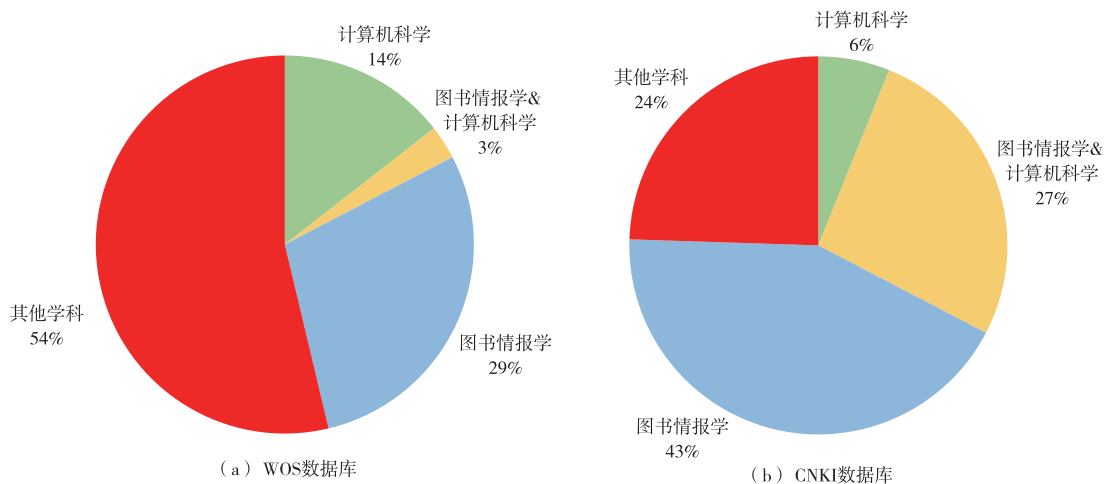
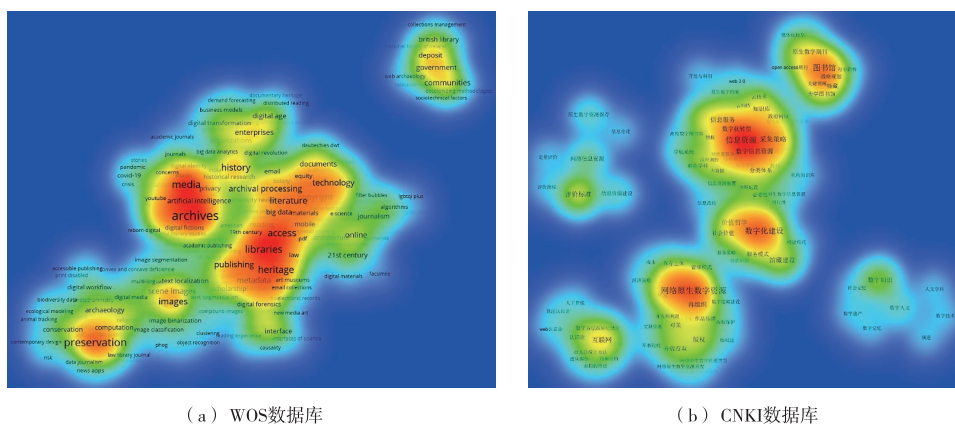


图2 国内外原生数字资源研究的学科分布

于该学科门类,而在中文相关文献中这一比例更是高达 70%。紧随其后的是计算机科学 (WOS Category: Computer Science; CNKI 专题: 计算机软件及计算机应用),该学科对原生数字资源研究亦贡献卓著,在中英文相关文献中所占份额分别达到了 33% 和 17%。最后,值得注意的是,我国在原生数字资源研究领域存在发文内容同质化现象,中文相关文献中 76% 都聚焦图书情报学/计算机科学,且在这两门学科之间呈现出了极为紧密的联系:27% 的中文相关文献同时涉及到这两门学科。

### 3 国内外热点主题评述: 共现分析与突现检测的视角

一篇学术文献的关键词集合是对其研究主题的高度凝练和集中描述<sup>[15]</sup>。因此,一个研究领域的主题概貌可以从既有文献所囊括的关键词当中窥探一二。不仅如此,一篇学术文献所给出的任意两个关键词之间必然存在着某种关联,而这种关联能够通过它们在既有文献中的共现频次予以表征:共现频次越高,关联越紧密<sup>[16]</sup>。在此,本文采用由荷兰莱顿大学科学技术研究中心 (CWTS) 开发的信息可视化软件 VOSviewer,对有关原生数字资源的中英文文献所包含的关键词分别进行共现分析,并相应地绘制关键词共现知识图谱,如图 3 所示,以便更为直观地洞察国内外学界在该研究领域的热点主题及其脉络联系。



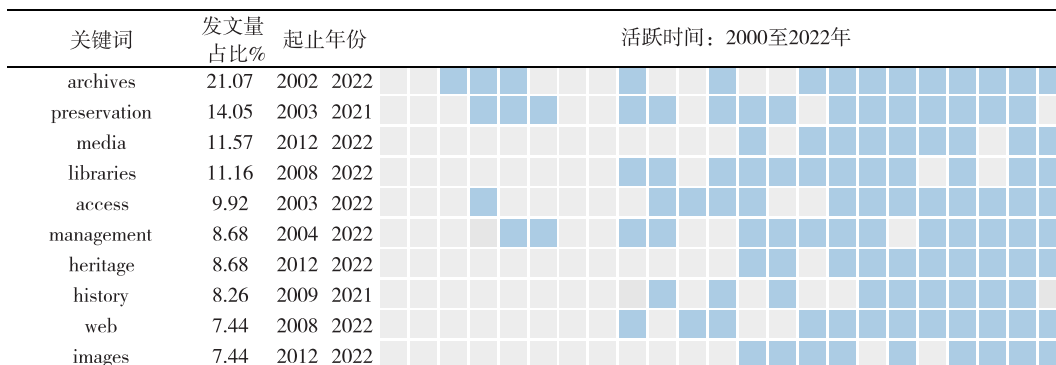
(a) WOS数据库

(b) CNKI数据库

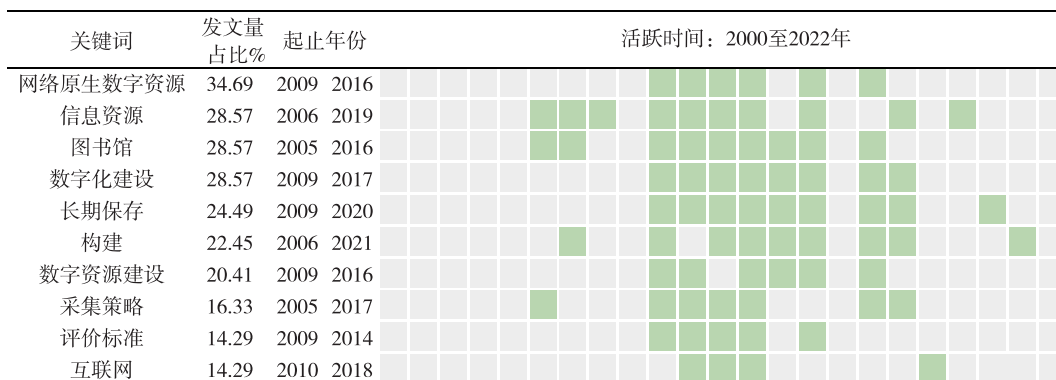
图 3 国内外原生数字资源研究的关键词共现知识图谱

综观国外相关研究,总共有三簇热点主题:①“archives (档案)”是其中最受欢迎的主题,包罗“media (媒体)”“artificial intelligence (人工智能)”“privacy (隐私)”等方面的内容;②“libraries (图书馆)”连同“literature (文学/文献)”一起构成了外延最广的一簇热点主题,所触及的内容覆盖“publishing (出版)”“heritage (文化遗产)”“access (可访问性)”等等;③还有一簇主题则以“preservation (长期保存)”为核心,围绕“archaeology (考古学)”“data journalism (数据新闻)”“news apps (新闻类移动应用)”等内容加以钻研。反观国内相关研究,共有四簇热点主题:“信息资源”“图书馆”“数字化建设”以及“网络原生数字资源”,但簇与簇之间的连结却显得相对松散。总体而言,国内外学界在原生数字资源研究领域存在不少共通

之处，特别是在馆藏化建设这一议题上<sup>[17-18]</sup>。然而，不难发觉，国外学界在原生数字资源研究领域内所积累知识的结构化程度相对较好，对应的关键词共现知识图谱能够比较完整地展现各个中心词的外延边界。相较之下，我国专家学者在过往探索中更多显露出的是主题集中的特点，即主题重复率较高。虽然这些探索也辐射了原生数字资源研究的诸多方面，但是每个方面都缺乏更为充分的延拓和更加宽泛的诠释。最后，作为一点补充，图4展示了国内外原生数字资源研究中出现频率排名前十的关键词及其活跃时间。从对比结果来看，相较于英文相关文献，有关原生数字资源的中文文献所包含的高频关键词的平均活跃时间更短（7.3年<11.7年），而平均发文量占比却更高（23.3%>10.8%），这两点观察与从图1和图3中得出的结论大体上保持一致。



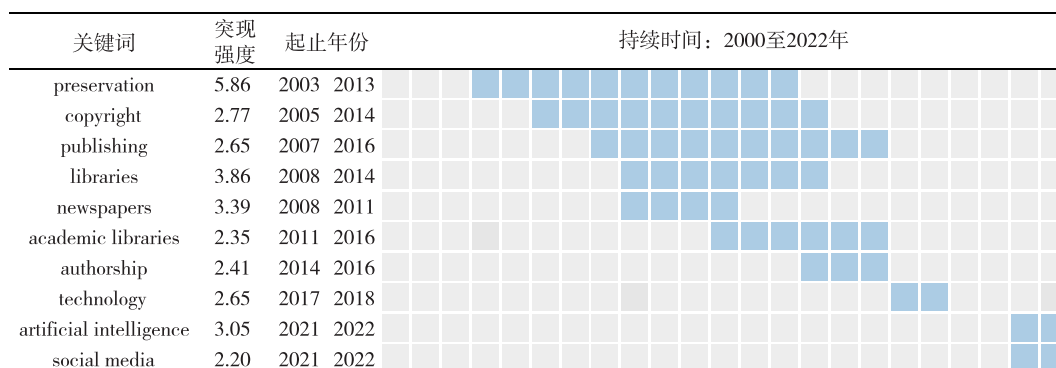
(a) WOS数据库



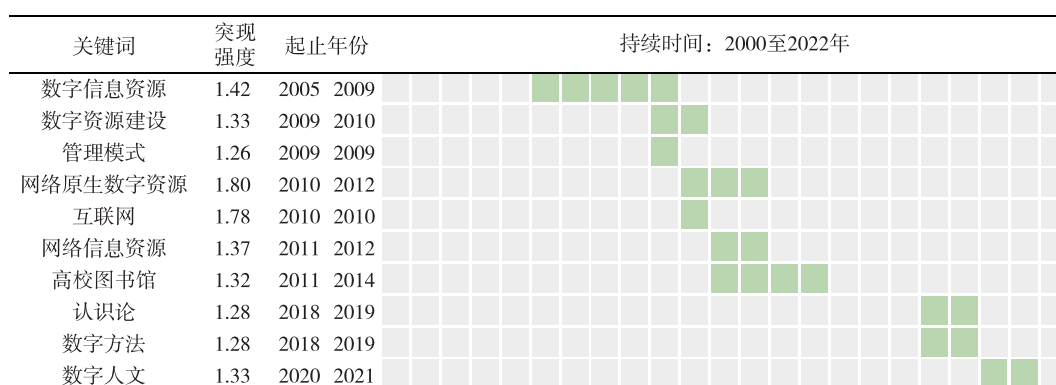
(b) CNKI数据库

图4 国内外原生数字资源研究中出现频率排名前十的关键词及其活跃时间

为了更为全面地解读原生数字资源研究的主题演进路径，本文进一步检测了2000至2022年间该研究领域中的突现词，以期更加详实地揭示不同时期骤然兴起的热点主题。顾名思义，突现词指出现频率在某一时间段内急剧升高的关键词，而发掘突现词无疑有助于更加充分地明晰一个研究领域的学术动态与未来趋势<sup>[19]</sup>。在此，本文采用由陈超美教授开发的信息可视化软件CiteSpace，对有关原生数字资源的中英文文献所包含的关键词分别进行突现检测，并从中挑出突现强度排名前十的关键词，如图5所示。



(a) WOS数据库



(b) CNKI数据库

图 5 国内外原生数字资源研究中突现强度排名前十的关键词及其持续时间

在国外相关研究中,“preservation (长期保存)”是突现强度最大、出现年份最早、持续时间最长的关键词,亦即原生数字资源开发利用所面临的首要任务<sup>[20]</sup>。自此之后,“copyright (版权)”和“publishing (出版)”陆续登场并盛行一时,结合后来涌现的“authorship (作者身份)”可以看出,知识产权保护长期以来一直是原生数字资源建设过程中难以回避的议题<sup>[8]</sup>。除此之外,不难察觉到三处热点变迁:①“academic libraries (学术图书馆)”逐步接替“libraries (图书馆)”成为了数字馆藏的关注焦点<sup>[21]</sup>;②“artificial intelligence (人工智能)”从一众“technology (技术)”中脱颖而出,已然成为盘活数字资产的不二法宝<sup>[22]</sup>;③新闻档案工作的主战场正在由“newspapers (报纸)”等传统媒体转向“social media (社交媒体)”等新媒体<sup>[23]</sup>。相较之下,我国在原生数字资源研究领域的热点更迭大致能够划分为两个阶段:①2005至2014年间浮现出许多表述相近的突现词,例如“数字信息资源”“网络原生数字资源”与“网络信息资源”,表明在该时期内“原生数字资源”这一提法尚未被国内学界普遍接受<sup>[10-11,24-25]</sup>,与此同时,我国专家学者开始涉足“数字资源建设”过程中可能遇到的“管理模式”“互联网”等具体问题,特别是面向“高校图书馆”这一应用场景<sup>[26-28]</sup>;②2018至2021年,我国原生数字资源研究迈入了一个全新阶段,诸如“认识论”“数字方法”“数字人文”之类的宏观主题相继崛起,我国专家学者开始探讨如何在全球数字化转型浪潮中更具理性、更富建设性地拥抱原生数字资源<sup>[13,29-30]</sup>。总的来看,相较于英文相关文献,有关原生数字资源的中文文献所包含的高突现性关键词的平均持续时

间更短(2.4年<5.7年),且平均突现强度更低(1.42<3.12)。换言之,我国在该研究领域所关注的热点主题相对单一。

## 4 研究结论与展望

原生数字资源是人类在与数字世界交互过程中形成的以数字形式为单一载体形态、大量累积、可获取利用、能够创造社会效益或经济效益的数字信息资源。眼下,全球数字化转型风起云涌,数字经济时代开幕在即,原生数字资源作为关乎数字经济发展的关键要素,其研究价值与意义不言而喻。面对这一形势,本文在文献回顾与总结的基础之上,选取发文情况与热点主题两个方面作为考察主线,较详细地评述了国内外原生数字资源研究现状,旨在为后续的探索实践提供前瞻性参考和阶段性借鉴。细数过往,国内外原生数字资源研究已逾二十春秋,其探讨的核心问题从何谓原生数字资源(What),到如何开发利用原生数字资源(How),再到何以加强原生数字资源建设(Why),可谓循序渐进且环环相扣。然而,值得注意的是,目前我国原生数字资源研究的发展态势并不容乐观,具体表现如下。首先,我国在原生数字资源研究领域的发文量仅达到了美国的约二分之一,且发文期刊国际显示度相对不高,绝大多数的研究成果发表于国内刊物。若长此以往,必将削弱我国在该研究领域的学术话语权。其次,有关原生数字资源的中文文献还存在非常明显的内容同质化现象,研究方向集中于图书情报学与计算机科学两门学科,在主题广度上远不及英文相关文献。最后,也是最重要的一点,近五年来我国专家学者对原生数字资源的研究兴趣大幅度趋淡,其发文量占全球比重面临着断崖式下滑,正在慢慢失去前期,特别是2009至2014年间积攒的科研优势。综上所述,我国原生数字资源研究似乎陷入了瓶颈,亟待“新鲜血液”的注入。

纵观我国数字化进程,21世纪第三个十年或将迎来新一波的原生数字资源研究热潮。2021年开春之际,纲领性文件《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》发布,专篇部署“加快数字化发展建设数字中国”<sup>[31]</sup>。随后不久,《“十四五”数字经济发展规划》《“十四五”国家信息化规划》等一系列规划密集出台,大力倡导“扩大优质数字资源供给,鼓励公共数字资源更大范围向社会开放”“提供方便快捷、资源共享的全国公共文化数字资源服务”,共同擘画“十四五”时期数字中国建设的美好蓝图<sup>[32-33]</sup>。结合上述政策文件的精神实质以及国内外原生数字资源研究现状来看,笔者认为在未来一段时间内,可以考虑从如下四个方面深入开展探索性工作。

(1) 资源转化与利用。数字化正在以不可逆转之势深刻地改变着人类社会。数据作为一种新型生产要素,行将走向世界舞台的中心。我国作为原生数字资源富地,坐拥海量数据的所有权和控制权,以致愈发迫切地需要推动这些资源向资产、甚至向资本进行转化,从而达成数字信息资源向数据要素的跃升,真正发挥这些资源在社会生产和再生产过程中的效力<sup>[34]</sup>。正如马克思所言:“使那些在原有形式上本来不能利用的物质,获得一种在新的生产中可以利用的形态”。

(2) 价值评估与定价。原生数字资源如何创造价值?中国信息通信研究院政策与经济研究所发布的《数据价值化与数据要素市场发展报告(2021年)》中提出了实现数据价值化的“三化”



框架, 即资源化、资产化、资本化<sup>[35]</sup>。其中, 资产化是铸就原生数字资源价值的必由之路, 主要包括两个阶段, 即资产价值评估和资产定价<sup>[36]</sup>。目前, 世界各国政府对数字信息资源的价值评估与定价都展开了积极探索, 但还未成熟到可供全面地推广与实施, 相关的理论、技术和方法皆有待进一步研究。

(3) 知识重组与创新。在数字经济时代, 知识与经济之间的相互渗透日益加深, 知识管理的作用空前凸显。与此同时, 知识管理也正面临着前所未有的挑战, 即如何有效地整合不同来源、分散无序、相对独立、多种模态的原生数字资源, 继而实现既有知识的重组, 并从中萃取出新的知识。对此, 最新的研究表明, 资源标引和实体/事理关联构建乃是知识重组的诀窍, 亦是利用数字信息资源来催生知识创新的基石<sup>[37]</sup>。

(4) 权属界定与流通。随着数据成为关键生产要素, 原生数字资源的权属问题也愈益突出。若权属不清, 则价值不明。该问题不但会阻碍原生数字资源开发利用和流通使用, 还会影响数字经济市场化的推进。从全球范围来看, 现有针对数字信息资源的权力规制可以分为欧美两大体系, 但两者皆未明确地对数据权属问题加以阐释。相较之下, 我国在这一方面的立法实践起步较晚, 法律制度体系尚不健全, 仍停留在对数据使用规范的强调, 亟需出台相关法律来界定数据究竟归谁所有<sup>[38]</sup>。

## 【参考文献】

- [1] 全国科学技术名词审定委员会. 图书馆·情报与文献学名词 [M]. 北京: 科学出版社, 2019: 91.
- [2] 人民网. 我国数字经济规模超 45 万亿元 (新数据 新看点) [EB/OL]. [2022-7-3]. <http://finance.people.com.cn/n1/2022/0703/c1004-32464202.html>.
- [3] 中国互联网络信息中心. 第 50 次《中国互联网络发展状况统计报告》[EB/OL]. [2022-8-31]. <https://www.cnnic.cn/n4/2022/0914/c88-10226.html>.
- [4] 求是网. 不断做强做优做大我国数字经济 [EB/OL]. [2022-1-15]. [http://www.qstheory.cn/dukan/qs/2022-01/15/c\\_1128261632.htm](http://www.qstheory.cn/dukan/qs/2022-01/15/c_1128261632.htm).
- [5] 陈京莲, 陈立军. 基于逻辑斯蒂模型的主题词被引频次和发文量增长的实证研究 [J]. 情报科学, 2019, 37 (1): 74-79.
- [6] 马费成, 宋恩梅. 信息管理学基础 (第 2 版) [M]. 武汉: 武汉大学出版社, 2011: 68-74.
- [7] 朱丹君, 段蓓虹. 构筑数字时代作品传播新格局——解决数字图书馆有关版权问题的必由之路 [J]. 图书馆, 2005 (3): 69-71.
- [8] Varian H R. Copying and copyright [J]. Journal of Economic Perspectives, 2005, 19(2): 121-138.
- [9] 方兴东, 钟祥铭, 彭筱军. 全球互联网 50 年: 发展阶段与演进逻辑 [J]. 新闻记者, 2019 (7): 4-25.
- [10] 陈红星, 张淑芳. 网络原生数字资源: 概念、特征与类型 [J]. 图书馆建设, 2010 (5): 1-4.
- [11] 夏南强, 李倩. 网络原生数字资源的类型、特点与开发 [J]. 图书情报工作, 2011, 55 (3): 47-51.
- [12] 吴吉义, 李文娟, 曹健, 等. 智能物联网 AIoT 研究综述 [J]. 电信科学, 2021, 37 (8): 1-17.
- [13] 成素梅. 人工智能的几个认识论问题 [J]. 思想理论教育, 2019 (4): 21-25.
- [14] Tzouganatou A. Openness and privacy in born-digital archives: Reflecting the role of AI development [J]. AI & Society, 2022, 37(3): 991-999.
- [15] 孙佳佳, 李雅静. 基于关键词价值细分的高价值热点主题识别方法研究 [J]. 情报学报, 2022, 41 (2):

118-129.

[16] 许鑫, 陈路遥, 杨佳颖. 数字人文研究领域的知识网络演化——基于题录信息和引文上下文的关键词共词分析[J]. 情报学报, 2019, 38(3): 322-334.

[17] 范亚芳, 曹作华. 网络原生数字资源馆藏化建设模式研究[J]. 情报理论与实践, 2012, 35(11): 85-88.

[18] Bury S J. Museum libraries and archives in the digital 21st century [M]//Giannini T, Bowen J. Museums and Digital Culture: New Perspectives and Research. Cham, Switzerland: Springer, 2019: 483-490.

[19] Kleinberg J. Bursty and hierarchical structure in streams [C]//Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Edmonton, Canada: Association for Computing Machinery, 2002: 91-101.

[20] Grigar D, Moulthrop S. Traversals: A method of preservation for born-digital texts [M]//Sayers J. The Routledge Companion to Media Studies and Digital Humanities. Abingdon, UK: Routledge, 2018: 351-361.

[21] Pinfield S, Cox A M, Smith J. Research data management and libraries: Relationships, activities, drivers and influences [J]. PLoS ONE, 2014, 9(12): e114734.

[22] Decker S, Kirsch D A, Venkata S K, et al. Finding light in dark archives: Using AI to connect context and content in email [J]. AI & Society, 2022, 37(3): 859-872.

[23] Clavert F. History in the era of massive data: Online social media as primary sources for historians [J]. Geschichte und Gesellschaft, 2021, 47(1): 175-194.

[24] 李宝强, 孙建军. 试论数字信息资源配置模式[J]. 情报资料工作, 2007(2): 44-48.

[25] 常娥, 魏彬. 网络原生数字资源优选与评价研究[J]. 情报杂志, 2012, 31(2): 163-167.

[26] 常娥, 袁曦临. 网络原生数字资源管理问题探析[J]. 图书馆建设, 2009(5): 27-30.

[27] 肖时占, 陈红星, 雷素芳, 等. 网络原生数字资源建设模式刍议[J]. 图书情报工作, 2012, 56(7): 58-61, 78.

[28] 郑雯, 毕艳芳. 论高校图书馆对网络原生数字资源的整合[J]. 图书馆论坛, 2012, 32(1): 94-98.

[29] 陈鹏. 数字方法的兴起: Web 在社会与文化研究中的重定位[J]. 哲学分析, 2018, 9(2): 108-118, 198.

[30] 许苗苗, 邵波. 我国数字人文发展的脉络、问题及启示[J]. 图书馆学研究, 2020(14): 2-10.

[31] 国务院. 中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要 [EB/OL]. [2021-3-13]. [http://www.gov.cn/xinwen/2021-03/13/content\\_5592681.htm](http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm).

[32] 国务院. 国务院关于印发“十四五”数字经济发展规划的通知 [EB/OL]. [2021-12-12]. [http://www.gov.cn/zhengce/content/2022-01/12/content\\_5667817.htm](http://www.gov.cn/zhengce/content/2022-01/12/content_5667817.htm).

[33] 中国网信网. “十四五”国家信息化规划 [EB/OL]. [2021-12-27]. [http://www.cac.gov.cn/2021-12/27/c\\_1642205314518676.htm](http://www.cac.gov.cn/2021-12/27/c_1642205314518676.htm).

[34] 李碧珍, 吴芄梅. 数字经济对社会生产与再生产过程的影响与重塑[J]. 当代经济研究, 2021(11): 14-19.

[35] 中国信通院政策与经济研究所. 数据价值化与数据要素市场发展报告(2021年) [EB/OL]. [2021-5-27]. [http://www.caict.ac.cn/kxyj/qwfb/ztbg/202105/t20210527\\_378042.htm](http://www.caict.ac.cn/kxyj/qwfb/ztbg/202105/t20210527_378042.htm).

[36] 尹传儒, 金涛, 张鹏, 等. 数据资产价值评估与定价: 研究综述和展望[J]. 大数据, 2021, 7(4): 14-27.

[37] 向安玲, 高爽, 彭影彤, 等. 知识重组与场景再构: 面向数字资源管理的元宇宙[J]. 图书情报知识, 2022, 39(1): 30-38.

[ 38 ] 童楠楠, 窦悦, 刘钊因. 中国特色数据要素产权制度体系构建研究 [J]. 电子政务, 2022 ( 2 ) : 12-20.

## Research on Born-digital Resources in China and Abroad: Review and Prospect

Wang Yongzhen Wang Xianwen

(WISE Lab, Institute of Science of Science and S&T Management, Dalian University of Technology, Dalian 116024, China)

---

**Abstract:** [ **Purpose/significance** ] This paper summarizes and discusses the current state of research on born-digital resources in China and abroad, with the purpose of providing a reference guide for subsequent studies. [ **Method/process** ] This paper selects Chinese and English literature on born-digital resources from 2000 to 2022 as the survey objects, and uses bibliometric methods and knowledge mapping tools to comprehensively investigate and compare the related publication outputs and popular topics in China and abroad. [ **Result/conclusion** ] The empirical findings presented in this paper show a noticeable gap in the number of publications on born-digital resources between China and other countries, especially the United States. Also, China's publications on born-digital resources are mostly published in domestic journals of low international visibility. In addition, there is content homogeneity among them, with fewer topics addressed than their international counterparts. Given these findings and the basic connotations of China's digital transformation, this paper provides Chinese academics and practitioners with four directions for future studies on born-digital resources, including transforming born-digital resources into assets that play a more positive role in social (re)production, evaluating the value of born-digital resources and developing a pricing mechanism accordingly, reorganizing the knowledge from multi-source heterogeneous born-digital resources to stimulate innovation, and clarifying the ownership of born-digital resources to promote their circulation in the market.

**Keywords:** Born-digital resources; Bibliometrics; Knowledge mapping; Visualization analysis

---

( 本文责编: 魏 进 )