

论数据的法律界定

许春明 杨欢欢*

目次

- | | |
|----------------------|-------------|
| 一、“数据”语义混乱阻碍学术对话场域统一 | 五、数据客体的构成要件 |
| 二、“数据”相近、相似概念的辨析及厘清 | 六、数据客体的分类 |
| 三、数据法律属性的确定 | 七、结语 |
| 四、数据法律概念的界定 | |

摘要 当前“数据”相关概念的语义混淆极大增加了学术对话成本。对此,应首先从横向专业领域区分与本源词“数据”“信息”“数字”相关的派生词,并从纵向区分三个本源词的关系,即信息是被记录的有待传播的内容,数字是随着信息技术发展带来的一种新的记录信息的手段,数据是信息记录的结果,其中数据元素是组成一份数据的最小单元。进一步地,从事实层面、价值层面、法律层面辨析数据与信息的关系,明确信息权益保护原始信息、数据权益保护增值信息。基于数据有用性、够用性、可用性和好用性的经济属性,数据的法律定义是“以数字化形式存在、能够产生增值信息以优化目标决策的数据元素集合及其处理结果”,并据此确定数据客体的构成要件包括:数据元素之间具有相关性(质量要件)、数据元素集合具有规模性(数量要件)、数据具有现实的可利用性(形式要件)、数据具有用以优化目标决策的目的性(实质要件)。以数据行为环节“信息价值增值程度”作为分类标准,将数据客体分为原始数据、衍生数据、数据产品三类。

关键词 数据产权 数据元素 权利客体 衍生数据 数据产品

一、“数据”语义混乱阻碍学术对话场域统一

(一)“数据”现有语义的释义概况

1. 国内外权威词典的释义

各大词典所收录的词语释义通常是已被广泛接受的规范性解释,但从实际考察来看,国内

* 许春明,同济大学上海国际知识产权学院教授,管理学博士;杨欢欢,同济大学上海国际知识产权学院博士研究生。本文系国家社会科学基金重大项目“新形势下我国参与知识产权全球治理的战略研究”(项目编号:21&ZD165)的阶段性成果。

外权威词典中对“数据”的解释也不尽相同。《汉语词典》中将“数据”解释为“电子计算机加工处理的对象”。〔1〕《辞海》(第7版)将“数据”定义为“描述事物的数字、字符、图形、声音等的表示形式”。〔2〕《牛津词典》将“数据”定义为“事实或信息,尤指经检验并用于发现事物或做决定的事实或信息”或“计算机存储的信息”。〔3〕《柯林斯词典》将“数据”解释为“可以将信息称为数据,尤其是当它以事实或统计数据的形式出现时”或为“可以被计算机程序存储和使用的信息”。〔4〕

2. 国内外有关数据法律、法规、政策文本的释义

国内外有关数据、数据要素的公开法律、法规、政策文本均有涉及“数据”的释义。2021年《数据安全法》第3条规定:“本法所称数据,是指任何以电子或者其他方式对信息的记录。”我国原技术监督局发布的标准将数据定义为:“数据是指信息的可再解释的形式化表示,以适用于通信、解释或处理。”〔5〕中国信息通信研究院《数据要素白皮书》指出,大数据时代,数据是基于二进制编码的、按预先设置的规则汇聚的现象记录。〔6〕2022年中国移动通信研究院《开启数据要素流通市场3.0时代白皮书》中将数据定义为“任何以电子或其他方式记录或识别的客观事物的符号”。〔7〕2019年美国《开放的、公开的、电子化的及必要的政府数据法案》(The Open, Public, Electronic, and Necessary Government Data Act)指出,数据为以任何形式或介质记录下来的信息,开放政府数据时特别指明数据需要满足机器可读的条件。

3. 不同使用语境下的“数据”释义

在传统语境下,数据是指数值,如海拔1000米、温度20℃等。在IT领域,数据概念扩大,数据是指网络空间里的所有东西,是网络空间的唯一存在,即电子数据,不仅包括数据资产“2023/06/11”等符号、字符、日期形式的数据,还包括文本、声音、图形、图像和视频等类型的数据,而且政府文件、住宿记录、网上购物记录、银行消费记录等也是数据。〔8〕在科学语境下,数据是指记录下来的事实,即以数字或文字等形式呈现的客观实体属性的值。〔9〕

综上所述,考察国内外权威词典、数据法规政策文本及不同使用语境下的“数据”释义,其概念层次和定义维度大不相同。“数据”概念在国内外尚未形成各领域和各学科的通用性定义。

(二) “数据”语义统一是数据确权研究的重要前提

当前学界对数据确权的研究百家争鸣,主要分为三方面内容:一是围绕数据赋权必要性问题,〔10〕

〔1〕 参见在线《汉语词典》, <http://cidian.aies.cn/MTIwNjY5.htm>, 2024年4月8日访问。

〔2〕 参见在线《辞海》, <https://www.cihai.com.cn/detail?q=%E6%95%B0%E6%8D%AE&docId=7583002&docLibId=72>, 2024年4月8日访问。

〔3〕 参见在线 Oxford Learner's Dictionaries, <https://www.oxfordlearnersdictionaries.com/definition/english/data?q=data>, 2024年4月8日访问。

〔4〕 参见在线 Collins Dictionary, <https://www.collinsdictionary.com/dictionary/english/data>, 2024年4月8日访问。

〔5〕 原国家技术监督局发布的国家标准《信息技术 词汇 第1部分:基本术语》(GB/T 5271.1—2000),第156页。

〔6〕 参见中国信息通信研究院:《数据要素白皮书(2022)》,2023年1月,第1—2页。

〔7〕 中国移动通信研究院:《开启数据要素流通市场3.0时代白皮书》,2022年,第6页。

〔8〕 参见叶雅珍、朱扬勇:《数据资产》,人民邮电出版社2022年版,第3页。

〔9〕 参见迟丽华等:《信息处理技术员教程》,清华大学出版社2005年版,第1页。

〔10〕 参见付新华:《企业数据财产权保护论批判——从数据财产权到数据使用权》,载《东方法学》2022年第2期,第133页。

探讨数据赋权的利弊；^{〔11〕}二是围绕数据保护路径问题，^{〔12〕}具体讨论数据产权的权利属性；^{〔13〕}三是数据产权的制度构建问题，具体探讨数据产权的主体、客体、内容等制度安排。^{〔14〕}

但不可忽视的问题是，在法学领域，“数据”一词的语义并未形成统一规范。在不同知识领域下，数据的含义并不相同，数据概念的不统一导致了观点的复杂性和冲突性。具体来说数据概念的混淆易产生以下三个问题。

第一，数据概念内涵的不统一，增加了学术对话成本。数据法律概念的统一是讨论数据确权、数据客体分类、数据许可与使用等内容的大前提，基于数据不同内涵而提出的学术观点难于对话讨论，“各自为营”的论证只会令数据确权保护的讨论更加无序。如，对数据形式的理解，有学者从技术领域出发，将数据理解为以0和1组成的比特形式，提出其所研究的“数据”限于计算机及网络上流通的在二进制基础上以0和1的组合而表现出来的比特形式。^{〔15〕}但也有学者认为除了以电子形式记录的数据，其他方式记录的也是数据。^{〔16〕}有学者将数据和信息相提并论，认为数据的本质就是信息。^{〔17〕}还有学者从数据形态角度提出，数据可能是数据产品，具体表现为数据库、某个平台等。^{〔18〕}由此可见，学者对同一数据概念存在不同观点，可能只是基于不同数据概念层次上产生的分歧，增加了后续讨论的沟通成本和对话成本。

第二，不同领域的数据内涵交错使用，模糊了数据的法律概念。有的学者在定义数据权利客体时，往往将数据资源、数据资产、数据要素等概念与数据等同或包含其中，^{〔19〕}这些与数据相近、相似概念所指向的对象在本质上不存在区分，仅是由于不同学科领域下术语表达形式不同及内容侧重点不同而已。有的学者将不同领域的数据概念替换使用，这会对法律领域下形成统一数据概念造成困扰和疑惑，导致难以认清数据产权的保护对象。如，在计算机领域下使用数据，则将其解释为能够被计算机输入、计算、输出的符号介质的总称，由具有一定意义的数字、字母、符号及模拟量等要素构成。在经济学领域下使用数据资产，将其定义为由实体控制的，预期能够给实体带来未来经济利益的，脱离了实物形态可复用的无形资产；^{〔20〕}或使用数据生产要素概念，将其定义为以电子形式存在，通过计算的方式参与到企业生产经营活动中来发挥价值的资源。^{〔21〕}乃至有学者直接将经济概念作为法律的规范性概念，在数据产权语境下界定权利客体时，认为数据产权的对象就是数据资源，即经过一定加工处理，能够用于特定目的、具有财产价值的数据集。^{〔22〕}

〔11〕 参见刘琳：《大数据时代商业数据财产权理论的勃兴与批判》，载《华中科技大学学报（社会科学版）》2022年第2期，第102—103页。

〔12〕 参见郝思洋：《知识产权视角下数据财产的制度选项》，载《知识产权》2019年第9期，第45—60页。

〔13〕 参见戴昕：《数据界权的关系进路》，载《中外法学》2021年第6期，第1561—1580页。

〔14〕 参见刘文杰：《数据产权的法律表达》，载《法学研究》2023年第3期，第36—53页。

〔15〕 参见梅夏英：《数据的法律属性及其民法定位》，载《中国社会科学》2016年第9期，第167页。

〔16〕 参见杨东、白银：《数据“利益束”：数据权益制度新论》，载《武汉大学学报（哲学社会科学版）》2024年第1期，第65页。

〔17〕 参见陈筱贞：《大数据权属的类型化分析——大数据产业的逻辑起点》，载《法制与经济》2016年第3期，第44—45页。

〔18〕 参见王利明：《论数据权益：以“权利束”为视角》，载《政治与法律》2022年第7期，第105页。

〔19〕 见前注〔16〕，杨东、白银文。

〔20〕 参见曾雪云：《数据资产深刻影响企业经济行为》，载《新理财》2022年第8期，第35—36页。

〔21〕 参见徐攀、李杰义：《企业数据资产入表路径：框架与实践》，载《财会月刊》2024年第7期，第58页。

〔22〕 参见于海纯、陈润恺：《数据资源持有人的法律解释》，载《科技与法律（中英文）》2024年第2期，第22页。

但需要注意的是,经济学家关注经济收益,而不如法学家那样严谨辨析数据相关概念,法学家不应直接将经济政策意义上的表达等同于法律对民事权益的规定,并将其作为数据权益构建的研究基础。^{〔23〕}混用不同领域下数据相近、相似概念的做法,由于在统一数据法律内涵时会存在学科侧重点的差异,最终导致无法形成真正意义上的数据法律概念。

第三,数据概念的不当扩大,致使数据权益内容发生不当叠加。在广义上理解数据还是在法律概念下理解数据,决定了数据产权的权益范围。当前不少学者主张数据产权结构性分置或构建权利束,实则是扩大了数据概念,使数据产权之下的权益不当叠加。如有的学者认为数据产权制度中,数据保护的对象不限于数据财产权益,还涉及人格权、人身安全、国家主权、国家安全等利益。^{〔24〕}有的学者基于产权结构性分置理论,将数据产权分置为个人数据财产权及企业数据财产权,其中个人数据财产权又分置为人格权财产性利益和新型财产权,企业数据财产权又分置为传统财产权和新型财产权。^{〔25〕}

虽然《中共中央、国务院关于构建数据基础制度更好发挥数据要素作用的意见》(以下简称《数据二十条》)同样提出了数据产权结构性分置的要求,但不同的是,《数据二十条》所提出的数据产权结构性分置是为满足不同数据生产主体的利益分配和利用者的使用需求。有的学者所提出的数据产权结构性分置的观点则是混淆了多种权益。如此一来,容易造成权利界限模糊和体系性不强的问题,权利客体边界模糊的困境就会变成法官司法适用的困境。^{〔26〕}换言之,数据产权结构的分置应为一个权益的多种利用、分配方式,而非多个不同类型权益的叠加。因此,统一数据概念、合理限定数据的保护边界,有利于区分数据上的多种权益,并厘清法律层面上需要保护的数据权益内容。

综上,数据范围的模糊不清影响后续数据法律属性的确定及数据权属的配置,^{〔27〕}厘清当前数据的相关概念,统一学术对话场域是进行后续研究和立法的重要前提。

二、“数据”相近、相似概念的辨析及厘清

“数据”相关概念混用主要有两个原因:一是在横向上,与本源词“数据”相关的派生词较多,词语之间需要专业区分;二是在纵向上,随着社会的发展及科学技术的进步,本源词“信息”“数字”“数据”密切相关且具有不同的时代内涵,语义识别和区分较为困难。换言之,“信息”“数字”“数据”三个本源词汇内涵差异不大,在具体使用时易相互替代而造成理解的混乱,加之“信息”“数字”“数据”相关的派生词也十分丰富,从而在更大的词汇量内造成用语的混乱。因此,需从纵向、横向分别梳理对比,对“数据”相关概念进行辨析以解决用语的相互替代和语义混乱问题。

(一) 横向区分:“数据”相近、相似概念的类型化

学科角度不同、研究领域不同是派生词语产生的主要影响因素。以此为基础,将“数据”相似概念以本源词和相关学科词缀为类型化标准,界分相关概念如下(详见表1):

〔23〕 参见程啸:《企业数据权益论》,载《中国海商法研究》2024年第1期,第56页。

〔24〕 参见陈兵、林思宇:《数据产权结构性分置的运行逻辑与实践机制——以统筹数据保护和利用为目标》,载《南开学报(哲学社会科学版)》2024年第1期,第39页。

〔25〕 参见熊波:《数据财产权的分置性理论构造与刑法保护》,载《清华法学》2024年第2期,第31页。

〔26〕 见前注〔16〕,杨东、白银文,第66页。

〔27〕 参见姬蕾蕾:《企业数据保护的司法困境与破局之维:类型化确权之路》,载《法学论坛》2022年第3期,第112页。

表 1 数据相似概念类型化表

词缀 本源词	XX 化	XX 技术	XX 资本	XX 资源	XX 资产	XX 经济	XX 产品	XX 产权	XX 主权	XX 治理
信息	信息 化	信息 技术	信息 资本	信息 资源	信息 资产	信息 经济	信息 产品	信息 产权	信息 主权	信息 治理
数字	数字 化	数字 技术	数字 资本	数字 资源	数字 资产	数字 经济	数字 产品	数字 产权	数字 主权	数字 治理
数据	数据 化	数据 技术	数据 资本	数据 资源	数据 资产	数据 经济	数据 产品	数据 产权	数据 主权	数据 治理
区分 标准	技术 用语		经济 用语					法律 用语	政治 用语	

三个本源词及十个词缀共组成了三十个派生词，上述词汇在各种语境及学科领域内均有使用，但因三十个派生词未达成共识性定义，在实际研究中常不加区分地被使用，进而造成词语混用和理解困难。^{〔28〕}由上表可知，“信息”“数字”“数据”所派生的相关概念可通过技术、经济、法律、政治等领域划分加以辨别，从而减少不同领域对同一研究对象的用词争议，真正聚焦到“信息”“数字”“数据”三个本源词的辨析与关系的厘清上。

（二）纵向区分：以信息产生、记录、传播、接收过程为标准

信息是指音信消息或通信系统传输和处理的对象，^{〔29〕}数字是指用来记数的符号，^{〔30〕}数据是指电子计算机加工处理的对象。^{〔31〕}从《汉语词典》释义中，无法简单地将“信息”“数字”“数据”三者区分开来。从三个本源词的产生和发展来看，信息是最早产生并存在的对象，数字和数据是围绕信息而产生的概念。在大数据时代，信息、数字、数据三者的关系日益紧密，这正是研究对象“数据”难以剥离并清晰定义的重要原因。

由于信息的客观特征表现为无体无形，其在产生、记录、传播、接收过程中需要以一定的记录手段被记录并借助载体被保存和传播。因此，要厘清信息、数字和数据三者的关系，需要对信息生产阶段进行梳理，以明确信息、数字和数据各自所指代的内容。

众所周知，从人类社会诞生起，信息就已经存在了，人们最初采用结绳记事的方法来记录信息。随着印刷技术、通信技术的发展，信息的记录、传播、接收过程也不断演变出新的方式。在此过程中，作为信息记录结果的数据一直存在，但并未成为生产力要素。对此，通过梳理信息产生、记录、传播、接收等过程的历史演变脉络，对比信息生产各个环节的内容，可以明确数据在信息生产环节中的定位，并进一步厘清“信息”“数字”“数据”之间的关系。

从信息本体、信息接收主体、信息接收方式的发展历程来看，信息作为本体，在人类社会发展过程中自始至终未发生改变，但伴随着科学技术的发展，信息的接收（读取）主体发生了改变，由人类转变为机器，人类大脑只能处理有限的信息，而机器凭借先进的技术条件可快速高效地处理大量信息。由此信息的接收方式也从人类的视觉、听觉、触觉转变为技术支持的编译器。

〔28〕 从研究领域的大类区分已足以辨析各概念间的关系，因此 30 个派生词的具体定义非本文的界定重点。
〔29〕 参见在线《汉语词典》，<http://cidian.aies.cn/Njg4NjQ=.htm>，2024 年 4 月 8 日访问。
〔30〕 参见在线《汉语词典》，<http://cidian.aies.cn/MTE5OTI0.htm>，2024 年 4 月 8 日访问。
〔31〕 参见在线《汉语词典》，<http://cidian.aies.cn/MTIwNjY5.htm>，2024 年 4 月 8 日访问。

从信息记录手段的发展历程来看,无体无形的信息需要以一定的方式被记录,才能使信息被接收主体以一定的方式识别和读取。在计算机语言出现以前,人类通过符号、文字、图片等视觉感知对象或语言、旋律、有声摩斯密码等听觉感知对象来记录信息;在计算机语言出现以后,人类对信息的记录手段转变为 0101 的数字形式以及波形等其他二进制的表示方式。信息通过不同的记录手段被固定表达,由此形成一条记录结果,该信息的记录结果即为一个数据元素。^[32]成千上万条信息被记录则形成海量的数据元素,众多数据元素按照一定的分类或单位被集合在一起即可组成一份份不同主题内容的数据。^[33]如以户为单位,甲市 A 区某一户的用电信息被记录则形成一个数据元素,A 区每家每户的用电信息被记录并集合则形成一份 A 区的居民用电数据;如以区为单位,则甲市 A 区的所有居民用电信息被记录为一个数据元素,甲市各区居民用电信息被记录并集合形成一份甲市的居民用电数据。由此可见,数据元素是组成数据的基本单元,同时,根据数据元素集合范围的不同,数据元素的大小是可变的,也是相对的。

从信息载体角度来看,因为信息本体是无体无形的,信息的记录手段是无体的,所以记录的信息(即数据元素)要被传播需借助合适的物理载体,信息载体的要求即为有体有形。因为有体物属于民法典中物权法规制和保护的对象,所以信息的物理载体如龟壳、书本、硬盘、磁盘等介质属于物权保护范畴,但需注意的是,信息或数据作为无形财产的价值是无法被物理载体的现实价值所体现和包含的。信息与信息载体、数据与数据载体是不同财产范畴内的问题,应根据信息产生、记录、传播过程进行剥离,合理界分信息、数据与其物理载体,不可混为一谈。

综上,可将信息生产的各个环节及内容表示为:信息(无体无形)被通过不同的记录手段(无体“有形”)形成记录结果“数据(元素)”(无体无形)并借助合适载体(有体有形)被存储、传播。记录手段“无体有形”的“形”^[34]是指记录手段以有一定形状的符号或波形等来表达和描述信息。

(三)“信息”与“数据”关系的进一步阐明

结合信息生产环节的剖析,“信息”“数字”“数据”三者的关系可表述为,信息是被记录有待传播的内容,数字是随着信息技术发展带来的一种新的记录信息的手段,数据是信息记录的结果,其中数据元素是组成一份数据的最小单元。但目前学界对信息与数据的混淆依然存在“信息包含数据说”“数据与信息等同说”“数据包含信息说”三种常见类型。^[35]不少学者将信息与数据等同看待,或并列使用“信息数据”这一概念。^[36]这在一定程度上导致信息权益保护与数据权益保护混为一谈。对此,可从事实层面、价值层面及法律层面分别阐明信息与数据的区分与联系。

[32] 为更好地理解和界定数据,本文引入了组成数据的最小单元即数据元素的概念,以区别于现有研究将数据与数据集合作为个体和总体的做法。从《汉语词典》收录的释义来看,数据元素是指组成数据的基本单位,是某种数据结构中能独立存在的数据单位。其大小随着数据类型的不同而不同。数据元素的概念更贴合大数据时代对一条条信息记录结果的表达,而无须创造新的概念去对其进行解释。

[33] 本文按照是否可以电子化的技术标准,将数据分为电子数据与非电子数据。

[34] 此处的“形”为形状而非物理形体。形状一般是指平面的形,是二维空间;形体是三维空间,是形状和体积的结合,是立体的。

[35] 参见袁曾、刘志鸿:《数据确权的逻辑厘清与差异化制度建构》,载《学术交流》2023 年第 12 期,第 50—51 页。

[36] 参见刘士国:《论客体对数据基本产权和分类确权的决定作用》,载《政法论丛》2024 年第 1 期,第 72 页;申卫星:《论数据产权制度的层级性:“三三制”数据确权法》,载《中国法学》2023 年第 4 期,第 44 页。

1. 数据保护与信息保护是两个不同的法律范畴

虽然数据之上信息权益与数据权益并存,但信息权益和数据权益分属不同法律范畴。信息权益可为现有法律所保护,因而在构建数据产权时,其所保护的数据权益并不涵盖信息权益。^{〔37〕}以信息网络环境为例,对个人信息的数字化记录结果形成“个人数据”,即个人信息的数据;对公共信息的数字化记录结果形成公共数据,即公共信息的数据。个人信息与个人数据、公共信息与公共数据为两个不同的保护对象,应分属不同的法律范畴:个人信息、公共信息等数据权益的内容应从《个人信息保护法》等规范内寻求法律保护和规制;个人数据、公共数据等新客体应建立新的数据产权制度予以保护和规制,同时个人数据、公共数据的保护要受到在先权利的约束,即个人数据、公共数据的收集、利用不能损害他人就其个人信息所享有的合法权利,不能妨碍他人对公共信息的合法收集、利用。结合上述信息与数据的联系与区别,信息保护与数据保护的关系可概括为两点:一是数据保护必然涉及信息保护;二是数据保护不延及信息保护,但受信息保护的限制。

2. 数据权益与信息权益的对象是两个不同层次的价值

在数据保护的现有研究中,学者提出了将数据作为汇编作品的著作权法保护,^{〔38〕}将数据作为方法发明的专利法保护,^{〔39〕}将数据作为商业秘密保护等多种现有法律保护路径,^{〔40〕}但上述对数据保护的探究存在一个共性问题,即混淆了数据上的数据权益与信息权益。^{〔41〕}即便是主张创设数据财产权的学者,也未能将数据权益与其他法律权益区分开来。如有学者主张数据财产权分置性结构,通过人格权财产权利益、新型财产权、传统财产权等分类模式保护数据财产,不仅扩大了数据权益的保护范围,也模糊了真正的数据权益。^{〔42〕}

根据上述信息与数据的关系,可以进一步明晰数据权益与信息权益的价值层次(详见图 1)。在事实层面,如前所述,对一系列信息的记录形成一系列数据元素,数据元素的集合形成一份数据,数据是信息记录的结果。因此,数据与信息不可分离。在价值层面,一份数据既包含直接价值也包含潜在价值。所谓直接价值是指数据本身所具备的价值,即可以直接读取和使用的价值;所谓潜在价值是指一份数据所具有的潜在增值或未实现的价值。直接价值体现在一份数据中每个数据元素记录的原始信息(非增值信息),即直接读取数据元素记录内容即可获得的浅层信息,如性别、年龄等个人信息内容。潜在价值体现在一份数据中隐含的深层信息(即增值信息),需要借助大数据技术通过积累数据元素数量并挖掘数据元素之间的相关性得以产生和体现。一份数据既记录了原始信息,也产生了增值信息;相对应地,数据的价值体现为浅层的直接价值和深层的潜在价值。在法律层面,有价值才有保护必要,对于原始信息的直接价值可以通过现有法律保护,但对于增值信息的潜在价值存在保护不足,因而提出创设数据权益保护。综上,数据权益保护的价值对象是深层增值信息的价值。

〔37〕 见前注〔23〕,于海纯、陈润恺文,第 23 页。

〔38〕 参见徐实:《企业数据保护的知识产权路径及其突破》,载《东方法学》2018 年第 5 期,第 56 页;崔国斌:《大数据有限排他权的基础理论》,载《法学研究》2019 年第 5 期,第 7 页。

〔39〕 参见北京市海淀区法院中关村法庭课题调研组:《大数据知识产权司法保护的调研报告——以大数据的产权、模式和伦理为视角》,载微信公众号“知识产权那点事”,2017 年 7 月 17 日。

〔40〕 见前注〔38〕,崔国斌文,第 5 页。

〔41〕 无论是著作权、专利权、商业秘密权,还是数据权,都可在广义上理解为一种信息产权。但实质上信息内容的深浅承载着不同的商业价值,虽然数据所包含的深层信息是一种广义的信息权益,但为更好表达和区分深层信息和浅层信息权益,将数据的深层信息权益称为数据权益,数据的浅层信息称为(狭义)信息权益。

〔42〕 见前注〔25〕,熊波文,第 37—42 页。

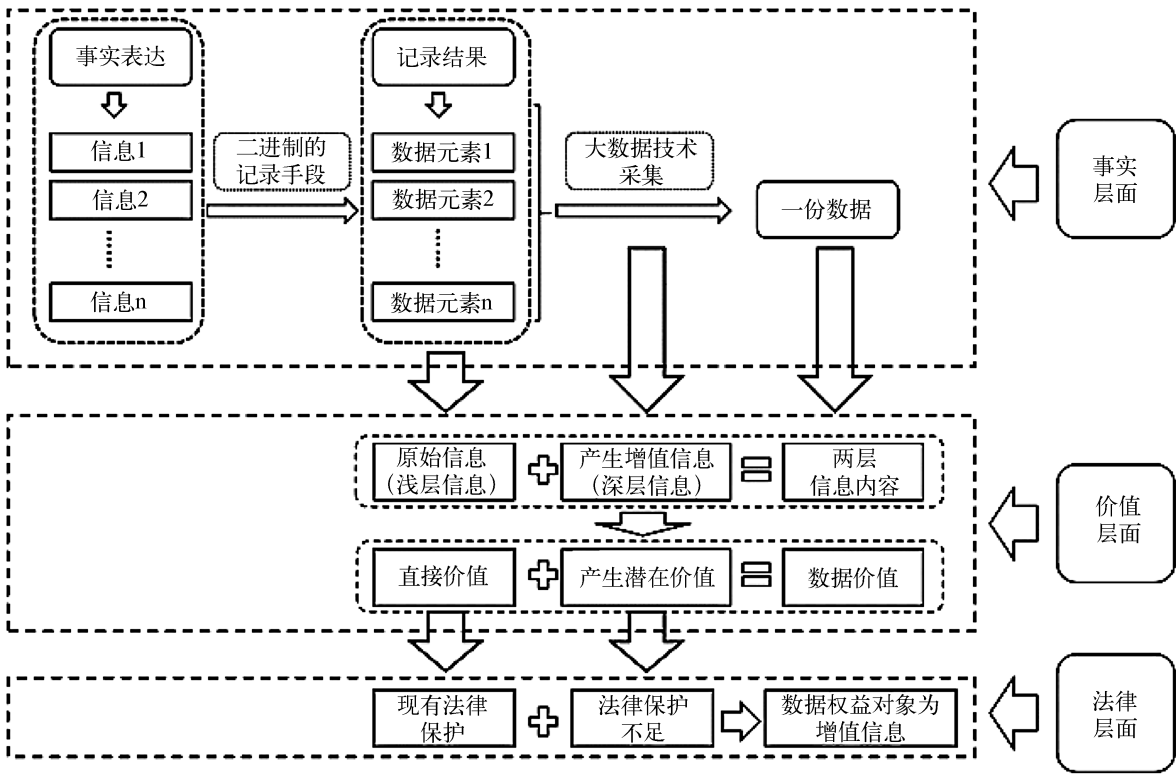


图1 数据与信息的辨析——以信息网络环境为例

所谓的数据权与专利权、著作权、商业秘密权的交叉，实际上为数据中所包含的原始信息内容（浅层信息内容）与专利权、著作权、商业秘密权保护客体的交叉，而非数据权益（即深层增值信息价值的保护）与著作权、专利权、商业秘密权的交叉。

三、数据法律属性的确定

一方面，数据作为一种客观存在，具有其本身的自然属性；另一方面，数据在生产、分配、流转等实践基础上构建的人与人之间的各种关系，使数据同时具备了经济属性。数据的法律属性既无法脱离自然属性而存在，又在一定程度上肩负着服务数字经济属性的制度功能。因此，确定数据的法律属性，应从数据的自然属性出发并结合其经济属性考量。合理地将数据的经济属性转化为相应的法律属性，并据此界定数据内涵，从而为数据经济的发展提供适配的制度支持和法律保障。

（一）数据的自然属性

数据的自然属性具体可以分解为四个方面。一是数据的物理属性，即数据以二进制的形式占据存储介质的物理空间，数据的物理属性使其可以被复制、传输等；二是数据的存在属性，即数据以被人类所感知的形式存在，数据通过 I/O 设备呈现进而被人类认识和理解，如果无法被人类感知也就无法确定数据的存在与否以及以何种内容存在；三是数据的信息属性，即数据通过解释后会有具体的含义（含义即信息），随意输入的字符如“34isnojsiDIn adde & 382shim”同样是以二进制形式存在于存储介质中并通过 I/O 设备被人类所感知，但其无法被解读出任何含义；四是数据的时间属性，即数据不会老化，只要更新数据载体，数据就能一直存在，在不同时刻数据都能保持一

致,数据的更新需求会使数据过时但不会使数据本身老化。^{〔43〕}

(二) 数据的经济属性

前文所述数字经济、数据资产、数据要素、数据资本等无一不是数字经济属性所派生的相关概念。数据的经济属性指数据的价值及使用价值属性,具体表现为数据是否有用、数据是否够用、数据是否可用、数据是否好用等四个方面,即有用性、够用性、可用性和好用性。

数据的有用性,即是否有用,表现为三个方面。一是数据本身具有价值,即数据描述了现实世界的事物;二是数据能满足用户的价值需求,即数据自身的价值能满足用户的某种应用需求;三是数据价值能实现,如果数据有价值且能满足数据的价值需求,但受限于技术水平无法处理数据或者所需数据受到法律法规的保护无法被收集,则数据价值无法实现,数据仍然是无用的。^{〔44〕}

数据的够用性,即是否够用,可分为三种情形。一是数据不够用,即数据不能满足用户某一应用需求,无法达到预设期望;二是数据够用,即数据能满足用户需求且达到预设期望;三是数据超够用,即数据超出了决策问题所需要的数据体量,带来数据资源的浪费。因此,对同一数据,在用户不同的应用需求和预设期望下,有时够用,有时不够用,即数据是否够用是相对的。若有用的数据不够用,则会影响数据的价值实现。^{〔45〕}

数据的可用性,即是否可用,包括两项内容。一是数据是否允许被使用,若数据侵犯隐私权、关涉国家安全则数据不被允许使用;二是数据是否能被访问,若数据无法被访问、被机器读取或不可编程等,则数据不可用。^{〔46〕}

数据的好用性,即是否好用,包括两项内容。一是数据是否质量高,若数据质量不高,则无法通过数据获得满意的处理结果或数据结论;二是数据的使用效果,通过对数据的处理分析及数据结论的应用是否能够满足某个用户的实际决策需求。^{〔47〕}

(三) 数据的法律属性

数据成为数据权利的保护客体,其应具有一定的可保护利益。从数据经济的角度看,数据产品能够为使用者提供统计、预测等服务,具有交换价值和较高的市场价值。^{〔48〕}数据的可保护利益对应到数字经济属性,即体现为数据的价值和使用价值。数字经济属性中的有用性、够用性对应数据的价值,数字经济属性中的可用性、好用性对应数据的使用价值。从国家政策导向出发,数据确权要有利于数据要素的流转和交易,服务于数据市场的发展,与数字经济密切相关。因此,数据法律属性为数字经济属性背书,故其应体现出数据的经济属性。

第一,对应“有用”的经济属性,数据应具有可被大数据技术所挖掘的深层信息。如电商平台上的消费者数据,其浅层信息内容是对消费者实际购买情况客观的现实描述,但借助大数据技术可通过收集消费者多次购买信息的记录分析挖掘得到消费者的消费偏好、消费水平等深层信息内容,而该数据分析结论可用于对消费者进行定制化推送等服务优化。

第二,对应“够用”的经济属性,数据应具有数量规模,即数据不是单一的数据元素,而应

〔43〕 见前注〔8〕,叶雅珍、朱扬勇书,第 4—5 页。

〔44〕 参见王霖:《企业数据法律保护的路径探索——评〈企业数据法律保护和规制研究〉》,载《价格理论与实践》2023 年第 6 期,第 98—99 页。

〔45〕 同上注,第 99 页。

〔46〕 同上注,第 100 页。

〔47〕 同上注。

〔48〕 参见李晓珊:《数据产品的界定和法律保护》,载《法学论坛》2022 年第 3 期,第 125 页。

是包含较大数量规模的数据元素的集合。数据成为第五大生产要素的政策背景是大数据处理技术的飞速发展,大数据技术是挖掘数据价值、发挥数据价值的重要手段,其能够应对来源多样性、一致性、复杂性、生成快速等超过传统数据库处理能力的巨量数据。^{〔49〕}因此,数据的法律属性应体现出其所集成数据元素的规模性。

第三,对应“可用”的经济属性,数据应具有可读取性且能够被允许访问。一方面数据应以可被计算机读取的二进制形式存在,如此才能使用大数据处理技术对其价值进行挖掘,提高其处理和分析效率。而模拟数据、非电子数据等不能被计算机识别的数据,其潜在价值无法通过大数据技术挖掘与实现,即不符合成为数据权利客体的条件。另一方面,数据读取应遵守《个人信息保护法》《数据安全法》的要求,对关乎国家安全的数据、个人隐私的数据即便具有可读性,因被法律所禁止而不被允许访问和读取,其仍无法满足数据“可用”的条件。

第四,对应“好用”的经济属性,数据应服务现实需要能产生积极效果,即对数据中深层信息的发掘利用,能帮助解决实际问题或为优化相关决策提供重要依据。如,通过对全国网络平台消费者购买数据的分析,可获知各地区消费者的消费偏好,为企业优化市场布局提供决策依据。

四、数据法律概念的界定

综合上述对数据相近、相似概念的厘清以及数据法律属性的分析,可将数据定义为“数据是指以数字化形式存在、能够产生增值信息以优化目标决策的数据元素集合及其处理结果”。

(一) 以数字化形式存在

“以数字化形式存在”对应着数据的“可用”属性,数据必须能够被计算机所识别利用才能产生法律所保护的数据利益。如果数据以模拟数据或非电子数据形态存在,则其不能被大数据技术所处理和分析。当模拟数据或非电子数据被盗用,利用人对数据的处理和分析方式离不开大数据技术,那么只要利用人将模拟数据和非电子数据转化为数字数据,就进入数据产权保护、救济的范畴内。

(二) 产生增值信息

“产生增值信息”对应数据的“有用”属性。一份数据价值包含直接价值和潜在价值。从数据价值产生的时间节点来看,无论数据元素是否被集成一份数据,数据元素所承载的原始信息价值并未发生改变,数据生产者在原始信息的生产过程中未付出实质性劳动;当数据生产者利用大数据技术集合数据元素后,一份数据才产生潜在价值,即增值信息价值。该增值信息不为每个数据元素所记录的信息内容所包含,而是在原始信息内容之外产生新的信息。具体来说,增值信息体现在彼此孤立的数据元素通过大数据技术集成增值,集成的数据元素之间的相关性、互斥性等关系类型可以通过技术分析挖掘,得到用于优化决策的目标信息,彼此孤立的数据元素通过服务于同一目标决策而产生数据利益。^{〔50〕}需要指出的是,数据的增值信息并非固定内容,其取决于目标决策所需,并依赖大数据技术的使用。

因为从浅层信息到深层信息发生的价值质变离不开数据生产者付出的大量资金、技术和劳

〔49〕 参见李国杰、程学旗:《大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考》,载《中国科学院院刊》2012年第6期,第647—648页。

〔50〕 所谓目标信息是指对数据元素集合检索、收集、分析、挖掘时我们意欲发现的事实、关系等,彼此孤立的数据元素通过服务于同一目标信息而产生数据利益。

动,所以增值信息的价值劳动来源于数据生产者,从而才产生了对数据权益保护的正当性。

(三) 能够用以优化目标决策

“能够用以优化目标决策”对应着数据的“好用”属性,即数据元素集合通过大数据处理获取的目标信息能够用以解决实际问题,为优化目标决策提供依据。精准的数据能够提高决策的效率和质量,良好的数据整合能够发挥 $1+1>2$ 的价值效应。^[51] 经济主体通过数据采集和处理能够从海量数据中不断提取出新的信息和知识,获取关于外部环境和内部组织的真实情况,实现真实信息的互联。通过对信息的解读和利用缓解市场信息不对称问题,从而驱动经济主体精准决策,促进数字经济的高质量发展。^[52] 如果数据元素只是一串无意义的字符或者数据元素关系混乱,既无法得到数据元素之间有效的正相关、负相关关系,也无法得到其他有助于理解和解决实际问题的数据元素关系,那么该数据元素集合就未产生潜在价值或其潜在价值不具有变现可能。因此,该种数据元素集合不仅无法成为生产要素意义上的数据,也无法成为法律意义上数据产权所保护的客体。

(四) 数据元素集合及其处理结果

“数据元素集合及其处理结果”对应着数据的“够用”属性。数据元素是组成数据的基本单位,一个数据元素是一条信息记录,但彼此孤立的数据元素不具备产生增值信息的可能。因此,法律概念的数据应指数据元素的集合,即对作为数据产权客体的数据具有数量规模上的要求。同时因为数据元素集合需要借助大数据处理,在数据开发和数据应用的过程中会衍化出对数据元素集合不同形式的处理结果,该结果正是数据潜在价值的实现形式,因此也应为数据产权保护客体所容纳。

五、数据客体的构成要件

根据数据的定义,可将数据客体的构成要件分为四项:一是质量要件,即要求数据有用;二是数量要件,即要求数据够用;三是形式要件,即要求数据可用;四是实质要件,即要求数据好用。

(一) 质量要件:数据元素之间具有相关性

数据客体的质量要件要求组成数据的各数据元素之间具有某种内在联系,此为数据具有潜在价值的必然要求。

1. 单个数据元素具有能被人类解读的含义

数据是由一定体量的数据元素组成的,如果每个数据元素都是诸如“34isnojsiDIIn”这类无意义字符,或者其他无法为人类所理解和认知的表达,则其不具有成为数据组成元素的特征。无意义的数据元素组合得到的数据无法被大数据技术处理分析以得到人类所需的增值信息,因此要求组成数据的单个数据元素应具有能被人类解读的含义。当然,在组成数据的海量数据元素中,如果仅有少量无意义的数据元素且不会对数据质量造成实质性影响的,则不影响数据满足成为权利客体的质量要件。

[51] 见前注[16],杨东、白银文,第 66 页。

[52] 参见戚聿东、刘欢欢:《数字经济下数据的生产要素属性及其市场化配置机制研究》,载《经济纵横》2020 年第 11 期,第 64 页。

2. 各个数据元素之间具有相关性

当每个数据元素都具有能够被人类所解读的含义时,还需要组成数据的各个数据元素之间存在某种潜在的关联性。对不同领域、不同对象的不同维度进行客观事实描述而形成的单一数据元素组成的数据具有异质性和零散性,各数据元素之间不具有可被大数据技术所分析和发现的潜在规律和信息联系。不同对象在不同领域、不同维度的数据元素集合只是形式上的集合,而非数据元素原始信息内容上的集合。因此基于该数据无法建立原始信息内容上的关联性,也就无法产生有效的增值信息。从优化经营策略、完善销售布局、便利人们生活等角度来看,该类数据质量较低,既不具有产生经济效益的潜在价值,也不具有产生社会效益的潜在价值。

(二) 数量要件:数据元素集合具有规模性

1. 单个数据元素不是数据权益保护对象

根据单个数据元素的原始信息价值,可将单个数据元素分为价值密度低的数据元素和价值密度高的数据元素。对价值密度低的单个数据元素来说,其所记录的信息能够产生的经济效益和社会效益微乎其微。如A在甲购物平台上的一件衣服在购买记录,B在乙商铺的一条支付记录,其所能被解读的信息就是该数据元素本身的含义,对A和B来说,该数据元素记录了他们的消费行为,对商家来说,该数据元素记录了他们的交易行为,此外无法产生更多的经济价值和社会价值。因此,单个彼此孤立的数据元素价值密度极低,不足以成为数据保护的客体。

对于价值密度高的单个数据元素来说,一个数据元素所承载的信息内容本身就具有利用价值。^[53]如一种配方、一篇文章作为一个数据元素时,该数据元素本身就具有财产性利益,且能够在现有知识产权法律制度下得到保护,因此其也不是数据权益所要保护的客体。

2. 数据客体是具有一定数量规模的数据元素集合

如上文所述,一方面价值密度高的单一数据元素原本就能从现有知识产权法律中得到保护,其不是数据权益所要保护的客体;另一方面价值密度低的单一数据元素不值得法律保护,其也不是数据权益所要保护的客体。有学者指出,达到实质规模数量是数据客体要件,要求数据包含海量数据元素数目,具有反映宏观市场现象的能力。^[54]数据权益所要保护的数据对其所集合的数据元素应具有量的要求,通过数据元素量的规模化实现信息增值,即数据权利客体是具有一定数量规模的数据元素集合。对于“一定数量规模”的衡量要看数据元素的集合是否通过各数据元素所含有原始信息的量的积累产生了新的信息内容,即“一定数量规模”要求数据元素量的积累引起由其组成的该份数据所包含信息价值的质变,能够在所有数据元素记载的原始信息之外产生新信息(即增值信息)。

(三) 形式要件:数据具有现实的可利用性

1. 数据能够被现有算力水平处理

一定数量规模的数据元素集合需要在现有算力能够处理的范围内。自计算机诞生起,数字化数据就不断产生,但是大量的数据元素并不能被充分处理和利用,因为算力无法达到。如果数据虽满足质量要件和数量要件,但受限于现有的技术水平而无法被处理,那么该类数据不具有被现实利用的条件,不能成为数据权益所保护的客体。类似地,在专利法领域,曾经有多人就“永动机”的设计方案申请发明专利,但是其违背了能量守恒这一基本的自然规律,不具有成为现实的可能,

^[53] 见前注〔8〕,叶雅珍、朱扬勇书,第8页。

^[54] 参见冯晓青:《大数据时代企业数据的财产权保护与制度构建》,载《当代法学》2022年第6期,第116页。

因此其不能够被授予专利权。^{〔55〕}

2. 数据以数字化形式存在且能为机器读取

数据需要以数字化形式存在才能够被机器读取,进而才能被大数据技术所处理、分析和应用。如果数据以非电子数据或模拟数据的形式存在,那么其不具有可读取性,人类就无法借助大数据手段分析和利用数据。因此,非数字化形式存在的数据不能成为数据权益的客体。

3. 数据的获取和利用需符合法律规定

《个人信息保护法》和《数据安全法》均对数据元素的获取、利用、处理等做出了限制和规范。对于符合数量要件、质量要件和实质要件的数据,若其关涉国家安全和个人隐私,则会被法律所禁止获取、处理或利用,即便其具有巨大的经济价值和社会价值也不能够被允许访问和使用。因此,只有符合法律规定能够被民事主体获取和利用的数据才能够成为数据权益的保护客体。

(四) 实质要件:数据具有用以优化目标决策的目的性

数字化形式、增值信息、数据元素的集合等限定都是为优化目标决策这一目的而服务的。换言之,质量要件、数量要件、形式要件等都是为实质要件而准备的。“能够用以优化目标决策”是数据成为权利客体的核心要件,该要件不要求数据已经完成目标决策的优化,仅要求数据具有潜在的优化目标决策的信息价值,即用于优化目标决策的数据元素之间的隐藏规律和各类相关关系可以是潜在而未明知的。

大数据技术通过挖掘数据元素之间的隐藏规律和各种相关关系可以辅助人们提高决策科学性和优化决策效果。成为数据产权制度下的数据需要具有潜在的优化目标决策的价值,从而使人们有可能将数据处理分析的结论运用到商业布局、社会资源配置等方面以实现数据的经济效益和社会效益。

此外,所谓“目标决策”不是绝对的,既不是唯一目标决策也不是所有目标决策,该目标决策基于不同数据利用主体的需求,决策内容是相对的。

六、数据客体的分类

(一) 数据客体的分类标准

数据客体的分类,按照主体、来源、结构、行业等不同标准存在多种分类。但在构建数据产权制度的背景下,目前的数据分类标准缺少法律内涵。从数据客体特征来考虑,数据是民事主体创造性活动的结果。所有受数据权益保护的数据均应符合最低创造性标准,即经过多环节的技术处理创造出与公共领域分散的数据元素形态截然不同的数据。这种创造性行为包括改变数据元素的存在状态(汇集)、改变数据元素的存在数量(筛选)、改变数据元素的存在秩序(分类)、改变数据元素的存在价值(分析)等。上述不同的数据行为对数据的价值增益有所区别。据此,结合上文界定的数据法律概念,可按数据行为各环节的“信息价值增值程度”作为数据客体的分类标准。

(二) 以数据行为环节确定数据客体类型

数据生产可分为数据元素产生、数据采集、数据开发、数据应用四个数据行为环节。为了明确各数据行为环节的“信息价值增值程度”不同,将对应行为主体在各环节的数据行为、生产投入、增值贡献及相应产物进行梳理和对比(详见表 2)。

〔55〕 参见王迁:《知识产权法教程》(第 6 版),中国人民大学出版社 2019 年版,第 266 页。

表 2 数据行为各环节情况对比表

数据生产环节	数据生产内容	行为主体	数据行为	生产投入	信息增值贡献	数据产物
数据元素产生	一条个体信息以数字化形式记录,形成一个数据元素	自然人、法人、非法人组织等个体	上网等	网络等各类活动	使原始个体信息得以记录	数据元素
数据采集	将零散、游离的数据元素汇集成数据元素集合,即一份数据	数据采集者	汇集性处理	技术、人力	原始信息→产生潜在增值信息(产生隐含价值-隐含信息)	原始数据
数据开发	对数据进行筛选,分析,挖掘	数据开发者	分析性处理	技术、人力、智力劳动	潜在增值信息→目标信息(发现价值-提取信息)	(狭义)衍生数据
数据应用	对数据开发结果进行创造性应用	数据应用者	应用型处理	技术、人力、智力劳动	目标信息→辅助优化目标决策(应用价值-应用信息)	数据产品(广义衍生数据)

1. 数据元素产生环节：数据元素

在数据元素产生环节,数据生产的主要内容是用户将个体信息以数字化形式记录形成一个数据元素,用户通过各种数字活动产生大量数据元素,该劳动投入的贡献是使一个数据元素记录了原始的个体信息内容。^[56]数据元素是数据的自然状态,零散、无序,游离于数字空间之中。一方面,该类数据^[57]产物因数据利用价值极低而无法纳入数据权益的保护范围;另一方面,数据元素记录的原始个体信息内容本身可以依据相应法律受到保护,如,涉及作品的信息内容可以依据《著作权法》对其加以保护,涉及商业秘密的信息内容可以从《反不正当竞争法》中寻求保护。

2. 数据采集环节：原始数据

在数据采集环节,数据采集者将零散的、游离的海量数据元素汇集成数据元素集合,即形成一份数据。数据采集者对数据元素的汇集性处理行为投入了大量的技术成本,该劳动投入的贡献是使一份数据基于数据元素所承载的海量原始信息可以产生潜在增值信息。数据采集环节的数据产物是原始数据,原始数据是可以获得数据权益保护的初始客体类型,其保护的财产利益为汇集性处理数据产生的潜在增值信息,即数据具有随时可以提取目标信息用于优化决策的潜在价值。

3. 数据开发环节：衍生数据^[58]

在数据开发环节,数据开发者根据决策需要对数据进行数量筛选、内容筛选以及内容类型化等处理,数据开发者借助大数据技术对数据进行分析性处理形成服务于决策需要的衍生数据(如指数型、统计型、预测型等数据),该劳动投入的贡献是开发原始数据中的潜在价值,从原始数据中提取出服务决策所需的具体目标信息。对衍生数据保护的财产利益为衍生数据中已经发现和释

[56] 个体信息包括个人信息或企业信息等个体用户信息。

[57] 此处的“数据”是指由一个数据元素组成的数据。根据数学的集合概念,集合中最少 0 个元素(即空集),最多为无穷个元素。因此,数据作为数据元素的集合,其可以仅包含一个数据元素。根据数据的定义理解,即由一个数据元素组成的数据。本文若无特别指出,数据均指非空、非单个数据元素的集合。

[58] 衍生数据包括狭义衍生数据和广义衍生数据,本文若无特指,衍生数据指狭义衍生数据。

放的信息价值。

4. 数据应用环节：数据产品

在数据的应用环节,数据应用者将数据与相应算法结合,形成面向不特定多数用户稳定调取数据分析结果的应用型数据产品,如淘宝公司的数据产品“生意参谋”。^{〔59〕}数据产品本质上也是一种广义的衍生数据,其系通过对原始数据或狭义衍生数据进一步技术加工处理,以应用软件或应用平台等形式对外向不特定用户提供服务的一种技术形态。其应用型的处理行为给数据产品使用者提供了简单易操作的数据分析服务,使用户能够根据决策需要选择数据分析范围、数据分析对象等筛选条件,从而获得决策所需的目标信息。该数据产品保护的财产利益为其能够给数据产品主体带来商业利益的普适性的市场应用价值。

狭义衍生数据与数据产品“信息价值增值程度”的不同体现在两个方面:

第一,服务对象范围不同。狭义衍生数据的市场服务面较窄、服务对象具有特定性,数据利用者加工原始数据产生的一份衍生数据往往以自用为主或供某几个主体使用。数据产品的市场服务面较宽,数据利用者为不特定多数主体,一份衍生数据可以反复被提供给不同的调用主体,数据服务具有普适性、一般性。

第二,数据利用者投入劳动不同,获得的权利内容不同。狭义衍生数据利用者往往需要投入技术成本等以获得衍生数据,或者通过委托他人进行技术分析以获得衍生数据,其对数据的信息增值做出了贡献,可以享有衍生数据的权利。而数据产品的利用者,仅通过数据产品主体已设置好的程序和操作进行数据分析结果的调取,其仅享有阅读、下载的权利,并未对数据信息价值增值做出实质性贡献,因此其不能享有对数据产品的权利。

综上,数据是通过一定的技术手段对数据元素进行汇集,使其从零散到集中而形成的数据元素集合(即原始数据),或通过一定的技术手段对汇集的数据元素集合进行筛选、分析、挖掘而形成的数据处理结果(即狭义的衍生数据),或将数据元素集合与相应算法结合形成稳定的不特定主体皆可用的数据产品(即广义的衍生数据)。其中,数据元素不是数据权益的保护客体,原始数据、衍生数据、数据产品是数据客体的三种类型。

根据上述分析,可将原始数据定义为:以数字化形式存在能够随时提取潜在信息用于优化目标决策的数据元素集合;衍生数据定义为:以数字化形式存在能够释放信息价值用于优化目标决策的原始数据的处理结果;数据产品定义为:以数字化形式存在能够重复服务于不特定主体根据设定操作调用优化目标决策所需信息的原始数据或衍生数据的处理结果。

七、结 语

随着数字技术的发展,数字化水平和能力不断提升,数字经济已成为推动中国经济增长的主要引擎之一。^{〔60〕}经过采集、处理、分析、消费、再处理,数据源源不断地产生,驱动着企业的经营决策,成为新的交易内容,为数据持有者提供巨大商业优势。在数据各生产环节中涉及众多数据生产者,由于数据权属划界不甚明晰,无法避免各主体之间产生使用、收益、处分等各项权能的冲突,数据产权制度的建立迫在眉睫。基于数据确权研究现状,本文从法律语境下界定数据概念,并进

〔59〕 参见杭州铁路运输法院民事判决书,(2017)浙 8601 民初 4034 号。

〔60〕 参见《白皮书:数字经济已成为推动中国经济增长主引擎之一》,载中华人民共和国中央人民政府网站 2022 年 11 月 7 日, http://www.gov.cn/xinwen/2022-11/07/content_5725115.htm。

一步探析数据客体条件,为数据产权具体制度设计提供逻辑起点。为建立完善的数据产权制度,未来研究还需要在权利主体、权利内容、权利限制、权利利用等方面进一步地开展,以构建体系化的数据产权制度,助力数据经济发展。

Abstract The current semantic confusion of the concept of “data” greatly increases the cost of academic dialogue. Therefore, it is necessary to first distinguish the derived words related to the basic terms “data”, “information”, and “digits” across professional fields, and then vertically distinguish the relationship of the three basic terms. Information is the content that is recorded and waiting to be disseminated, digits are a new means of recording information brought about by the development of information technology, and data is the result of recording information, with data elements being the smallest unit composing a set of data. Furthermore, through analyzing the relationship between data and information at the factual level, the value level, and the legal level, it is important to clarify the protection of original information for information rights and the protection of value-added information for data rights. Based on the economic attributes of data, such as usefulness, sufficiency, availability, and usability, the legal definition of data is “a collection of data elements and their processing results that exist in digital form, capable of generating value-added information to optimize target decisions”. Based on this, the constitutive elements of data objects include: the correlation between data elements (quality requirement), the scale of the data element collection (quantity requirement), the practical usability of the data (formal requirement), and the purposefulness of the data to optimize target decisions (substantive requirement). Using the “degree of value-added information” in the data behavior process as a classification standard, data objects are classified into three categories: raw data, derived data, and data products.

Keywords Data Property Rights, Data Element, Object of Right, Derived Data, Data Product

(责任编辑:曹博)