

大数据确权问题研究

徐 莉

桂林电子科技大学 广西桂林 541004

摘 要: 目前大数据商业化的利用已成趋势。大数据商业化利用的前提就是大数据确权, 本文在明晰大数据和数据确权的相关概念之下, 立足于基本法理理论与平衡商业发展和用户利益的现实需求, 明确了大数据的财产性属性, 厘清了大数据交易链条上的各类主体, 从而确定各主体间的权利义务关系以及各个主体享有的具体权能, 至此在大数据确权的理论体系基础上, 提出现实意义的大数据确权方式: 首先, 在现有的“提交权属证明+专家评审”模式下, 提出建立个人数据银行进行个人数据的管理和交易, 其次提出利用区块链的去中心化、唯一确权性、不可篡改等内在优势解决大数据确权的技术方案, 并期待在此基础上进一步解决大数据确权理论与实际操作中遇到的问题。

关键词: 大数据; 数据确权; 所有权; 区块链

国际数据公司发布的 2022 年指南显示, 在 2021 年全球对于大数据技术总投资为 2176.1 亿美元。国务院在 2016 年 12 月发布了《国家信息化“十三五”规划纲要》中明确指出, 推进大数据产业发展, 要优先开展数据资源共享开放行动。但在推动数据资源共享开放的过程中, 各主体之间抢夺数据、大数据侵权、大数据纠纷等问题频发。发展大数据产业, 之所以要优先解决大数据归属问题, 是由于大数据确权问题是大数据产业发展的基础和前提。数据权利的主体有哪些? 数据权利的内容包含哪些权能? 这些最基本的问题不解决, 不但会造成大数据开发利用中不断产生纠纷, 权责归属模糊, 难以保障数据安全和个人隐私, 相反会反向制约大数据共享开发, 与“规划”的目标背道而驰。

1. 大数据的概念

最早提出大数据概念的是 Victor Male Schemberg 和 Kennis, Kookyer 在共同编写的《大数据时代》一书中这样描述的: 大数据是通过对“所有数据”的分析处理而得到的。这里的“所有的数据”是指人类所能获得所有数据的总和。著名大数据研究机构 Gartner 从大数据的属性出发, 给大数据下了如此定义: 大数据是指经过新的处理而具备数量多、增长率高、数据资产多样化等特性, 也就是所谓的 3V 特征。在此基础上, IBM 公司对大数据的特征做了更进一步的拓展, 给出了 4V 特征: 高容量、高增长率、数据种类多样化、数据价值化。也有部分学者认为在法律的世界中, 对于什么是大数据必须明确加以确定, 进而提出了更为准确的定义, 例如有学者认为: 大数据是大量不特定的主体按照确定的目的进行挖掘和加工的。这个概念包含了行为要件、行为对象来、对象性质等内容, 在此法律视角的定义下: 有关“大数据”和“个人隐私”、“个人信息”、“商业秘密”、“(一般) 数据”等各范畴均已界定与区分, 这也是明确“大数据确权”的前提。

2. 数据确权的内容

2.1 大数据的法律属性

确定大数据的法律性质是数据确权的内容之一, 也是数据确权的前提条件。依据大数据产生和利用方式的不同, 可以将大数据划分为个人数据和整体数据。其中, 因个人(如自然人、法人和其他组织)在互联网上的行为而产生的数据, 属于应归于个人信息的个人数据(personal data)。而整体数据是由个体数据集合而成, 通常蕴含着巨大的商业价值, 因此这才是真正意义上大数据。所以这里更多的讨论的是整体数据的法律性质。

2.1.1 汇编作品

关于汇编作品, 我国《著作权法》第 14 条明确了这一点: “汇编若干作品、作品的片段或者不构成作品的数据或者其他材料, 对其内容的选择或者编排体现独创性的作品, 为汇编作品”。因此有

些学者主张: 大数据的法律属性可以视作汇编作品。但解析相关法条发现, 汇编作品要求一具有独创性, 然而大多数的大数据只是依据一些技术手段将信息进行简单集合, 没有体现出著作权法中所谓的“独创性”, 因此如果依据著作权法第 14 条对大数据进行保护, 大部分的大数据作品并不在其保护范围中, 这样一来, 这部分大数据便无法得到保护, 这恰巧与我们数据确权的目的相违背。

2.1.2 信息资产

我们在上文讨论大数据的概念时提到, 大数据在特定语境下是一种信息资产, 利用大数据实际上也就是利用这种信息的价值, 因此, 有学者提出, 信息资产并不属于传统财产权客体, 如果将大数据定性为信息资产, 现有的法律制度无法给予保护, 因此需要增加信息资产这一新的财产权客体, 并建立信息资产财产制度。

从立法的角度上来说, 每出现一个新的事物就需要建立一种新的制度, 那我们的法律将会变得非常庞杂且不具系统性, 因此当务之急是在现有的法律制度框架下进行解析和保护, 而不是理解寻求新的法律制度来进行专门的保护, 这无非就是“饮鸩止渴”式的自救模式。而新制度的建立虽然十分有效, 但也非一日之功, 其他的保护方式还需要继续探索, 直至新的法律制度建立。

2.1.3 民法上的集合物

基于上述讨论, 关于大数据的法律属性, 既不可以简单的定性为汇编作品, 也不能一味寄希望于建立新的信息资产制度来进行保护, 因此需要基于目前民法理论来寻求保护途径。那么大数据具体数据属于哪一种财物呢? 首先数据属于一般财物, 而并不是公共财物, 因为大部分财物仍为特定人享有, 大数据即使具备持续复制的特性, 也无法算作公共财物。其次, 大数据在民法上应该属于一个集合。集合体不是简单的几个独立的物体相加, 而是由于它的“集合”(set)而产生新的特性, 从而成为一种新的物。大数据的本质虽然从微观上讲是由无数个个体数据组成的整体数据, 但因为它的“集合”特性, 使得这一总体数据具备了构成新事物的独立价值, 也就是作为“集合物”(complex)。

2.2 大数据的权利主体

在大数据应用日益丰富的大环境下, 随着技术的不断发展, 对大数据的权利主体进行认定也变得更加困难。究其原因, 还在于大数据的不断应用, 使得大数据的流通链条变得更长也更复杂。从最初数据自己采集自己使用, 到现在数据的所有者、采集者、控制者、使用者等主体的不断分离。而在这个长长的数据链条中, 更多的主体付出了劳动、贡献了技术和智慧, 使得大数据的价值也不断的释放。而最初的数据的创造者作用被弱化。因此很多挖掘数据的主体权利变得更加重要。因此, 确定大数据权利主体的核心是明确数据的创造者、控制者、处理者和使用人之间的权利关系。目前主要有

四个认定标准对数据权属进行认定：流汗原则、创造者归属原则、最佳利用者原则、效益原则。在数据全生命周期内，不能机械地采用某一原理对数据进行权属认定，而应着重于对数据的“所有权”、“使用权”和“收益权”进行平衡划分。有序地推动大数据交易和流通，注重平衡大数据产业发展和数据权利人利益，应当综合考虑四项标准。

3. 大数据确权方式

经过上述大量篇幅，已经明确了大数据确权的相关理论知识，那在此理论的基础上，如何在实践中实现大数据确权，并不是那么容易，大数据交易链条不断变长，涉及的主体也增多，这使得在实践中确认具体的权利人和侵权人变得更加困难。

3.1 传统确权模式

“提交权属证明+专家评审”模式是目前贵阳大数据交易所提出的传统确权方式。首先是大数据控制人将大数据权属证明提交交易所；其次是由大数据交易所组织专家，根据请求对数据进行评估；最终评审结果将由大数据交易所所在专家评审结束后公布。此种模式看似流程清晰，但却存在一定的问题：一是专家评审受主管影响，不够公平；二是提交资料无法永久保存，不利于后续审计。三是全程采取纸质或者电子方式，存在被篡改的可能。因此从技术角度探讨出一套可靠的、可操作的大数据确权方法，正是解决这些现实弊端的当务之急。

3.2 创新性确权模式

3.2.1 个人数据银行

目前也存在一些个人数据保存和管理的项目和平台。例如，由美国国会图书馆发起的“国家数字信息基础架构保存计划”其中的个人数字归档可以进行个人数据保存管理。另外还有日本信息通信技术企业富士通（Fujitsu）也建立了 Dataplaza 数据交易市场，成为各主体的数据交易活动平台；但这些平台管理存在着数据产权模糊、数据管理散乱、数据难以开放和流通等问题，造成其价值难以发挥。因此学者提出建立个人数据银行，大数据是一种财产，那么个人数据就是个人财产，就和我们放在各个银行的存款一样。从这点上来看，个人数据也可以采用目前银行的资产管理模式来进行管理和保存。这样不但可以实现个人数据的保存，还可以解决个人数据的流通困难等问题，给个人数据的权利人带来收益。个人数据银行就是以银行个人货币资产管理的模式及架构为基础建立的，以保护保护个人数据产权、知情权、隐私权和收益权为核心的系统。其中个人数据银行主要包括数据确权、汇聚、管理、交易和增值等服务功能，目前也有部分学者对此进行了技术上的论证，甚至研发了一部分产品。这说明这个方式从理论和技术上都具有现实意义。

3.2.2 基于区块链的大数据确权方案

个人数据银行解决方案，更多的是从管理的角度来解决数据确权的问题，并没有真正解决数据确权的源头问题。因此考虑到大数据“量大且多”的特征，如何准确高效的完成大数据确权，是一个很大的挑战。有学者针对这一挑战提出了可以结合热门的区块链技术来加以解决。

区块链是以比特币为代表的数字加密货币体系的核心支撑技术，它的最大优势就是我们常提到的“去中心化”，具体是指利用数据加密、分布式共识和经济激励等技术手段在不需要相互信任的分布式系统中实现点对点交易。区块链技术的出现解决了高成本、低效率以及数据存储不安全等传统中心化机构中存在的问题。伴随着比特币的迅猛发展，区块链技术的研究和应用逐渐热起来，甚至被认为是具有里程碑意义的第五次创新，仅次于大型主机，PC，Internet，移动社交网络。

大数据交易运用区块链技术之后，具体是在大数据交易环节，采用 B L S 短签名方案，这样每一个市场主体都能拥有唯一的“ID”。

基于区块链的唯一确权和不可更改性，因大数据产生的各类纠纷中，溯源难题便迎刃而解。在这个环节中还可以引入审计中心、水印中心。基于区块链的本质特征：唯一确权、不可篡改、可追溯，大数据在注册上链和查询阶段可以实现对确权结果和相关证据的链上保存，确保确权结果的完整性和不可篡改性。这样就解决了原来“提交权属证明+专家评审”模式的主观性问题，弥补了个人数据银行的数据确权的源头性问题。基于区块链技术的大数据确权是目前最值得期待的解决方案，它完全可以满足公平、完整且没有欺骗性的大数据确权要求。甚至在此基础上，可以进一步讨论大数据交易生命周期中的可追责性。但由于目前区块链研究仍处于起步阶段，未来是否可以实现区块链与大数据确权，以及如何更好的将双方进行融合，仍待进一步探索。

4. 结论

大数据确权制度的确立，让要求“后手权利不能凌驾于前手权利之上”在法律上有据可依。换言之，在整个数据生命周期的链条上，个人数据无论经历过多少次的流转、交易，后续主体对个人数据的使用权限都不得超过用户最初授权的范围。这就对很多平台将挖掘、整理、分析的数据并再次交易进行限制，明确企业的责任。在大数据时代背景下，国家的数据产业是否可以快速健康发展，关系到国家以及人民的重大利益。而大数据确权也是一个重要的前提和开端，能够很好的处理大数据的交易、使用和个人隐私。

在技术高速发展的今天，个人数据保护应该从最初的限制，转变为既要安全，又要发展，这是一个非常重要的问题。在大数据环境下，数据从产生、收集、处理、交易、使用、在交易直到消亡，这期间经历了无处次的复制和分享。整个数据生命周期所经历的这个链条，已经变得比较漫长了。而在这些链条的各个节点，参与到数据生命周期主体也变得更多，所以我们上述所探究的大数据确权，大数据权利体系也会随着不同主体的参与，各个主体之间的数据关系而不断丰富和发展，所以在数据确权的基础上，相关的资料收集、传递、保管和使用行为等方面的法律规范，也应该随着时代的变化而变化。

参考文献：

- [1]维克多·迈尔·舍恩伯格，肯尼恩·库克耶.大数据时代：生活、工作与思维的大变革[M].盛杨燕，周涛，译.杭州：浙江人民出版社，2013：8，127.
 - [2]Gurrin C, Smeaton A F, Doherty A R. LifeLogging: Personal big data. Foundations and Trends in Information Retrieval, 2014, 8 (1): 1-125
 - [3]周林彬,马恩斯.大数据确权的法律经济学分析[J].东北师大学报(哲学社会科学版), 2018(2): 31
 - [4]刘凯湘.民法总论[M].2版.北京：北京大学出版社，2008：59.
 - [5]彭云.大数据环境下数据确权问题研究[J].现代电信科技, 2016, 46(5): 18.
 - [6]费孝通.乡土中国[M].北京：北京大学出版社，2012：158-159.
 - [7]王玥雯.新时代大数据的概念及应用研究[J].江苏科技信息 2021(3): 8: 22
 - [8]袁勇,王飞跃.区块链技术发展现状与展望[J].自动化学报, 2016, 42(4): 481-494.
 - [9]彭云.大数据环境下数据确权问题研究[J].现代电信科技, 2016, 46(5): 17-20.
 - [10]林华.大数据的法律保护[J].电子知识产权, 2014(8)
- 作者简介：徐莉，1991年08月，女，江苏连云港，桂林电子科技大学，汉族，硕士研究生，法律（非法学）。