

数据流通理论

数据资源权利配置的基础

高富平*

摘要 数据的价值在于认知世界,大数据和人工智能改变了获取和分析数据的手段,使人类有了新的知识和智慧供给方式。各数据控制者掌握的碎片化的数据只有不断流动、汇聚、结合用于各种目的的数据分析,才能实现其价值,满足智能分析需求。数据流通实现了数据社会化利用,形成数据经济的内核。数据流通本质上是数据使用许可,涵盖一对一许可、一对众许可和相互许可三种方式,三种许可全面构筑了数据社会化利用的模式。个人数据和非个人数据均具有可流通性,关键在于法律上认可数据控制者具有许可使用的法律能力。在没有法律明确赋权的情形下,基于数据的事实控制使数据控制者具有事实上的数据使用权,可以开启数据流通之门。关键在于需要法律明确数据控制(使用)合法性判断规则和数据流通利用的责任规则。

关键词 数据流通 数据可流通性 数据使用许可 数据经济

信息和通信技术发展到今天,尤其是当网络被深度和普遍应用,人、物(自然界、机器等)、组织的活动或运行数据被各种网络系统、传感器和智能设备记录下来,形成描述客观世界的数据。今天,我们不仅可以全息地数字化记录(被称为数据化)人类所处的物理和社会环境、物或人的轨迹或行为,而且具有处理和分析海量数据所需要的能力,这便是大数据技术。^{〔1〕} 在本

* 华东政法大学法律学院教授、数据法律研究中心主任。本文为国家社科基金重大项目“大数据时代个人数据保护与数据权利体系研究”(项目编号:18ZDA145)的研究成果。

〔1〕 1998年,美国计算机科学家 Mashey 在一次演讲中首先使用“big data”概念。《自然》杂志在2008年9月推出了有关大数据的封面文章《大数据:从数据中提取内涵》,“大数据”才成为IT行业中的热门词汇。现在人们多用大数据来描述如今信息技术及其应用。IBM将大数据概括为4个“V”,即数据量(Volume)大,数据形态多样(Variety),数据处理速度(Velocity)快,数据价值(Value)密度低。

文中,数据或大数据仅仅指由网络设施、连网设备所记录的关于人、物、组织的行为或运行轨迹的数字记录(此类数据亦被称为机器生产的数据)。此类数据有两个重要特征,一是实时持续产生;二是可以溯源到特定的对象(即人、物或组织)。这两个特征使得我们可以对数据描述对象的行为(性能)特性、规律等进行分析。随着算法、算力和算能的提高,数据挖掘分析(大数据分析)演绎为人工智能。人类通过各种算法分析直接由机器生产的数据,洞察和分析客观世界,发现知识(规律)和辅助决策。换言之,人工智能开启了一种新的知识和智慧形成或供给方式,使人类由过去的主要依赖人力(观察记录)收集数据、分析客观世界,形成知识和智慧,转变为依赖人力和智能机器两种方式形成和供给人类知识和智慧。由此,人类社会进入大数据时代(本文称为数据时代),开启数据驱动的经济(Data-driven economy),〔2〕或称为数据文明时代。〔3〕数据时代(数据文明、数据经济)的主要标志是,数据成为社会基础资源和经济活动的要素,成为比土地、资本、劳动力等更为核心的要素,堪称“石油”。〔4〕事实上,现在人们已经开始重视并利用数据资源,尽可能多地获取和控制数据,并利用各种数据分析工具分析数据(包括人工智能),应用于科学研究、社会治理、商业活动等。〔5〕

数据之所以被视为资源,原因就在于其分析价值,单个数据可以直接描述对象的某个或某类特征,但海量数据相互联系,就能够抽象出数据对象背后的普遍特征,并通过其透析客观世界或分析对象的规律、特征,预测未来。但是每个社会主体所掌握的数据是有限的,要形成足够大、满足各种使用目的的数据集,〔6〕就必须利用他人掌握的数据(掌握数据的主体,称为数

〔2〕 2013年,软件和信息产业协会发布了《数据驱动创新》白皮书,提出并定义了数据驱动创新;OECD于2015年发布了研究报告《数据驱动创新:为增长和福祉的大数据》;2014年,欧盟议会、欧盟理事会、欧洲经济和社会理事会和地区理事会联合发布一份通讯《走向繁荣的数据驱动经济》,明确提出“数据是未来知识经济和社会的核心”;2017年,欧洲政治战略中心发布了《进入数字经济》。伴随这些文件的使用,数字经济逐渐成为描述大数据应用下的社会经济形态的主流词汇。

〔3〕 涂子沛认为:“数据正在改变所有那些组成文明的要素”,“就如同农耕之于古代文明,工业革命之于现代文明,数据将催生一种全新的文明形态”。他称之为数文明。参见涂子沛:《数文明——大数据如何重塑人类文明、商业形态和个人世界》,中信出版集团股份有限公司,2018年10月第1版,前言,第XX页。

〔4〕 *Regulating the Internet Giants: The World's Most Valuable Resource Is No Longer Oil, but Data*, *Economist*, May 6, 2017. p.7.

〔5〕 例如,人类行为在互联网/移动互联网上留下数字痕迹的大数据分析或知识发现已经运用于社会科学研究的诸多学科。较为典型的包括影响力与关注度、社交媒体内的子网络、痕迹背后的动因等。参见陈峰:“数字痕迹:人类行为大数据的生成、算法和知识发现”,《图书馆学研究》2018年第22期,第57页。

〔6〕 数据集(big datasets)是设备、传感器、互联网交易、电子邮件今天和明天产生的大规模、多样、复杂的、纵向和(或)分布式数据集(datasets)。National Science Foundation, Solicitation 12-499: *Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)*, 2012, <http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf>. 最后访问日期:2019年9月20日。

据控制者),〔7〕同时也要让他人利用自己的数据资源,即实现数据的社会化利用,而不仅仅是自我利用。数据流通即是实现数据社会化利用和实现数据资源价值的必然路径,正是有社会化数据流通和利用,才能形成数字经济。

数据流通要求承认数据控制者享有权利(本文称为数据使用权),否则无法构筑数据资源的流通秩序。相反,如果没有数据流通或不存在流通必要,也就没有赋权的必要。因此,对数据流通的揭示是数据赋权研究的重要理论基础。笔者在《数据生产理论——数据资源权利配置的基础理论》一文中已经揭示,数据并不是天然存在的,而是被生产出来的,原始数据的生产者和原始数据的汇集者(数据集的生产者)均应当配置以权利,由此使数据得以流通利用,从而满足数据分析、人工智能的需要(以生产出知识和智慧)。〔8〕本文将进一步阐释数据流通的含义和必要性,以及数据的可流通性、数据流通形式和法律规范,以勾勒数字经济的基本形式,同时也为数据赋权提供进一步理论支撑。

一、数据流通:数据价值实现方式

流通的基础含义是流动,但作为一种资源的流动,主流的形式是交换或交易,而交易需要相应的法律制度安排。由于至今在法律上对于数据上是否存在以及存在什么法律权利(产权)还未有共识,因而数据是否具有流通性仍然是一个问题,〔9〕研究文献寥寥。〔10〕讨论数据产权问题可以为数据流通提供制度基础,但产权制度需要针对不同的数据类型或数据流通内容进行设计,否则会泛化,不具有实际意义。本文试图揭示数据流通(事实),以数据流通需求“驱

〔7〕 数据控制者是国际社会在个人信息保护立法中提出的一个概念,它指实际控制使用数据的主体;而与此相对的概念是数据主体,特指个人数据描述的对象为自然人(因为人是主体)。数据控制者被广泛接受,用来表示实际拥有或控制数据的主体,至于该主体享有什么权利则不去讨论。

〔8〕 高富平:“数据生产理论——数据资源权利配置的基础理论”,《交大法学》2019年第4期,第17页。

〔9〕 从笔者查阅文献来看,人们只讨论数据流动(data flow or flow of data)很少使用数据交易(data transaction)、数据转让(data transfer)之类术语。重要原因是不存在基于产权的数据交换或交易。在实践中多用数据分享(data sharing)来描述数据流通和利用事实。数据分享回避了产权问题,又能描述数据被不同主体使用的事实。本文旨在将流通概念引入数据资源的配置利用以勾勒数字经济基本框架。

〔10〕 截止2019年9月25日,中国知网检索,除有几篇行业呼吁性文章(邬贺铨:“大数据的共享与开放面临哪三大挑战?”,《财经界》(学术版)2017年第20期,第7页;张敏翀:“数据交易:趋势与对策”,《信息通信技术》2017年第4期,第6页;张峰:“数据流通释放数据价值”,《转件和集成电路》2017年第7期,第1页;车品觉:“数据如何在高信息量社会中安全流通”,《西部大开发》2019年第6期,第1页),只有少数研究个人数据的学者关注到个人数据流通(如米新丽、卫洪光:“论个人数据的使用、流通与监管”,《河南财经政法大学学报》2018年第5期,第6页),很少有系统研究数据流通理论和制度的文献。在学位论文中,有一篇研究科学数据共享硕士论文(孙燕华:“科学数据共享中的知识产权保护与数据使用许可”,兰州大学农学专业学位论文,2015年,第25页);有一篇研究如何保护旅客隐私信息的同时完成高铁与其他交通方式间旅客出行数据互联互通的工科博士论文(喻麒睿:“高铁共享汽车数据流通机制及关键技术研究”,中国铁道科学研究院博士研究生学位论文2019年版,第4页),尚未有数据流通或数据分享或数据许可的法律学位的论文公开。

动”产权制度设计,为产权制度提供“经济基础”。

(一)数据价值的实现

数据具有价值,而价值需要被界定、确认(保护)和实现,围绕数据价值实现的活动构成了数字经济的内容。法律制度无非是为每一位价值创造者配以权利,使其可以进行商业交易,让渡权利,实现价值(经济利益)。因此,要揭示数据流通,就必须揭示数据价值的实现路径。

作为一种资源,数据的价值在于认知世界,为人类提供知识和智慧。人类历史即是不断描述、分析、探索自然世界和人类社会,形成认知、知识,创造出各种文明的过程。这一过程可以简化为获取数据和分析数据的过程。在大数据出现之前,人类主要通过肉眼、仪器等,观察、记录、描述客观世界,通过统计分析、演算、逻辑推理等方法分析特定对象,形成各种认知。所采取的方法主要是数据分析,包括定性数据分析、验证性数据分析(证实)、探索性数据分析(发现)。伴随大数据时代的来临,人、组织、机器、自然界等活动或运行被实时地以数字化方式记录下来,形成描述这些对象的海量数据(被称为大数据),并逐渐产生了从海量数据中获取另一类知识的分析方法——数据挖掘。^{〔11〕}数据挖掘也被称为大数据分析,大数据分析逐渐演进为人工智能。人工智能成为人类发现和获取知识和智慧(辅助决策)的新手段。^{〔12〕}

数据产生和利用方式的改变并没有改变数据的基础价值。无论过去还是现在,数据的价值均在于分析其描述的对象,获取对对象的认知,分析其规律,预测其行动趋势等。这里的分析对象,可以是单体,也可以是复合体在特定时间段的行为或轨迹。比如,既可以是一个人,也可以是一群人的行为特征或预测其行为倾向等数据;既可以是一台机器,也可以是若干台机器运行数据。小到查阅、阅读数据,中到简单归纳统计分析,再到复杂的算法模型下的智能分析,都是为了从数据中提取有用的信息,获取知识,做出决断或辅助决策。为了构建清晰的数据经济的分析工具,笔者将从数据中获取知识或智慧的价值视为数据的使用价值(使用价值等同于分析价值),即数据被分析利用后所产生的价值。一旦我们将数据的价值定位于此,那么所有的数据流动即是围绕该目的展开。

数据作为一种资源,其独特性在于可重复利用,而且可以在不断的结合汇集和演算分析中产生新的价值,而且这个过程并不会终止,可以不断演进下去。与物质商品不同,数据的使用不会减损它的价值,反而会获得新信息或知识,赋予新的含义;数据使用的过程也是不断流动

〔11〕 数据挖掘是指从海量数据中通过算法搜索隐藏于其中的信息的过程。数据挖掘通过统计、在线分析处理、情报检索、机器学习、专家系统(依靠过去的经验法则)和模式识别等诸多方法来实现。相对地讲,数据挖掘是从海量数据中找到隐藏的规则或知识,而数据分析的数据样本要小,且分析目标比较明确;数据挖掘是机器从数据样本集中发现知识规则,而数据分析是人的智力活动结果。

〔12〕 学者指出:智能制造的核心要义便是在两化融合的基础上构建智能分析优化系统“工业大脑”,对大数据进行智能化分析进而实现智能决策。参见于洪、何德牛、王国胤、李劼、谢永芳:“大数据智能决策”,《自动化学报》中国知网网络出版;2019-04-22;http://kns.cnki.net/kcms/detail/11.2109.TP.20190422.1029.008.html,最后访问日期:2019年9月20日。

以及与其他数据结合的过程。数据并没有因不断结合被消耗,反而是在积累中成为更大的和更有用的数据(集)。因而数据具有积累性(Cumulative)与非消耗性(Non-deteriorating),不会在使用中减损、毁坏,不会在流动中丧失价值,反而会增值。数据自身不仅不因使用而损坏,而且可以无损地流动。有学者将之定义为无损性或非消耗性,即“对于数据提供方和数据使用方来说,数据流转与传统民法中物的流转方式不同,它可以通过多次的复制提供,且不因为多次的提供和交易而减损自身的内容。”^[13]数据和数据价值的非消耗性决定了数据适合流动,决定了数据流动之后并不妨碍原数据控制者对相同数据的利用。

数据的价值不是固定的,而是可变的和多样的。这既取决于它与哪些数据结合以及与大规模数据结合,也取决于采取什么样的算法以及用于什么样的目的或场景。相同的数据可以发展出不同的应用场景,每个场景之间可以互动互通,也可以各自独立;同样的数据,运用不同的算法或者不同的分析目的,得出的结论完全不同。因此,数据在不断交换、聚合、匹配中产生新价值。数据流通可以说就是不断发现其价值的过程,让数据变得有生命力或有活力。现在出现了许多数据聚合平台,其所提供的服务本质上是将数据提供商和企业应用联系在一起,让数据在平台上进行相互交汇和连接。

数据来源具有多样性,每个数据控制者均可能合法地采集和获取相同对象的数据;同样一个主体或物可以被很多设备记录,形成多数据源。每一个数据源是对同一主体或物的描述,多少都具有相关性。但是,每个数据都是有缺陷或不完整的,要形成一个对象的足够多的数据,就需要把这些数据关联起来,如此才能对同一对象作出更准确的分析。大数据的价值和能力在于它的多维度性。来源于不同渠道的对于相同对象的数据相关联系,才能具有分析价值。数据来源(渠道)越多,相同的分析方法得出的精确度就越高。数据之所以要流动就是因为数据是分散、孤立生成和存在的,单个的数据并没有多少使用价值。针对同一对象的数据可以被若干不同的主体采集(称为数据的生产),而每个数据采集者(数据控制者)所掌握的关于该对象的数据总是有限的、不完全的。在大数据时代,数据量足够大已经不是问题。但是,数据不是光大就有价值。数据要有多样性和多维度,只有通过不同维度收集和采集,才能发现数据的互联关系。而数据的多维度的实现方式就是数据的流通和汇集。为了对某个对象作出更加全面或精准的分析,就需要掌握更多关于该对象的数据。显然,仅凭单个主体的力量无法形成关于特定对象的数据集,无法实现大数据分析,而需要从其他数据控制者处获取有关该对象的数据。于是,数据流通指数据从数据控制者(数据提供者)到数据使用者(数据需求者)之间的流动。数据流通不包括个人提供数据给数据控制者的情形,此种情形被称为采集或直接获取数据。

数据流通的目的在于实现数据的分析价值(使用价值),但是数据流通环节体现的是数据

[13] 王磊:“从数据属性视角看数据商业化中的使用规则”,《网络信息法学研究》2018年第1期,第4页。

的交换价值。数据交换价值的基础是数据关联性、准确性和可用性(数据质量),^[14]并受数据供给(市场)的影响。在大数据背景下,任何数据都具有潜在的分析价值,其价值取决于它与多少数据结合、汇集,形成多少个特定分析目的的数据集。每一次流动意味着数据潜在使用价值的实现,而每一次数据分析都是若干被汇集数据使用价值的实现。因此,数据流通体现的是该数据在最终的数据分析价值中的市场价格,即数据的交换价值。每次数据流通旨在实现数据的交换价值。对于数据控制者而言,实现其控制的数据价值的方式是流通数据,向别人提供数据而实现数据的价值。承认数据的交换价值是构建数据流通的前提。

综上,数据的价值在于分析利用,分析利用需要不断流通,数据流通即意味着数据使用。在某种意义上,没有数据的流通,就没有数据价值的实现。数据在流通中被使用,数据的使用伴随数据流通,数据需要流通是数据的天然属性。我们所需要做的是在数据资源或数据经济概念下理解和构建数据流通制度。

(二)以数据流通为基础的数据经济的建构

大数据和人工智能改变了获取和分析数据的手段,但没有改变知识或智慧来源于数据的事实,这意味着数据的价值早已有之。在人类社会进步的长河中,数据一直扮演着文明“火种”的角色,数据也一直被人类分享和利用,只是一直没有进入商业流通的范畴。这是因为数据一直处于自由利用状态,任何人对于采集和获取一手的数据都具有先用权,数据一旦公开即进入任何人都可以自由使用的状态;通过数据分析形成的智力成果在满足一定条件时可获得知识产权保护,其成果的商业利用受到产权制度调整。^[15]但是,数据一直处于公共领域,未进入商业流通。即使将数据加工成为人们可方便利用的产品(即数据库),其保护问题至今也悬而未决,更多依赖法律救济,而不是赋权。^[16]在人类进入数据时代,数据的获取和利用方式的改变,使我们有必要重构数据利用秩序,因而需要勾勒其数据价值实现——流通——的路径。

随着5G应用,客观世界的数字化进程不断推进,人类将进入全面依赖数据的时代。在数据时代,数据生产和供给成为一种生产(经济)活动,成为需要社会化和市场化配置利用的资源。^[17]在这样的社会,每个主体都是数据的生产者,也是初始控制者。这决定了每个数据控

[14] 数据质量影响着数据的利用,是交易双方关注的核心问题。数据产品定价的基础是数据品种、时间跨度、数据深度、数据完整性、数据样本及数据实时性等。参见张敏:“交易安全视域下我国大数据交易的法律监管”,《情报杂志》2017年第2期,第6页。

[15] 参见薛铁成:“综述与评鉴:人工智能创作作品的现状及法律保护路径初探”,《中国海洋大学学报(社会科学版)》,2019年05期,第6页。

[16] 以“淘宝公司诉安徽美景信息科技有限公司反不正当竞争一案”为例,“生意参谋”作为数据产品并未承认其财产权赋权,而是构成了竞争性的财产利益,通过反不正当法的救济实现了对于数据产品的保护。参见杭州互联网法院(2017)浙8601民初4034号。

[17] 中央财经委员会办公室副主任韩文秀在中国共产党十九届四中全会新闻发布会上透露,中国共产党十九届四中全会已经提出要“健全劳动、资本、土地、知识、技术、管理和数据等生产要素按贡献参与分配的机制”,这将加快数据的资产化和资本化进程。

制者对相同数据进行控制的正当性。但是通过网络、传感设备记录、采集形成的单个数据本身并没有多少价值,只有汇聚多源头、多维度的足够大的数据,才能进行智能分析,每个碎片化的数据才变得有意义。数据的价值在于分析描述对象(人、物、组织)的特征和规律,预测行为或趋势。当数据量小时,缺乏联动性,价值还未显现,而当样本足够多时才能洞察规律或做出预测,体现数据的价值。因此,碎片化的数据需要聚集汇合才能用于分析,只有被分析使用,数据才能实现其价值。如果将人工智能看作是数据分析利用的方式,那么数据就是人工智能的原料。数据需要不断流通来“喂养”人工智能实现其价值。

在没有流通制度设计的情形下,采取各种爬虫技术手段从数据生产者(数据控制者)处爬取数据是获取数据的主要手段。^[18] 随着数据价值被重新发现,任何数据拥有者都希望尽可能控制和获取数据,以防范他人抓取,即使自己不利用或没有能力利用数据,也不愿意向他人提供数据。数据流通旨在为原始数据生产者实现其数据价值提供通道,此时需要原始数据的生产者愿意将自己生产和控制的数据拿出来,供他人使用。在理性人的假设下,商业性流通是最有效的方式,即承认原始数据生产者对数据的控制权,通过交换或交易获取他人的数据或对价,可以激励数据生产和数据提供(分享),最大化地实现数据的价值。

实际上,数据生产并不能直接用于智能化分析,还必须按照特定目的,收集汇聚、清洗整理、分类归集,形成可用的数据资源(称为数据集),为数据分析提供原料。笔者在《数据生产理论——数据资源权利配置的基础理论》一文中提出,数据经济在原始数据生产者、数据集的生产者和数据分析者之间展开,贯通他们之间的就是数据流通。^[19] 也就是说,数据流通发生于两个层面,一是在原始数据生产者与数据集的生产者之间,前者为后者提供原料,后者加工处理成为可供数据分析使用的数据集(数据产品);二是数据集生产者与数据分析者之间,前者为后者提供数据分析的原料,后者“生产”出知识或智慧。没有数据生产,就不能汇集加工成为满足数据分析的数据集;没有数据集的生产和供应(流通),就难以满足大数据分析和人工智能对具备一定品质的数据的需求。由此,两类数据流通共同构筑数据经济的产业链。如果说任何一种经济活动本质上体现为特定经济资源的市场化配置和利用,那么数据流通决定着数据经济的基本框架。因此,数据流通是数据经济的核心,构筑数据流通秩序就是构筑数据经济的基础制度。

二、数据流通的含义和形式

数据通过流通实现数据的价值,因而数据流通包括了所有向他人提供数据的行为,这样的

[18] 爬虫技术从网页数据采集,逐渐发展为 app 数据采集、网络数据采集,以致于爬虫成为高效获取大数据的最便捷的途径。于是诞生大量的专业爬虫和数据挖掘公司,如 diffbot, import.io, scrapinghub, apify 等,国内也有神箭手、八爪鱼、造数等专业数据抓取工具。

[19] 高富平,见前注[8],第 11 页。

行为在实践中存在不同的表现方式。

(一)数据流通的含义

在笔者看来,数据流通是向他人提供数据或使他人接触或使用数据的行为。

数据流通实现数据的使用,这种使用可以是简单的阅读或识读,也可以是对大量数据进行演算分析。数据不同于其他物质商品,其他物质商品的使用以实际占有为前提条件,而数据的使用可不以控制该数据为条件,查询、知晓、阅读、(利用数据)通信等均可以实现数据的使用。数据的价值取决于用于什么用途、实现怎样的目的,因而数据的价值取决于利用,而不取决于其本身。同样的一个数据与不同的数据结合、与不同规模数据的结合所产生的价值是完全不同的。数据价值的多样性、可变性决定了数据需要与其他数据结合(即流通)产生新价值,使数据流通成为实现数据价值的重要方式。对于数据资源来讲,流通即使用,使用即流通,流通即是数据的生命,流通是数据社会化利用的方式。

由于数据的占有(控制)不完全具有排他性,^[20]同时可以为多人同时占有(控制),这导致数据的使用也不具有竞争性,同时可以为多人同时使用。数据这种可并存共用性决定了数据流通并非以数据拥有者(提供者)放弃数据(即数据转让)为典型,而以允许他人使用数据为常态。数据流通即使用决定了数据的流通本质是允许他人使用数据,因而,数据使用许可成为数据流通的基本内涵。因此,本文以数据使用许可(简称为数据许可)来描述数据流通。

数据流通利用是社会运行不可或缺的要害,只是在法律正式确认和规范之前多以非正式方式存在或以隐蔽的方式存在。例如,在许多的商业合作或交易中都内含着同意使用某些数据资源的内容。大量的数据获取和使用则遵循自然法则,比如利用爬虫工具获取他人系统中产生的数据非自愿“流通”方式。一旦数据流通被法律认可,那么数据使用许可合同就成为数据流通基本的法律形式。数据许可合同的基本功能首先在于明确数据流通的方向或来源,确立数据流通的秩序,使数据流通可追溯;其次,数据许可合同可以明确数据流通双方的权利、义务和责任,将数据流通纳入法律调整范围。

从法律的角度看,由于数据控制者并不享有法律意义上的财产权(至于其拥有什么权利,本文暂不作探讨),因此其许可使用并没有明确的授权,而只是可以阻却数据使用者对数据的违法使用。^[21]因此,数据许可合同在数据流通中显得非常重要,需要清晰地描述数据控制者

[20] 数据具有有限排他权,参见崔国斌:“大数据有限排他权的基础理论”,《法学研究》2019年第5期,第8页。

[21] 这里在广义上使用许可,将许可视为违法性阻却。在法律一般禁止的情况下,只有依法申请有权行政主管部门颁发许可才可以合法从事某种行为。行政许可使法律禁止的行为合法化。在私法领域奉行法不禁止即自由,任何个人行为不受其他私人的控制或干涉。在私人享有财产权的情形下,每个人都有不作为义务,如果要从事某种行为,就得获得财产权人的授权或许可。典型地,知识产权人可以许可他人使用其专有的智能成果,使不能从事的行为具有法律基础。在这种情形下,许可是基于明确的财产权做出。但是,在私法领域也存在大量的许可行为,即许可人并没有法律上财产权,但是基于对某种财产利益的合法控制也可以许可他人使用该财产利益,给他人创设行为合法性基础。本文即是在这样的广义上使用许可一词的。

(数据提供方或许可人)与数据使用者(数据接受方或受许可人)之间有关数据使用的权利和义务关系。数据使用许可的基本内容是数据控制者许可数据使用者按照一定的条件使用特定范围的数据。数据许可合同的关键内容即是要明确数据的范围、使用目的、使用条件和安全保障义务等。其中,数据使用条件包括数据使用范围和方式、是否支付对价(使用费)、数据是否存储、是否允许再许可等。数据许可可以是商业性的,也可以是非商业性的,它可以支撑数据的自由使用(开放数据),也可以支撑数据交易,实现数据的市场化利用。

由于数据可共用和易再流通性,数据流通面临许多不可控的因素。传统通过物品占有来控制物品的流向和使用方式在数据流通中荡然无存,需要引入新的机制使数据流通具有一定的可控性,才能构筑数据流通的秩序。这既决定了数据使用许可合同在数据流通中发挥着非常重要的作用,同时也需要建立数据流通安全的服务机构予以支撑,比如数据交易服务平台。总之,数据许可既实现了数据的流通,又保障数据的流通的秩序。

(二)数据流通的方式

既然数据流通本质上是数据许可使用,我们那么应当就从数据许可使用方式角度勾勒数据流通形式。从数据利用行为的角度,数据使用行为可以包括数据查询、数据调用、数据缓存、数据拷贝或复制、数据交换、数据演算分析等方式。这些方式构成数据许可使用的内容,是对使用范围的限定,主要由当事人来自由约定。从法律规范上来看,可以从数据许可主体的角度,将数据流通分为一对一许可、互为许可和一对多(众)许可,这三种许可构成了数据的三种流通方式。

1. 一对一数据许可

一对一数据许可是指数据拥有者(提供者)仅向特定对象提供数据,允许其使用数据。一对一数据许可是常见的数据流通方式。它可能内含于企业之间的业务合作中,一方许可另一方使用特定范围的数据;也可以是单独数据许可使用合同,比如开放 API 协议。^[22] 在开放 API 协议中,一个系统的运营者向另一个主体开放数据接口或信道,以让该主体的程序能够调用该系统形成和存储的数据,甚至在许多情形下可能不需要转移和存储数据,只需要允许其处理或运算利用数据。实践中,网络公司为他人提供广告服务,广告公司购买所谓的“广告位”,实现精准推送或营销,^[23]大多数情形即是采纳数据一对一许可调用方式利用数据。在数据开放过程中,为确保数据安全,数据提供者也可以只开放接口,允许数据分析者运算数据,而不移转数据。因此,开放系统接口许可特定对象在特定时间使用特定数据是最为常见的数据流通方式。

2. 互为数据许可

两个以上的数据拥有者相互进行数据使用许可即为互为数据许可,这是共同利用各自产

[22] 开放 API 是服务型网站常见的一种应用,网站的服务商将自己的网站服务封装成一系列应用编程接口。

[23] 鞠宏磊、李欢:“精准广告相关隐私问题的规制原则与策略”,《编辑之友》,2016年第6期,第2页。

生的数据的一种行为,这种数据许可利用本质上是相互许可方式,亦可以称为数据共享。数据共享的基本特征,一是主体限定在特定范围之内,至少两个主体;^[24]二是特定主体之间相互使用各自拥有或控制的数据,存在相互许可使用机制。从计算机学科的角度来看,数据共享就是使位于不同地方、使用不同计算机和不同软件的用户能够读取他人数据并进行各种操作运算和分析。在数据共享中,每个共享人既是数据提供方,也是其他主体提供的数据的接受方或使用方。数据共享可以使特定范围的主体更充分地使用已有数据资源,减少资料收集、数据采集等重复劳动和相应费用。共享数据可以视为该范围主体的共同数据资源,因此数据共享的原理是相互让渡自己的数据使用权,实现数据的共享共用。^[25]数据共享可以通过数据交换、联通的方式实现,也可以通过共同建设共享平台(数据池)的方式实现,二者的区别在于是否涉及数据集中。但不管怎样的方式,首先需要建立一套统一的、法定的数据交换标准,规范数据格式,使数据按照统一的标准进行交换、流通、被共享主体调用。

3. 一对众数据许可

一对众数据许可是指数据拥有者对不特定主体进行的许可使用,其根本特征在于数据使用人具有大众性,是面向社会需求者的数据许可。一对众许可大致分为自由使用许可和有条件使用许可两类。

自由使用的许可即将特定数据明确为开放数据(Open data),不设任何条件且由不特定社会主体随意取用。开放数据本质上也属于一种数据许可,只是数据控制者不再设置任何条件提供给社会主体使用数据,使用的权利源自于法律。在实施时,可以采取标准化的数据开放许可条款或数据开放政策。相对地,有条件使用许可是数据拥有者向不特定需求方许可使用数据,但是限定了数据的使用条件,包括使用目的、使用主体资格、使用对价等。有条件数据许可本质上是一种数据交易,它通过市场机制将数据资源配置给需求者,实现数据的社会化利用。由于这样的市场化的一对众的数据许可必须是可控的、有序的,因而数据交易通常需要借助专门数据交易平台实现。美国数据经纪人(Data broker)、我国上海数据交易中心建设的数据交易平台等都属于为数据交易提供服务的机构。一旦通过市场化的方式在数据提供方与需求方之间实现匹配,那么双方之间也需要许可协议,而这种许可可能是一对一的许可,也可能是一对众的标准化许可协议。只有后者才是真正的一对众许可。

因此,从主体的角度来看,数据有序利用的秩序不外乎三种:一对一许可、相互许可和一对众的自由利用许可,只是促成这样的机制有市场化和非市场化之分。所谓的市场化,是将数据作为“商品”,通过公开数据交易磋商,匹配数据供需求双方。下表清晰地地区

[24] 两个主体之间数据共享也称为数据交换(data exchange),交换双方互为数据的提供方和接受方。当两个以上主体相互许可使用时,才是真正意义上的共享。

[25] 数据共享不仅应用于公共事务领域,而且在科学研究、商业领域也广泛应用。集团公司、关联公司甚至具有业务关系的公司之间共享企业数据、客户资源、个人数据的情形很普遍,本质上属于数据共享的一种方式。

分它们之间的差别。

表 1 数据有序利用的三种类型

类型	是否市场化	协议类型	对价	使用条件
一对一许可	非市场化	数据许可协议或许可条款	取决于约定	有限制条件
	市场化	个性化许可协议	有偿	有限制条件
一对众许可	市场化	标准化数据许可协议	有偿	有限制条件
	非市场化	自由利用许可(开放数据许可)	无偿	无限制条件
相互许可(共享)	非市场化	数据共享协议	取决于约定	取决于约定

数据流通需要一定的秩序,而许可协议便是创制这一秩序的工具。英文“Data sharing(正确译文应为数据分享)”一词实际上涵盖所有让他人使用数据的方式,即一对许可、相互许可(数据共享)和一对众许可。在某种意义上,数据流通、数据开放和数据分享三个词均可以用英文 Data sharing 来表达。数据许可方式是准确描述和规范数据流通(分享或开放)的工具,只是因为数据可流通或可交易未被普遍接受,在 Data sharing 中并未得到强调。本文的数据流通或许可方式恰恰是在考虑市场和非市场化两种方式的基础上,全面描述了数据社会化利用的实现方式。

三、数据的可流通性

从数据描述的对象的角度来看,数据是关于世界万物的数字化,其种类无法枚举。由于人是主体,关于人的数据的利用受到关注,所以个人数据(在我国亦称为个人信息)概念首先被创制出来。随着物联网应用,人类进入万物互联的数据化时代,大量无关于人的数据的利用也受到关注。于是,从是否与人有关或有联系,人们区分出个人数据和非个人数据,并以此来概括所有的数据。这里对数据可流通性的探讨亦采用此二分法。

(一)个人数据的可流通性

个人数据或个人信息与个人之间的联系在于,该信息与个人有关,并且是用来识别或了解某个人。这种可以识别某个人的客观事实并不当然地让个人拥有或控制个人信息。^[26]“标识自己和识别个人(被识别)是社会的必然现象。”^[27]个人信息具有公共属性,关涉他人和社

[26] 高富平:“论个人信息保护的目——以个人信息保护法益区分为核心”,《法商研究》2018年第1期,第1页。

[27] 高富平:“个人信息保护:从个人控制到社会控制”,《法学研究》2018年第3期,第16页。

会利益,个人信息的使用就不完全是个人“说了算”,^[28]因此,应当是全面平衡个人利益、使用者利益(社会利益)和公共利益,使每种个人信息的使用行为具有法律上的正当性。^[29]这就意味着个人数据具有可流通性或可为人使用的特征,只是要满足一定的合法性要求。个人数据流通利用可能会危害个人尊严、自由、甚至平等,因而需要建立个人数据流通利用的规则,确保个人信息流通利用行为不侵犯数据上主体的权利。在某种意义上,个人数据保护法就是规范个人数据流通利用的法律。^[30]

不过,个人数据保护法是以个人为中心构建的,只考虑了数据脱离个人的一次流动利用,而没有考虑脱离个人的数据流通。当个人向他人提供数据或者个人行为被设备采集或记录,就形成了个人数据与个人(数据主体)的分离,产生了个人数据(信息)。与主体分离后的数据为数据控制者使用,形成首次数据使用关系,完成个人数据的首次流动。但只有当数据控制者将所拥有的个人数据又提供给了他人使用时,才是我们讲的数据流通。从数据取得的角度,数据流通不是从数据主体处直接采集数据,而是从数据控制者处间接获取数据(使用权)。

在某种意义上,世界各国的个人数据保护法均承认个人信息具有可流通性,只是需要征得信息主体的同意。个人数据保护和利用的基本原则是合法、正当和必要原则,使用者在首次向个人采集个人信息时,要告知收集范围、使用目的和方式等,之后的使用均受首次告知目的限定。^[31]由于数据控制者许可他人使用数据的行为显然超出了初次收集时确定的目的或用途,因此,个人数据流通需要征得数据主体的同意,这是个人数据流通的普遍规则。^[32]但问题在于,在数据需求方(数据使用者)不能直接事先接触数据主体的情形下,获得数据主体的同意几乎是不可能的。于是为了解决数据流通的合法性基础,就要求数据控制者在初始采集个人数据时获取数据主体的同意,同意数据控制者将数据提供给第三人。但是,第三人是谁、使用目的、方式等并不清晰,这样的同意是否为有效的数据主体同意,在法律上始终存疑。因此,主体同意作为个人数据流通(间接取得)的合法性基础难以支撑个人数据的流通利用,甚至成

[28] See Priscilla M. Regan, *Legislating Privacy: Technology, Social Values and Public Policy*, Chapel Hill, NC: University of North Carolina Press, 1995.2, p.211.

[29] 参见高富平:“个人信息使用的合法性基础——数据上利益分析视角”,《比较法研究》2019年第2期,第8页。

[30] 只是欧盟 GDPR 没有用流通,而用了流动(movement)或自由流动(free movement)。

[31] 目的限定原则是个人数据保护法的基本原则。1980年OECD《隐私保护和个人数据跨境流通的指南》正式提出“目的特定化原则(Purpose Specification Principle),即个人数据收集的目的应当在收集时确定,随后的使用限制在实现该目的的必要范围内,或者用于实现其他与该目的不冲突的目的和每次更改时确定的目的”。之后该原则被各国所吸收。在欧盟 GDPR 中被拆分为“目的限制原则”和“最小范围原则”两项(参见 GDPR 第5条(b)和(c)项)。我国的《个人信息安全规范》表达为最少够用原则:除与个人信息主体另有约定外,只处理满足个人信息主体授权同意的目的所需的最少个人信息类型和数量。目的达成后,应及时根据约定删除个人信息。

[32] 我国亦采这样的规则,例如,《网络安全法》第42条明确规定“未经被收集者同意,不得向他人提供个人信息”。

为数据流通的法律障碍。^[33]

2016年出台的《网络安全法》第42条的但书条款在个人同意之外,为个人数据流通提供了另一种合法性基础,即经过处理使个人信息达到“无法识别特定个人”且“不能复原”。只不过对该条款存在不同的理解,尤其是如何判断数据处理达到了“无法识别特定个人且不可还原”。这里至少有两种理解:其一,这里的处理只是去除身份识别,使数据中不存在直接与个人关联的身份数据或者唯一性识别符;其二,处理之后的数据不具有识别能力,包括不存在与其他数据结合可以关联到某个人。前者只是排除唯一识别信息,而后者则排除了一切可以关联到个人的识别因素,排除结合识别个人的可能性。由于识别个人是个人数据的根本属性,因而多在第二种含义上理解“但书条款”,且认为“去识别”+“不能复原”的结果是将个人数据转化为非个人数据。笔者认为,这样的解释可能较为符合第42条但书条款文义,但是并不符合个人数据流通利用的实践,我们应当在第一种含义上理解但书条款或在立法中创制一条个人数据流通规则。

个人数据的价值即在于识别分析数据对应的主体的行为趋势或倾向,这是大数据给经济和社会发展带来的红利。如果去掉所有的识别因素,不再关联到个体,那么数据就不再具有任何个体分析价值。因而,需要在维护主体权利和保障数据安全的前提下实现个人数据的流通利用。也就是说,所流通的数据必须具有指向或关联到某个体的可能性,即经过处理后的数据仍具有可关联和可交换匹配性,实现数据的汇集和分析利用。因此,为实现个人数据流通的数据处理,只应当去除个人数据(集)中直接识别、显著关联个人且唯一性的识别符。^[34] 标识符是网络环境对用户分类的便捷工具,其对个人权利或隐私的危害不在其本身,而在于用于何种目的和如何使用。直接或唯一性标识符具有明显的指示和关联效果,允许其自由流通会大大增加个人隐私被侵害的风险,甚至会给个人人身和财产安全带来风险。去除这些识别符之后增加了个人识别的困难(难度),减少了危害个人安全的风险,但是不能根本改变大数据环境下这些数据对个人的识别能力。不管采用什么样的技术手段,只要处理后的数据还有识别个人的颗粒度,就仍然具有识别个人的可能性。因此,不同方式进行去识别(匿名化、假名化等),只是个人数据安全风险管理的一种措施,并不因此使数据根本上丧失识别性因而转化为非个人数据。进而无论假名化还是匿名化处理后的去个人标识的数据本质上仍然属于可识别个人的数据,仍然应受个人数据保护法的调整。个人数据保护法需要对去识别技术措施本身以及经过相应技术措施处理后的数据的流通利用进行规范。总之,笔者认为,个人数据本身亦应当具有可流通性,只是需要相应的制度设计。这已不是对《网络安全法》第42条但书条款进行解释

[33] 参见任龙龙:“论同意不是个人信息处理的正当性基础”,《政治与法律》,2016年第1期,第5页。

[34] 作为计算机语言的用语,标识符一定是在一个处理域内具有唯一性。标识符也普遍地应用于个人数据处理中,用来区分用户或个人。标识符属于识别个人的信息。但并非所有识别个人的信息都可以作为标识符。能够唯一关联到个人的社会身份信息、网络设备ID(或数字身份)都可以成为标识符。

所能解决的,而需要个人信息保护法才能解决。对此本文不再细论。

(二)非个人数据的可流通性

非个人数据是与个人数据相对应的概念,因而一般采排除法来定义。即个人数据以外的数据都属于非个人数据。《欧盟议会和理事会关于欧盟非个人数据自由流动框架条例》^[35]将非个人数据定义为“个人数据之外的数据”(参见第3条)。显然,用这种排除法来定义其范围并不能准确揭示非个人数据的内涵和外延。笔者认为,非个人数据是对人以外的物质世界和社会要素进行数字化记录或描述的数据,包括但不限于关于物质天体、机器设备、组织机构等的数据库。非个人数据之所以为非个人数据,其根本点在于其不具有识别分析个人的能力,而只是用来分析物、机器、组织行为的能力或价值,只是对事件或现象的描述,而不能直接追溯或关联到人。

非个人数据大致分为两类,一类是“无关个人的数据”,另一类是“去识别”+“不能复原”的数据。前者从一开始就与个人无关或者不与个人发生联系;后者专指那些曾经是个人数据,但是被技术处理后不具有识别个人能力的数据库。

物联网的出现和应用是大数据产生的重要原因,^[36]而源自于客观世界和机器的数据大多数与人无关。无关个人的数据主要是关于物、组织和事件的数据。物包括自然物也包括人工物。比如传感器收集的室外温度、湿度、风力等,关于构筑物、建筑物构造、位置等数据,关于机器运行状况、能耗指标等的记录,如此等等都是关于物的数据。同样,关于组织机构的基本情况、运行情况的数据构成关于组织的数据。另外,大量数据是关于事件的,而非直接描述特定对象(物、组织和人),事件是数据的主题。事件大致分为自然灾害类(比如地震或地质灾害、洪涝火灾等)、事故灾难类(安全事故、环境污染、生态破坏等)、公共卫生类、社会安全类(比如群体性事件、刑事案件、影响市场稳定的事件等)等。

无关个人的数据的可流通性几乎不存在问题,因为非个人数据上不存在主体权利,不受数据库主体的权利掣肘,其流通性只遵循合法性原则。数据控制者所获取的数据只要合法,法律不限制或不禁止数据流通利用,也不侵害他人的财产权益(如商业秘密),那么数据控制者就可以许可他人使用数据。因此,无关个人的数据基本上遵循谁控制,谁流通,谁流通,谁负责原则。具体参见下文关于数据控制者流通数据的法律基础的讨论。

那些原来属于个人数据但被技术处理后不具有识别个人能力的数据库,亦属于非个人数据库,

[35] 欧盟委员会于2017年9月13日发布《欧盟议会和理事会关于欧盟非个人数据自由流动框架条例建议案》,该建议案于2018年获得通过,即REGULATION (EU) 2018/1807 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807>,最后访问日期:2019年9月20日。

[36] 目前较为公认的物联网的定义是:通过射频识别、红外传感器、全球定位系统、雷射扫描仪等信息感测设备,按约定的协议,把任何物品与互联网连接起来,进行信息交换和通讯,以进行智能化识别、定位、跟踪、监控和管理的一种网络。

可以按照非个人数据流通规则自由流通。问题在于个人数据处理到什么程度才属于非个人数据,可以不适用个人数据保护法而流通。按照笔者观点,个人数据既可因去除显著性的身份标识符而成为可流通的个人数据,也可因不具有识别个人的能力而作为非个人数据流通。

需要进一步指出的是,个人数据与非个人数据的划分具有模糊性,二者之间可以相互转换,比如货物运输中的定位数据,是货物数据,可以用于智能化管理,但仍具有可识别到货主(人)的可能性。同样,电表、水表等的状态数据,也具有这样的可关联到人的特性。许多与机器或物有关的数据,只有在切断与人联系(去身份)之后,才真正成为非个人数据。在许多情形下,某个数据是否具有识别个人的能力关键在于数据应用的场景、所结合的数据多少、使用方式等,我们很难孤立地看待某个数据是否属于个人数据。这意味着个人数据与非个人数据的划分并非是数据的现状或事实分类,而是基于管理上的分类,在实践中具有可变性。数据可流通性及其如何流通不能简单地从数据本身来判断,而应纳入数据应用的具体场景,在特定数据治理和风险控制体系下才能安全地实现。

四、数据流通的法律基础

既然数据流通是数据控制者许可他人使用数据的行为,那么数据控制者必需具有“许可权”,即许可他人使用数据的法律能力。缺失许可权就缺失流通的法律基础。正如所有权是物(商品)流通的法律基础、知识产权是智力成果商业利用法律基础一样,数据流通也应当具有相应法的律基础。这便是人们一直在探讨的数据产权问题。本文认为,即使在法律上还没有明确甚至不能明确数据控制者享有什么权利,数据控制者基于对数据的实际控制也可以成为数据流通或许可使用的法律基础。法律的作用仅在于确认和保护这种事实,同时规范其数据利用行为,防范数据利用的风险。

(一)基于事实控制的“数据使用权”

由于数据的特殊性,数据不可能成为某个主体支配的“物”,不可能设立所有权。最主要的一个原因在于数据使用就意味着流动,意味着相互匹配混合,数据在流动中实现价值,因而不能清晰地区分“你、我、他”之间的数据,从而形成清晰的产权边界。在理论上,数据初始生产者或者数据主体只要不向外提供数据,那么是可以独占和独享数据的,但这也意味着无法实现数据的社会化利用的价值。数据一旦流动,走上社会化利用之路,便很难保持其独立性,因而很难通过特定、独立的数据来界分其权利边界。^[37] 在法律上我们很难赋予数据控制者对数据

[37] 业界研究者提出数据资产登记制度来解决政府数据资源变资产,解决流通问题(参见余志兰、彭敦北:《建立政府数据资产登记制度,促进政府数据资源交易流通》,《通信企业管理》2017年第5期,第66—69页),但登记是否可行需要实践检验。

的支配权,^[38]甚至即使能够赋予法律权利,因数据的高度流动性,这种赋权也没有意义。但是,在某个时点数据还是有控制者的,特定主体对特定范围的数据具有某种程度的控制,只要法律承认和保护数据控制者对数据具有事实上的控制,那么也能以此构筑数据利用秩序。

要取得数据财产权就首先必须事实上控制数据,不控制数据就谈不上保护。在前网络时代,谁记录数据,谁便控制数据,数据不被记录就意味着丧失控制。在网络时代,数据因为各种系统、智能终端而采集和记录下来,因而架设这些网络基础设施的主体就成为数据事实上的控制者。再加上各种数据安全措施的采取,使得数据为特定主体控制成为可能。因此,网络和各種技术手段使对数字形态的数据的事实控制具有可能性。

在没有法律确认或赋权的情形下,数据控制者对所控制的数据享有使用权,这种使用权是基于其合法获取并控制而产生的。因此,基于事实控制即享有数据使用权。^[39]而现在的问题是,基于事实控制的数据控制者是否可以许可他人使用。对此,作者的答案是肯定的。基于对数据的合法控制,数据控制者不仅可以自己使用,而且也可以许可他人使用,只要这种许可使用不侵犯他人权利,也未违反禁止性的法律规定。一旦法律认可数据控制者具有允许他人使用数据的许可权,那么在每个数据控制者处沉睡的数据就可以被开放出来,在与他人的数据结合中发挥其价值。一旦承认数据控制者可以许可他人使用数据,那么数据控制者的数据使用权事实上就丰满为一种财产权,具有上升或转化为法律权利的可能。本文对此不展开论述。

许可他人使用数据意味着数据具有了交换价值,即数据的财产化就需要法律出面干涉数据这种流通利用行为。由于数据控制者基于事实控制形成的数据使用权的客体范围具有非固定性、不确定性,因而其数据控制(取得)合法性、数据可许可使用性的判断是一件非常复杂的事情。如果没有一定的规则指引数据流通双方,那么就会导致数据流通利用充满不确定性,面临巨大的法律风险。目前我国的数据流通利用现状基本上就处于这样的状态,一方面法律禁止违反法律向外人提供数据的行为,严重的可能面临刑事风险,^[40]另一方面对于什么是合法的数据流通行为却没有明确的法律规范。在这样的情形下,没有人愿意向外提供数据,也没有

[38] 胡朝阳教授以个人信息为例指出了个人信息财产权赋权的困境,认为基于“外部性内在化”原理赋予信息主体以个人信息权并赋予数据控制者以大数据财产权虽有助分别规制其双重外部性(数据安全风险等负外部性,分享经济价值实现等正外部性),但网络大数据背景下其双重外部性规制彼此交互影响而面临两难困境。参见胡朝阳:《大数据背景下个人信息处理行为的法律规制——以个人信息处理行为的双重外部性为分析视角》《重庆大学学报(社会科学版)》,Doi:10.11835/j.issn.1008-5831.fx.2019.06.001,最后访问日期:2019年9月20日。

[39] 民法中存在两个层次的财产秩序,一个是类型化的财产权,即所谓的法定财产权,法律赋予此类财产权利人明确的权利,权利人不仅可以自主利用,而且还可以自由处分;另一类是非类型化的财产利益,仅仅是一种事实存在,包括物的占有状态和其他合法利益(法益)。参见高富平:《信息财产—数字内容产业的法律基础》,法律出版2009年版,第188页。

[40] 2015年刑法修正案(九)确立了“侵犯公民个人信息罪”,2017年《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》进一步细化了“情节严重”的标准。

人敢从他人处获取数据。为了减少数据流通的不确定性,法律应当提供两个方面的规则:一是数据控制合法性判断规则;二是数据流通利用的责任规则。

(二)事实控制的合法性

法律的作用在于界定权利,明确交易规则,以降低交易成本和风险。在货物交易中,物权法为交易双方提供了一套公示规则和判断转让者是否有权利及权利是否清洁的规则。在数据许可使用中,法律无法清晰界定数据控制者的权利,数据控制者也无法拿出一份证明文件,证明其权利及其权利的清洁性。^[41]于是,我们只能依赖数据控制的合法性来替代传统的权利公示和判断规则。在数据许可中,降低交易风险的主要手段是数据提供者要证明其数据控制事实是合法的,数据使用者则要设法判断其合法性。一旦满足合法性要件,那么数据许可使用就具有合法性,数据流通风险就在当事人的可控制范围之内。

任何人都不能让渡大于自己的权利。在没有明确法律界定的情形下,数据控制者对数据的事实控制应当满足以下条件,数据许可使用才满足合法性要件:

其一,数据来源合法。数据来源合法指采集、收集和以其他方式获取数据是合法的。只有直接和间接获取数据不违反法律的禁止性规定,不侵害他人的权益,数据控制者对数据控制(使用)才具有合法性基础。《民法总则》第111条明确规定:“任何组织和个人需要获取他人个人信息的,应当依法取得并确保信息安全,不得非法收集、使用、加工、传输他人个人信息,不得非法买卖、提供或者公开他人个人信息。”《关于强加网络信息保护的决定》《消费者权益保护法》《网络安全法》等法律对个人信息的收集和使用作出规范,因而在取得个人数据的环节必须守法,不遵守法律会导致侵权责任、行政处罚甚至刑事责任,导致数据持有丧失法律基础。同样的原理也适用于数据上可能存在的其他权利,如商业秘密等。只要采取法律允许的数据取得方式,同时不侵犯个人信息法益,那么数据控制者就可以合法地拥有数据并使用数据。

其二,数据可流通性。数据可流通是指数据本身具有可流通性或者可以许可他人使用数据。合法取得并不一定能够随意地使用数据。这是因为数据上存在的利益多样性,即使合法取得的数据也并不能消灭或忽视数据上存在的合法利益(这是数据区别于物质商品的一大特征)。因此,数据使用环节仍然需要维护数据上所承载的利益主体的利益。如前所述,非个人数据除了存在商业秘密外受到的限制较少,而个人数据往往受到数据主体权利和利益的制约,并不当然地具有可流通性。个人数据是否可以流通取决于流通(许可使用)的目的是否危害数据主体权利或者是否要受主体权利控制。比如合法获取的基因数据可以用于科学研究,但不能用于包括识别个人目的或效果在内的使用(这个目的仍然受数据主体控制,需要征得同意);再如合法取得的用于特定目的的个人数据若超出个人同意范围之外,则原则上不具有可流通性,数据控制者不能自主许可他人使用。因此,当数据上存在位阶高于数据控制者的利益时,

[41] (美)詹姆斯 R.卡利瓦斯、迈克尔 R.奥利弗:《大数据商业应用》,陈婷译,人民邮电出版社 2016 年版,第 151—152 页。

就制约了数据的可流通性。这一原理同样适用于数据上存在公共利益或公共安全的情形,即使数据控制者合法取得数据甚至自己完全可以自主使用数据,国家可以基于公共利益、公共安全或国家安全而限制数据流通使用。因此,数据控制者的许可使用受到许多限制,只有在法律允许的范围之内或者在遵守法律规定前提下才能实现数据流通。

(三)数据流通利用的责任规则

数据本身是非常特殊的资源,除了数据上可能并存各种利益需要保护或协调之外,数据本身还存在安全和合规风险,数据流通必须实现流程可控和责任可追溯,否则数据流通会转嫁或扩大数据上的风险,进而使数据因流通利用失序,危害个人和社会安全。数据一旦进入流通,就会处于不断的流通中,必须清晰界定和界分每一个流通阶段数据流通主体之间的责任。否则一旦出现侵权或违法责任,就要追究流通链条上每一个当事人或者让参与其中的当事人承担相应的责任,这也是数据流通主体不愿意面对或不能承受的风险。因此,为促进数据流通,必须合理界分和分配数据流通中数据上的风险和责任。

数据流通的风险和责任主要来自两个方面:一是源自于数据控制合法性判断带来的风险;二是源自数据安全风险。由于数据并不存在清晰的产权边界,其所谓的“数据使用权”存在一定的模糊性,我们借助数据控制(使用)的合法性判断许可使用的合法性,无论从自证清白的角度,还是从相对方审查判断的角度来看,都存在一定的困难。如果细查每一项数据的合法性来源,不仅可行性存疑,而且需要花费大量时间和成本,反而会妨碍数据流通。在这种情形下,法律应当向参与数据流通的每个当事人配置合理的责任,使其在其能力范围内对自己的流通行为承担责任,并合理地切割数据流通链条中参与者相互之间的责任,避免不当的责任转嫁或连带。在这方面,笔者认为应当区分数据的侵权责任和违法责任。对于数据的获取和流通是否违反法律的强制性规定,流通双方对各自的行为独立负责,不牵连到相对人之外的前手或后手。数据提供者要对所提供数据的合法性负责,对数据合法性承担瑕疵担保责任,对数据使用者使用的合法性进行必要的判断和控制,对于超出合同约定之外的使用不承担责任;数据接受(使用)者要对数据来源合法性进行必要的慎审调查,对自己使用行为的合法性承担责任,以期合理分配当事人之间在数据控制和使用等合法性审查方面的责任。在数据收集和使用侵害他人的商业秘密和数据上的人格利益或个人信息保护利益时,应当按照侵权责任法的基本原理认定数据提供者和数据使用的侵权责任,只有在当事人知道数据获取或使用行为存在侵权事实时,才需要对流通相对人的侵权行为承担责任。同样,在数据安全责任方面,我们亦应当确立数据流通参与者独立安全责任的原则,每个主体对于各自数据的存储和流通环节中的安全承担独立的责任。在发生数据泄露等不安全事件情形下,数据提供者只有在疏于审查直接使用者资质的情形下,才承担相应的责任,而不应随意地牵连数据流通链条上的其他当事人对安全事件承担法律责任。在缺失清晰权利边界的情形下,是否侵权、是否违反法律均要对行为的正当性、合法性进行判断,其界定本身非常复杂。法律只能做出原则性规定,而详细的规则主要依赖司法实践的探索和提炼。

数据流通区别于物品流通的显著特点就在于数据上存在合法性和安全风险,在数据流通者通过流通实现数据价值的同时,要做好数据合规和风控。只有每个参与者各自严格把控风险,才可能避免合法性风险和安全风险因数据流通而扩大。强化数据流通主体的数据治理能力和数据流通风险管控能力显得非常重要。在这方面,数据安全风险评估是数据流通的主体风险防范的重要机制,在从事数据流通前应当针对拟进行的数据流通对于数据保护的影响和安全进行评估,只有达到隐私安全和信息安全标准的数据才能进入流通。同时,为做到数据流通责任的可界定和可追究,需要数据流通主体签署详尽的数据许可使用合同,全面地界定和明确双方的权利和义务,使数据流通和使用行为有章可循,同时数据流通参与者要建立完善的数据治理机制,保存数据流通和使用各环节的记录,供数据主体查询和执法、监管机构审查,使相互之间的责任可以得到界分。

五、结 论

在人类社会进入数据时代后,数据成为社会基础资源和经济活动的要素,这意味着数据需要社会化流通利用;数据流通要求赋予数据权利,以使数据流通利用成为可能。清晰地勾勒和描述数据流通有利于界定在哪些层面、对何种性质数据进行赋权。因此,本文对数据流通内涵、形式和法律制度需求的论述旨在为设计数据产权制度、构筑数据流通或数据经济的秩序提供理论基础。

数据价值最终表现为数据分析(现在多表现为智能分析),智能分析产生的知识或智慧的价值在于为人类各项活动的开展提供智力支持。数据流通利用的制度设计即是围绕该目的展开。为此,本文将数据流通定义为原始数据在不同数据控制者之间的流通,使数据不断与其他数据结合匹配成为数据智能分析原料的过程。数据流通的对象包括数据集,但不包括智能分析产生的结果(作品、知识或智慧)提供或被应用于社会活动。同时,数据流通指数据生成或生产出来之后从数据控制者(数据提供者)到数据使用者(数据需求者)之间的流动,不包括个人提供数据给数据控制者的情形(这一过程是数据分离或生成过程)。这样意义上的数据流通就是数据控制者处分数据(使用权),从而实现数据资源社会化配置和利用。而数据赋权即是明确数据控制是否享有或享有怎样权利的问题。

数据流通是向他人提供数据或使他人接触或使用数据的行为,在法律上表现为数据使用许可。一旦承认数据控制者对数据使用具有控制权,那么任何数据流通或使用就取决于数据控制者的意思。数据控制者既可以放弃对数据后续使用的控制,允许人们自由利用数据,也可以设定数据使用条件,继续控制数据的使用。在后一种情形下,数据被视为财产,需要法律界定数据控制者对数据的利用控制可以到什么程度(权利内容、效力等)。

从数据许可主体的角度来看,数据流通分为一对一许可、互为许可和一对多(众)许可,这三种许可使用大致可以涵盖所有的数据流通利用方式,只是每一类许可使用的性质、方式等内

容并不完全相同。在某种意义上,只要承认数据控制者享有许可使用的权利,数据流通利用方式取决于数据控制者的意愿,他们便可以创制丰富的数据流通方式。这里的三类划分只是一种概括。

数据流通需要解决两个问题,一是数据的可流通性问题,二是数据流通的法律基础。数据的可流通性主要存在于个人数据,因为违反数据收集时所确定的使用目的的流通需要获得数据主体的同意,这会增加数据流通的交易成本,实际上让数据流通变得不可行。为此笔者提出,去除个人数据(集)中的直接识别、显著关联个人且唯一性的识别符之后的个人数据,即使仍保留一定的可识别性,其仍然可以流通。

数据流通的法律基础实际上是要回答数据控制者是否有权许可他人使用数据的问题。为此笔者提出,数据控制者基于事实控制即可以产生这样的权利,只是法律一方面需要明确控制者对于数据的事实控制所要满足的合法性要件,另一方面要建立数据流通利用的责任规则,使数据流通责任可追溯,风险可控,从而确保数据流通利用的安全。

Abstract: The value of data is to interpret or analyze the world and provides knowledge or wisdom. Big data and artificial intelligence, however, provide a new way for human being to form knowledge or wisdom with intelligent analysis. Data sharing is necessary for data value realization in that intelligent analysis is based on data sets about one object which are collected and aggregated from fragmented data held by different controllers. Data sharing is in essence the core of data economy for achieving social utilization of data. Data sharing is implemented through data use licensing, including one to one licensing, one to mass licensing, and cross licensing. These three licenses draw the outline of social utilization of data. Data is sharable or transferable regardless personal data or non-personal data. The key issue is that the data controller's ability to use data is affirmed by law. Even without expression in law a data controller has data use right to give the data licensing based on the control fact, which will open the door of data sharing. However, it is necessary to establish the legal rules for judging the legality of data control and use as well as assuming accountability.

Key Words: Data Sharing; Data Sharability; Data Use Licensing; Data Economy

(责任编辑:王锡铨)