

数据知识产权：将知识产权扩展到网络空间

叶雅珍^{1,2}, 朱扬勇^{1,2,3}

1. 复旦大学计算与智能创新学院, 上海 200438;
2. 上海市数据科学重点实验室, 上海 200438;
3. 上海数据研究院, 上海 200011

摘要

数据知识产权作为数据产权制度的重要内容而广受关注。虽然数据知识产权实践如火如荼, 但是否需要设立数据知识产权、如何设立数据知识产权在学术界尚存争议。分析讨论了数据形式的知识产权的存在性, 给出了可赋数据知识产权的3种数据集类型, 从新的保护客体、新的甄别方法、新的保护手段三方面论述了新设数据知识产权的必要性, 提出了一个数据知识产权的定义, 以及知识产权从物理空间向网络空间扩展形成数据知识产权的框架, 为数据知识产权推进提供了理论支持。

关键词

数据产权; 知识产权; 数据权益保护

中图分类号: TP399

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2025051

Data intellectual property is the extending of intellectual property rights from reality to cyberspace

YE Yazhen^{1,2}, ZHU Yangyong^{1,2,3}

1. Computation and Artificial Intelligence Innovative College, Fudan University, Shanghai 200438, China
2. Shanghai Key Laboratory of Data Science, Shanghai 200438, China
3. Shanghai Data Research Institute, Shanghai 200011, China

Abstract

Data intellectual property has garnered significant attention as a crucial component of the data property rights system. Despite the rapid development of data intellectual property practices, there remains academic debate regarding whether it is necessary to establish data intellectual property rights and how such rights should be structured. In this context, the existence of intellectual property in the form of data is analyzed, and three types of datasets that could be endowed with intellectual property rights are identified. The necessity of creating novel data intellectual property rights is discussed from the perspectives of newly protected subject matter, innovative identification methods, and enhanced protection mechanisms. A definition of data intellectual property is proposed, and a conceptual framework is constructed in which intellectual property extends from the reality to cyberspace, thereby providing theoretical support for the advancement of data intellectual property.

Key words

data property rights, intellectual property rights, data rights protection

0 引言

随着数字经济的发展和数据要素市场构建的不断深入，数据的资产性、财产性已获得广泛关注和认同，有关数据财产的保护及数据产权制度的构建等议题竞相提上日程。由于目前有关数据财产权利的法律体系构建及具体立法方案还有待进一步论证，在现行法律框架下，构建数据知识产权保护规则不仅是贯彻落实国家政策的重要举措，也是建立数据产权制度的重要内容，具有重要意义，得到各方重视。

数据知识产权的探索建设工作正在积极开展。特别是在2021年中共中央、国务院印发的《知识产权强国建设纲要（2021—2035年）》及国务院印发的《“十四五”国家知识产权保护和运用规划》等文件的指导和部署下，国家知识产权局抓紧落实，于2022年11月确定了8个首批数据知识产权试点地区，至2024年再新增9地，目前试点地区已扩大至17个省区市。各地结合自身特点积极开展数据知识产权探索工作，陆续出台各自的数据知识产权登记办法，各地颁发证书总量已高达2万多份并在持续增加中，且在司法裁判、资产入表等场景中成功应用，取得了可喜进展。各试点工作的有效经验为数据获得知识产权保护提供了重要参考和实践支撑。

一些数据的知识产权是与生俱来的，一些则是知识产权数字化生产方式带来的，然而，更多的网络空间中的创造性劳动产生的大量数据需要赋予知识产权来加以保护。本文通过对原始数据的创造性劳动和

数据再生产的创造性劳动进行讨论，分析了数据形式知识产权的存在性；给出了可赋予数据知识产权的3种数据集类型，即知识产权数据化形成的数据集、知识产权数字化生产形成的数据集和以生产数据为目的的智力劳动形成的数据集；从新的保护客体、新的甄别方法、新的保护手段3个方面论述了新设数据知识产权的必要性；给出了数据知识产权的定义，提出了知识产权从物理空间向网络空间扩展形成数据知识产权的框架，为数据知识产权推进提供了理论支持。

1 数据知识产权的存在性

已经有很多知识产权以数据的形式存在，换一种说法就是已经有很多数据是具有知识产权的。除此之外，针对更一般的数据，还有哪些数据集是可以被赋予知识产权的？这是本文要探讨的内容。知识产权的核心是“具有创造性劳动的智力成果”^[1]，那么数据是否含有创造性劳动的智力成果呢？有两个问题：第一个是原始数据的知识产权问题，即是否能够以及如何能够在原始数据产生时获得数据知识产权？第二个是再生产数据的知识产权问题，即是否能够以及如何能够对原始数据进行加工创造后产生的新数据赋予知识产权？^[2]

1.1 原始数据的创造性劳动

以照片为例。一张纸质照片经扫描后的数据照片和一张用数码相机拍摄的数据

照片，二者在本质上是一样的并无不同。一张纸质照片如果因为摄影师的独特视角、构图、后期处理等创造性劳动而获得知识产权，那么一张用数码相机拍摄的数据照片也应该因为摄影师的独特视角、构图、后期处理等创造性劳动而获得知识产权。进一步，如果拥有知识产权的纸质照片扫描后的数据照片仍然延续了原有知识产权，那么直接用数码相机拍摄的有创造性劳动的数据照片更应被赋予知识产权。拍摄照片是原始数据的生产，因此，存在一些原始数据是创造性劳动的智力成果。

再以计算机软件为例（注意，计算机软件也是数据）。当计算机软件达到相关要求后，它就是可获得相应知识产权保护的数据产品。值得注意的是，在当前的低代码环境下，软件开发所写的代码数量越来越少，但在此基础上形成的计算机软件仍然可以获得软件著作权等有关知识产权的保护；由此可见，汇聚了他人的数据并融入创造性内容而形成新的数据产品的情形是可以获得知识产权保护的。这个例子有两层含义，一层含义是由程序员自己开发的那部分软件，可以看作原始数据的生产；另一层含义是作为数据产品的计算机软件融入了其他程序员已经开发好的软件。这就是说，融合已有数据的新数据生产可以看作一种创造性劳动。

网络小说也是原始数据的创造性劳动。与从实物形态发展而来的数据产品不同，网络小说是没有纸质出版物的数据产品，并不是以完整作品形式进行发表的，而是创作者直接在网络上进行内容的实时创作，并按照章节以在线方式更新发布，读者在线浏览阅读，小说的发展方向和更新与否乃至篇幅长短等在一定程度上由小说的热度来决定（无人问津则停更，反响热烈则长续，甚至有篇幅高达千万字的小说作

品），这种情形在传统纸质图书中是很少见的^[3]。更重要的是，网络小说是纯粹在网络空间中创造出来的数据产品，是典型的创造性劳动成果，受到了知识产权的保护。

1.2 数据再生产的创造性劳动

通常，一份研究报告或者一篇学术论文是在现有文献基础上开展分析、总结、推理和创新的，从而得出新的结论，形成创造性的智力成果。这与对网络数据进行加工，从中提取、整理、分析并形成新的数据产品本质上是一样的，即两者都有创造性劳动的存在。

数据库汇编是数据再生产形成知识产权的典型例子。数据库汇编通过对已有数据的收集、选取、整理、编排等形成一个具有特定组织结构和价值用途的数据集合。这一过程中，有关数据再生产的劳动整体上可获得版权或专有权利的保护^[4]。对在数据选取与编排方式等方面体现出独创性的数据库，相关组织会提供版权保护；而对那些内容未达到独创性要求的数据库，在欧盟还专门有《欧洲共同体关于数据库法律保护的指令》（简称《数据库指令》）给予数据库所有者为期15年的特殊权利保护^[5]，旨在有效防止竞争对手对数据进行不公平提取和使用行为，从而鼓励人们对数据库开展实质性投资^[6]。

中国知网、企查查等是数据库汇编这一数据再生产行为的实践代表。中国知网作为知名学术数据库，不仅汇聚了期刊论文、会议文献和学位论文等学术资源，还通过数据加工方式构建了包括引文索引、学科分类等在内的知识组织框架，开发了查重检测、学术影响力评价等增值服务功能，这些创造性工作使其形成了受法律保护的数据库编排架构，具备了学术知识服

务能力和商业价值。同样，企查查作为商业信息平台，不仅汇集整合了企业工商登记、司法判决、知识产权等多维数据，更是利用相关技术，构建了股权穿透分析、关联图谱等可视化商业情报系统，开发了风险预警、信用评估等决策服务功能，这些创造性工作使其形成了专业的商业情报服务能力和市场竞争优势。

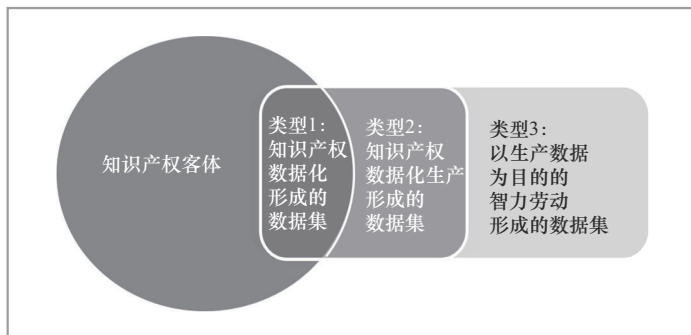


图1 具有知识产权的3种数据集

2 具有知识产权的3种数据集

图1展示了具有知识产权的3种数据集。

2.1 类型1:知识产权数据化形成的数据集

一部分传统知识产权正通过计算机软件技术转化为数据形式，并受其原有知识产权保护，使其在网络空间受保护地合法流通。以音乐作品为例，作曲者、词作者等在创作完成音乐作品后享有著作权，并会以初始版权人身份将音乐作品的部分或全部版权转让给音乐出版商，音乐出版商在获得版权人的机械复制权等相关授权许可后，会将音乐作品制作成唱片加以发行^[7]。而以iTunes Store^[8]、Spotify为代表的数字音乐服务商/平台在从音乐作品版权人及录音版权人（唱片公司）等处¹获得相关授权许可²后，将传统音乐产品（如磁带、黑胶唱片等）进行数据形式的标准化处理得到音乐数据产品。得益于音乐版权制度的发展和不断完善，音乐作品法定许可的数字环境适用性得到了扩张，使音乐数据产品沿袭了原有知识产权并可在市场上合法流通，从而形成了如今高达百亿美元^[9]的产业规模。

类似的情形还有很多。例如，现有很

多图书馆在开展数字化工作，将传统的绘画、书籍、影视作品等馆藏数字化为电子数据^[10]。它们大多从已有的实物作品发展而来，是一类具有知识产权的数据集。

2.2 类型2:知识产权数字化生产形成的数据集

知识产权的生产过程也正在经历着深刻变革，利用数字化设备或计算机软件技术直接在网络空间中创造生产得到具有知识产权属性的作品变得越来越普遍，即知识产权生产过程也被数字化，并直接得到数据形式的结果。

随着数码相机及数据摄影设备的普及，摄影作品的创作方式发生了根本性变化，创作者可以直接以数据形式进行创造和呈现，快速生产出具有知识产权属性的创作成果，使摄影作品的传播、版权保护等全面进入数字化模式。

计算机软件是更典型的例子。程序员在开发计算机软件时，会调用各种各样的库函数、模块代码、开源代码等，在此基础上结合自己编写的代码最终开发出一款计算机软件。该软件中倾注了开发者的智慧和劳动，被视为著作权法保护范围内的客体对象，可以通过软件著作权等保护开发者相关权益。此外，由代码、文档、角

①如果涉及音乐作品的公开表演，还需要向表演权管理组织申请相关权利许可。

②2018年，美国出台了《音乐现代化法案》，对在数字环境下的音乐作品制作录音的法定许可申请程序加以完善，从按作品许可改为一揽子许可，并专门设立一个新的非营利性的“机械权许可组织”，负责管理一揽子法定许可下的授权和许可费收支，以及收集音乐作品及产品的权利管理信息，对永久下载的音乐数据产品采用“贯穿式许可”，即由唱片公司等录音版权人代向数据音乐服务商/平台收取音乐作品法定许可费，从而使音乐数据产品服务更便利和高效地获取音乐版权授权。

色设计、背景设计、音乐、脚本等多元素共同构成的网络游戏，是基于计算机软件技术在网络空间中的创造，凝结了游戏开发者的智慧和劳动^[11]。

另外，由科学家创造性生成的科学数据，以及开展了有关数据预处理、汇编、分析处理的数据集，乃至专门构建的数据集等，都是依托计算机软件技术在网络空间中倾注智慧而产生的成果，具有一定的创造性特征，大部分属于数据形式的智力成果。

基于数字化方式直接生产的数据集，与前者在网络空间的存在形式并无差别，皆为数据形式，因此其中满足知识产权的保护条件的数据集获得了相关知识产权保护（如网络小说获得著作权保护，计算机软件可获得包括著作权、专利权和商业秘密等在内的多种知识产权的适度保护等）。

2.3 类型3：以生产数据为目的的智力劳动形成的数据集

除前述在网络空间中已获得知识产权保护的数据集外，其他以数据形式存在的需要纳入知识产权保护且符合条件的数据集，应被赋予知识产权保护。这正是当前数据知识产权受到广泛重视的原因。

温州市大数据运营有限公司开发的“信贷数据宝居民信用贷款能力分析数据”（登记号：DIP2024000000439）就是一个典型案例。该数据产品基于授权使用的公积金、社保、不动产等公共数据，通过创造性的智力劳动，构建了居民可支配收入能力评估模型，为金融机构提供贷前业务授信额度测算依据，以及贷中还款风险评估服务，已应用于20多家金融机构。作为浙江省首批实现数据资产入表的数据知识产权，该产品严格遵循“姓名+身份证号”的用户授权机制，在保障隐私安全的前提

下，将原始公共数据转化为具有商业价值的信用评估服务，不仅验证了公共数据授权运营的创新模式，也体现出了数据知识产权对于以生产数据为目的的智力劳动形成的数据集作为创新客体保护的重要意义和作用。

类似地，温岭市聚焦草编非物质文化遗产保护，登记数据知识产权9件。以浙江舜浦工艺美术品股份有限公司登记的“帽子洗标工序效率等级情况数据”（DIP2023000001762）为例，该数据集通过收集分析草编帽子的物料名称、效率等级、洗标工序等关键数据，实现了对草编工艺、流程、质量、效率的监测和改进。该数据集以生产数据为目的，凝结了企业对传统工艺进行数字化改造的智力劳动成果。作为全国首个非遗领域数据知识产权登记案例，温岭市获得了5 000万元融资授信，为传统非遗技艺的数字化传承和产业化发展提供了示范，充分展现了数据知识产权在激活传统文化价值、赋能传统产业升级方面的重要作用。

3 新设数据知识产权的必要性

无论是原始数据生产，还是数据再生产，均可能蕴含创造性智力劳动。因此，研究更广泛的由创造性智力劳动形成的数据集，并赋予此类智力成果以知识产权，不仅具有逻辑上的必然性和法理上的正当性，更是实现数据价值化的关键。那么，是否需要新设数据知识产权？答案是肯定的。这主要是因为数据知识产权的保护对象、甄别方法以及保护手段均具有显著的新颖性和独特性，现有知识产权体系难以充分覆盖和完全适用^[12]。

以照片为例，数据照片和传统纸质照

片在劳动、传播、权益认定与保护等方面存在明显差异，见表1。

数据知识产权的核心在于知识产权客体的表现形式，即数据集。这导致数据集的知识产权属性甄别方法、保护手段发生了变化，因此，需要新设数据知识产权。

3.1 新的保护客体

知识产权领域出现了新的保护客体——数据集。

首先，对数据的创造性智力劳动是需要保护的。创造性劳动是版权保护的核心原则。欧盟《数据库指令》超越了该原则，放宽了创造性要求，使其在版权界备受争议。然而，技术的进步及经济的发展势必对版权及有关智力劳动成果的相关权利范围扩张提出新要求，正如欧盟《数据库指令》通过引入“辛勤原则”^[6]来适应新兴数据市场竞争和挑战，虽然实际效果有限并未达到预期^[13]，但也为数据相关的知识产权保护工作做了有益探索。不论未来数据领域的知识产权保护范围及制度建设如何发展和构建，创造性劳动必须基于“人类创作/发明原则”，即数据产品中必须具有人类实质性贡献，这点是判断能否给予数据产品版权保护的先决标准。在AI发展

迅猛的当下，数据产品（数据作品）含有的AI生成（创作）的浓度越来越高，只有在人类作者对其表达元素有足够的控制时，该产品（作品）方能受到版权保护，这点受到了绝大多数国家政府相关机构的普遍认同^[14]。

其次，除了针对数据的创造性智力劳动之外，一些具有商业秘密的数据集也是需要知识产权保护的。比如，科研机构在研究过程中收集、整理和呈现的科学数据，虽然其中有一大部分是对客观事实的记录描述，不一定满足创造性要求，但因其具有较高的实用价值和商业价值，可以通过商业秘密或特殊权利获得保护。再如，企业用于新产品开发的数据集、市场调研报告及内部运营数据等，这些数据自身具有极高的商业价值，并被施加了合理的保密措施，可通过商业秘密保护数据内容。

当然，既有原创性（创造性）又有实用性或商业价值的数据集，肯定是可以被赋予知识产权的。例如，原创设计用于实验设计、算法实验和算法效果评估的相似点集数据挖掘实验数据集^[15]，在创造过程和成果上都具有独创性，并兼具实用性和经济价值，应被赋予知识产权，在满足一定条件下可以受到版权、数据库特殊权利、商业秘密、专利权等多种知识产权的

表1 数据照片和纸质照片的差异

维度	数据照片	纸质照片
劳动	通过数字设备大量产生,随时生产,成本低廉	胶片损耗、冲印困难,成本高,数量小
传播	大规模网络传播	实物照片分发
共享	大规模同时共享	分时共享
保存持久性	可方便转储,持久	材料损耗
权益保护	需赋予数据知识产权	著作权(版权)
市场监管	尚未健全	已成熟

保护，其最符合的是数据知识产权的保护条件。

人们在工作学习中广泛利用计算机软件技术开展各类创造性劳动，如使用计算机办公软件编写 Word 文档、PPT 文稿、Excel 表格等数据形式的文件内容，以及使用计算机绘图软件开展各类数据形式的技术图纸设计和艺术作品创作工作等。这些作品总体上具备了知识产权要求的条件。

更一般的数据集能否被赋予知识产权，取决于其是否属于人类创造性劳动的数据形式智力成果，是否具有实用性或商业价值。虽然直接纳入知识产权保护的原始数据仍然不多，但经过创新处理、技术优化或商业化运营的数据集，则可能因其具备独创性而受到版权保护，或因其实用性和经济价值而受到商业秘密、专利或其他形式的知识产权保护。未来，随着数据产权体系的发展，对一般数据的知识产权界定可能会更加明确，从而为数据市场提供更清晰的法律保障。

3.2 新的甄别方法

新的知识产权保护客体（数据集）需要新的甄别方法。甄别一份研究报告或一篇学术论文是否存在抄袭或剽窃行为是判断其是否具有原创性的重要标准。显然，界定一个数据产品是否存在创造性劳动，也需要对其在数据的选择、编排、组织或处理方式等方面是否存在抄袭或剽窃行为开展甄别。但是，针对数据的甄别要难得多。

以论文创新性甄别为例。论文创新性甄别主要靠同行评议。同行专家通过阅读论文、复现试验结果等手段来判别一篇论文是否有创新性、是否存在抄袭行为。然

而，一个数据集有时可能会大到超过一个传统图书馆的纸质藏书总量，显然，甄别数据集的创新性和是否抄袭不可能采用人工阅读的方式，而是要通过人工智能等技术来加以甄别。这是传统知识产权开展实质性审核所欠缺的。

又如，对一个 PB 级的数据集进行知识产权甄别，人工根本无法进行，可能需要计算机集群处理一周甚至更长时间，而这需要大量的算力予以支撑，技术层面的软硬件问题值得关注。当前缺乏数据集“创新性”标准量化体系，特别是对数据重组、特征提取等衍生作品的原创性进行判定的标准尚未建立。法律规制层面，有关数据的权属问题及其合理使用范围等问题正在逐步解决。这些问题的复杂性远超传统文本抄袭检测，需要“技术-标准-法律”三位一体的协同创新方可有效应对。

3.3 新的保护手段

新的知识产权保护客体（数据集）需要新的保护手段。由于数据非常容易复制并快速传播，如何保护数据知识产权是个非常困难的问题。当前，数据要素市场建设成果寥寥的一个主要原因是数据权益保护方法的缺失。例如，当某位数据购买者在数据市场购入某一数据集后，其可迅速将该数据集转售给多个第三方，这种行为严重侵害了数据原创者的合法权益。这类问题通过传统的知识产权保护手段，如取证、追溯、诉讼等，往往难以得到有效解决，其中如何电子取证、如何快速阻止传播等问题需要采用新型存证方式和追踪方式来加以解决和实现。但这些新技术新方法又不适用于传统知识产权保护。因此，需要建立新的知识产权保护手段，即建立专门的数据知识产权保护手段。

4 将知识产权扩展到网络空间

4.1 现有数据知识产权概念分析

有关数据知识产权的概念内涵，目前各数据知识产权试点地方、社会团体以及有关学者都进行了积极探索，多地政策法规文件关于数据知识产权概念的表述见表2。这些表述有共性的内容，也有分歧，说明在实践层面关于数据知识产权的界定尚未达成共识。

在共性内容方面，虽然各地关于数据知识产权概念的表述有所不同，但都涉及权利客体、权利主体、权益内容等核心要素，总体来讲，数据知识产权保护的是满足一定条件的数据集合的合法权益，给予法律上的初步确认。这些条件包括：

- 数据收集路径必须合法合规；
- 经过一定规则或算法处理加工的；

- 具有实用价值或商业价值的；
- 具备智力成果属性的；
- 尚未公开的。

在分歧方面，上海市的概念中还强调了“实质性加工和创新性劳动”这一要件，这是与其他地方的概念的不同之处。基于此，有关数据知识产权的审核模式主要分为两派。

一是以浙江省为代表的形式审查模式。形式审查是知识产权审查的一种方式，其聚焦于申请材料的程序性审查，仅对文档材料的形式要件（包括文件完整性、格式规范性及基本法律要件）进行合规性核查，不涉及对数据集内容实质创新性的专业评判。这种审查方式具有审查周期短、行政成本低、操作流程标准化等特点。为保障审查质量，通常采用“自愿登记、形式审查”的基础框架，同时设置“初步公示、异议程序”的监督机制，还要求申请者前置完成数据存证及/或公证程序。这样既确保了审查效率，又通过程序制衡维护了登

表2 多地政策法规文件对数据知识产权概念的界定

出处来源	类型	数据知识产权概念	发布时间
浙江省产品与工程标准化协会 《数据知识产权质押服务规程》 (T/ZS 0270—2022)	团体标准	数据知识产权为自然人、法人和非法人组织对其合法处理数据加工后形成的数据产品和服务,依法享有的财产权益 ^[16]	2022年3月
杭州市《数据知识产权交易指南》 (DB3301/T0403—2023)	地方标准	数据知识产权是权利人对经过一定算法加工、具有实用价值和智力成果属性的数据,依法享有的专有的权利 ^[17]	2023年5月
《山东省数据知识产权登记管理 规则(试行)》	地方性政策文件	数据知识产权是指权利主体对于依法依规获取,经过一定规则处理形成的,具有实用价值、智力成果属性及非公开性的数据集合,享有的自主管控、加工使用、经营许可和获得收益等权益	2023年10月
《浙江省数据知识产权实务指引 (试行)》	地方性政策文件	数据知识产权是数据处理者对依法获取、经过一定规则处理、具有实用价值和智力成果属性的数据集合所享有的权益	2024年8月
《上海市数据产品知识产权登记 存证暂行办法》	地方性政策文件	数据产品知识产权,是指自然人、法人或者非法人组织对其合法获取的数据资源,经过实质性加工和创新性劳动后形成的具有智力成果属性和商业价值的数据加工集合、数据加工产品、数据技术算法等数据产品享有的权益	2024年11月

记公信力。

二是以上海市为代表的实质性审查模式。实质性审查是另一种知识产权审查方式，是在形式审查的基础上对数据集实质创新性内容进行审核。显然，要对数据集开展实质性审核是非常困难的，上海市的做法是针对数据产品赋予知识产权，审查数据产品是否存在“实质性加工”和“创新性劳动”，重点考察是否有应用场景、是否属于智力成果。为了确保实质性审查的质量，申请登记的项目要经过人工智能初审和人工复审两道程序，部分有争议的项目还要经过专家三审；此外，还引入了数据产品登记机构参与部分实质性审查工作，从而满足市场主体对可流通交易的数据加工产品和服务的登记需求，也为破解数据知识产权理论争议提供了实践样本。

上述两派总体来讲都认可新设数据知识产权的必要性，并给出了甄别方法的原则和保护原则，但两派在概念界定和登记实践等方面仍存在显著分歧，尚未形成统一标准。特别是，上海市强调了实质性审核要求，对数据集智力劳动的实质性审核非常困难，需要技术创新来予以支撑。

另外，更值得注意的是，现有定义在适用范围方面存在局限性，当前两派的定义都主要针对“类型3：以生产数据为目的的智力劳动形成的数据集”，而有关“类型1：知识产权数据化形成的数据集”和“类型2：知识产权数字化生产形成的数据集”的均未涉及，显然现有的数据知识产权的相关定义具有不完备性。

综上分析，亟须重新给出一个兼具各方共识性和理论完备性的数据知识产权定义。

4.2 定义数据知识产权

虽然数据知识产权在保护客体上存在

特殊性，对应的法律框架还需进一步完善，但其本质上符合知识产权保护的基本条件。定义数据知识产权需要考虑两个方面：知识产权的符合性和涵盖3种类型的数据集。

（1）知识产权的符合性

首先，数据知识产权是一类知识产权，定义数据知识产权必须要满足知识产权的要求，即数据知识产权必须符合传统知识产权的基本特征。获取数据知识产权，其客体所需满足的一定条件与传统知识产权客体所需具备的基本核心条件高度契合。主要体现在如下4个方面。

- 合法性要求：要与传统知识产权保护中的“合法性”要求一致，并要求保护对象不得违反法律或公序良俗。因此，作为数据知识产权的客体，数据集的生产、收集和加工路径、手段必须合法合规。只有通过合法合规的途径获得的数据集，才能作为合法权益受到保护，这为得到数据知识产权保护提供了正当性基础。

- 创造性或智力成果属性：要与著作权保护中的“独创性”、专利权要求发明具有“创造性”等要求相符，数据集合需经过一定规则或算法进行实质性加工处理，并具备“独创性”或“创造性”的智力成果属性。

- 实用性或经济价值属性：要与专利权要求发明具有“实用性”、商业秘密需证明信息的经济价值等知识产权保护要求相一致，数据集要能在实际应用中展现出显著的商业优势或技术创新，其实用性或经济价值得到认可。

- 秘密性：要与商业秘密的“秘密性”要求一致，也与专利的“新颖性”保护逻辑相通。因此，尚未公开的数据集合更符合知识产权保护的本意，也是为了维护数据所有者的商业秘密和竞争优势。

（2）涵盖3种类型的数据集

当前各地数据知识产权的表述主要是针对类型3的。如前所述，类型1、类型2、类型3的数据集都具有知识产权。因此，数据知识产权的定义必须涵盖这3种类型的数据集。

- 类型1：知识产权数据化。这类数据集的知识产权属性体现在数据产生之前，表面上看，数据化只进行了表现形式的转化，其甄别方法没有变化，但是，表现形式转化成数据后，保护对象和保护手段都发生了变化。

- 类型2：数字化生产。一个知识产权的数字化生产，如数码相机拍照、网络小说创作，其产生的数据集客体在保护对象、甄别方法、保护手段等方面都需要新的规制和技术的支持。

- 类型3：以生产数据为目的。如上讨论，这类数据集是各界主要关注的知识产权问题。以生产数据为目的，保护数据集生产者的合法利益，数据集的生产、收集和加工的路径、手段等的甄别都是新的课题，保护对象和保护手段都是新的。

综上，定义数据知识产权如下。

定义（数据知识产权）：数据知识产权是指智力创造的以数据集为表现形式的成果；数据知识产权，又称数据知识财产权，是关于人类在社会实践中创造的以数据集为表现形式的智力劳动成果的专有权利。

数据知识产权的核心是知识产权体系中出现了一种新的表现形式——数据集，即出现了保护客体为数据集的情形。这导致需要新的甄别方法和新的保护手段。因此，数据知识产权概念定义的要点是以数据为表现形式的知识产权。

显然，该定义完全具备知识产权的符合性，并且该定义也完全涵盖了3种类型的数据集。在数据知识产权定义的基础上，

再研究数据知识产权的甄别方法和保护手段。

4.3 知识产权向网络空间扩展

“以数据为表现形式”体现出知识产权的一种新的形式，或者说一种新的知识产权客体形式。这种形式的知识产权客体是网络空间中的数据集，这与传统知识产权不同，可以看成知识产权从物理空间向网络空间的扩展（如图2所示）。

由于数字技术的进步，一些知识产权（如音乐、影像和图书等知识产权作品）首先被数字化成网络空间的数据集，曾经的五大唱片公司及音乐版权机构状告Napster公司事件^[18]，向人们展示了数据形式的知识产权保护与传统知识产权保护的巨大不同。随后，一些数字化设备生产的数据集依然能够按照传统的知识产权方法进行甄别，例如数码相机生产的照片、网络小说等，但其权益保护问题非常突出。当前，更多的非传统概念的含有智力劳动的数据集（不是一幅画、一首音乐等传统概念，而是一张数据表格、一个算法实验数据集）等待着知识产权保护。这些问题使我们不得不研究知识产权概念向网络空间的延伸，形成专门针对网络空间中含有智力劳动的数据集的知识产权，即数据知识产权。

可以说，数据价值越被认可，关于数据权益的保护就越受重视。其中，数据集的知识产权的存在性已经有了相当长的历史。因此，将知识产权向网络空间延伸，形成数据知识产权，是非常有意义和有必要的。

类型1、类型2和类型3是三大类可被赋予知识产权的数据集，共同构成了数据知识产权的基本范畴。值得注意的是，类型1和类型2都是符合数据知识产权定义

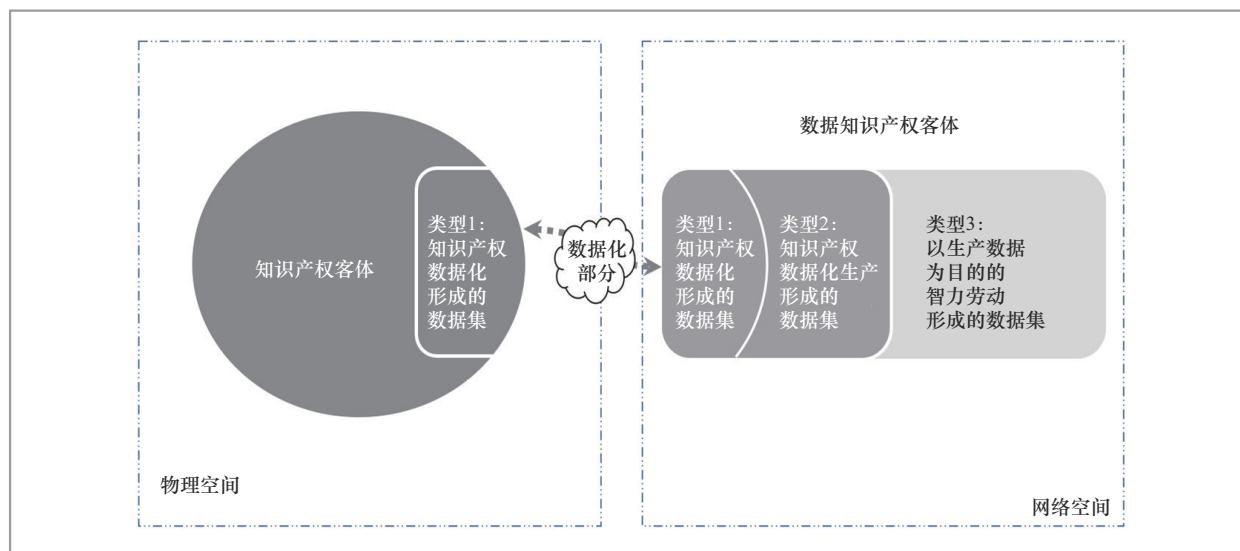


图2 知识产权向网络空间扩展框架

的，将这些数据集重新赋予数据知识产权，纳入数据知识产权管理，将有利于这些数据集的流通使用、权益保护，更好地发挥其价值。数据知识产权的客体是网络空间中的数据集，其创造来源是物理空间的人或组织，其价值实现最终是货币。实际上，数据知识产权构成了物理空间与网络空间融合的知识产权体系。

5 结束语

数据价值凸显，需要法规保护，需要数据产权制度。在数据产权制度尚未建立时，数据知识产权先行就成为一种必然选择。在网络空间中，有一部分数据蕴含着创造性劳动，具有智力成果属性，应该被赋予知识产权进行保护。在现有产权制度框架下，通过既有方式对可赋予知识产权的数据集进行保护，虽然能在一定程度上满足当前实践的部分需求，但是仍然具有片面性，并非长久之计。数据知识产权作

为知识产权在网络空间的延伸，是一种数据产权形式。在国家政策的指引下，有关数据知识产权的实践探索工作正在有序开展，为数据知识产权制度的建设提供丰富的实践经验和重要参考。

本文从新的保护客体、新的甄别方法、新的保护手段3个方面论述了新设数据知识产权的必要性，给出了数据知识产权的定义。后续工作将围绕新的甄别方法、新的保护手段两个方面开展：①进一步加强包括数据创造性劳动甄别在内的理论研究，深入探讨数据的智力成果属性及其法律保护的边界；②总结各地试点的成功经验，积极推动数据知识产权法律制度的完善工作，逐步构建物理空间与网络空间融合的知识产权体系，以确保数据知识产权的持续有效保护。

参考文献：

- [1] World Intellectual Property Organization. What is intellectual property? [R].

2021.

[2] 叶雅珍, 朱扬勇. 数据知识产权: 一种可流通交易的数据权[J]. 大数据, 2024, 10(2): 192-195.

YE Y Z, ZHU Y Y. Data intellectual property is a negotiable and tradable data property[J]. Big Data Research, 2024, 10(2): 192-195.

[3] 叶雅珍, 朱扬勇. 数据资产[M]. 北京: 人民邮电出版社, 2021.

YE Y Z, ZHU Y Y. Data assets[M]. Beijing: Posts & Telecom Press, 2021.

[4] 吕炳斌. 数据知识产权登记: 商业秘密模式抑或数据库模式[J]. 知识产权, 2024, 34(6): 62-79.

LV B B. Data intellectual property registration: trade secret model or database model[J]. Intellectual Property, 2024, 34(6): 62-79.

[5] 单晓光. 数据知识产权中国方案的选择[J]. 人民论坛·学术前沿, 2023(6): 38-47.

SHAN X G. The Chinese option for data intellectual property rights[J]. Frontiers, 2023(6): 38-47.

[6] 克里斯托弗·梅, 苏珊·K. 塞尔. 知识产权批判史[M]. 马腾, 译. 北京: 商务印书馆, 2024.

MAY C, SELL S K. Intellectual property rights: a critical history[M]. Translated by MA T. Beijing: The Commercial Press, 2024.

[7] 刘家瑞. 论美国数字音乐版权制度及启示[J]. 知识产权, 2019, 29(3): 87-104.

LIU J R. Law and policy of digital music copyrights in the United States[J]. Intellectual Property, 2019, 29(3): 87-104.

[8] WALDFOGEL J. Music file sharing and sales displacement in the iTunes era[J]. Information Economics and Policy, 2010, 22(4): 306-314.

[9] IFPI. Global music report 2023[R]. 2023.

[10] 叶雅珍, 刘国华, 朱扬勇. 数据产品流通的两阶段授权模式[J]. 计算机科学, 2021, 48(1): 119-124.

YE Y Z, LIU G H, ZHU Y Y. Two-step authorization pattern of data product circulation[J]. Computer Science, 2021, 48(1): 119-124.

[11] 王春晖, 戚湧, 董玉鹏. 数字知识产权理论与实践[M]. 北京: 人民邮电出版社, 2024.

WANG C H, QI Y, DONG Y P. Theory and practice of digital intellectual property[M]. Beijing: Posts & Telecom Press, 2024.

[12] 肖冬梅. “数据”可否纳入知识产权客体范畴?[J]. 政法论丛, 2024(1): 137-148.

XIAO D M. Can “data” be integrated into the category of objects of intellectual property rights? [J]. Journal of Political Science and Law, 2024(1): 137-148.

[13] 孔祥俊. 商业数据保护的实践反思与立法展望: 基于数据信息财产属性的保护路径构想[J]. 比较法研究, 2024(3): 72-92.

KONG X J. Legislation for commercial data protection: reflection and outlook: a conception based on information property attribute of data[J]. Journal of Comparative Law, 2024(3): 72-92.

[14] U. S. Copyright Office. Copyright and artificial intelligence, part 2: copyright - ability[R]. 2025.

[15] 朱扬勇. 相似点集挖掘实验数据集[M]. 上海: 上海科学技术出版社, 2021.

ZHU Y Y. Experimental dataset for similar point set mining[M]. Shanghai: Shanghai Scientific & Technical Publishers, 2021.

[16] 数据知识产权质押服务规程: T/ZS 0270—2022[S]. 杭州: 浙江省产品与工程标准化协会, 2022.

Code for pledge service of data intellectual property: T/ZS 0270—2022[S]. Hangzhou: Zhejiang Association for Standardization of Products and Engineering, 2022.

[17] 数据知识产权交易指南: DB3301/T 0403—2023[S]. 杭州: 杭州市市场监督管理局,

2023.
Data Intellectual Property Transaction
Guide: DB3301/T 0403—2023[S]. Hang -
zhou: SCJG of Hangzhou, 2023.

[18] HONIGSBERG P J. The evolution and
revolution of Napster[J]. University of
San Francisco Law Review, 2001(36):
473-508.

作者简介



叶雅珍（1985-），女，博士，复旦大学计算与智能创新学院教师，上海市数据科学重点实验室数据资产研究室主任，复旦大学数据产业研究中心主任助理，上海数据研究院特聘研究员，浙江省数据知识产权研究基地智库专家。专著《数据资产》作者，在数据资产研究与实践方面创新地提出“数据资产化框架”“数据产品运营的两阶段授权模式”“盒装数据产品形态”“数据交易的标的形态及其权利授予体系”“基于测度空间的数据资产量化定价模型”等。主要研究方向为数据科学和数字经济，近期主要研究方向为数据资产、数据知识产权、数据流通和数字化转型等。



朱扬勇（1963-），男，博士，复旦大学计算与智能创新学院教授，复旦大学数据产业研究中心副主任，上海数据研究院学术副院长。《大数据》期刊编委会副主任，农业大数据产业技术战略联盟副理事长兼首席科学家，大数据协同安全国家工程研究中心室副理事长。国际数据科学倡导者，提出数据界和数据学、特异群组、数据自治、数据资产、数据财政等概念和体系。发表学术论文200多篇，出版《数据学》《旖旎数据》《特异群组挖掘》《数据自治》《数据资产》等专著，《大数据技术与应用丛书》主编、《大数据资源》和《大数据技术》主编。主要研究方向为数据科学和数字经济，近期研究重点为数字化转型、数据财政、数据资产、数据自治与数据跨境等。

收稿日期: 2025-05-13