



数据合规专题

第八期

数据爬取的合规应用（下）

（2025年3月）

正如《数据爬取的合规应用（上）》中所述，数据爬取技术的应用可能会在不同程度上涉及民事、刑事、行政等方面的法律风险。因此，建议行为人在使用数据爬取技术的同时，做好全流程的评估及监管工作。

一、评估爬取的数据性质及爬取行为的合法性

司法实践中，法院认为互联网环境下数据获取（处理）方式包括：对公众开放且不需要授权的数据处理、需要授权但已获得授权的数据处理，以及需要授权但未获得授权的数据处理（或者超出授权处理权限的处理），后两者存在未经授权或超越授权的问题。因此，在数据爬取前，行为人应当有效识别数据性质以及数据爬取行为的可行性。

1、公开数据的爬取

公开数据通常能够由社会不特定公众通过一般途径获得，包括无须任何条件即可被公众获得的数据，如政府部门在官网上主动公布的政务数据、企业在年度报告中主动公示的企业数据以及个人信息主体自愿在网络上公开的个人信息等；以及公众在完成一定条件后可以获取的数据，如完成手机软件注册登录程序或向权利主体支付特定费用



后可以获得的数据。若相关数据只对特定人群开放，其他主体无法通过一般途径获得的，则该等数据属于非公开数据，通常获取门槛较高。

因此，在数据爬取过程中，行为人可以在符合获取条件的前提下爬取公开数据，但应避免抓取非公开且未经授权或超越授权的数据，尤其是涉及个人信息、知识产权、商业秘密及国家安全等方面的数据，以避免出现相关法律风险。

2、非公开数据的爬取

若行为人确需爬取相关权利人的非公开数据的，应当获得权利人的合法授权，如：与权利人签订合同、取得权利人出具的授权书等，并按照合同、授权书等授权文件中的授权范围、期限及条件等约定或要求爬取相关数据。

值得一提的是，对于已设置防护措施或 robots 协议的权利主体，无论其数据属于公开数据还是非公开数据，行为人在爬取时均不得规避其系统防护措施或径直违反 robots 协议。实践中，许多网站运营者通过设置 IP 限制、身份认证等反爬虫技术措施及防护机制，以防止网站数据被他人非法爬取。若数据爬取行为人在未获得前述网站运营者授权的情况下，利用技术手段破坏或绕过该等防护机制的，则该爬取行为可能被认定为具有不正当性。此外，部分网站还可能设置 robots 协议，以将该网站允许或者禁止爬取的数据内容告知数据爬取行为人，行为人在进行数据爬取时应当遵守 robots 协议的规定。虽然 robots 协议为网络运营方的“单方技术声明”，不具有强制性，但在司法实践中，司法机关仍会将是否违反 robots 协议作为判断爬取



行为正当性的重要参考因素。部分情况下，robots 协议的效力可能因其限制内容是否具有正当性而产生争议。若数据爬取行为人认为 robots 协议不合理的，可以事先就该等不合理部分的内容与被爬取网站运营者协商解决，或在其效力得以确定后再进行爬取，而不建议径行违反 robots 协议爬取相关数据。

二、做好数据爬取及后续数据处理中的监管工作

1、遵守法律法规规定、符合权利人要求

数据爬取是数据收集的一种方式，因此行为人须严格遵守法律法规关于数据收集的相关规定，如：遵循“合法、正当、必要”的原则，根据权利人的授权范围限制爬取的数据类型及数量等，避免出现收集数据不正当、超出必要限度、超出授权许可范围等违法违规风险。

在后续的使用、存储、传输及公开等数据处理过程中，行为人亦应满足法律法规规定及权利人要求，在授权期限及范围内正确使用爬取数据，并确保所爬取数据的完整度、真实性及安全性等，不得超出权限或采取其他不当方式处理数据。

2、不侵犯他人合法权益

除严格遵守法律法规规定并符合权利人要求外，行为人还应在数据爬取及后续处理过程中尽可能采取积极有效的相关举措将对权利人的损害降至最小，以免对权利人的正常生活或运营产生不利影响，从而侵犯权利人的合法权益。

在数据爬取过程中，为更及时、广泛地获取相关数据，行为人可能会增加爬取数据的频率、次数及数量，但该行为却会在不同程度上



对权利人的生活或网站运营产生不利影响，如：行为人过频访问或爬取用户数据，导致用户界面频繁出现授权弹窗，影响用户体验甚至无法正常使用手机软件；或者，行为人过频、过多地爬取相关平台或网站的数据，导致其服务器过载或崩溃，进而影响其正常运营甚至产生财产损失等。

此外，若行为人在爬取相关数据后，违反法律法规的规定或权利人授权文件的要求使用数据，如：篡改、删减所爬取的数据导致权利人名誉受损、产生错误判断或直接发生财产损失，或者通过简单搬运、抄袭他人成果，进而实质性替代他人产品或服务，同样会因违法违约或构成不正当竞争等情形而侵犯权利人的合法权益。以爬取企业数据为例，司法实践中，部分法院在认定数据爬取行为的正当性及合法性时，通常会考虑爬取行为是否对被爬取网站服务器的正常运行产生额外负担导致系统负载过高、是否造成网站运营者较大经济损失或者对其在市场上构成实质性替代等因素。

因此，在数据爬取及后续处理过程中，行为人应在遵守法律法规规定、符合权利人要求的同时，采取对权利人损失最小的积极举措，正确爬取及使用相关数据。

三、做好事后补救工作并建立日常合规管理体系

除事前评估及事中监管外，数据爬取行为人还应当重视事后的补救及应对工作。若行为人收到权利人的侵权通知或通过自查自纠发现存在或可能存在侵权情形的，应当第一时间停止爬取行为并核实是否存在侵权事实。若确实构成侵权的，行为人应当立即删除相关数据或



采取权利人认可的其他合理举措，通过积极应对来防止事件的进一步发展。

此外，在日常管理中，为防患于未然，行为人可以建立数据爬取的内部评估及管理制度、完善数据爬取的内部审批流程，做好各环节的留痕及实时监管工作。同时，行为人还可以结合自身的数据爬取情况制定行之有效的应急方案，并加强对内部人员的日常培训，以便能够有效降低法律风险的发生或在侵权事件发生时能够及时有效应对。

北京道商律师事务所

2025年3月【 】日