

# 大数据有限排他权的基础理论

崔国斌\*

---

**内容提要：**人类社会进入大数据和人工智能时代，大规模数据集合的知识产权保护问题日益突出。诞生于“小数据”时代的知识产权法满足了数据产业的基本需求，但还是留下了一些空白。处在公开状态的没有独创性的大数据集合缺乏具体的法律保护手段。为了避免这一领域的市场失败，同时避免限制公共领域的行动自由，应当为耗费实质投入并达到实质规模的大数据集合设置有限排他权，即公开传播权。这一保护机制既能够满足数据行业的需求，又能够兼顾后续数据利用者的利益，不会损害著作权法等法律所维护的公共政策。

**关键词：**大数据 知识产权 公共领域 公开传播权

---

## 引言

人类社会进入大数据和人工智能时代，数据逐渐被视为重要性超过石油或土地的战略资源，是文化生活、企业竞争、科技创新和民主对话的基础。有效规范大规模数据（大数据）的收集、加工、传输、存储、使用、分享和转让等行为，明确数字经济参与方的权利和义务，对于保证数据资源的充分供给和有效利用至关重要。在这一方面，现有知识产权制度存在空白地带。法院依靠反不正当竞争法中模糊的“商业道德”保护条款来平衡数据市场主体的重大利益，不再让人满意。社会期待更清晰的理论指引和更明确的产权规则。在设计更具体的大数据或大数据集合产权规则时，决策者需要回答一系列问题。比如，大数据集合保护的正当性基础究竟是保护基于劳动的自然权利、避免市场失败还是维护商业道德，保护大数据集合是否会损害公共领域的行动自由，数据规模达到何种程度才能够获得保护，究竟应该设置何种权利内容与权利限制，采用何种立法模式，如何处理大数据产权与其他用户权利的冲突，等等。笔者试图围绕其中的重点问题，为此类新型的产权保护

---

\* 清华大学法学院副教授。

本文为清华大学自主科研计划课题（2015THZWYY01）的成果之一。

提供一套相对完整且具有解释力的基础理论,为决策者和后续研究人员提供指引。本文并非行业实证研究,对后续可能的实证研究纠正本文提出的制度设计方案,持开放态度。

“数据”这一概念掩盖了大数据知识产权问题的复杂性和多样性。在不同的背景下,“数据”含义其实并不相同。在科学意义上,数据含义比较窄,它是指记录下来事实,是以数字或文字等形式呈现的客观实体属性的值。<sup>[1]</sup>而在商业活动中,数据的含义被大大拓宽,既包括传统意义上的科学数据、个人信息、商业统计数据等,也包括那些已经成为知识产权法保护客体的内容,比如商业秘密(营业秘密、技术秘密)、作品(文字作品、美术作品与视听作品等)、录音录像制品,等等。<sup>[2]</sup>在讨论大数据集合的知识产权问题时,笔者选择忽略单个数据条目本身的内容和属性,即无论它是科学数据、个人信息、交易数据还是文字作品、图片或视听作品,也不论它体现了用户、商家还是网络平台的知识产权、隐私权、肖像权等权益,均将它视为一个内容中性的数据黑点。每个黑点本身内部内容的保护和利用所引发的问题,由其他法律来处理。只有一个企业收集了海量的黑点形成数据集合之后,才有大数据知识产权保护方面的问题。

即便是大数据集合的知识产权保护,在商业秘密、著作权保护机制之外,空白地带也仅仅是处于公开状态的整体上没有独创性的大规模数据集合的保护问题。此类大数据集合与《欧盟数据库指令》中的“数据库”(database)有些重叠,但并不完全相同。该指令中的“数据库”特指“独立作品、数据或其他材料经系统地或有序地排列后的集合”。<sup>[3]</sup>这一定义并未对数据库的规模提出明确要求,并且暗示数据的编排须满足一定的结构化要求。而本文并不强调大数据集合本身的结构化,但是对大数据集合的数据量有实质要求。如后文所述,这一数据规模通常远远超出普通作品的数据库规模,很可能也超出《欧盟数据库指令》意义上大部分数据库的规模。因此,这里的大数据集合与《欧盟数据库指令》意义上的“数据库”不完全重叠。为了避免不必要的误解,本文没有使用“数据库”这一术语来表述大规模的数据集合。

## 一、知识产权制度保护大数据的有限空白

在讨论大数据集合的知识产权问题时,如果我们只是满足于宽泛的“数据”概念的新奇感,不了解现有知识产权制度提供的保护可能性,则很容易产生意外发现全新领域因而可以随意挥洒的错觉。<sup>[4]</sup>在民法总则立法过程中,曾经出现过将“数据信息”与商业秘密、作品等并列,当作全新知识产权客体的意见。<sup>[5]</sup>后来,立法者又将它从知识产权客体

[1] 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;《科学数据管理办法》(国办发〔2018〕17号)第2条。

[2] 参见方巍等:《大数据:概念、技术及应用研究综述》,《南京信息工程大学学报(自然科学版)》2014年第5期,第406页。

[3] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases, Art. 1.

[4] 比如,过去就有意见认为数据库尚未有知识产权保护,是全新领域。See James Gibson, *Re-reifying Data*, 80 Notre Dame L. Rev. 163, 169 (2004).

[5] 参见《中华人民共和国民法总则(草案)》(2016)第108条,载全国人大网 [http://www.npc.gov.cn/zgrdw/npc/lftz/rlyw/2016-07/05/content\\_1993427.htm](http://www.npc.gov.cn/zgrdw/npc/lftz/rlyw/2016-07/05/content_1993427.htm), 2019年9月3日访问。

中移除，单独规定。<sup>〔6〕</sup>实际上，数据信息既可以是商业秘密，也可以是作品，也可能二者皆非。民法总则的起草者实际上并没有理顺数据信息与现有知识产权客体的交叉重叠关系，进退失据。

很多学者在讨论大数据的产权问题时，忽略现有知识产权框架，尤其是商业秘密、著作权等机制对各类数据的保护，提出非常宽泛的数据财产权保护方案，并忽略数据与现有知识产权的交叉重叠问题。<sup>〔7〕</sup>当年欧洲立法者在制定其数据库指令时，就很可能对已有知识产权法律、合同和技术保护措施的有效性认识不足，从而制定出了过度保护的特别权利规则。<sup>〔8〕</sup>

为了避免重蹈欧盟的覆辙，我们有必要先了解大数据保护领域最为重要也非常成熟的商业秘密保护、著作权法、反不正当竞争法的实践，准确识别出现有制度之外可能的空白地带，为后续的讨论奠定基础。

### （一）作为商业秘密的大数据集

我国没有专门的商业秘密保护立法，而是在反不正当竞争法的框架下保护商业秘密。与该法配套的最高人民法院的司法解释已相当完善。学界对于商业秘密是否属于财产性权利（绝对权）一直存在争议，但大多持肯定意见。<sup>〔9〕</sup>美国最高法院也认为商业秘密是宪法意义上的财产性权利。<sup>〔10〕</sup>笔者接受主流意见，将商业秘密保护视为一种相对独立的财产性权利的保护机制，与后文提到的著作权保护机制并列。确认这一点，对于理解后文关于数据排他权属性的讨论有重要意义。

大数据集合之所以容易作为商业秘密来保护，有很多原因。首先，大数据集合所包含的单个数据本身常常就是秘密数据。比如，物联网上各种私人设备产生的单个用户数据、医院产生的个人病历数据、电子商务平台后台生成的单个用户的交易数据、物流和航空公司用户的个人行程数据等，都因为收集者采取了保密措施而使公众无法通过公开渠道获得。数据收集者之间的数据交换通常都是通过保密渠道进行。单个用户的数据具有秘密性，这些数据的集合自然也具有秘密性。司法实践中，早就有法院沿着这一思路认定此类数据集构成商业秘密。<sup>〔11〕</sup>

其次，即便单个数据条目不具有秘密性，大数据集合作为一个整体，最初大多存储在私人空间中，为保密措施所覆盖。公共领域不存在同样的大数据集合。第三方可能可以手工或自动（网络爬虫）收集数据条目以制作大数据集合，但是这一过程耗时费力，并非易事。在原始的数据收集者采取技术措施（比如IP地址屏蔽）阻止第三方反复访问自己网站并下载公开的数据条目时，第三方收集完成大数据集合的过程就更加困难。因此，该大数

〔6〕 民法总则第127条规定：“法律对数据、网络虚拟财产的保护有规定的，依照其规定。”

〔7〕 参见龙卫球：《再论企业数据保护的财产权化路径》，《东方法学》2018年第3期，第59页以下；程啸：《论大数据时代的个人数据权利》，《中国社会科学》2018年第3期，第121页以下。

〔8〕 See Miriam Bitton, *Exploring European Union Copyright Policy through the Lens of the Database Directive*, 23 Berkeley Tech. L. J. 1411, 1448 (2008).

〔9〕 参见蒋志培等：《〈关于审理不正当竞争民事案件应用法律若干问题的解释〉的理解与适用》，《法律适用》2007年第3期，第26页；Jacqueline Lipton, *Balancing Private Rights and Public Policies: Reconceptualizing Property in Databases*, 18 Berkeley Tech. L. J. 773, 817 (2003).

〔10〕 See *Ruckelshaus v. Monsanto Co.*, 467 U. S. 986, 1004 (1983).

〔11〕 比如，周慧民与衢州万联网络技术有限公司侵害商业秘密纠纷上诉案，上海市高级人民法院（2011）沪高民三（知）终字第100号民事判决书。

据集合整体上很可能“不为公众所知悉”，也不属于“无需付出一定的代价而容易获得”的信息，<sup>[12]</sup>依然可以作为商业秘密得到保护。这使得收集者能够阻止他人从整体上接触和利用自己采取保密措施的大数据集合。

国内有学者对此提出否定意见：“对于现实中大多数的数据信息而言，信息制作者采集的信息本身大多来自公有领域，是任何人均可以从公开渠道直接获取的，显然，将各地为公众所知的信息汇编之后形成的成果认定为具有秘密性是荒谬的。”<sup>[13]</sup>这一意见在国内有一定的代表性。可能是受这一思路指引，国内诸多大数据权利人原本可以主张商业秘密保护，却舍近求远，寻求反不正当竞争法第2条的保护。<sup>[14]</sup>上述意见实际上忽略了数据集合与它所包含的数据条目自身的秘密性的差别。单个数据条目来源于公共领域，并不妨碍大数据集合本身取得商业秘密法上的秘密性。只要组合信息作为一个整体，无法轻易从公共领域中获得，则该组合信息依然可能具备所谓的秘密性。<sup>[15]</sup>比如，汇集众多客户的客户名册，即便其中的单个客户的名单和联系信息存在于公共领域，该客户名册整体依然可能构成商业秘密。<sup>[16]</sup>

最后，大数据集合很多时候并非现有分散的数据条目的简单汇总，而是经过数据收集者的补充、筛选、格式调整、数据删节、加密等措施，这使得大数据集合中数据条目信息的存在状态与公共领域的分散数据形态有很大差别，更加符合商业秘密保护的秘密性要求。比如，很多医疗机构会对医疗数据进行匿名化处理，并选择按照商业秘密机制来保护这些数据并对外发放许可。

以上分析表明，大数据集合能够获得有效的商业秘密保护，数据行业的实践也证明了这一点。欧盟的调研报告表明，数据集合的确大多处于私人控制之下，共享存在诸多困难。<sup>[17]</sup>这从反面表明，商业秘密机制是大数据集合保护和交易的最为重要的制度基础。

当然，现有商业秘密机制能够有效地保护处于秘密状态的大数据集合，并不意味着它刚好充分考虑了大数据时代各种数据集合背后的公众或后续数据利用者的利益，是一种完美的保护机制。实际上，数据类型千差万别，商业秘密保护既可能不足，也可能过度。现在的确有不少研究人员担心，商业秘密保护导致很多数据集合很难为公众获取或接触，认为有必要建立数据公开的激励机制，甚至考虑采取措施限制数据收集者获得商业秘密保护。比如，有人建议强制要求医疗机构向公共机构提供匿名化的医疗数据，由政府建立公共的数据集合供自由获取，以方便后续的医学研究。<sup>[18]</sup>美国国会也于2007年立法要求医药厂商

[12] 参见《最高人民法院关于审理不正当竞争民事案件应用法律若干问题的解释》（法释〔2007〕2号）第9条、第11条。

[13] 参见芮文彪等：《数据信息的知识产权保护模式探析》，《电子知识产权》2015年第4期，第97页。

[14] 比如，淘宝（中国）软件有限公司诉安徽美景信息科技有限公司不正当竞争纠纷案，浙江省杭州市中级人民法院（2018）浙01民终7312号民事判决书；北京微梦创科网络技术有限公司诉北京淘友天下技术有限公司不正当竞争纠纷案，北京知识产权法院（2016）京73民终588号民事判决书。

[15] See Vermont Microsystems, Inc. v. Autodesk, Inc., 88 F.3d 142 (2d Cir. 1996); Catalyst & Chem. Servs., Inc. v. Global Broung Support, 350 F. Supp. 2d 1 (D. C. 2004).

[16] 参见《最高人民法院关于审理不正当竞争民事案件应用法律若干问题的解释》（2007）第13条。

[17] See European Commission Study on Data Sharing Between Companies in Europe: Final Report, European Commission, Apr. 24, 2018, available at <https://publications.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en>, last visited on May 1, 2019.

[18] See Michael Mattiolid, *Disclosing Big Data*, 99 Minn. L. Rev. 535, 577-579 (2014).

公开部分临床试验数据（临床一期的数据除外）。<sup>[19]</sup> 欧洲航空监管部门也一直要求空客公司共享部分数据。<sup>[20]</sup> 在大数据时代，立法者确实有可能需要对大数据集合的秘密性要求和保护力度等作适当的调整，限制（比如强制许可）甚至剥夺（比如强制披露）部分大数据集合的商业秘密保护，以实现更精细的利益平衡。不过，进一步的讨论已经超出本文的范围。

对数据收集者而言，如果自己的商业模式要求它对无数用户提供数据，则商业秘密保护机制的效果将大打折扣。首先，数据收集者要采取有效的保密措施，就必须与所有接触者签署保密协议。这会产生很高的交易成本。其次，在数据接触者众多的情况下，收集者即便可以让每个人都签署保密协议，也很难有效监督他们以保证保密协议得到有效执行。<sup>[21]</sup> 最后，在某些情况下，如果数据收集者向太多人提供数据，即便所有人都接受保密协议，法院也有可能认定该数据事实上丧失秘密性。<sup>[22]</sup> 如果数据收集者的商业模式要求它无限制地向所有公众开放提供数据，对于这部分的公开数据，商业秘密保护制度将彻底失去用武之地。

## （二）作为作品的集合或大数据集合

除了商业秘密保护外，著作权法对于大数据集合的保护也很重要。著作权法不保护不具备独创性的单纯事实或数据。对它们进行简单加工或汇总，也不能使大数据集合获得独创性。但是，如果数据收集者对于数据的选择、整理和编排作出独创性贡献，则能够从整体上对该大数据集合享有类似汇编作品的著作权，可以禁止他人从整体上复制该数据集合。在公众希望获得体现了收集者独创性的大数据集合而不是单纯数据本身时，著作权保护能够阻止他人的复制和传播行为。如果公众希望得到的是不具有独创性的数据本身而不是数据选择或编排的结果，则著作权法无法阻止他人单纯复制该大数据集合中的数据并对外提供的行为。

在没有作出独创性贡献时，数据收集者不能对大数据集合寻求著作权保护，这似乎意味着著作权法与大数据集合的保护不再相关。实际上并非如此。如果数据条日本身构成作品，数据收集者依然可能以迂回的方式从著作权保护中获益。比如，社交网站用户撰写的文字、拍摄的照片或视频内容等，很可能构成著作权法意义上的作品或录像制品而获得保护。数据收集者可以先从用户那里获得独占性或排他性授权，然后再利用该授权阻止他人传播自己的数据集合。大众点评网就这么做过，也的确获得了有的法院的支持。<sup>[23]</sup>

在大数据集合中的数据条日本身构成具有独创性的作品或表达时，我们将这些“作品或表达”视为大数据集合中的一个“黑点”，忽略这个黑点的内部内容或结构，这时无数

[19] See Jeffrey M. Drazen, Stephen Morrissey & Gregory D. Curfman, *Open Clinical Trials*, 357 *New Eng. J. Med.* 1756, 1756 (2007).

[20] 参见前引 [17]，欧盟调研报告，第 66 页。

[21] See Kristen Osenga, *Information May Want to Be Free, But Information Products Do Not: Protecting and Facilitating Transactions in Information Products*, 30 *Cardozo L. Rev.* 2099, 2117 (2009).

[22] 参见前引 [9]，Lipton 文，第 818 页。

[23] 参见上海汉涛信息咨询有限公司诉爱帮聚信（北京）科技有限公司著作权侵权纠纷案，北京市海淀区人民法院（2010）海民初字第 4253 号民事判决书。新浪微博也曾经修改《微博用户服务协议》第 1.3 条和第 1.5 条，禁止用户授权第三方使用微博内容。参见新浪微博管理员关于微博内容版权归属的说明，<https://www.weibo.com/1934183965/FlZJcql1bL? type = comment>。这一努力引发争议，最终新浪改变了措辞，但依然限制用户授权第三方抓取数据。最新版的协议参见 <https://weibo.com/signup/v5/protocol>，2019 年 9 月 3 日访问。

“黑点”的集合与狭义或科学意义上的数据的集合并无本质差别。在大数据产权意义上，我们关注的是数据收集者为收集这些数据“黑点”而付出成本，而非“黑点”本身的创作成本。比如，Youtube网站所收集的用户创作的视频集合可能被视为广义的大数据集合，但该集合中的用户视频并非狭义的数据，而很可能是受保护的作品。在讨论大数据集合产权保护时，我们不关心每个用户为创作每个视频所付出的成本，而只是关注Youtube网站收集这些视频所付出的成本。为了避免不必要的误解，再次说明一下，后文在讨论问题时没有刻意区分狭义的非作品类数据和上述作品类数据，但默认的大多是前者。在某些情形下，相关讨论可能要适当变通才能适用于用户创作的作品类数据的集合。

此外，著作权法保护的技术措施对于数据收集者也有重要意义。著作权法禁止他人未经许可破坏权利人为保护著作权而采取的技术措施或提供用于破解该技术措施的专门工具。在大数据集合中的单个数据或数据集合整体上构成作品的情况下，如果收集者采用技术措施保护该数据集合，则该技术措施受著作权法保护，这没有疑问。在大数据集合整体和局部均不构成作品的情况下，数据收集者保护该数据集合的技术措施，则很可能不受著作权法保护。这是因为我国著作权法规定的受保护的技术措施应该是为保护著作权而采取的技术措施，在大数据集合整体和局部均不构成作品或邻接权客体的情况下，保护它的技术措施并非著作权法意义上的技术措施。不过，数据收集者可以轻松消除这一不确定性——它只要将不受著作权法保护的数据和部分受著作权法保护的内容混在一起，再采取技术保护措施，就可以利用著作权法阻止他人对该技术措施的破解。<sup>[24]</sup> 著作权法降低了数据收集者的技术措施被破解的风险，在一定程度上强化了收集者对大数据集合本身的控制。

### （三）有限空白地带的产权保护建议

认真分析关于商业秘密和著作权的现有保护规则之后，可以发现这一领域并不像想象的那样是一片全新的法律领域，相反，它早已被现有知识产权法覆盖得差不多，宽泛的大数据集合立法并无栖身之所。不过，有限的空白地带的确还存在，那就是处于公开状态的非独创性大规模数据集合的保护问题。此类数据集合既不能作为商业秘密得到保护，也不能作为著作权法上的作品受到保护。

这一空白并不是最近才被揭示出来，相反，在数十年前就已经是知识产权法领域众所周知的存在。比如，始于1960年代的北欧国家版权法的“目录规则”（The Nordic Catalogue Rule）被认为是最早的对不具独创性的公开数据集合进行保护的法律尝试。<sup>[25]</sup> 随后，欧共体于1988年在公开的政策文件中讨论此类数据库的知识产权保护立场，并于1992年发布了数据库保护的建议草案。<sup>[26]</sup> 欧洲最终于1996年通过了其数据库指令，确立版权和特殊权利并行的立法模式。1996年世界知识产权组织也组织外交会议讨论《数据库条约草案》（The WIPO Draft Database Treaty of 1996）。同一时期，美国国会也有过轰轰烈烈的数据库特殊保护的立法尝试，提出过多个版本的数据库特殊保护立法草案。不过，美国立法草案最

[24] See Pamela Samuelson, *Mapping the Digital Public Domain: Threats and Opportunities*, 66 *Law & Contemp. Probs.* 147, 160 (2003).

[25] See J. H. Reichman, *Legal Hybrids Between the Patent and Copyright Paradigms*, 94 *Colum. L. Rev.* 2432, 2492-2493 (1994).

[26] See F. W. Grosheide, *Database Protection-The European Way*, 8 *Wash. U. J. L. & Pol'y* 39, 47 (2002).

终因为科学界和网络行业的强烈反对而被搁置。<sup>[27]</sup>现在,美国依然只是在司法创设的“热点新闻学说”(hot news doctrine)的框架下,对不受版权保护的实时事实消息提供非常有限的保护。<sup>[28]</sup>20世纪90年代世界范围内所热议的数据库与今天讨论的大数据集合在结构上和规模上可能有很大差别,但二者所折射出的制度空白却没有本质差别。现在学者们对大数据产权问题的讨论,在广度和深度上实质超越1990年代经典论文的并不多见。<sup>[29]</sup>

像美国一样,我国知识产权界过去也差不多与世界同步讨论数据库专门立法问题。<sup>[30]</sup>不过,这同样也没有促成中国对没有独创性的公开数据集合进行专门立法。从1990年代开始,法院先是适用民法通则的一般条款(比如广西电视节目表案),<sup>[31]</sup>然后更多地利用反不正当竞争法的一般条款(第2条)保护公开的数据集合。比如,北京阳光数据公司诉上海霸才公司案、上海汉涛公司诉爱帮网案、上海钢联诉纵横公司案、上海汉涛公司诉百度地图案、新浪微博案(淘友诉微梦)、淘宝诉美景案,等等。在这些案例中,数据收集者大多耗费一定的资源,却并没有作出独创性贡献,因此无法获得著作权保护。同时,这些数据收集者选择了直接向大众提供数据内容的商业模式,法院通常认为此类数据本身或数据集合无法作为商业秘密得到保护。

利用反不正当竞争法的原则条款保护处在空白地带的大数据集合,法院必须个案判决,有很大的不确定性。比如,美国波斯纳法官反对利用非法侵占学说(misappropriation)保护数据集合,认为这一反不正当竞争法上的学说过于模糊,缺乏清晰的界限,即便在很少情形下适用,也应该有联邦立法作为基础。<sup>[32]</sup>我国反不正当竞争法第2条的“商业道德”原则条款同样非常模糊,对法院确定具体的保护客体范围、禁止行为类型、可能的权利限制、保护期限等,均无法提供明确的指引。我们不能总是指望法院在重要的数据案例中依据这一条款实现精细的利益平衡。在大数据和人工智能时代,数据所涉及的财产利益的重要性急剧提升,投资者对大数据集合的确定性期待也随之增加。现在应该是对大数据集合的产权保护模式给出明确答案的时候了。

为了填补上述制度空白,笔者建议先对大数据集合提供有限排他权保护,即阻止他人未经许可向公众传播收集者付出实质性投入收集的实质数量的数据内容的权利。这里的数据数量和实质投入要求与传统作品相比有数量级上的差别。这里的有限排他权大致包含著作权法上的发行权、广播权、信息网络传播权等权利(以下简称“公开传播权”)。<sup>[33]</sup>除了公开传播权外,大数据集合的权利人不享有复制权、演绎权或其他权利。从现有的大数据

[27] 参见前引[24], Samuelson文,第159页。

[28] See Int'l News Serv. v. AP, 248 U. S. 215 (1918).

[29] See J. H. Reichman & Pamela Samuelson, *Intellectual Property Rights in Data*, 50 Vand. L. Rev. 51, 73 (1997).

[30] 参见郑胜利、崔国斌:《数据库保护的立法现状与基础理论》,载郑胜利主编:《北大知识产权评论》第1卷,法律出版社2002年版,第94页以下。

[31] 参见广西广播电视报社与广西煤矿工人报社电视节目预告表使用纠纷案民事判决书,载最高人民法院知识产权审判庭编:《最新知识产权司法文件精选(一)》,中国标准出版社1999年版,第569页以下。

[32] See Richard A. Posner, *Misappropriation: A Dirge*, 40 Houston Law Review 621, 638 (2003-2004).

[33] 公开传播权并无准确定义。这里仅仅为方便起见采用较广义上的表述,用以指代上述各项权利。其中最重要的应该是信息网络传播权,广播权可能只是在特殊场景下才有价值。有些学者在更广范围上使用公开传播权这一概念,甚至覆盖朗诵、广播、展览、出租等行为。参见郑成思:《知识产权法》,法律出版社2003年版,第351页。也有学者在更窄意义上使用,排除了发行权等。参见陈绍玲:《论著作权法中的公开传播权》,《华东政法大学学报》2015年第2期,第41页。

集合产权保护的案例看,绝大多数收集者通过诉讼试图阻止的都是他人的公开传播数据的行为,而非单纯的数据复制或其他数据利用行为。如后文所述,限制复制行为很可能过度限制后续的数据利用行为,过度增加社会成本。现阶段,我们对相关数据行业的发展规律的了解还很有限。谨慎起见,先仅保护公开传播权是合理的选择。

立法者将来应进一步明确大数据保护的客体范围、权利内容、权利归属、保护期限、权利限制、与其他权利的衔接等规则。本文重点关注大数据集合有限排他权的基础理论,限于篇幅,无法对这些制度设计层面的问题展开深入讨论。

## 二、大数据有限排他权与公共领域的利用自由

笔者建议的大数据集合有限排他权,虽然只是最低限度的保护,但依然是对著作权法所赋予的公共领域行动自由的限制。众所周知,著作权法有所谓思想与表达二分原则,即著作权法保护作品中的独创性表达,但排除狭义的数据、事实、抽象思想和其他非表达性要素。<sup>[34]</sup>虽然学者们对这一原则的正当性有不同的理解,但很少直接质疑它的必要性。对大数据集合的非独创性内容提供额外的产权保护,不能回避的问题是,它是否实质损害这一原则所维护的公共政策。透过这一分析,我们也能更好地理解大数据集合产权保护的合理的客体范围与权利边界。

### (一) 著作权法排除数据的原因

著作权法思想与表达二分原则排除数据或事实的原因,大致包含以下几个方面:

首先,著作权法通常将事实和数字,比如各种历史事实、科学实验数据、电话号码信息等等,视为作者发现的某种客观事实,而不是作者的创作内容。<sup>[35]</sup>既然不是作者创作的,按照著作权法的逻辑,自然不应该被作者垄断。如果大数据集合的数据条目是用户创作的作品,那也是用户的贡献,而非数据收集者的创作。

其次,在数据规模较小时(即“小数据”时代),著作权法仅仅保护作品中具有独创性的内容,通常能够提供充分的创作激励。毕竟,绝大多数作品都是作者在收集了非独创性事实和数字的基础之上,添加更多的独创性或个性化内容之后才完成的。后来者虽然可以复制事实和数字,但并不能直接复制著作权人的个性化表达,依然需要先付出相当的努力或资源去创作独创性的表达,然后才能与权利人竞争。因此,单纯复制在先作品中的事实和数字,并不能够让后来者获得太多的竞争优势。<sup>[36]</sup>在个性化表达和数字混在一起时,保护作品中的独创性表达,不保护数字或事实,通常足以保护著作权人的创作积极性。因此,在著作权法之外,我们并不需要额外机制鼓励著作权人公开作品中包含的事实和数字。

再次,狭义的数据或事实是作品的表达基础(building blocks)。<sup>[37]</sup>个性化表达通常建立在数据、事实或技术信息引用的基础之上。禁止利用公共领域的数字或事实(非商业秘

[34] See TRIPs Article 9 (2) (1995); Harper & Row, Publishers, Inc. v. Nation Enterprises, 471 U. S. 539, 556 (1985).

[35] See Feist Publications, Inc. v. Rural Telephone Service Co., 499 U. S. 340, 345-350 (1991).

[36] 参见前引[32], Posner文,第636页。

[37] See Jane C. Ginsburg, *Copyright, Common Law, and Sui Generis Protection of Databases in the United States and Abroad*, 66 U. Cin. L. Rev. 151, 154 (1997).

密), 会显著增加后来者的表达成本, 损害其创作积极性, 妨碍艺术和科技进步。<sup>[38]</sup> 具体而言, 其一, 后来者需要为自己作品中所要利用的数据或事实信息寻求许可或者独立收集该数据, 这将显著增加交易成本或创作成本。这类成本很可能使得部分后来者放弃利用该数据创作有价值的作品。在数据来源单一时, 后来者的单独收集或获取许可的成本可能更加高昂, 问题会更加突出。其二, 在先作品中公开的具体数据信息常常缺乏个性特征, 后来者很难识别哪些处于公共领域, 哪些受产权保护。如果法律禁止利用在先作品中的数据信息, 后来者会担心动辄得咎, 预防侵权的成本会显著增加。为了避免出现上述负面现象, 排除事实和数据的保护, 就成为著作权法的自然选择。

最后, 在数据信息公开之后提供额外的产权保护, 还存在证明抄袭的信息成本问题。著作权法承认独立创作例外, 即被控侵权人如果独立创作(没有抄袭), 则即便作品内容相似, 也不构成侵权。<sup>[39]</sup> 事实和数据常常缺乏个性化特征, 处在公开状态可以自由接触, 权利人要证明他人抄袭该事实信息而不是独立发现该事实信息, 难度很大。理论上, 立法者可以放弃独立创作或独立收集抗辩, 即原告无需证明被告存在抄袭行为, 只要数据信息相同就构成侵权, 从而降低原告维权的难度。这类似于专利法保护技术方案的模式。不过, 这意味着社会要建立起事先审查数据信息新颖性甚至创造性的机制, 告知公众哪些事实或数据信息受保护, 哪些不受保护。否则, 公众无法判断现实生活中随处可得的事实和数据是否可以自由利用。这类审查制度将产生巨大的社会成本。单条数据信息在表达层面(而非技术应用层面)的价值有限, 很难支撑起类似专利制度的管理成本。因此, 过去的共识是, 在表达层面, 保护不具有独创性的公开的数据信息, 在制度上是不可行的。著作权法的独创性要求、思想与表达二分等规则, 刚好避免了上述难题。

## (二) 数据规模增加之量变与质变

在网络时代, 数据收集的规模和所耗费的成本都达到空前的程度, 相关的投资激励机制与维权成本等都发生了很大变化。如我们所知, 技术意义上的大数据的规模巨大, 文字类数据信息通常以 GB 字节为基本处理单位, 甚至达到 TB (1024GB) 乃至 PB (1024TB) 字节以上规模。<sup>[40]</sup> 不过, 特殊数据集合立法对数据集合规模的要求要比技术意义上“大数据”的量级低很多。比如, 很有价值的全国法律法规数据库也不过 1GB 左右; 100 万级别的用户基本信息库可能也不超过 1G; 100 万级别的司法案例数据库估计也就 100G。<sup>[41]</sup> 即便如此, 这些数据集合的规模与传统的单个作品的容量规模还是有多个数量级上的差异。日常生活中常见的有限数据的集合, 比如一周电视节目表、单位电话号码本、NBA 或 CBA 某个赛季的比赛安排表、有限的作品汇编、单个科研项目产生的有限实验数据之类的数据集合, 远未达到本文所述的大数据集合的数据规模要求, 应该交由传统知识产权法决定它们的保护问题。如果被拒绝保护, 则没有必要提供额外的产权保护机制。

著作权法排除数据规则体现的是“小数据”时代的社会现实, 已经跟不上大数据时代的行业需求。在讨论大数据产权保护时, 很多学者还是沿着著作权法思想和表达二分的传

[38] See *Telstra Corp. v. Desktop Marketing System*, (2001) F. C. A. 612, para. 83.

[39] See *Mazer v. Stein*, 347 U. S. 201, 218 (1954).

[40] 参见孟小峰、慈祥:《大数据管理:概念、技术与挑战》,《计算机研究与发展》2013年第1期,第147页。

[41] 假定每个用户的基本信息数据为 1K, 每个案例的数据量为 100K, 就可以粗略估计出上述数字。

统思路,强调数据总是应该留在公共领域。<sup>[42]</sup>在他们看来,无论规模大小,数据或公开技术信息等本质上是不能保护的,此类保护机制与专利法或著作权法的逻辑存在根本矛盾,不可能运作良好。<sup>[43]</sup>在“小数据”时代,这类意见大致是正确的,但是在大数据时代则未必。对于今天的学者而言,正确认识“小数据”到“大数据”的质变过程,对于理解大数据时代的数据集合保护思路至关重要。

如前所述,在“小数据”时代,学术界普遍认为数据并非人为创造,不应受到保护。在大数据时代,这一理由的说服力非常有限。数据集合的规模急剧增大,继续坚持这些大数据作为一个整体也存在于客观世界中、并非收集者的创作的意见,就显得非常勉强了。在数据条目为前文所说的作品类数据,即网络用户或第三方创作的著作权内容(文字、图片或视听作品等)时,大数据集合实际上是无数作品的集合。我们充其量可以说,收集者对该大数据集合没有独创性贡献,但并不能说它是收集者发现的客观存在的数据或事实。

著作权法在排除数据或事实时,假定对作品独创性表达的保护能够顺便促进数据或事实信息的供给,但在大数据集合问题上这一假定并不总是成立。现在,海量数据的收集工作本身常常耗资巨大,同时,数据收集者不再需要以某种体现独创性的方式对数据进行深加工,就可以直接向用户提供数据。这是因为数据检索技术的进步使得很多信息的个性化整理、分类和编排变得不再重要。<sup>[44]</sup>用户可以方便地在非结构化的数据集合中找到自己所需的具体信息。<sup>[45]</sup>对于用户而言,重要的是数据本身,而非收集者本身的独创性贡献。这时,著作权法通过保护作品中的独创性表达激励数据收集者的策略就不再有效。即便大数据集合中用户提供的数据信息条日本身构成作品,也只是该用户能够主张著作权,数据收集者并不能够直接从中受益。在大数据集合需要耗费实质性投入的情况下,我们不提供额外的产权保护,则很可能无法保护数据收集者的投资积极性。这在后文有进一步论述。

在“小数据”时代,著作权法拒绝保护数据或事实对保证后来者创作新作品或表达新观点有重要意义,在大数据时代则未必。大数据集合内是海量的数据,大数据有限产权保护对后来者的创作或表达自由的负面影响急剧下降。其中的原因至少有三方面:

首先,数据量急剧增加后,普通公众出于创作作品之目的或表达思想情感之需要,在自己的作品中复制和传播该海量数据的必要性也迅速降低。限制后来者复制并传播相同内容的海量数据,对于后续作品的创作行为也不会产生实质影响,毕竟绝大多数作品中所直接引用的数据量有限,远未达到大数据集合的规模。“小数据”时代的事实和数据,或者大数据时代体现海量数据的分析结果的数据(依然是“小数据”),才是后来者再创作的重要基础。

其次,随着数据集合内数据数量的迅速增加,限制传播该数据集合所引发的表达或言论自由方面的关切反而会下降。大数据集合更接近于工业原材料,而不是典型的言论表达的媒介。理论上,大数据集合整体上的确能够传递非常复杂的信息,有言论价值。<sup>[46]</sup>大数

[42] 参见前引[29], Reichman等文,第89页。

[43] See Dennis S. Karjala, *Misappropriation As a Third Intellectual Property Paradigm*, 94 Colum. L. Rev. 2594, 2603 (1994).

[44] 参见前引[32], Posner文,第635页。

[45] 参见前引[4], Gibson文,第182页。

[46] 关于数据的言论属性,可以参考 Jane Bambauer, *Is Data Speech?*, 66 Stan. L. Rev. 57 (2014); Tim Wu, *Machine Speech*, 161 U. Pa. L. Rev. 1495 (2013).

据集合的产权保护，当然会限制它们所体现的言论价值。不过，对外传播海量数据，属于一种典型的商业言论。为了保护数据收集者的利益对此类商业言论的限制，并不容易产生所谓的“寒蝉效应”，因此其正当性也强于限制普通言论。与限制传播“小数据”信息相比，限制对原本就公开的大数据集合的传播，对公众表达或言论自由的影响要小得多，因为除了典型的商业机构外，一般公众并不通过直接传播“海量数据本身”的方式发表言论。

最后，在数据条目实际上构成作品时，大数据的排他权保护对于言论自由的影响就更加可以忽略不计。言论自由保护领域的主流意见认为，著作权法的相关学说已经足以保护言论自由价值。<sup>[47]</sup>正如美国最高法院金斯伯格（Ginsburg）法官在 *Eldred v. Ashcroft* 案中指出的那样，思想和表达的二分学说与合理使用学说是版权法内在的自由言论保护机制，只要立法者停留在版权法的传统框架内，就没有第一修正案（言论自由）的问题。<sup>[48]</sup>在大数据集合的内容是海量的受著作权法保护的内容的集合时，大数据集合的排他权的范围很窄，仅仅限制公众传播包含海量作品的集合整体，而不限制公众利用其中的单个作品本身。从言论自由保护的角度看，既然传统著作权法限制利用单个作品可以接受，那么大数据的排他权就更可以接受了，因为这一保护所影响的行为类型与公众范围更窄。

前文还提到，“小数据”时代的事实或数据信息的个性化程度较低，保护事实或数据信息会遇到抄袭认定的难题。在大数据保护领域，这将不再是突出的问题。将来法律禁止抄袭和传播的数据量必须达到实质性的量。抄袭者辩称自己独立收集相关信息或通过变通数据信息掩盖抄袭行为的可能性不是很大。因此，认定抄袭通常也不是问题。

### （三）制度成本的容忍度提高

在大数据时代，数据集合的数据规模急剧增大，数据挖掘技术迅速进步，数据的用途也迅速增加。海量数据的利用价值一般会大大超过普通作品。这一价值增值在产权制度设计方面有重要意义。

按照德姆塞茨（Demsetz）的产权理论，在资源价值增值的社会背景下，公众对于资源产权化所产生的制度成本的容忍会增加。他当时所描述的例子是印第安人土地私有化的例子。欧洲与北美之间的皮毛贸易导致皮毛的价值上升，设置土地私有产权保障海狸皮毛供给变得更有效率，社会因此接受了土地私有产权制度的安排。<sup>[49]</sup>

大数据集合的产权保护实际上遵循类似的逻辑。在“小数据”时代，公众对于数据或事实使用的交易成本的忍耐较低。比如，写一篇文章就可能要引用很多事实信息，针对这些事实信息获得许可，很可能导致作者放弃引用。因为平均而言，获得许可的成本可能远远超过文章本身的价值。在大数据时代，数据交易大多是在商业机构之间而不是在权利人与普通公众（消费者）之间进行。商业机构能够忍受的交易成本相对较高。同时，大数据商业价值增值的效果应该明显超出数据集合制度的管理成本和许可谈判的交易成本，否则人们关于大数据时代的各种美好期待就不可思议了。在大数据规模和价值增加导致产权成本相对而言不再高昂的情况下，向大数据集合提供有限的产权保护，然后由交易双方谈判以实现数据资源的最有效利用，应该是合理的选择。

[47] See Neil Weinstock Netanel, *Locating Copyright within the First Amendment Skein*, 54 *Stan. L. Rev.* 1, 3 (2001).

[48] See *Eldred v. Ashcroft*, 537 U. S. 186, 221 (2003).

[49] See Harold Demsetz, *Toward a Theory of Property Rights*, 57 *Am. Econ. Rev.* 347, 352 (1967).

当然，大数据集合的后续利用多种多样，如果产权保护过度限制了后续使用者的选择，则又可能产生新的不效率。很多人反对数据特殊保护的一个重要理由在于，它增加了后续增值产品和服务进入市场的障碍。<sup>[50]</sup>比如，医疗数据等就可能为各种数据挖掘应用提供可能性，被用于各种医学研究、社保评估、保险政策研究、广告投放等等；<sup>[51]</sup>网络图片数据可以被收集起来用于测试和训练数据挖掘、人脸或图像识别等技术。如果大数据集合产权过宽，覆盖所有后续数据利用行为（比如复制、演绎等），则的确会导致后续数据利用的成本上升。如果过多的使用者受到上述类似的负面影响，则有可能导致过宽的产权保护本身整体上没有效率，因而失去正当性。

将来，随着大数据价值和社会容忍的制度成本的提升，决策者会在大数据集合的产权框架内外，逐步引入精细的权利限制机制。产权制度内在的权利限制包括独立收集例外、合理使用抗辩等，产权制度外部的权利限制包括用户数据自主权、<sup>[52]</sup>反垄断法、消费者保护等公共利益保护法律的限制。权利限制规则与权利保护规则相互作用，使得整个产权机制能够实现精细的利益平衡。

### 三、大数据有限排他权与市场激励

在说明大数据集合的产权保护不会实质危害著作权法所强调的公共领域的行动自由之后，本文将从正面考虑大数据集合产权保护对于激励数据收集者的必要性。基本分析思路是，在大数据集合的内容必须公开时，如果他人的搭便车行为会导致数据收集行业出现市场失败，而现有的法律或非法律机制不足以消除该失败，则额外的产权保护是必要的。

#### （一）劳动并非当然需要产权激励

大数据集合的最低限度的产权保护，在一部分学者看来，其正当性可能不言而喻。比如，有意见认为，企业花费大量成本收集数据，如果不提供产权保护，则数据企业就“没有经济动力从事此类活动”；企业收集数据后对数据集合享有权利，“这种法律效果与民事主体合法建造房屋而自该房屋建造完毕之时取得房屋的所有权是相同的”。<sup>[53]</sup>习惯于有形财产的民法学者可能倾向于接受“劳动成果应受保护”这一直觉性原则的指引。

不过，在知识产权领域，劳动学说或自然权学说的论证思路的说服力或号召力非常有限。中国决策者实际上并不当然接受劳动或自然权学说，而更多地接受功利主义的指引。<sup>[54]</sup>如果市场本身提供了有效的替代性激励机制，比如，自然的领先时间足以保证领先者获得合理的投资回报等，额外的产权保护就不是合理选择。过去相当长的时间里，集成电路布图设计的抄袭和模仿相当困难，同时这一领域技术发展神速，这导致集成电路布图

[50] 参见前引〔29〕，Reichman等文，第119页以下。

[51] See Marc A. Rodwin, *Patient Data: Propert, Privacy & the Public Interest*, 36 Am. J. L. & Med. 586, 590 - 592 (2010).

[52] 数据自主权的代表性立法是欧盟《通用数据保护条例》（General Data Protection Regulation, 简称GDPR, Regulation (EU) 2016/679）第20条。

[53] 参见前引〔7〕，程啸文，第117页以下。

[54] 参见崔国斌：《知识产权法官造法批判》，《中国法学》2016年第1期，第144页以下。

设计方面的知识产权保护在很多人看来是多余的。<sup>[55]</sup>

对于单个数据或少量数据，在绝大多数情况下，现有的“商业秘密+著作权”的保护模式已经能够为数据收集者提供一定的激励，额外的产权保护不合理地限制了公共领域的行动自由，于是劳动或资本投入本身并不能当然地成为大数据集合获得额外的产权保护的理据。比如，一项技术数据的获取可能耗费巨大的劳动或天文数字的投资，但这并不当然意味着在这一数据被公开之后还有必要在著作权或商业秘密之外给予额外的产权保护，以阻止他人对该公开数据的利用和传播。同样的道理，大数据集合也不能因为收集者付出劳动或投资，就当然具有获得额外产权保护的正当性。

正因为如此，过去人们对于数据库是否真的需要额外产权保护一直有很大争议。美国反对额外的产权保护机制的意见认为，并没有证据表明，缺乏额外的产权保护会阻碍新数据库的产生。<sup>[56]</sup>在美国反不正当竞争法（非法侵占学说或热点新闻学说）和合同法的框架下，数据收集者能够获得相当的保护，没有必要制定特殊权利立法。<sup>[57]</sup>欧洲虽然有数据库指令，但主流意见大多怀疑它为数据库提供了过度保护，实际经济效果并未符合立法者的预期。<sup>[58]</sup>

其实，即便确有必要为劳动成果提供产权保护，劳动学说通常也不能说明究竟提供何种程度的产权保护是合适的。<sup>[59]</sup>比如，大数据集合作为劳动或投资成果，在获得商业秘密或著作权保护之后，是否还要获得额外的保护？究竟应提供复制权还是传播权保护？应设置何种权利限制？劳动学说或自然权学说基本上无法为回答这些问题提供具体指引。

## （二）大数据领域的市场失败

大数据集合是否需要额外的产权保护，关键在于收集者是否为数据收集付出了实质性的劳动或资金，而在于现有的产权保护是否能够避免市场失败，即是否能够避免数据收集者因为他人的抄袭而没有足够的市场领先时间收回投资。<sup>[60]</sup>

如前所述，对于单个数据或小规模数据集合，商业秘密和著作权机制在绝大多数情况下已经提供了充分的激励机制，因此立法者并不愿意在此基础上提供额外的产权激励机制。不排除个别情况下，比如对于必须及时公开的天气预报信息，现有保护机制提供的激励机制不够充分，但为了照顾绝大多数情形，立法者依然会选择不对此类数据提供额外保护。为了激励此类信息的供给，社会可能通过政府投入等替代机制来弥补产权激励机制的不足。

随着数据集合规模逐渐增加，数据收集工作耗费的成本迅速增加，而现行的著作权法或商业秘密保护法无法阻止他人复制公开的数据集合并对外提供。数据收集者很可能因此无法获得足够回报，从而产生市场失败。在传统的数据库时代，这就已经是很现实的问题。

[55] See Leon Radomsky, *Sixteen Years after the Passage of the U. S. Semiconductor Chip Protection Act: Is International Protection Working*, 15 Berkeley Tech. L. J. 1049 (2000).

[56] See Daniel J. Gervais, *The Protection of Databases*, 82 Chi.-Kent L. Rev. 1109, 1167 (2007).

[57] 参见前引 [37]，Ginsburg 文，第 152 页。

[58] See *Intellectual Property: Evaluation of EU Rules on Databases*, European Commission IP/05/1567, Dec. 12, 2005, available at [europa.eu/rapid/press-release\\_IP-05-1567\\_en.pdf](http://europa.eu/rapid/press-release_IP-05-1567_en.pdf), last visited on March 21, 2019.

[59] See Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs*, 84 Harv. L. Rev. 281, 286 (1970).

[60] 关于市场失败的概念可以参考 Pamela Samuelson et al., *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 Colum. L. Rev. 2308, 2314 (1994).

比如,在非常有名的 ProCD 案中,ProCD 花费了 1000 万美元来编制电话号码数据库,而被告 Zeidenberg 仅仅花费 150 美元购买该数据库,就可以对外提供类似服务。<sup>[61]</sup>再比如,有学者以法律数据库为例指出,将公共领域的法院判决、法律和政府文件等数字化并制成可检索的数据库,需要耗费实质性的资源。即便对公共领域保护极度关心的人也不会怀疑,如果法律不保证当事人对数字化结果的某种独占性控制以收回投资,则没有公司会承担该实质性投入。<sup>[62]</sup>

在大数据时代,数据规模进一步增大,原始收集者和复制者的成本差距可能进一步拉大。大众点评网、百度搜索引擎、新浪微博等网络服务商常常耗费高额成本收集海量用户或第三方产生的数据,生成庞大的数据集合以提升用户利用其网络服务的体验。网络公司的市值动辄上百亿,很大程度上是建立在数据资源的基础上。他人通过数据复制所获得的竞争优势哪怕是 1% 的市场份额,也是价值惊人。如果不直接保护此类公开数据,社会只接受丛林规则,数据收集者为了维持自己的商业模式,就必须不断收买随时会出现的真实的或伪装的复制者以消除他们的搭便车行为。最终,数据收集者的动机会因此受损,市场失败很可能出现。这一点,从过去关于规模较小的数据库产权保护的文献看,并无太大争议。<sup>[63]</sup>即便是最反对扩张知识产权领地的意见也支持给予某种程度的数据库产权保护。大家争论的不是是否提供保护,而是多大程度的保护才能避免市场失败。<sup>[64]</sup>随着数据规模的增大,额外产权保护对公共领域表达自由的影响降低,为公开的大数据集合提供额外的产权保护的正当性也随之增强。

当然,以上观察仅仅表明,限制他人向公众传播大规模数据集合的行为,对避免市场失败而言是必要的。进一步限制公众对数据的复制或演绎等行为,并不具有直觉上的正当性。如前文所述,如果将产权保护范围进一步扩大,使得收集者能够控制其他类型的数据利用行为,对公共领域的行动自由的限制增加,则需要更充分的理由。

在上述分析过程中,我们忽略了数据收集者可能具备的特殊市场优势(比如品牌优势、支配地位、网络效应、商业模式等),假定数据收集者和其他竞争者处在充分竞争的市场,数据集合本身是企业的核心资源,因此容易得出竞争者的搭便车行为会导致市场失败的结论。现实中的部分数据市场可能与这一假想情形存在较大差别,数据收集者可能拥有巨大的品牌优势、市场支配力或网络效应,甚至具有跨行业的交叉补贴的能力。比如,在前面提到的大众点评网的例子中,大众点评网可能已经取得市场支配地位,用户有很高的忠诚度(粘性),同时点评数据可能有一定的网络效应,后来竞争者的抄袭可能很难给市场领先的大众点评网造成实质损害;证券交易所本身处于支配地位,从证券市场的组织上获得充分回报,收集整理证券交易信息是其主营业务的内在需要,并非出于从数据利用中获得回报之目的;<sup>[65]</sup>医药公司为了验证已经受到专利保护的技术方案而获取临床试验数据集合,而该专利保护通常能够保证数据收集者收回投资并得到合理回报,<sup>[66]</sup>等等。对这些企

[61] See ProCD, Inc. v. Zeidenberg, 86 F. 3d 1447, 1449 - 1450 (1996).

[62] 参见前引 [24], Samuelson 文, 第 152 页。

[63] 参见前引 [29], Reichman 等文, 第 137 页。

[64] 参见前引 [4], Gibson 文, 第 232 页。

[65] 参见前引 [32], Posner 文, 第 629 页。

[66] See Peter Yu, *The Political Economy of Data Protection*, 84 Chi. -Kent L. Rev. 777, 784 - 785 (2011).

业而言，好的商业模式比知识产权保护更重要。它们凭借数据之外的资源，可能就可以维持竞争优势，将抄袭者远远地甩在后面。向公众提供数据是它们主营业务自动产生的副产品，可能并不需要额外的产权激励。如果我们只是将目光放在上述数据收集者所具有的特殊市场优势上，则很可能得出额外的产权保护没有必要的结论。

不过，在讨论大数据产权保护的必要性时，假定数据收集者处在充分竞争的市场中、没有特殊优势，才是合理的选择。毕竟，数据收集者所处的行业或采取的商业模式五花八门，它们中的绝大多数一开始并没有因为自身的品牌、网络效应或商业模式等因素而获得稳定的竞争优势，从而能够利用市场领先时间获得充分回报。常识告诉我们，诸如收集用户内容的视频网站或社交网络，科学期刊论文、法律法规与案例数据库，类似谷歌图书馆项目的作品数据库，大型地图导航系统的数据库，维基百科之类的网络知识库等数据集合，从开始经营到取得市场支配地位，有遥远的距离。在它们取得竞争优势之前，产权保护对于维持正常的商业竞争秩序、保护投资者的预期就已经很重要了。

另外，数据收集者取得的网络效应或优势地位并不能够保证其不被竞争对手超越。网络服务行业的竞争非常激烈，领先可能只是暂时的，后来者居上的例子比比皆是。即便是垄断市场上的寡头也需要利用财产权来保护自己的投资预期。在大数据成为用户关注的最核心资源时，如果没有额外的产权保护，则后来的竞争者随时可以通过抄袭让领先者失势。如果大数据市场完全是弱肉强食的丛林，朝不保夕，则很少有数据收集者愿意进入这一高风险市场来玩勇敢者的游戏，或者数据收集者会在技术措施等自我保护方面过度投入，从而导致市场失败。

当然，品牌号召力、支配地位、网络效应、商业模式等的特殊市场优势，对于具体的大数据市场的监管可能具有重要意义。如果有证据证明，在特定情形下，数据集合保护的确可有可无，而披露数据对于维护公共利益有重要意义时，决策者可能会基于公共安全、消费者保护或反垄断等考虑采取措施限制甚至剥夺数据收集者的数据产权。前文提到的药物临床试验数据、空客公司飞机安全数据等，就可能被强制要求披露。

### （三）数据条目著作权保护的不足

处于公开状态的整体上不具备独创性的大数据集合，如前所述，无法获得商业秘密保护，也无法直接获得著作权法保护。不过，如果数据集合所包含的数据条目内容本身具有独创性，则可以作为作品获得著作权保护。数据收集者虽然不享有该条目的著作权，但可以利用合同机制从用户手里获取独占性或排他性授权，然后利用该授权阻止他人对外提供同样的大数据集合内容。这是否意味着大数据集合本身的额外产权保护就不再必要呢？

即便数据条目构成作品，对于数据收集者而言，通过协议从用户那里获得数据条目的独家授权也会产生实质性的交易成本和许可费成本。相当一部分的网络服务商可以通过网络点击合同等简便的方式收集用户的著作权许可，貌似不会耗费太多资源，但事情并不如此简单。以视频网络或社交媒体网站为例，用户提交的内容复杂多样，价值千差万别，用户给予独占授权的意愿也有很大差异。部分用户为创作高质量视频而实质投入，可能并不愿意将独占性权利授予网络服务商。网络服务商要求用户免费地授予其独占性权利，属于非常重要的合同条款。仅仅通过网络点击合同中的格式条款呈现，未必满足相关法律所要

求的合同条款提示要求。<sup>[67]</sup> 发生争议后, 法院可能认定该条款未被订入合同, 不构成合同内容。<sup>[68]</sup> 在实体法上, 如果网络服务商处于支配地位, 独家授权条款则有可能被认为是服务商滥用优势地位追求“显失公平”结果的工具, 因而有被撤销的可能。<sup>[69]</sup> 另外, 普通用户也大多会认为这并非公平合理的条款, 企业坚持此类条款可能损害自身商誉, 得不偿失。因此, 为了获得可靠的独占性授权, 网络服务商很可能需要和不同用户进行个性化的许可谈判。在网络服务商具备市场支配地位时, 尤其如此。这意味着实质性的交易成本和许可费支出。如果网络服务商不愿意负担, 则很可能只能通过格式条款谋求在有限范围内的非独占地使用用户提交的作品, 而由用户保留其他权利, 包括授予第三方(包括网络服务商的竞争对手)使用该内容的权利。

值得一提的是, 很多时候, 网络服务商阻止他人从整体上复制并传播自己收集的大数据集, 就足以保护自己的利益, 并不需要从无数用户那里获得对于具体内容的独占性控制权。比如, 美国著名的照片分享网站 Flickr 就不寻求独占授权, 而是许可用户保留所有权利或按照开放许可协议(creative commons)对外许可。<sup>[70]</sup> 在前面提到的大众点评网案中, 也有理由相信, 大众点评网在意的应该是竞争对手百度公司系统性地复制并对外提供它收集的用户内容, 而不是自己是否真的取得了对于单个用户点评内容的独占权。著作权法对数据条目的著作权保护并非为数据收集者量身定做的间接保护机制。如果法律直接规定数据收集者对数据整体有直接的控制权, 数据收集者就无需再和每一个用户进行许可谈判, 从而大大节省社会成本。其中的道理与著作权法保护广播组织权类似, 即依靠广播组织对广播节目信号的控制权而不是依靠它对广播节目内容本身的著作权来保护广播组织的投资。

从社会角度看, 迫使网站(数据收集者)超出自己的需求从用户手中获取使用数据条目的独占性授权, 也未必符合社会公共利益。部分数据收集者被迫越过成本障碍获取海量用户数据或作品的独占性控制权后, 很可能产生很强烈动机去维护自己获得的著作权独占使用权, 限制后续的数据复制和利用行为, 藉此获取额外的市场利益以弥补它为获取用户授权而支付的交易成本。这会增加公众后续创新的成本。而在笔者建议的有限产权机制下, 数据收集者并不能限制他人除公开传播数据之外的利用行为, 对数据的后续利用行为的影响很小。

可能有人会认为, 即便数据收集者没有集中这些数据条目的著作权, 用户个人也可能基于著作权而威胁后续的数据利用行为。也就是说, 数据收集者集中用户作品著作权的行为没有使事情变得更糟糕。但实际上, 相当部分网络用户通过网络平台公开自己的作品或数据, 并非出于直接的营利目的, 因此很可能不在乎后续第三方的数据收集和利用行为。比如, 大众点评网的用户很可能并不介意第三方对其数据的各种利用(包括通过信息网络传播)。这些用户就有限的数据利用行为提起诉讼, 能够获得的利益相对有限, 因而很可能会选择容忍很多网络侵权行为。在网络用户创作意愿不受实质影响的情况下, 在用户不发

[67] 参见消费者权益保护法(2013)第26条、合同法第39条。

[68] 参见韩世远:《合同法总论》, 法律出版社2018年版, 第923页。

[69] 参见民法总则第151条、合同法第154条。

[70] See Judith Harbert, *Understanding How Copyright Works on Flickr*, available at <https://www.thesprucecrafts.com/understand-how-copyright-works-on-flickr-4106956>, last visited on March 20, 2019.

出侵权通知前，容忍侵权行为客观上增进了社会福利。<sup>[71]</sup>而数据收集者“被迫”获取独占授权之后实现一定的规模效应，很可能比个人用户更愿意去寻求著作权保护，从而增加后续利用数据的社会成本。

#### （四）合同或技术措施保护的不足

在商业秘密或著作权法机制外，公开向公众提供数据的数据收集者如果无法获得额外的产权保护，还可能会利用合同条款或技术措施来限制接触该数据的用户范围，并限制该用户的后续传播行为。比如，数据收集者可能要求用户先注册，接受限制性合同条款后再接触相关数据，未注册用户则不能获得该数据。注册协议条款通常禁止用户大量复制数据并对外提供。如果用户未经许可大规模下载或对外提供数据，则要承担违约责任。在著名的 ProCD 案中，数据收集者就是这么做的；<sup>[72]</sup>在欧洲法院判决的 RynaAir 案中，权利人也如此要求。<sup>[73]</sup>如果数据收集者对合同约束信心不足，还可以采取技术措施来保护数据。比较常见的做法是监控用户或访问者的网络行为，限制用户的大规模下载行为。如果有违反，就禁止用户继续访问。比如，新浪微博、百度贴吧、领英<sup>[74]</sup>等都通过 Robots 协议和必要的技术监控措施等阻止大规模复制。

不过，在法律不提供额外的产权保护的情况下，数据收集者的合同约束或技术措施的保护有明显的局限性。合同约束不够有效，原因是多方面的：首先，随着用户数量的增加，协议监督就变得困难。很多时候，数据收集者很难察觉注册用户下载数据后向不受合同约束的第三方提供数据的行为。一旦该第三方合法获取上述数据，后续传播就很容易失控。其次，即便数据收集者能够有效地向第三方传递合同约束，该合同约束也可能因违反知识产权法所维护的公共政策而得不到支持。<sup>[75]</sup>比如，《欧盟数据库指令》第 15 条规定，不得利用合同排除某些合理使用。之所以如此，是因为当合同约束能够有效地向第三方传递时，合同权利与对抗所有人的财产权之间的界限就不再清楚，合同约束对公共领域的侵蚀就十分明显。

技术措施保护也不能完全让人满意，原因在于它总是有可能被破解，不能保证一直有效。比如，对于 IP 地址限制之类防止大规模复制的技术措施，竞争对手会通过 VPN 等欺骗手段伪造众多的 IP 地址，躲过权利人的技术封锁，以尽可能多地复制数据。如果没有法律干预，数据收集者和竞争对手很容易陷入丛林规则主导的无休无止的消耗战。百度和奇虎之间就进入过类似的数据抓取与反抓取的技术措施升级的对抗状态，直至最后对簿公堂。<sup>[76]</sup>

为了避免丛林规则引发的恶性循环，著作权法直接对技术措施提供保护，禁止第三方破坏此类技术措施或者向公众提供破坏此类技术措施的工具。在著作权法的框架之外，其他法律也可能禁止公众破坏类似的技术保护措施。比如，我国在计算机或网络安全方面的法律法规也禁止公众未经许可破坏技术措施去访问权利人的网络系统并大量获取数据。<sup>[77]</sup>

[71] See Tim Wu, *Tolerated Use*, 31 Colum. J. L. & Arts 617 (2008).

[72] See ProCD v. Zeigenberg, 86 F.3d 1447 (1996).

[73] See Ryanair Ltd v. PR Aviation BV, C-30/14 (CJEU, 2015).

[74] See hiQ Labs, Inc. v. LinkedIn Corp., 2017 WL 3473663 (N. D. Cal., 2017).

[75] 参见前引 [21], Osenga 文, 第 2117 页。

[76] 参见北京百度网讯科技有限公司、百度在线网络技术(北京)有限公司与北京奇虎科技有限公司不正当竞争纠纷案,北京市第一中级人民法院(2013)一中民初字第2668号民事判决书。

[77] 参见网络安全法第 27 条。

在美国,网络公司也会利用《计算机欺诈与滥用法案》来阻止未经许可访问受技术措施保护的网路内容。<sup>[78]</sup>

从公共政策的角度看,权利人过度依赖技术措施保护,会显著降低数据的流动性和可及性,使得原本不受法律保护的数据获得保护。<sup>[79]</sup>互联网企业采取技术措施保护数据资源,也使得网路市场朝更加封闭的纵向一体化模式发展。<sup>[80]</sup>这不仅压缩了公共领域的空间,限缩接触数据的公众范围,同时也使得公众获取数据的过程变得更为繁琐,增加了数据收集者推广业务的成本。当年美国《数据库投资与知识产权反盗版法案》(Database Investment and Intellectual Property Antipiracy Act of 1996)引发广泛争议的议题之一,就是数据库上的技术措施问题。该法案建议禁止提供规避技术措施的装置,使得数据收集者能够通过技术措施限制公众的合理使用行为,引发很多批评。

当然,这里并不是说法律提供额外的产权保护后,所有权利人都会放弃采取技术措施。实际上,只要通过技术措施强化数据控制的额外收益超过采取措施的成本,权利人就有动力采取该措施。<sup>[81]</sup>法律提供了更强的产权保护,只是削弱那些不担心公众接触和利用数据的权利人而不是所有权利人采取技术措施的积极性。即便如此,额外的产权保护依然值得提倡。

#### 四、大数据排他权的立法模式选择

在商业秘密保护机制和著作权法基础之上,对大数据集合提供额外的最低限度的产权保护,理论上可行的立法模式有多种:反不正当竞争法模式、特殊立法模式或著作权法框架下的邻接权模式等。在立法者作出最终选择之前,法院暂时依赖反不正当竞争法原则条款提供保护,有一定的合理性。但长远而言,特殊立法或邻接权保护所代表的财产权模式(绝对权或排他权)应该是更合理的选择。

有些学者试图在上述知识产权法的框架之外寻求解决方案,比如参考物权法,将数据想象成抽象物,然后设置大数据物权。<sup>[82]</sup>传统物权法机制相对于现代知识产权法而言,并没有充分考虑数据这类无形物的非竞争性(non-rivalry)以及背后的公共利益需求,因而没有建立起类似知识产权法的精细的利益平衡机制。<sup>[83]</sup>以物权机制作为模板来制定大数据集合的产权保护规则,不过是将知识产权法重新发明一遍,很容易陷入过度保护的泥潭,不值得认真对待。

##### (一) 权利法与行为法的模式差异

在数据集合的保护模式选择上,一直有财产权(权利法)模式与反不正当竞争法(行

[78] 在 Craigslist 案中,法院就支持原告利用上述法案阻止被告未经许可从其网站抓取公开数据的主张。See *Craigslist Inc. v. 3Taps Inc.*, 942 F. Supp. 2d 962 (N. D. Cal., 2013).

[79] 参见前引[21], Osenga 文,第 2120 页以下。

[80] 参见胡凌:《商业模式视角下的“信息/数据”产权》,《上海大学学报(社会科学版)》2017年第6期,第12页。

[81] 参见前引[24], Samuelson 文,第164页。

[82] 参见周林彬、马恩斯:《大数据确权的法律经济学分析》,《东北师大学报(哲学社会科学版)》2018年第2期,第33页。

[83] 参见前引[54],崔国斌文,第144页。

为法) 模式选择的争议。比如, 欧洲在制定其数据库指令时, 一开始希望采用反不正当竞争法模式, 但后来在利益集团的游说下走向了宽泛的财产权模式。在美国, 虽然数据库保护特殊立法并未成功, 但也一直有这两种模式选择的争议。

在传统意见看来, 反不正当竞争法与权利法保护模式的差异显而易见。传统的财产权观念强调人对物的绝对支配。即便没有物理形态的物, 也要想象出一个抽象物。<sup>[84]</sup> 而反不正当竞争法通常只限制竞争者具体的竞争行为, 而不是像权利法模式那样将排他权表述为对客体(大数据集合)的全面支配。在美国, 以反侵占或反盗用为代表的反不正当竞争法模式只是限制不公平的抄袭行为, 因此法院将保护延伸到不应保护的客体(比如事实)的风险大大降低。对权利人较弱的保护, 对公众而言就意味着更多的自由; 如果仅仅限制竞争对手的低成本的数据利用行为, 就足以使得数据收集者获得足够的市场领先时间以避免市场失败, 则无需采用宽泛的产权模式。<sup>[85]</sup>

在笔者看来, 反对利用权利法模式的意见, 大多还停留在传统的财产权观念。现代的被广泛接受的霍菲尔德财产权学说强调财产权(property right)调整的是人与人之间的关系, 而不是人对物的支配。<sup>[86]</sup> 在法律现实主义者看来, 财产权是单纯的权利束(a bundle of rights), 在个案中需要根据公共政策来定义每一个权利束。<sup>[87]</sup> 放弃人对物的绝对支配的思路后, 财产权概念有了更大的弹性和适应性。但与此同时, 财产权的确失去了它本来的或者说传统的核心含义。法院不再能够依据财产权的概念直接推断所争议的财产权的具体内容, 而是要在个案中基于公共政策来判断。<sup>[88]</sup> 对这种法律现实主义的财产权观念, 美国学者亨利·史密斯(Henry Smith)提出过批评意见, 认为这实际上使得财产权的内容可以被任意塑造, 可以解释任何事情, 实际上也等于什么都没有解释。他认为, 传统财产法强调人对物的支配关系, 是通过模块化应对现实复杂性的法律机制的一部分, 可以降低公众的信息成本, 有特别的意义。因此, 我们不能忽视财产法对于人与物(支配)关系的强调。<sup>[89]</sup>

史密斯对于传统财产权建立在有物理边界之上的绝对排他权理念的合理性的解释, 以及对于财产权概念泛化的批评, 都很有道理。不过, 随着现代社会普遍接受各种排他程度各异的无形财产的观念, 比如专利权、著作权、商标权、商业秘密权都被视为财产权, 甚至著作权中的某一项排他权都被视为可转让的财产权, 传统财产权概念已不可避免地瓦解了。现在要重新限缩公众对财产权概念的使用, 并为新出现的各种“排他性利益”创设新的概念标签, 几乎成了不可能的任务。排他权机制的模块化组合中, 本来就有一般性的较宽的排他性(exclusion)机制到具体行为管制(governance)机制的过渡, 不同社会的策略有很大差别。<sup>[90]</sup> 我们现在要做的只是在观念上接受: 只要该管制或排他的内容可以被概括

[84] See Kenneth J. Vandavelde, *The New Property of the Nineteenth Century: The Development of the Modern Concept of Property*, 29 Buff. L. Rev. 325, 328 (1980).

[85] 参见前引[43], Karjala文, 第2599页。

[86] See Wesley Newcomb Hohfeld, *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 26 Yale L. J. 710, 721 (1917).

[87] See Henry E. Smith, *Property as the Law of Things*, 125 Harv. L. Rev. 1691, 1697 (2012).

[88] 参见前引[84], Vandavelde文, 第365页以下。

[89] 参见前引[87], Smith文, 第1697页以下。

[90] 同上文, 第1705页。

为权利人排除公众对某种客体的某种利用行为，权利人可以自由转让这一利益，则无论管制或排他的范围宽窄，其实都可以被视为一种财产权利。无视现代社会的商业实践，单纯以排他或保护范围的宽窄来区别反不正当竞争法模式和财产权模式，已经没有太大意义。

理论上，反不正当竞争法通过对竞争关系的强调、对特定行为类型的限缩和对个案中反竞争效果的关注，可以使得它所禁止的竞争行为范围看起来比财产权模式要窄很多。不过，在保护大数据集合的具体操作层面，二者的差异几乎可以忽略。在仅仅保护大数据集合的公开传播权的情况下，侵权者几乎总是与权利人直接竞争，是否再强调竞争关系，对侵权认定影响不大。对侵权行为的严格限定（仅仅限于公开传播行为），使得无论是权利法模式还是行为法模式，保护范围差别很小。

部分学者担心财产权模式导致对数据的保护过于宽泛，<sup>[91]</sup>多半低估了财产权模式本身的弹性和适应性。其实，大数据集合产权范围的宽窄并不单纯取决于法律对客体范围和排他行为的表述，同时还要依赖权利限制或侵权抗辩条款。比如，著作权在很多人看来是一种财产权利，但实际上它并不支持权利人对作品的绝对支配。众所周知，著作权法不禁止第三方通过独立创作获得完全相同的作品，此即独立创作抗辩。此外，思想表达二分、合理使用例外等一系列权利限制规则，也大大压缩了排他权的范围。在大数据集合产权立法中，如果立法者设置严格的客体要件和有限的排他权内容，同时设置宽泛的权利限制，比如独立收集例外、合理使用例外、保护期限等，也会使得大数据集合产权受到实质限制。此时，财产权模式下的大数据排他权并不一定比反不正当竞争法提供的保护更宽泛。

在实务层面，反不正当竞争法和财产权两种模式的真正差异并非排他范围的宽窄，而是二者在定义排他权、权利限制、转让或许可权的方式、个案中反竞争效果权衡、救济方式等方面的差异，使得二者对特定商业实践的契合程度不尽相同。数据产业最重要的需求有两方面，即禁止未经许可的公开传播和促进数据许可或转让等商业交易。在反不正当竞争法的框架下，他人的公开传播行为可以被描述为一种不正当竞争行为，因此阻止他人公开传播的目的比较容易实现。但是，在反不正当竞争法的框架下，建立复杂的权利内容、权利限制、保护期限、数据利用的许可或转让规则，则不是十分名正言顺。这看起来是在反不正当竞争法内部建立一个局部的财产权体系，显得不伦不类。比较而言，在数据特殊产权模式下，权利人能够阻止他人未经许可传播大数据集合的行为，同时也可以对外转让该排他权。从观念上，将这种保护范围很窄的排他权理解为一种财产权，并通过具体立法将权利内容标准化，能够很容易套用现有的财产权交易规则来规划数据集合的交易，更符合数据市场的现实需求。

## （二）特殊立法或邻接权模式

在确定按照权利法（排他权）模式立法保护大数据集合后，将相关立法放在现有知识产权法的框架下，自然是合理的选择。大数据集合一开始大多以商业秘密形式存在，自然进入知识产权法的视野。大数据集合作为一种无形客体，外在形式上也与作品十分类似。同时，数据集合具有所谓的非竞争属性，可以多人同时使用而不相互妨碍，在这一方面与普通作品或技术方案等知识产权客体并无本质差别。因此，很多北欧国家过去就通过放松

[91] 参见前引[13]，芮文彪等文，第99页。

著作权法的独创性标准来保护传统的数据集合。<sup>[92]</sup> 在大数据集合被公开之后，禁止他人公开传播，与著作权法禁止公开传播作品的做法接近。另外，知识产权领域过去也一直将对于缺乏独创性的数据库的保护视为一种典型的知识产权问题加以讨论，并努力在知识产权法的框架下提供解决方案。这在世界范围内均大致如此。<sup>[93]</sup> 在这一背景下，立法者将大数据集合作为一种知识产权客体来处理，顺理成章，可以有效降低社会的学习和沟通成本。

大数据集合排他权作为一种知识产权，单独立法在理论上并没有什么障碍。国内有不同意见认为：“特殊权利保护的立法模式未获普遍认可，且与我国现有的知识产权法律体系不相协调，自然不能成为我国保护数据信息的立法选择。”<sup>[94]</sup> 这一结论并不令人信服。在前文我们已经讨论了，虽然欧盟的数据库特殊权利过于宽泛因而受到广泛批评，但这并不意味着较窄的特殊权利保护立法本身不可行。实际上，当初美国立法者也尝试单独立法保护数据库。通过特殊立法，将大数据集合与传统的作品和小规模数据库区分开来，同时规定较窄的特殊权利内容（限于公开传播权），与现行知识产权法体系并不冲突，在立法技术上是可行的。

在操作层面，如果立法者发现单独立法过于麻烦，将大数据集合保护规则放在著作权法的邻接权框架下也是可以考虑的方案。部分邻接权客体原本就不需要满足独创性要求，如广播组织制作的节目信号。保护大数据集合也主要是为了保护投资而非独创性劳动。这与广播组织权等邻接权的立法目的并无本质差别。数据收集者的权利内容（即公开传播权）与普通邻接权也类似。从立法技术角度看，将大数据集合当作邻接权客体，可以避免在单行立法中大量重复著作权法的既有条文。比如，权利的归属（包括共有规则）、许可使用规则、侵权归责原则、损害赔偿规则、侵权救济、权利限制、行政执法规则（如果有的话）等，都可以直接借用。另外，将数据纳入著作权法框架，熟悉相关业务的版权行政管理部门能够名正言顺地成为行政主管部门。这比由其他并不熟悉版权行业的行政部门来主管大数据集合的保护事务更为合适。

最后，对于数据集合的保护放在著作权法的框架下，也有比较法上的成例。德国就是在著作权法的框架下保护数据库。<sup>[95]</sup> 国内其实早就有意见认为，“以邻接权保护数据库从理论到实践都是契合的。”<sup>[96]</sup> 从数据库到大数据集合，在立法技术上仅一步之遥，不存在明显障碍。

## 结 论

虽然世界已经进入大数据时代，诞生于“小数据”时代的知识产权法还是表现出了很强的适应性，在大数据产权保护方面只是留下些许空白。真正需要立法者关注的只是那些处在公开状态没有独创性的大数据集合。为了维护公共领域数据利用的自由，同时又避免

[92] See Jane G. Ginsburg, *No Sweat Copyright and Other Protection of Works of Information after Feist v. Rural Telephone*, 92 Colum. L. Rev. 338, 372 (1992).

[93] 参见前引〔29〕，Reichman 等文，第 54 页。

[94] 参见前引〔13〕，芮文彪等文，第 99 页。

[95] 参见德国著作权法第 87 条。

[96] 参见林华：《大数据的法律保护》，《电子知识产权》2014 年第 8 期，第 84 页。

大数据领域的市场失败，应当采用简明的“大数据有限排他权”思路。这一保护机制具有足够的制度弹性，既满足数据行业的现实需求，又能够兼顾公众后续利用数据的利益。将来，随着信息技术的进步，大数据集合的价值进一步上升，数据收集、流转和利用等环节的分工进一步细化，拓展大数据集合的权能也有可能是合理的选择。

笔者并不指望为大数据集合的知识产权问题找到终极意义上的解决方案，也不奢望本文在有限篇幅中提出的制度设计细节都能获得认可，但是笔者确实希望在一定程度上消除我国在大数据集合产权保护方面理论研究的混乱，为后来的研究人员或决策者提供一个初步但可靠的宏观分析框架。在这一分析框架下，后来者经过更深入的理论分析或更可靠的实证研究，可能会在具体问题上得出不同结论，但这不影响本文整体分析思路的价值。

---

---

**Abstract:** Human society is entering the era of big data and artificial intelligence, and intellectual property protection of large-scale data sets has become increasingly important. While intellectual property law born in the era of “small data” roughly meets the basic needs of big data industry, it does leave a blank area, that is, the lack of specific legal protection for big data that has no originality and is open to the public. In order to fill in this gap, avoid market failures and protect the freedom in public domain, Chinese legislature should entitle those who have made substantial investment in collecting substantial amount of data a limited exclusive right, i. e. a right of communicating the data to the public. This *sui generis* protection is able to meet the actual need of data industry, does not conflict with public policies maintained by copyright law, and can guarantee the freedom of downstream data users.

**Key Words:** big data, intellectual property, public domain, right of communication to the public

---

---