

# 反不正当竞争法视角下数据抓取行为违法性的认定

刘继峰 张 雅

**摘 要:** 作为数字经济时代的关键生产要素,数据成为互联网企业争相抢夺的重要资源。实践中爬虫技术与反爬虫技术的博弈生动体现了数据抓取与数据保护的矛盾关系。我国当前对于数据抓取行为的竞争法规制主要依照反不正当竞争法第二条一般条款。在这一条款的具体运用中,需要遵循社会整体利益最大化原则平衡数据保护与竞争权益之间的冲突、数据开放与信息保护之间的冲突;还需要通过对竞争关系、行为正当性、数据必需设施抗辩等实践中争议较大且严重影响违法性认定的因素进行考察,明确反不正当竞争法视角下数据抓取行为的违法边界和正当性空间。

**关键词:** 数据抓取;正当性;不正当竞争;数据垄断

中图分类号: D922.294

文献标识码: A

文章编号: 1009-2447(2021)04-0102-10

数字经济时代的到来催生了以数据为代表的新型生产要素。<sup>[1]</sup>随着算法的优化和人工智能技术的进步,以数据为输入的数据驱动正在逐步替代以人的经验分析为输入的流程驱动,成为当今企业创造业务增长点、提升运营效率、保持可持续发展的重要手段。数据资源越来越成为互联网企业核心竞争力,对数据的获取和运用甚至直接决定企业经营的成败。正是由于数据资源的战略经济意义,实践中出现了大量数据抓取与反抓取的较量。数据抓取关涉数据持有者、数据抓取者和社会公众三方的权益,但利益关系较为复杂,甚至存在一定的利益冲突:既要保护数据抓取者自由竞争的权利、避免其恶意获取数据资源的“搭便车”行为,又要考虑数据持有者的成本投入和知识产权、防止其拒绝数据共享违法实施反竞争行为,同时还要关注用户个人信息保护的需要。这些盘根错节的利益纠葛大大延伸了问题的深度与广度,使得个案中数据抓取行

为的法律性质通常具有高度不确定性,严重影响数据的共享与保护,阻碍数据驱动型产业的创新发展。

作为域外反公共数据抓取行为第一案<sup>[2]</sup>,hiQ诉LinkedIn阻止其访问用户公开数据案自2019年一审判决结果公布后就引起了广泛讨论。案中,LinkedIn认为,用爬虫技术对用户资料进行抓取会严重威胁用户隐私;hiQ则辩称,LinkedIn垄断公共数据是为了排挤竞争对手,会阻碍互联网的开放与创新。最初,美国法院认为“不禁止公司抓取可在互联网上公开访问的数据”,即LinkedIn不应禁止竞争对手hiQ从LinkedIn用户公开的个人资料中收集用户信息。然而,利益关系并没有这么容易厘清。2021年6月14日,美国最高法院要求下级法院重审本案。一时间,对于数据抓取与数据垄断的认识、数据抓取行为违法边界的探讨再次引发广泛的关注。

在我国司法实践中,惯常的做法是寻求反不正

**基金项目:** 国家社会科学基金重点项目“完善权利与权力配置关系及运行制约法律机制研究”(20AZD107)

**作者简介:** 刘继峰,男,黑龙江富锦人,中国政法大学民商经济法学院教授、博士生导师,研究方向为竞争法;张雅,女,河南南阳人,中国政法大学民商经济法学院博士生,研究方向为竞争法。

当竞争法第二条一般条款的保护，以违反“商业道德”“诚信原则”之名打击违法数据抓取行为。在这一规则运用中，如何正确认识一般条款的适用条件与抗辩理由、明晰数据抓取行为的违法边界和正当性空间，于数字经济发展而言具有时代意义。

## 一、爬虫与反爬虫的技术博弈

“爬虫”只是一种形象化的表述，在互联网领域表达的是潜入并获取信息的行为。了解爬虫的技术原理对于理解数据抓取法律关系具有重要意义。

### （一）爬虫的基本原理及技术应用

数据的“抓取”（Scraping）或“爬取”（Crawling）等<sup>[3]</sup>主要通过网络爬虫（Web Crawler）技术实现。网络爬虫又称为网络机器人（Web Robot）或网络蜘蛛（Web Spider），源于1993年的万维网漫游者（World Wide Web Wander）<sup>[4]</sup>，是当代互联网信息技术发展的伟大成就之一，是一种按照指定规则检索、收集特定目标范围互联网数据的自动运行程序或脚本。<sup>[5]</sup>该程序或脚本会通过自动向网站发送系列特定检索指令实现对网站内系列网页信息的全部收集。<sup>[6]</sup>也就是说，爬虫基于HTTP协议，可以实现有针对性的、精确的数据获取，其抓取的内容是所有基于HTTP协议传输的信息。

这种技术的应用性极强，能够打破信息交流壁垒，为用户提供海量的目标信息资源。例如，搜索引擎获取信息的关键性技术之一就是网络爬虫<sup>[7]</sup>，搜索引擎利用爬虫技术在网页间建立系统性关联，通过对网页信息的系统性获取，建立起针对这些网页内容的搜索服务。<sup>[8]</sup>具体来说，每一条信息在万维网上都有统一且唯一的地址URL<sup>①</sup>（Uniform Resource Locator），即网络地址。首先建立一个等待抓取的URL队列，从一个或多个初始网页的URL开始，按顺序不断抽取每一个URL，访问并解析其对应的网页，抓取该网页中的所有URL再存入待抓取队列进行循环抓取，直到队列中所有的URL全部抓取完毕或满足系统的一定停止条件时结束。<sup>[9]</sup>随着商业模式的创新和技术的升级，爬虫也逐步向智能化、支持多协议的方向发展。当今社会对于爬虫

技术的使用已经不仅限于搜索引擎领域，新闻媒体、社交软件、打车软件、导航软件等领域也都已实现了普遍应用。爬虫技术已经全方位渗透进了我们的生活。

### （二）反爬虫技术的防御

爬虫技术能够实现高效的数据汇集与收取，其本身是中立的，只是数字时代一种更便利的信息收集工具<sup>②</sup>，爬虫并不等于害虫。但是，爬虫带来的超负荷访问量可能严重影响网站服务器的正常运转，用户信息、管理员信息、著作权信息等隐私内容也常常面临泄漏风险。<sup>[10]</sup>而且，在以数据为核心竞争资源的互联网时代，海量数据意味着竞争优势，于是大量的反爬虫（Anti-Spider）技术和措施开始被应用。<sup>[11]</sup>常见的反爬虫手段可以分为两大类。

一是Robots协议。全称是“网络爬虫排除协议”（Robots Exclusion Protocol），是1994年荷兰工程师Martijn Koster首次提出和发起的一项针对网站所有者和数据爬取者的技术措施。由网站所有者根据自身需要在网站程序开头写入robots.txt文件，载明该网站中禁止爬虫抓取的网页，在爬虫程序访问网站时对其进行“允许”或“不允许”的提示。通常来说，遵守协议的爬虫会首先读取要爬取网站的robots.txt文件，对于禁止访问的数据不予收集。Robots协议作为一种最为简单、基础的技术措施得到了互联网行业的普遍通行和遵守<sup>③</sup>，但这种“君子协议”作用有限，效力尚无定论<sup>④</sup>，因此更多是一种没有强制力的惯例和声明。

二是爬虫识别技术。它包括一系列技术性的反爬虫手段。由于爬虫通常是以模仿用户浏览的方式来抓取数据，网站服务器采取反爬虫措施时必须区分爬虫访问和真实的用户访问。因此，产生了IP检测、Cookie检测、接口频率控制等访问频率识别技术，对网站设置访问频率阈值。由于爬虫访问频率远高于用户访问频率，一旦对该网站的访问频率达到阈值，就会触发屏蔽请求、禁止访问、输入验证码等保护机制，有效识别并拦截爬虫抓取。此外，还有HTTP请求信息验证、建立用户轨迹模型、用户画像等手段，也可以识别是否是真正的浏览器用户在访问，对爬虫起到一定的防御作用。

爬虫技术与反爬虫技术都具有两面性。爬虫技术既是“窃取资源之矛”，却也是“利益共享之矛”；反爬虫技术既是“利益保护之盾”，却也是“数据垄断之盾”。只看到数据抓取是窃取商业资源的手段而予以否认，势必会导致互联网头部企业基于对数据的控制力形成数据垄断；只看到数据垄断的危害而一味放开爬虫范围，必会导致个人、国家、社会公共利益受损。由此观之，以“利益共享之矛”刺破“数据垄断之盾”，以“利益保护之盾”防御“窃取资源之矛”才是正解，正确理解爬虫技术与反爬虫技术的博弈原理，才能实现数据保护与数据共享的双赢。

## 二、数据抓取中的权益冲突与平衡

相较于法律条文呈现的具体规定，权益平衡问题关注的是法律所确定的规范性行为模式背后的社会意义，而不是某一个案的胜负得失。<sup>[12]</sup>数据抓取行为并不天然具有违法性，对其合法与否的价值评价通常来源于该行为所触及权益的保护或损害。因此，违法边界的划定前提是明确多重权益冲突时的平衡之法。数据爬取行为通常涉及信息所有者、信息抓取者和社会公众三方的权益，跳出其一对一、一对多复杂的权益对抗关系之外，可以将其间冲突系统归纳为数据保护与竞争权益之间的冲突、数据开放与信息保护之间的冲突两类。

### （一）数据保护与竞争权益之间的冲突

实践中数据抓取行为的纠纷主要发生在市场主体之间。我国目前主要从反不正当竞争法的角度对数据抓取行为的合法性进行评价。这是因为相关市场的后进入者通常受限于数据获取的困难而难以发展和创新。数据的获取和运用又在一定程度上影响着企业经营的成败，故而数据抓取成为被广泛应用的技术。在诸多数据抓取不正当竞争案件中，纠纷产生的根本原因在于数据保护与竞争权益之间的冲突。

对于数据持有者而言：一方面，实施数据保护具有正当性。前期的数据收集不仅需要付出大量的人、财、物、技术等成本投入，还需要足够的用户量提供支撑。数据收集不仅不是免费的，而且每条数据信息都具有价值。根据数据获取的难易程度

和数据的可替代性来看，获取路径多、门槛低、不容易被单一渠道垄断占有的数据具有更强的可替代性，更容易收集，成本投入更少，其价值就更小。而数据抓取者利用网络爬虫企图抓取的自然是高价值数据，极易触及数据持有者的数据保护机制（包括但不限于技术手段、法律手段等）。因此，如果没有足够的利益驱动，数据的跨主体共享是很难实现的。另一方面，可能侵害其他市场主体的竞争利益。强大的数据保护手段之下，数据的自由流通受限，尤其是在一些特殊的商业模式或情景下，数据获取的途径非常有限，有价值的信息通常被锁定在特定人员、特定路径之中，相关市场上的其他竞争者获取门槛也较高。<sup>[13]</sup>例如，社交软件具有极强的用户粘性，先进入市场的竞争者往往利用这一优势掌握大量的用户数据信息，而后进入者由于缺乏广大的用户群体，难以获取同前者媲美的海量数据资源。

对于数据抓取者来说：一方面，需要保护其竞争利益。市场中头部互联网平台往往利用自身的竞争优势获取巨额流量和海量信息，形成对数据的支配和控制，达到数据垄断的目的。数据垄断提高了相关市场的进入门槛，极易形成赢者通吃的局面。掌握海量信息的数据持有人往往凭借先发数据优势操纵市场，使其他市场主体丧失公平、自由的竞争机会。基于竞争利益的驱动，数据持有者拒绝共享的行为极为常见，在hiQ诉LinkedIn阻止其访问用户公开数据案中，一审法院就认为LinkedIn禁止hiQ收集其用户公开信息的行为，是为了排挤竞争对手，影响了hiQ的竞争利益。<sup>⑤</sup>另一方面，要避免其“食人而肥”“搭便车”等反竞争行为。如前所述，数据收集需要付出极大的前期成本，对数据持有者数据权益的保护能够让越来越多的企业意识到数据资源的无形性财产作用，进一步考虑数据效益增值，调动企业的创新自主性，促进社会资源的优化配置。倘若放任数据抓取行为肆意蔓延，只会打击数据驱动型企业技术创新的积极性。在谷米诉元光<sup>⑥</sup>、新浪诉脉脉<sup>⑦</sup>、大众点评诉百度<sup>⑧</sup>这三起著名的国内数据抓取行为不正当竞争纠纷中，法院无一例外地将利用网络爬虫技术大量获取并无偿使用数据持有人数据的行为评价为具有反不正当竞争法意



义上的不正当性，并适用该法第二条进行规制。

数据持有者和数据抓取者之间数据保护与竞争权益之间的冲突平衡，是数据抓取行为违法边界划定需要解决的核心问题。

## （二）数据共享与信息保护之间的冲突

数据爬取中所蕴含的数据自由流通在数字经济时代具有重要价值。<sup>[14]</sup>数据开放的重要性已经在世界范围内得到了广泛认可。在2019年OECD发布的报告中指出，加强数据的获取和共享（Enhancing Access to and Sharing of Data, EASD）有助于最大限度地提高数据重复利用的社会和经济价值，可以提升数据对数据持有者以及二级数据用户的价值，为国民经济和社会整体带来额外的积极溢出效益。根据数据的范围和数据的开放程度，就公共部门数据而言，数据获取和共享大概可产生相当于国内生产总值（GDP）0.1%至1.5%的社会和经济效益，在还包括私营主体数据的情况下，可产生相当于GDP的1%至2.5%的社会和经济效益（在一些研究中，可达到GDP的4%）。<sup>[15]</sup>欧盟认为数据流通有助于实现数据驱动型增长和创新<sup>[16]</sup>，并在《通用数据保护条例》（General Data Protection Regulation, GDPR）和《非个人数据自由流通条例》（Regulation on the Free Flow of Non-Personal Data）中都明确将“个人数据自由流通”作为立法目标。2021年6月最新颁布的《中华人民共和国数据安全法》第七条亦规定“保障数据依法有序自由流动”。事实上，无法开放共享的数据如同一个个信息孤岛，而数据抓取正是孤岛之间互联互通的桥梁和纽带，数据共享会产生1+1远远大于2的效果。

然而，尽管经济、社会效益对数据自由流通的需求与日俱增，数据共享依然难以发挥其潜力，个人、企业和政府经常面临数据访问的障碍，在数据开放的过程中不可避免会产生泄露个人隐私数据以及侵犯商业和非商业利益等一系列风险。如果说数据保护与竞争权益之间的冲突只涉及信息所有者和信息抓取者双方之间的关系，那么数据共享与信息保护之间的冲突则更为复杂，关涉信息所有者、信息抓取者、社会公众等多方的权益。目前大量App要求获取用户的位置、通讯录等个人信息，甚至随着AI技术和人脸识别技术日渐成熟，面部信息也

成为新的利益增长点。可以说，在信息化时代，以简单的二进制信息单元0、1的形式表示的数据，不仅承载着用户的个人隐私，还承载着企业的商业机密，甚至国家的机密信息。但用户往往无法控制数据信息的产生、储存、应用和转移，一旦关键数据被抓取，将可能导致无法预估的损失。美国最早的软件销售商之一Egghead曾将用户的信用卡信息储存在服务器中，但由于网站缺乏安全性保障，导致多达370万个信用卡号码数据被盗取，公司声誉一落千丈，最终宣布破产，被亚马逊公司收购。

虽然数据共享必然会为信息保护带来风险，信息保护也会在一定程度上限制数据共享，但二者不是非此即彼的关系，其权益冲突并非不可调和。

## （三）权益平衡之根本——社会整体利益最大化原则

在进行上述两对权益的冲突化解时，不能采取非黑即白的“一刀切”思路，各方主体的权益都应当受到平等保护。因此，应当遵循社会整体利益最大化原则使各方权益得到最大化实现。

利益是经济政治社会活动运转的内驱动力。社会整体利益是指全社会公众的共同利益，具有公共性、社会性、长期性的特点，并不是单个主体利益的简单叠加，个人或一方利益的最大化也绝不是社会整体利益最大化。相较于局部性经济效率概念——帕累托最优，社会整体利益最大化是全局性的经济效率概念。在数据抓取行为中多方主体权益冲突时，平衡的标准在于将整体损害降至最低。

针对数据保护与竞争权益之间的冲突，应当注重发挥二者的协同作用。数据保护所体现的创新政策和竞争权益保护所体现的竞争政策都是在世界范围内获得公开承认的基本公共政策，如何看待和处理二者之间的关系是一个很复杂的问题。但实际上，数据保护和竞争权益保护具有共同的价值目标——促进技术创新和提升消费者福利（Consumer Surplus）。调和二者之间的冲突应从这一突破点入手，在考虑数据持有者的成本投入和数据抓取者竞争权益的基础上，提高“数据所有权”的确定性、保护企业对企业（B2B）数据协议的合同自由，通过“合意+法定”的方式进一步明确数据抓取的违法边界，推动数据保护与竞争权益协同发挥作用。

针对数据共享与信息保护之间的冲突，应当明确数据共享必须获得相关授权。数据“开放性”不存在单一的最佳水平，数据的开放与共享不应当是无限的。我国司法实践中，通过新浪诉脉脉采取数据抓取行为不正当竞争一案，法院确立了“在Open API开发合作模式中，第三方通过Open API获取用户信息时应坚持‘用户授权’+‘平台授权’+‘用户授权’的三重授权原则”<sup>⑩</sup>，即被抓取的数据需为用户授权数据平台存储于自身服务器中，第三方的数据抓取要同时得到被抓取方（数据平台）和用户的三重授权。但是，此三重授权的做法无疑提高了数据共享的成本。笔者认为，针对涉及私益的信息，数据持有者必须获取用户（个人、企业）授权，而不应该被数据“实际控制人”所左右；针对涉及公共利益的信息，数据持有者必须依照法律法规规定或获得相关政府部门授权。如近日国家互联网信息办公室会同有关部门对《网络安全审查办法》进行修订，征求意见稿中明确规定“掌握超过100万用户个人信息的运营者赴国外上市，必须向网络安全审查办公室申报网络安全审查”。

### 三、数据抓取行为违法性判定因素考量

正如前文提到，爬虫协议和爬虫识别技术被广泛应用于规范数据抓取行为，二者本质上都是通过代码来实现目的。可以说，在网络空间中，一定意义上“代码”就是“法律”。<sup>[17]</sup>然而这一“法律”的地位和效力并没有得到真正的法律认可，这些由数据持有者采取的单方措施虽然在互联网领域普遍存在，但终究只是一种“代码规制”。随着技术的迭代升级，爬虫与反爬虫技术被不断突破更新，这种循环博弈依然无法对数据抓取行为的滥用形成有效防御。法律是维护社会公平正义的最后一道防线，数据抓取行为最初进入法律视野就是在竞争法领域，如1999年的eBay诉Bidder's Edge案<sup>⑪</sup>等。在我国，无论是反垄断法还是反不正当竞争法，都并无针对数据抓取的具体法律条文或规则，法官判决时通常依据一般条款进行自由裁量。<sup>[18]</sup>但一般条款只能提供行为正当性评价的基本原理，并不能创设权利。<sup>[19]</sup>因此，在规则运用过程中，明确数据抓取

行为违法性判定因素十分必要。抓取数据的目的是获取更大的竞争优势，提升产品或服务的质量或类型，故竞争关系、行为正当性是两个核心要素，另外，也应给予抓取人以一定的抗辩机会。

#### （一）竞争关系的判定

在反不正当竞争案件纠纷中判断经营者之间是否具有竞争关系是认定行为人是否构成不正当竞争行为的前提。在数据抓取不正当竞争行为认定中需要考虑数据抓取者与数据持有者是否具有围绕数据的竞争关系。该要素的判定包括两个方面，一是双方互联网经营者的身份判定，二是竞争关系的判定。

##### 1. 数据抓取行为中经营者的身份认定

关于数据抓取行为中经营者的身份认定，一方面，理论界的传统观点采取“主体资格论”，认为反不正当竞争法范畴下的经营者是指在市场竞争中具有同业竞争关系的，并且依法取得营业资格的市场主体；另一种观点是“行为标准论”，认为应当以是否具有市场经营活动行为作为判断经营者身份的依据。根据我国反不正当竞争法第二条的规定“本法所称的经营者，是指从事商品生产、经营或者提供服务的自然人、法人和非法人组织”可以看出更侧重于强调市场主体的行为，而不是资格。原国家工商总局出具的多份文件也确认了反不正当竞争法中“经营者”应根据市场经营行为认定。将经营者限定于取得营业资格的主体，不仅不利于互联网经济的发展，也无益于消费者权益的保护。另一方面，关于是否要求经营者的行为具有“营利性”也存在争议。在谷米（“酷米客”）诉元光（“车来了”）<sup>⑫</sup>一案中，被告元光公司主张自己和原告均不是反不正当竞争法意义上的经营者，因为双方开发的软件属于公益性质，不存在任何市场交易，用户只需注册、登录即可实现免费在线查询。但是法院认为“判断某相关市场主体是否系经营者，并不以其所提供的某项商品或者服务是否具有营利性为标准”。这里，笔者支持跨越“营利性质”的传统观念，将数据视为一种生产经营要素，获取此要素者即为经营者。互联网领域的发展日新月异，网络经济活动也有别于传统经济活动的显著特点。用户粘性和流量才是评价商品或服务竞争力的核心指

标，过分强调经营者的“营利性”便无法实现反不正当竞争法评价体系的与时俱进。

## 2.数据抓取行为中竞争关系的广义认定

关于数据抓取行为中竞争关系的认定，传统的竞争关系判断思路是“基于商品功能、效用”，即经营同类商品（服务）或者替代性商品（服务）的经营者之间具有竞争关系。然而，互联网竞争领域对数据资源的多样化应用，商品（服务）的差异性壁垒逐步被打破，不同行业间业务交叉的情况时有发生。即使数据抓取者和数据持有者分属不同行业领域，但其共同的目标是争夺用户资源和流量，已然具有了实质性的竞争关系。在我国司法实践中也采取了广义竞争关系的解释。例如，在爱奇艺诉电视猫MoreTV不正当竞争一案中，法院认为“不能将竞争关系狭义地理解为经营同类商品（服务）或者替代商品（服务）经营者间的同业竞争关系”<sup>[20]</sup>。竞争关系本质上是“市场环境下经营者之间的相互制约关系”<sup>[20]</sup>，因此，任何为争取自身交易机会削弱他人竞争优势增强自己竞争优势的行为都是具有竞争关系的表现。<sup>[21]</sup>

### （二）行为的评价

竞争行为的正当性判定是反不正当竞争法的核心问题。<sup>[22]</sup>因此，数据抓取不正当竞争行为违法判定中对于行为正当性的认定是重中之重，行为的正当性是反不正当竞争法第二条中“商业道德”和“诚信原则”的适用基础。就数据抓取行为而言，被抓取数据的公开性、原始性、授权情况、来源合法性等是在进行行为正当性评价时常产生争议的因素。

#### 1.数据的公开与否

是否为公开数据在一定程度上直接决定了抓取行为的正当与否。非公开信息数据自不必说，通常涉及个人隐私、商业机密、国家安全机密等，其“非公开”的本质意味着数据持有者对该数据的“不开放”态度，非经数据持有者同意或授权的抓取行为具有当然的不正当性。公开数据的抓取行为性质则一直争议不断，例如，在hiQ诉LinkedIn阻止其访问用户公开数据案中，法官认为，尽管hiQ的数据抓取存在涉及LinkedIn用户隐私问题，但没有证据表明用户的实际隐私期望，LinkedIn也未能提供

证据证明hiQ在五年内收集的数据对自己造成了经济或其他损害，因此，hiQ抓取LinkedIn上公开用户数据的行为合法。但该案件近期已发回重审，美国法院对于公开用户数据的态度可能发生反转。笔者认为，公开数据获取的正当性评价并不是单一的，需要根据数据的性质进行具体认识。当数据信息获取于用户时，正当性评价的首要价值取向是维护用户数据隐私。因此，单方面未经用户授权同意的抓取行为应认定为具有不正当性；而公开用户数据之外的信息，正当性评价最重要的考量因素应当是保护数据的自由流通性，故而，针对该类数据的抓取不应受到限制。

#### 2.数据的加工与否

就实践情况来看，数据抓取的内容不外两类信息：一种是平台从他处直接获取的一手原始数据，一种是经过分析加工的分析数据。抓取对象是否为一手数据信息也有可能影响抓取行为的正当性。通常来说，平台在进行一手数据收集后，会根据公司经营战略、产品需求等对数据进行分析、加工。经过分析加工的数据属于一定意义上的“劳动成果”，具备无形财产的属性，对该类数据的爬取行为则可以认定为“未经许可使用他人劳动成果”。虽然这种劳动成果不属于法定权利，未经许可使用的行为不能构成反不正当竞争法严格意义上的“搭便车”行为，但数据抓取方应当充分尊重被抓取方在信息的分析、加工过程中的辛勤付出。并且，在数字经济时代，数据所蕴含的价值难以估算，如果不加限制地允许抓取，将会严重阻碍企业的创新动力。因此，擅自抓取经过分析加工的数据具有明显的行为不正当性，而针对平台收集的原始数据的抓取正当性认定应当结合其他要素综合分析。

#### 3.数据的授权与否

一方面，“授权”不仅包括签订数据获取协议等明示授权，还包括默示授权。例如，数据抓取者在写有Robots协议的网站爬取数据的过程中没有收到禁止访问提示，则视为获得了抓取授权。另一方面，“授权”不仅包括数据平台授权，还包括用户授权，尤其是在爬取涉及用户个人信息的数据时，是否获得用户授权是抓取行为正当性评价的重要依据之一。上述hiQ诉LinkedIn案的判例意味着允许第



三方在用户不知情或未经用户同意的情况下访问数据持有者平台上的公开用户数据。但在我国司法实践中，与之相似的案情却有完全相反的判决结果。在新浪诉脉脉案中，脉脉在没有获得微博平台授权、也未经脉脉未注册用户许可的情况下抓取新浪微博用户的职业信息、教育信息。法院最终判定抓取非脉脉用户信息的行为具有不正当性，同时确立了针对公开用户数据的抓取需依照“用户授权+平台授权+用户授权”的三重授权原则。

相比于hiQ诉LinkedIn案，该案更强调尊重数据持有者和平台用户的数据开放意愿，但同时也阻碍了数据的自由流通。这是因为数据持有者出于自身竞争利益的考虑自然会在授权抓取方面层层设限，三重授权原则下正当数据抓取的难度系数将大幅提升。事实上，用户公开数据保护的本质是个人信息保护，用户有权选择是否向他人提供自己的信息。因此，结合抓取对象和抓取方式的综合效果考量，“用户授权+用户授权”（被抓取的数据为用户授权数据平台存储于自身服务器中的，同时数据抓取者也要得到用户授权）模式下的公开数据抓取行为应是正当的。

#### 4. 数据来源的合法与否

数据持有者本身数据来源的合法性是否会影响数据抓取行为正当性的评价，也是一直以来饱受争议的问题。实践中，非法获取数据资源的情况并不少见，例如，未经用户同意通过监控位置信息、监听、缓存等非法手段收集用户数据；利用网络爬虫技术违法获得其他平台数据资源。从源头上来看，上述平台对于非法获取的用户数据不具有正当权益，那么对于这类数据的抓取行为是否允许呢？笔者认为，在这一判断过程中，前端的数据来源合法性和后端的数据抓取正当性应当切割开。数据抓取行为是一个基于数据抓取者意志而进行的完整的单方行为，因此对于该行为的评价必须聚焦于行为本身违法性的判断，数据来源合法性并不成为影响因素。正如，抢劫他人因盗窃所得的财物时，抢劫和盗窃是两个不同主体分别实施的不同行为，对抢劫行为的认定和评价与受害人（盗窃者）的违法行为无关，只与行为人（抢劫者）的主客观情况有关。同样的，数据来源的合法性评价是针对数据持有者的，而数据抓取行为的正当性判断是针对数据抓取

者的。也即是说，即使数据持有者的数据来源缺乏正当性，但这并不是数据抓取者以不正当的方式削弱其竞争优势的理由，充其量会影响案件违法程度的判断，而不影响违法与否的判断，对抓取行为正当性的认定不产生影响。

#### （三）数据必需设施的抗辩

必需设施原则是反垄断法中认定拒绝交易行为违法性的重要原则之一。该原则的设立是为确保市场的自由竞争而附加给掌握必需设施经营者的一项强制性义务。在互联网领域，具有垄断地位的平台经营者在进行反数据抓取的过程中，极易触发反垄断法上的拒绝交易行为。因此，数据是否能够作为必需设施开放的问题与数据反抓取行为正当性认定紧密相关。目前，欧盟和美国的反垄断执法机关对于数据作为必需设施仍然保持谨慎的适用态度。2020年11月，国家市场监督管理总局发布《关于平台经济领域的反垄断指南（征求意见稿）》（以下简称《征求意见稿》），其中第十四条规定，分析是否构成拒绝交易，可以考虑“控制平台经济领域必需设施的经营者拒绝与交易相对人以合理条件进行交易”的情形，并规定了相关数据是否构成必需设施的认定规则：“一般需要综合考虑数据对于参与市场竞争是否不可或缺，数据是否存在其他获取渠道，数据开放的技术可行性，以及开放数据对占有数据的经营者可能造成的影响等因素。”该《征求意见稿》首次回应了数据垄断中的重要问题，但在2021年2月印发的《国务院反垄断委员会关于平台经济领域的反垄断指南》中，删除了《征求意见稿》内认定数据是否构成必需设施的款项，同时将“该平台占有数据情况”纳入认定相关平台是否构成必需设施的判断因素中。由此可见，我国法律层面在数据必需设施的应用上也持保留态度。但这只是基于现实应用困境考量的立法技术选择，并不意味着“数据能够构成必需设施”是一个伪命题。

大型互联网企业的数据垄断行为不仅会阻碍其他能提供更优质信息服务的企业进入市场，也会逐步因为缺乏竞争失去技术创新的外在动力。<sup>[23]</sup>虽然目前缺乏法律上的正当依据，但面对实务需求，也有学者提出可以在“个案分析的基础上审慎考虑必需设施原则的适用”<sup>[24]</sup>。那么，假使在今后的司法

实践中确实存在数据能够构成必需设施的情况时，对于数据抓取行为的影响应当如何认定呢？

数据垄断行为的实施通常离不开反爬虫技术的应用。在hiQ诉LinkedIn案中，hiQ向美国加利福尼亚北区联邦地区法院诉称LinkedIn拒绝数据抓取的行为属于拒绝开放必需设施的垄断行为。虽然最后法院认为hiQ没有准确界定本案相关市场而反对对其垄断指控，但引发了公众对数据抓取与数据必需设施之间的关系猜想。必需设施原则的适用本身需要考虑五个核心要素：一是由具有市场支配地位的经营者的控制；二是市场竞争所必需的；三是具有不可复制性；四是具有可开放性；五是经营者拒绝开放必需设施不存在合理理由。能够构成必需设施的数据当然也满足上述五个要素。

数据的获取有权利模式和义务模式。就前者而言，双方协商确定可获取的目标、范围等，协商的内容为获得合法数据的依据。就后者而言，如果数据构成必需设施，则数据开放的义务产生。数据抓取采取权利模式还是义务模式，不是经营者主体的市场份额决定，而是由数据的性质决定。当然，即便抓取的数据构成必需设施，也不意味着数据抓取行为就是合法的，只是证明数据权利人拥有向对方开放数据的义务。至于在开放的数据基础上，能否获得数据权利，还需要双方协商交易的条件。尽管欧盟《数字市场法》规定了守门人数据共享的义务，以减少其所收集数据的独占性控制。但是，对于跨平台企业所享有的数据权利是什么尚在探索之中，至多是赋予其访问权。跨平台企业对平台数据能否聚合，属于数据的可携带权的范畴，不是企业的自由获取权的范畴。

所以，数据抓取是未经许可而获得他人的数据，至于数据是否属于开放性数据，对违法性的判断不发生直接的影响。

数字经济时代的技术创新离不开数据的自由流通，数据抓取行为是实现数据流通的重要手段。数据抓取行为和爬虫技术本身并不天然具有违法性，但数据抓取行为关涉多方主体的权益。权益的矛盾与冲突、抓取对象和抓取方式都影响着行为正当性的评价，它们共同构成反不正当竞争法第二条中“商业道德”和“诚信原则”的判断基础。正确认

识数据抓取中的权益冲突与平衡，明确数据抓取行为违法性判定因素，才能为企业的竞争行为提供规范指引，实现数据保护和数据共享的双赢，达到社会整体利益的最大化。

## 注释

- ① URL由三部分组成：资源类型、存放资源的主机域名、资源文件名。也可认为由四部分组成：协议、主机、端口、路径。
- ② Sandvig v. Sessions, No. 1:16-cv-01368, Dkt. 24 (D. D. C. Mar. 30, 2018)。
- ③ 百度诉奇虎360搜索引擎违反Robots协议构成不正当竞争案，北京市第一中级人民法院民事判决书（2013）一中民初字第2668号。
- ④ eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1060-63 (N. D. Cal. 2000)。
- ⑤ LinkedIn Corporation v. hiQ Labs, Inc. Supreme Court of The United States Order List June 14, 2021。
- ⑥ 深圳市谷米科技有限公司诉武汉元光科技有限公司不正当竞争纠纷民事判决书，广东省深圳市中级人民法院（2017）粤03民初822号民事判决书。
- ⑦ 北京淘友天下技术有限公司等与北京微梦创科网络技术有限公司不正当竞争纠纷二审民事判决书，北京知识产权法院民事判决书（2016）京73民终588号。
- ⑧ 北京百度网讯科技有限公司与上海汉涛信息咨询有限公司其他不正当竞争纠纷二审民事判决书，上海知识产权法院民事判决书（2016）沪73民终242号。
- ⑨ 北京淘友天下技术有限公司等与北京微梦创科网络技术有限公司不正当竞争纠纷二审民事判决书，北京知识产权法院民事判决书（2016）京73民终588号。
- ⑩ eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058 (N. D. Cal. 2000)。
- ⑪ 深圳市谷米科技有限公司诉武汉元光科技有限公司不正当竞争纠纷民事判决书，广东省深圳市中级人民法院（2017）粤03民初822号民事判决书。
- ⑫ 北京爱奇艺科技有限公司与上海千杉网络技术发展有限公司、悦观网络技术（上海）有限公司其他不正当竞争纠纷一案民事判决书，上海市浦东新区人民法院民事判决书（2015）浦民三（知）初字第143号。



## 参考文献

- [1] 郭贺铨. 大数据时代的机遇与挑战[EB/OL]. (2013-02-16) [2021-06-25]. <http://theory.people.com.cn/n/2013/0216/c40531-20495578.html>.
- [2] 夏梦妍. 反垄断法视角下的反用户公开数据抓取行为——以美国hiQ诉LinkedIn案为出发点[J]. 中南法律评论, 2021(2):58.
- [3] Zachary G, Mark L. Robots Welcome? Ethical and Legal Considerations for Web Crawling and Scraping[J]. Wash. J. L. Tech. & Arts, 2018(13):280.
- [4] Seyed M, Mirtaheri, et al. A Brief History of Web Crawlers[EB/OL]. (2014-05-05) [2021-06-19]. <https://arxiv.org/pdf/1405.0749.pdf>.
- [5] Big Data: What Is Web Scraping and How to Use It, Towards Data Sci[EB/OL]. (2018-02-02) [2021-07-08]. <https://towardsdatascience.com/big-data-what-is-web-scraping-and-how-to-use-it-74e7e8b58fd6>.
- [6] Andrew S. Twenty Years of Web Scraping and the Computer Fraud and Abuse Act[J]. B. U. J. Sci. & Tech. 2018(7):373-374.
- [7] 李慧敏, 孙佳亮. 论爬虫抓取数据行为的法律边界[J]. 电子知识产权, 2018(12):59.
- [8] Javed M. How Do Internet Search Engines Work? Scientific American[EB/OL]. (2004-11-29) [2021-07-09]. <https://www.scientificamerican.com/article/how-do-internet-search-en/>.
- [9] Junghoo C. Crawling the web: Discovery and Maintenance of Large-scale Web Data[D]. L. A.: Stanford University, 2002:188.
- [10] 梁雪松, 张容. 网络爬虫对网络安全的影响及其对策分析[J]. 计算机与数字工程, 2009(12):86-88.
- [11] 陈利婷. 大数据时代的反爬虫技术[J]. 电脑与信息技术, 2016(6):60-61.
- [12] 尼尔·麦考密克. 法律推理与法律理论[M]. 姜峰, 译. 北京: 法律出版社, 2005:147.
- [13] 李慧敏, 孙佳亮. 论爬虫抓取数据行为的法律边界[J]. 电子知识产权, 2018(12):65.
- [14] 许可. 数据爬取的正当性及其边界[J]. 中国法学, 2021(2):166-188.
- [15] OECD. Enhancing Access to and Sharing of Data—Reconciling Risks and Benefits for Data Reuse across Societies[EB/OL]. (2019-11-26) [2021-07-08]. <https://www.oecd.org/going-digital/enhancing-access-to-and-sharing-of-data.pdf>.
- [16] European Commission. Staff Working Document—Guidance on sharing private sector data in the European data economy[EB/OL]. (2018-04-25) [2021-07-09]. [https://www.sohu.com/a/287300751\\_468714](https://www.sohu.com/a/287300751_468714).
- [17] William J M. City of Bits: Space, Place, and the Infobahn [M]. Cambridge: The MIT Press, 1995:111.
- [18] 苏青. 网络爬虫的演变及其合法性限定[J]. 比较法研究, 2021(3):89-104.
- [19] 刘晓春. 数据抓取的边界在哪里? ——以公开的用户生成内容为例[N]. 人民法院报, 2020-03-19(7).
- [20] 孔祥俊. 反不正当竞争法的创新性适用[M]. 北京: 中国法制出版社, 2014:114.
- [21] 孔祥俊. 反不正当竞争法新论[M]. 北京: 人民法院出版社, 2001:47.
- [22] 孔祥俊. 论《反不正当竞争法》的竞争法取向[J]. 法学评论, 2017(5):18-31.
- [23] Maurice S, Allen G. Big Data and Competition Policy[M]. Oxford: Oxford University Press, 2016:127-261.
- [24] 王健, 吴宗泽. 论数据作为反垄断法中的必需设施[J]. 法治研究, 2021(2):102.

## The Identification of Illegal Data Capture from the Perspective of Anti-unfair Competition Law

LIU Jifeng ZHANG Ya

**Abstract:** As a key production factor in the age of digital economy, data has become crucial resources that internet businesses compete for. The practical game between crawler technology and anti-crawler technology vividly demonstrates the contradiction between data capture and data protection. China's current competition rules against data capture are mainly in accordance with the second general provision of anti-unfair competition law. In the specific application of that provision, the principle of maximal overall social interest need to be followed to balance the clash between data protection and competition rights, and that between data sharing and information protection. Investigation needs to be conducted into those factors, namely, competition relationship, behavior legality, and defense for necessary data facilities, which are controversial and considerably affect illegality identification in practice. Thus, from the perspective of competition law, the illegal boundary and legal space can be confirmed.

**Key words:** data capture; legality; unfair competition; data monopoly