

· 技术 / TECHNOLOGY ·

# 数据权属和使用边界研究

顾立平<sup>1,2\*</sup>, 樊舒<sup>3</sup>

1. 中国科学院文献情报中心, 北京 100190
2. 中国科学院大学, 北京 100490
3. 武汉大学, 湖北 武汉 430072

**摘要:** 作为数据技术的新矿产, 数据产权是各个机构使用、开采、提炼、拥有、转售和出借的重要战略议题。数据产权涉及数据管理服务的主体、数据接入的合作方, 以及享受数据服务的公众等不同利益关系人, 其中核心关键是数据权属。在什么情境下如何合法合理合规地使用数据, 是近年数据政策的焦点, 也是实践上争论最多的一类问题。本文梳理欧洲、北美和亚太地区的数据政策以及实践案例, 建立数据权属和使用边界的分析框架, 包括: (1) 如何充实数据管理服务在国际国内法律法规的可论述空间的角色; (2) 如何兼顾已有商业合同中的许可协议的行为约定; (3) 如何发挥开放共享下的许可协议的可操作弹性。本研究成果可以提供长期观测国外国内国家级数据中心的政策和发展。

**关键词:** 数据产权; 数据保护; 数据管理服务

doi: 10.11871/j.issn.1674-9480.2018.03.005

## Research on Data Property Rights and Use Boundary

Gu Liping<sup>1,2</sup>, Shu Fan<sup>3</sup>

1. National Science Library, Chinese Academy of Sciences, Beijing 100190, China
2. University of the Chinese Academy of Sciences, Beijing 100049, China
3. Wuhan University, Wuhan, Hubei 430072, China

**Abstract:** As a new mineral resource of data technology, data property rights are an important strategic issue for various institutions to use, mine, extract, own, resell and lend. Data property rights involve the main body

基金项目: 大数据知识资源集成与服务的信息政策与法律法规议题 (Y17571001)

\*通讯作者: 顾立平 (gulp@mail.las.ac.cn)

of data management services, partners of data access, and the public who enjoy data services. The core is data property rights. In what context, how to use data legally and reasonably is the focus of data policy in recent years, and it is also the most controversial issue in practice. This paper reviews data policies and practical cases in Europe, North America and Asia-Pacific, and establishes an analytical framework for data property rights and use boundaries, including: (1) How to enrich the role of data management services in the space that can be discussed in international and domestic laws and regulations; (2) How to balance the behavioral agreements of existing license agreements in commercial contracts; (3) How to develop and operate flexibility the license agreement under open sharing. The results of this study can provide long-term observation of data policies and development of foreign and domestic national data centers.

**Keywords:** data property rights; data protection; data management services

## 引言

大数据时代的到来给数据保护带来了全方位的挑战，发展大数据无法避免的关键问题之一就是对于数据的保护。这一问题的妥善解决，才能为使得大数据的发展基础更为巩固<sup>[1]</sup>。

数据本身是指一种无关于形式的信息记录，是对事物的描述和记录。根据不同的需要对于数据的分类方式也各不相同。针对不同的生产主体即数据持有者来说，数据可以划分为个人数据、企业数据和公共数据等<sup>[2]</sup>。根据数据的加工处理程度或其所处的发展阶段划分，数据又可以划分为原生数据和衍生数据。原始数据是指未经加工处理的最早产生的裸数据，是不依赖于现有数据而产生的数据。而衍生数据是指原生数据被记录存储后经过算法加工、计算、聚合而形成的数据<sup>[3]</sup>。

目前，在全球信息化和知识经济大发展的背景下，大数据产业经济的发展对于数据，无论是原生数据还是衍生数据的依赖程度越来越高，掌握数据就是掌握优先权。为了推动数据共享、开放获取、科技创新，国家社会经济发展战略把投入大量财力所产生的成果，通过“互联网+”、“大数据”、“云计算”等信息技术得到加倍效果。于此，开放资源的利用在新技术服务中显得更加重要。即使，开放获取的原生数字资源，以类似计算机软件许可协议的方式，通过诸如知识共享许可协议的途经，对作品归属和使用方式有所规范，但指称对象并非程序或者数据，依然是

数字化资源的一种类型。因此，人们难以完全界定“是否有权进行数据与文本挖掘”、“并且产生衍生数据”、“进而获得数据资产”的合法合理性。其中的“双重悖论”是：数据本身不受保护，所以不能视为资产；如果不是资产，那么数据不应有所限制；反之亦然<sup>[4]</sup>。在这种情况下就产生了数据共享与知识产权保护冲突和协调问题。

针对目前的发展现状可以得出：在已经有一些开放许可协议和商业协议约定的情况下，数据不知道属于谁，由谁管理、谁能决定开不开放即归属权不明确问题，成为阻碍科技创新的最大障碍。数据权属即数据产权的归属问题，数据产权涉及不同利益关系人，其中核心关键是数据权属。在什么情境下如何合法合理合规地使用数据，是近年数据政策的焦点。明晰数据归属问题，对于提升数据使用效率，实现数据资源的合理配置，保护数据不同利益关系人的合法权益，促进开放共享具有重要的现实意义。因此，厘清数据权属关系，了解数据使用的限制和范围，建立数据权属和使用边界的分析框架是文章的主要目的。本文主要内容按照问题产生、问题成因和解决方案三个部分展开阐述。

## 1 问题产生

### 1.1 数据的流动性

新信息技术的发展和应用使得全球数据量呈现出指数级、爆发式的增长，数据的流动属性和资源属性不断增强<sup>[5]</sup>。衍生数据往往是在数据共享、机构合

作过程中将数据作为交换使用过程中产生的, 其归属问题相较原生数据更为复杂, 需要根据实际情况具体区分。而存储在数据知识库中的数据在机构之间交换使用之后、是否可以对第三方机构或和社会进行开放, 具体的开放原则或者实施方案如何制定, 这些问题的存在说明数据的流转使得数据权属分散难以界定, 需要重新进行数据确权和使用边界的规定。除此之外, 企业之间数据流转速度更快, 企业之间在合作过程中获取用户数据, 如用户交易信息、个人信息本身就具有一定的便利性, 这些数据的归属问题也是企业发展必须解决的问题之一。如何合理利用用户信息、如何有效处理科研数据、如何管理数据资源, 如何处理在市场应用中面临法律道德纠纷问题等都是数据流动日益频繁引起的与数据权属和使用边界研究有关的问题。

## 1.2 数据产权的模糊性

数据是否作为一种产权, 需要“具体情况具体分析”, 因为数据的所有权、控制权、发布权以及重使用权等, 既与众多新型的 IT 技术有关, 也与知识产权法、合同与条例有关, 同时涉及到公司、学者、律师、图书馆员、公众、政府部门、立法机构之间的相互对话。然而, 无论数据是否属于知识产权, 它们往往被视为具有知识产权的属性, 即使“开放数据”也需要经过科研人员、科研单位乃至资助机构等其他利益相关者的授权。在这种情况下, 出现了一些开放数据许可协议来帮助明确数据产权的归属问题, 国际组织制定的开放数据许可协议的类型主要包括两种: 知识共享许可协议和开放数据共用许可协议, 而开放数据共用许可包含 3 种类型, 即: 公共领域贡献与许可 (PDDL)、开放数据共享署名许可 (ODC-By) 和开放数据库许可 (ODbL)<sup>[6]</sup>。

对于科学数据的开放共享问题, 不同的研究机构都规定了相应的确权方法。如 NSF<sup>[7]</sup> 要求只有符合任何的政策规定和协议内容时, 才能给予受资助者保留自身知识产权的权利。ESRC<sup>[8]</sup> 则要求受资助者参见英国 1988 年《著作权、设计和专利法案》, 遵守所有法律条款。另对受资助研究的知识产权有两方面的

规定: 一是就受资助的研究而言, 科学数据的知识产权和责任归于研究机构 (除非有其他规定)。在特定的情况下 ESRC 有权保留知识产权。二是就与他人合作进行的研究而言, 研究机构应当以正式的协议承认所有参与研究的组织和个人的贡献和权利<sup>[9]</sup>。在研究开始时, 各方必须就此协议进行协商。对数据知识产权归属的规定, 即使对数据生产者贡献的直接承认, 又确保了数据得到最充分的利用<sup>[10]</sup>。

## 2 问题成因

### 2.1 数据产权的依据主要来自数据源的管理

在数据流转过程中涉及多个利益主体, 对数据产权进行合理界定的前提是要明确数据产权主要来自数据源的管理。数据源的管理, 是指数据管理以及数据权益管理, 包括采集、整理、清洗、转换、存储、分析、重组、备份、权益归属、许可授权、共享、开放, 销毁或者停止维护, 乃至发布发表等的质量控制与使用边界的管制措施。数据权益是指: 在合作协议下, 数据提供者和接受者之间没有披露、使用或者重制的限制来交换数据, 除非其他条款另有规定。这类协议可以是商业合同、声明承诺, 或者开放许可协议<sup>[5]</sup>。

数据源通常由创建数据的个人或者团队进行管理, 人们在使用他人的数据集、数据记录、社交网络种子文件等无法控制的数据资源时, 都需要从其他利益相关人那里获得访问权、发布权或者重使用权等。

### 2.2 数据源的管理混乱造成数据归属的不确定性

企业、单位、机构通过自身发展业务的需要, 利用相对于互联网技术上的优势, 以及直接获取终端用户的使用行为的记录的便利性等, 往往能够不经利益相关者的许可, 而获得这些数据资源, 进而绝大多数的企业甚至研究机构, 把数据资源视为数据资产, 而这种情况就使得数据使用时会遇到各种问题, 尤其是投入市场应用时所面临的法律道德纠纷问题。资产是个人或组织控制的有价值资产, 企业资产有助于实现企业的目标。数据以及由数据产生的其他衍生信息也

已经被归入企业资产的一部分。离开高质量的数据,很难有企业仍然可以高效运行。数据资产能够帮助企业正确决策,提供相应有效的辅助信息。市场领导者正利用数据资产,通过丰富的客户资料、信息创新使用和高效运营取得竞争优势。数据资产可以帮助企业深入了解市场和用户信息,掌握用户需求,提供更为优质的产品和服务,降低风险,实现企业效益的最大化。因此,随着企业的不断发展壮大,对于数据的需求和数据的依赖性不断提升,造成数据归属问题的不确定性问题也日益凸显。

目前优质数据依靠“以物易物”进行交易,短期看来,人们可以“各凭本事”、各自为政、各取所需;然而,长期来看,则不利于人们的数据资产的合法化,以及合法化交易,容易衍生知识产权方面的争议和争端,也不利于发展人工智能等的新兴科技及其应用。

### 2.3 数据源的管理混乱造成使用边界的不稳定性

以人工智能的技术发展和应用为例,人工智能需要优质数据进行智能进化,而阻碍这种进化的原因并不是数理知识不够,也不是计算资源不足,恰恰是数据使用边界的不稳定。

人工智能的主要核心技术之一是深度学习,深度学习的主要技术之一是模式识别,包括:特征向量空间、浅层特征表示、特征抽取与结构化特征等;核心问题是模式判断的准确性与复杂性。优质的数据可以训练良好的模型,也可以让良好的模型发挥实际作用,但是拙劣的数据不仅浪费计算资源,也不容易从良好模型里得到新的应用。高值数据的合理利用能够使得科学研究取得事半功倍的效果,而数据源管理混乱造成的数据使用边界不稳定问题,使得技术发展和创新极易陷入造成资源浪费、无法市场化、无法盈利、无法投入使用、无法对数据进行有效管理的现实困境之中。

## 3 解决方案

### 3.1 以“服务条款”和“用户同意”是无效的解决方案

服务条款,是用来说明、解释并且约束信息服务

提供方,在国家法律法规框架下,实现信息传递、交流、保存和管理的职责以及包括终止服务在内的信息服务体执行方式。用户同意,是协助各类用户(内容用户、系统用户、终端用户等)了解包括服务条款在内的使用规范,以及涉及到著作权限制与例外等的国家法律法规框架下的信息自由使用空间,在该服务范围内的边界。数据权属与使用边界的核心问题是数据资产,以及数据源的管理,也就是“谁有什么权利去规范人们如何利用”的范畴。但是,仅仅依靠“服务条款”和“用户同意”作为解决数据权属和使用边界问题的方案不能实现对现存风险的规避,存在巨大的潜在隐患。

相比之下,数据授权许可协议是数据开放共享的权利规范,能够为进一步发挥科学数据的巨大科学价值、经济价值、社会价值起到保驾护航的作用。数据授权许可证是一整套规范,涉及使用者被允许进行的行为活动:浏览阅读、下载存储、复制传播、内容重用、内容挖掘、信息抽取、内容汇编、重制衍生、附加许可、商业使用等<sup>[10]</sup>。目前,数据知识库所讨论的授权许可主要着重于元数据和数据的许可证协议(license),通常采用知识共享协议<sup>[11]</sup>。因此,各个权益方机构在开展数据管理的过程中,制定明确的权责机制,有利于相关工作的开展。在这一过程中需要解决数据所有权、知识产权归属等问题,包括科学伦理审核批准、数据客体权利、符合数据保护法、符合信息自由法、知识产权协议、其他保密协议等<sup>[12]</sup>。明确的权重机制有助于提高数据管理效率,在管理实践中尤为重要<sup>[10]</sup>。

### 3.2 数据产权政策是建立数据资产交易市场的关键

2018年4月26日,欧盟的官方网站上公布了一份名为《关于欧洲企业间数据共享的研究》(Study on data sharing between companies in Europe)的报告<sup>[13]</sup>。根据该报告所述,欧盟委员会致力于发展欧洲的数据经济。欧盟一直致力于建设欧洲数据经济,自2014年起,便采取一系列相关政策关注数据在促进经济增长、创造就业等方面的作用。为进一步建立一个政策框架以促进欧洲企业间的数据共享,欧盟委员会联合

everis<sup>[14]</sup>开展了一项研究来评估欧洲经济区内企业间信息共享的经济价值与潜力。这项研究从2017年7月开始,直到2018年2月完成,研究在欧洲31个国家内的129家不同规模的、分布于数据生成驾驶、智能农业、智能制造、电信运营商、智能生活环境到智能电网和计量设备六大行业的公司进行了调查访问<sup>[13]</sup>。通过调查显示造成数据共享困难的主要障碍包括“数据所有权”的不确定性和数据访问的不确定性以及数据位置的不合理限制和数据的自由流动。因此,充分理解数据合法使用的法律框架,了解管理数据收集和治理的各种法律规则以及遵守相应的许可协议和数据产权政策,积极向数据使用者传递这种理念,保持双方信任关系的稳定,才能在数据资产交易市场中占据有利地位<sup>[13]</sup>。

欧盟《一般数据保护条例》(GDPR)于2018年5月25日起正式施行,该条例是近三十年来数据保护立法的最大变动,旨在加强对欧盟境内居民的个人数据和隐私保护<sup>[15]</sup>。此外,它还将通过统一数据和隐私条例来简化对跨国企业的监管框架。根据“使用他人的数据集或者数据记录需要经过授权许可”的这个道理,欧盟GDPR立法的主要目的,主要为了保护终端用户的权利,并且制衡全球性互联网企业收割本地终端用户信息并且作为数据资产的能力,我国企业和科研团队今后容易遭遇市场进入障碍或者巨额罚款。

我国的网络安全法(2017年)和数据管理办法(2018年)也有类似GDPR的条文,但是属于指导性政策,而不是针对具体领域进行管制,我国企业、科研单位、高校以及其他利益相关者,应该共同努力,探索如何保护各自应有的数据产权,并且率先建立数据交易市场。

### 3.3 呼吁以“数据产权政策”解决具体问题

数据权益的归属一直是科学数据管理实践中的重要问题之一。明确各方机构所拥有的数据优先处理权,便于各个机构开展数据管理实践,能够有效避免其他利益相关方的数据权益冲突,提高管理效率<sup>[10,12]</sup>。

尽管在实践中,对于数据权属的问题通常采用许可协议进行界定,部分许可协议清晰地界定了数据

使用的条件、限制期间以及数据再利用的目的,除了提供法律保护,这些许可协议也是确保数据使用者间信任的工具。但是,在大多情况下,虽然存在模板协议,但是合同需要根据每个数据使用者进行具体调整。这些细致的协议通常会导致一个耗时、耗财、冗长和复杂的过程。具体案例中,适用许可协议的不确定性,以及为了每个客户而准备、调整许可协议的过程繁琐且成本高昂<sup>[13]</sup>。因此,良好的数据产权政策可以完善我国的数据治理。

随着我国的网络安全法(2017年)和数据管理办法(2018年)的颁布,留给了更多可以参与、完善、和优化实施方案的空间。因此,在相关问题上,人们需要寻找各个利益相关方的代表,组织圆桌会议,也需要邀请具有情报分析能力和跨领域的专家学者们,从不同角度提出问题 and 解决方案,共同努力。

## 4 结语

数据权属问题涉及技术、经济、法律、安全等多个方面,数据确权具有一定的复杂性。但是,对于数据权属和使用边界问题的关注,提供有效的政策分析框架,最终督促建立合理有效的数据产权政策时解决问题最直接的方案。数据产权政策的建立和完善有利于实现数据资源的合法保护和合理利用,提升数据管理能力,为更大范围的开放共享奠定基础,推动科技创新,加快创新型国家的建立。对于我国来讲,参考国外的相关政策研究,结合实际情况出台具体有效的数据产权政策,形成多元化治理体系是目前亟待解决的现实问题,需要各方共同努力与协作。

## 参考文献

- [1] 王融.大数据时代:数据保护与流动规则.北京:人民邮电出版社.2017.12-13.
- [2] 崔丽莎.大数据权属及保护制度研究 [EB/OL].[2018-09-05]. [http://www.sohu.com/a/203070639\\_65507](http://www.sohu.com/a/203070639_65507).
- [3] 王融.无处安放的数据权属 [EB/OL].[2018-09-05]. [www.199it.com/archives/532779.html](http://www.199it.com/archives/532779.html).

- [4] 顾立平.数据权益:跨越开放获取的边界.2016. 2015, 59(5): 53-59.
- [5] 付伟,于长钺.数据权属国内外研究述评与发展动态分析[J].现代情报,2017,37(7):159-165. [13] 《欧洲企业数据共享研究报告 (Study on Data Sharing Between Companies in Europe)》.[EB/OL]. [2018-09-05]. <https://publications.europa.eu/en/publication-detail/-/publication/004fcf02-49c7-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-69800191>.
- [6] 黄如花,李楠.开放数据的许可协议类型研究[J].图书馆, 2016(8):16-21. [14] ereris. About us [EB/OL]. [2018-09-05]. <https://www.everis.com/global/en/about-us>.
- [7] NSF. Dissemination and sharing of research results [EB/OL]. [2018-09-05]. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>. [15] 《欧盟数据保护通用条例 (General Data Protection Regulations, GDPR)》.[EB/OL]. [2018-09-05]. <https://www.gdpr.net/>.
- [8] ESRC. Research Data Policy [EB/OL]. [2018-09-05]. [http://www.esrc.ac.uk/\\_images/Research\\_Data\\_Policy\\_2010\\_tcm8-4595.pdf](http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf).
- [9] 张瑶,顾立平,杨云秀,等.国外科研资助机构数据政策的调研与分析——以英美研究理事会为例[J]. 图书情报工作,2015,59(6):53-60. 收稿日期: 2018年3月20日
- [10] 顾立平.科学数据开放获取的政策研究[M].北京: 科学技术文献出版社,2016. 顾立平: 中科院文献情报中心, 副研究员, 博士, 硕士生导师, 主要研究方向为数据政策。 E-mail: [gulp@mail.las.ac.cn](mailto:gulp@mail.las.ac.cn).
- [11] 孙铁楠,顾立平,宋秀芳,等.学科数据知识库的政策调研与分析——以生命科学领域为例[J].现代图书情报技术, 2015,31(12):13-20. 樊舒: 武汉大学信息管理学院, 硕士研究生, 主要研究方向为信息检索和用户信息行为。 E-mail: [530945415@qq.com](mailto:530945415@qq.com).
- [12] 杨云秀,顾立平,张瑶,等.国外科研教育机构数据政策的调研与分析——以英国10所高校为例[J]. 图书情报工作,