

# 开源大数据 锻造法律服务新渠道新产品的创新引擎

## ——法律服务领域制定实施蓝海战略反内卷

程立人

**[摘要]**内卷化及法律服务新渠道、新产品的开发是当前法律服务领域亟待解决的问题。解决问题需要政府推动，开源大数据、大数据库及大数据软件技术，找准法律思维和数据库思维的结合点，制定实施蓝海战略，在法律监督数据和手段供给、破产重整与大数据追逃、招商引资等方面发力，由此开辟一个全新的、非竞争性的服务空间，为全面依法治国、优化营商环境提供法律支持。

**[关键词]** 法律服务；大数据；数据库；开源

根据 360 百科和百度百科查询，法律服务，简称“法服”，是指律师、非律师法律工作者、法律专业人士(包括法人内部在职人员、退、离休政法人员等)或相关机构以其法律知识和技能为法人或自然人实现其正当权益、提高经济效益、排除不法侵害、防范法律风险、维护自身合法权益而提供的专业活动。本文以开源大数据、创新法律服务新渠道新产品为主线，就进一步夯实全面依法治国的数字化基础、做大做强做优法律服务业进行初步探索。

### 一、开源法律法规相关大数据是拓展法律服务渠道及空间的不二利器

“开源”这一词语，更多的用于软件领域，“开源”的目的是鼓励软件开发人员开放源代码、资源共享，以推动整个行业的发展。目前，法律服务与大数据的结合中存在数据量收集整理不足、数据库及数据开发程度不足、从事大数据开发的相关技术人才不足等导致的困扰，启动“开源”行动的要求更为迫切，意义更为重要，内涵更为丰富。将“开源”这一模式，移植到法律服务业，将为这一领域增加新鲜血液的补给，创造无限的机会。开源法律法规相关大数据，就是要将数据库思维和法律思维、大数据与法律服务有机结合，通过对大数据的整合、筛选、关联、比对、分析、挖掘等技术手段，创造法律服务新渠道新产品。就如何开源法律相关大数据，建议措施如下：

(一) 开源法律法规相关大数据，实现平台数据的资源共享和批量获取。目前的法律相关数据平台，公权力机关建立的主要有：全国人大常委会办公厅的国家法律法规数据库，最高法在互联网设立的中国裁判文书网等；社会组织设立的有：北京北大英华科技有限公司及北京大学法制信息中心的北大法宝，无讼网络科技(北京)有限公司的无讼系列网站及 APP 等。国家法律法规数据库开通仅有一年半时间，其作为法律法规数据的“国家库”“基础库”，具有极高的权威性和规范性，但目前仅仅提供已收录的法律法规预览、打印、下载等功能，存在法律法规收集不全、数据量与开发程度不足、数据文件无法批量导出等问题。法律法规等相关文件的产生、修改、废止的所有可公开的资料信息，应有一个便宜、高效获取的途径。与其他行业的大数据开源程度比较，明显看出法律服务行业远远落后。以金融行业为例，Pycharm(一种 Python 集成开发环境)通过安装第三方模块 Tushare，可随时导入你需要的股票或全部股票的日 K 级别、分钟级别、实时、分笔的股价数据及指数信息，以便于进行数

据分析。未来的法律服务应用软件也要增设这样的第三方模块，法律服务管理平台也要增设这样的开源数据接口，让有需求的机构、个人轻松获取其要的全部或部分法律法规相关大数据，为其搜索、汇集法律相关大数据减负。

北大法宝、无讼等对法律相关大数据的提供实行的是收费模式。这种收费模式一定程度上影响其自身发展，也影响了整个法律服务业的发展。只有开源，整个行业才有前景，蛋糕才能做得更大，大家能分到的利益比单打独斗、封闭运行时更多。各方要以开放共享的理念，改变“数据孤岛”现象，合力建设法律法规数据的“国家库”“基础库”，形成以提升法律服务效用为核心的全面开放、协同共生、共建共享的法律服务数字化生态共同体。比如说北大法宝，其已建成法律法规数据库、司法案例数据库、法学期刊数据库等多方面数据库，数据量大且数据开发程度相对较高，如果其走开源路线或是政府通过购买服务的形式支持其开源，将加速推动我国法律相关大数据的建设进程。

(二) 开发、开源法律法规相关大数据库，提供相关数据分析服务或提供数据库远程操作接口。世间万物都能用数据描述，数据大致可以分为三类。第一类是用于传播的媒体数据，如图片、视屏，这类数据只需要存储和读取，一般不需要做处理；第二类是数字类数据，其价值很高，目前是传统大数据处理与分析的主要方面；第三类是自然语言数据(NLP)，NLP是目前大数据处理和分析的新兴领域<sup>[1]</sup>。法律法规相关大数据在以上三类均有海量数据。如果只是将数据收集后简单堆砌存放，在海量数据中查找数据会很耗时、困难，所以需要—个工具来帮助管理数据，那就是数据库。最常见的数据库模型主要是两类，即关系型数据库和非关系型数据库。当前，关系型数据库应成为重点建构、研究方向。

前面提到的法律相关数据平台均已有关系型数据库的雏形，具备可扩展的升级潜力。比如说中国裁判文书网：网站的主体功能为检索、查看、收藏、分享、下载等；快捷检索支持关键词联想推荐，包括案由、关键词、审理法院、当事人、审理人员、律师、律所、法律依据八个类型的信息；高级检索可以实现多个信息项组合检索，包括全文、案由、案件名称、案号、法院名称、法院层级、案件类型、审判程序、文书类型、裁判日期、审判人员、当事人、律所、律师、法律依据等共计15个信息项；开通“批量收藏”、“批量下载”等功能。但该平台提供的服务距大数据分析、数据挖掘的现实需求有差距，主要问题在：尽管提供“批量下载”功能，但批量的获取渠道并不简捷，且下载下来的只是一个一个的文件，不是以数据库方式存放的数据，不利于进行数据分析；数据库中的表、字段需进一步扩充、优化等。目前系统中检索字段已包括审理法院、审理人员、律师、律所等，仅凭这几个字段就可以统计出律所与法院、律师与审理人员的“交集”。如果再增加一个字段表示案件受理、胜败诉情况，就可从“交集”和裁定、判决结果中进行进一步的数据分析，挖掘律师与法官的可能有不当交往的“痕迹”。如果对“法律依据”字段扩展成多个可做数据处理的字段，就可以核查同一法院、同一法官对类似案件审理中适用的法律口径是否有重大出入的问题。类似的可扩展数据分析、挖掘的情形不胜枚举。

以大数据库为基础的人工智能系统在审查合同、智能阅卷、法律文书自动生成等方面也有巨大的应用空间，可以帮助律师有效减少投入、降低成本。如2018年2月，由以色列法律科技公司LawGeex研发的法律AI平台与20名经验丰富的律师进行了一场人机竞赛，比赛内容是4小时审查5项保密协议(NDA)，并确定30个法律问题，包括仲裁、保密和赔偿。比赛的结果是，人类律师的平均准确率达到85%，而AI的准确率达到了95%；AI在26秒内完成的任务，人类律师平均需要92分钟<sup>[2]</sup>。国内的幂律智能(PowerLaw AI，一家专注于法律领域的初创人工智能公司)开发的合同生命周期管理(CLM)产品MeFlow、智能合同审查产品MeCheck、幂律AI开放平台也在这些领域有不俗表现。

大数据的收集整理、数据库的建立是一项耗时费力的基础性工作，处理、分析数据可能只需要几分钟、几行指令代码，但数据库的建立用时可能要以月、年来计量。如何建库、如何设计库中的表格、如何确定表中的字段，这是一个要重点研究、讨论的问题。而且，这种关系性数据库的建立与研究，伴随法律理论前沿的纵深推进，能同时推进法律理论的研究，也能为立法工作提供支撑和参考。2021年2月24日，中共中央政治局委员、全国人大常委会副委员长、中国法学会会长王晨在国家法律法规数据库开通仪式上的讲话中指出：“建设国家法律法规数据库，是学习贯彻习近平法治思想、推进全面依法治国的一项基础性工作，是“十三五”国家政务信息化工程的一项重要任务。”他同时指出，建设国家法律法规数据库是一项长期的系统工程，目前只是迈出了第一步，今后还有大量艰巨繁重的任务。这表明，法律法规数据库的建立、完善，已入国家的长远规划和顶层设计范围之内。我希望，这类数据库能以开源模式运行管理。对现有的数据库，如果不适合彻底开放，可考虑进一步提供、优化相关数据分析服务（如北大法宝），或提供数据库远程操作接口，提供给客户可对数据库进行处理的权限，共同开发、用好数据库。

（三）开发、开源法律相关大数据软件技术。2017年，国务院印发《新一代人工智能发展规划》，明确提出推进社会治理智能化，其中包含建设智慧法庭、建立人工智能法律法规、培育“人工智能”+“经济、社会、管理、标准、法律”等的横向复合型人才等要求。目前，国内从事法律服务软件开发的人数较少，同时拥有法律、数据库思维及技能的人数更少。解决这一问题的途径在于，加强法律相关大数据软件技术的开发、开源，鼓励一批有意愿研究法律服务和法律理论的律师、法官、检察官等，通过软件开源获取代码，凭借其从业经验、工作需要和法律、计算机方面的知识技能，拷贝、修改、创新程序代码，并以开源共享的方式不断改进，以推进法律法规相关软件行业的发展。开源软件的意义及其成效，常见诸网站、期刊等，此处不再赘述。

## 二、蓝海战略的制定实施是法律服务新渠道新产品的创新引擎

幂律智能在无讼研究院网站中撰文《法律圈“内卷”有多严重》指出：法律圈是内卷化程度较高的行业之一，势头堪比北京的沙尘暴，漫天黄沙，天昏地暗。历史表明，产业在不断被开创和扩展，其条件和边界也不是一成不变的，企业个体可以重塑这些条件和边界，不必在给定的市场空间内残酷竞争。蓝海的开拓者不将竞争作为自己的标杆，而是遵循另一套完全不同的战略逻辑，即“价值创新”（value innovation）。这是蓝海战略的基石。之所以称为价值创新，原因在于它并非着眼于竞争，而是力图使客户和企业的价值都出现飞跃，由此开辟一个全新的、非竞争性的市场空间<sup>[3]</sup>。不远的将来，为客户创造价值的法律服务新时代必将到来，商业世界的场景理论、蓝海战略将会在法律服务领域得到大量应用。以下，将根据个人视野所能及、主流研究重点聚焦内容不重复论述的考量，结合蓝海战略重建市场边界的基本法则，就如何创新法律服务新渠道新产品谈几点设想：

（一）拥抱监督，助力监督，创新有利于法律监督的数据信息和手段。“法律监督”是人大行使监督权的一种类别，同时也是宪法和法律赋予检察机关的重要职责。人大法律监督主要针对带有普遍性、典型性的问题，个案监督由检察机关负责。单就一事、一案去判定对与错，可能看不出有什么问题。但如将多件案件的办理信息以数据库的形式存放，通过关键字段的关联比对，会发现类似的案子判决结果畸轻畸重、引用法律依据不同等一系列问题。这种操作模式能让人大找准必须予以监督的共性问题，检察机关迅速排查出需纠正的违法个案，有利于提高法律监督的成效。

2022年2月7日晚，司法部通过官网发布《司法部有关部门负责人就全国律师诚信信息公示平台建设相关情况答记者问》。其中提及：“目前，全国律师总人数达到57.6万，律师事务所超过3.6万家。”而现有员额法官为12.8万左右，律师人数是员额法官人数的4.5倍。由于人数不对等，工作压力不均衡、角色定位的不同，律师要勇于去补人大、法官、

检察院工作上的短板，助力法律监督，创新有利于法律监督的数据和手段，优化自身的执业环境，优化当事人的诉讼环境，让大家都相信法律，愿意用法律渠道解决问题，法律服务市场也将因此扩容。在2017年全国司法体制改革推进会上，无讼创始人蒋勇说：“公检法对以大数据、人工智能为代表的科技革命的重视和积极拥抱，为司法改革打开了广泛的发展空间。相比之下，律师行业还有很大的进步空间，我们要迎头赶上，为司法改革作出应有的贡献。”<sup>[4]</sup>无讼网站“产品和服务”下的“赋能机构”模块中，“拥抱监管”成为介绍该模块的第一项标题。拥抱监管，也要助力监督，法律工作者要群策群力，创新法律监督工作方式方法，努力实现习近平总书记提出的要求，“努力让人民群众在每一项法律制度、每一个执法决定、每一宗司法案件中都感受到公平正义。”同时，这种创新可能达到“重设客户的功能性或情感性诉求”的效果，人大、检察院系统有可能将这一块业务纳入法律服务采购项目、法律顾问范围。

（二）紧盯执行，助力执行，共建逃废债的大数据追逃新模式。近年来，人民法院全力攻坚，执行工作取得显著成效，但与党中央提出的“切实解决执行难”目标和人民群众期待相比还有差距。法院的生效判决无法执行，被群众戏称为盖有官印的“白条”，严重影响公权力部门的形象。执行生效判决的财产部分，关键是要找到可执行的财产，重点是要追回那些出于逃废债目的而违法处置的财产。重大财产的处置，多数会在相关大数据管理系统中“留痕”，运用简单的指令代码和基本的法律常识足以搜寻到其违法交易的痕迹，经确认后可将违法的交易予以撤销，将变相处置的财产追回并归位到应予以执行的财产范围内。

《最高人民法院关于限制被执行人高消费的若干规定》第二条规定：“人民法院决定采取限制高消费措施时，应当考虑被执行人是否有消极履行、规避执行或者抗拒执行的行为以及被执行人的履行能力等因素。”《最高人民法院关于公布失信被执行人名单信息的若干规定》规定的纳入失信被执行人情形包括：有履行能力而拒不履行生效法律文书确定义务的；以虚假诉讼、虚假仲裁或者以隐匿、转移财产等方法规避执行的等。以上规定的出台，实际上反映了法院系统在执行中的困境：从理论上来说，法院如果认为被执行人实际上有履行能力，就应该找到、追回实际归属其的可执行财产，而不是简单化处理将其纳入“黑名单”；从实际工作情况来看，逃废债现象较为普遍，应接不暇的追逃案件，仅凭法院系统一家无力承担，所以只能简单化处理，“执行难”的问题长期无法破局的根源在于执行的人手不够。建议依据法律法规相关规定，对失信被执行人现有财产、以前一段时间及今后的财产处置、现金流出等相关信息予以公开，使其成为“玻璃人”，与其债务相关的债权人委托的律师、法官均拥有查询、处理、分析该信息的权限。这种操作模式，会让逃废债行为通过大数据处理和无数双眼睛的盯梢中曝光，会让法院借用外脑、外眼精准锁定可执行财产，破解“执行难”的问题。有能力的律所、律师要抓紧培育这方面的大数据处理能力，“跨越时间参与塑造外部潮流”，在即将到来的大数据追逃潮流中勇立潮头。

（三）盘活僵尸企业，护航企业退出，运用大数据在企业破产或非诉重整中探索出一条“护企暖企”途径。当前，已有相当一部分企业陷入债务危机。危机解决的途径主要是两条，一是向法院申请企业破产，在破产、重整、和解三种方式中争取到一种最符合自身发展的方式；二是争取通过非诉重整途径解决。直接破产是最便捷却也是最不得已的方式：破产后的企业厂房、设备只能以“白菜”价格处置，导致各方面的利益受损，失业人数增多，寒了企业家愿意继续投资的心；另外，这也可能是个别居心不良的企业主逃废债“优选”的方式，对银行信贷资产安全运作形成极大威胁。有实力的律所，要重新界定有企业债务处理需求的买方群体，从仅仅靠法院指定破产管理人来开展债务处理业务，向积极开拓非诉重整业务方面同时争取破产管理人业务转变。要学习中和正道集团在这方面的经验，积极提高自身的影响力，协调好债权人与债务人各自诉求，本着共赢的理念让企业度过暂时性困难，达到共同止损的目的。律所或个人如果作为破产企业管理人，要运用大数据追逃模式，考虑自编、购

置软件来辅助破产业务的开展,将作为破产管理人接管的破产企业待分配资产这一块蛋糕做大,提高各方面债权人的受偿比例和管理人的自身报酬。

### 三、政府及其部门的积极作为是做大做强做优法律服务业的关键推手

法律服务行业要想做大做强做优,政府及其部门需要进一步积极作为,发力推进,创造条件,创新举措。以下,就政府及其部门如何推进法律服务和优化营商环境,提三条建议,供参考。

(一) 试点破产业务或非诉重整业务。鉴于当前较为复杂的经济形式和中小城市律所复合型人才缺乏的现状,建议政府先行先试,组织机关中的法律、CPA、大数据处理人才等,集中攻关,在破产业务或非诉重整中找出一条“护企暖企”的可复制模式,保护好一部分有潜力的企业和企业家继续创业的信心。

(二) 将招商引资业务纳入法律服务范畴。招商引资已成为各级政府重中之重的工作。在经济高质量发展要求下,各地招商引资形式正由大包大揽转向精准发力,切实增强招商引资实效性。让律师参与招商引资工作,既起到防范法律风险的效果,又因为政府、大中型企业基本上都有法律顾问或公职律师、公司律师,通过这些律师的互动为政府和企业搭起一座招商引资的桥梁。

(三) 做好开展大数据处理的基础工作。根据作者本人多年来从事部门大数据处理的经历,数据获取难度、数据质量及兼容性存在不少问题。建议政府数据资源管理部门,要尽力汇集各部门的大数据库动态信息内容。如果该部门的数据库后台不归本级管辖,可通过定期从上级部门导入、更新数据的方式获取静态数据。由于各部门的数据格式要求不统一,对数据比对、查询、关联造成困扰,要组织各部门对数据进行净化处理,统一用于关联比对的字段的数据格式。按照法律法规及政策规定,开放数据库对外接口,合理分配账号及处理权限,供相关机构和个人开展大数据处理和个案查询。

党的十八大以来,以习近平同志为核心的党中央从坚持和发展中国特色社会主义的全局和战略高度定位法治、布局法治、厉行法治,引领我国社会主义法治建设发生历史性变革、取得历史性成就,开创了全面依法治国新局面<sup>[5]</sup>。全面依法治国最广泛、最深厚的基础是人民。当前,人民群众办事依法、遇事找法、解决问题靠法的观念不断增强,对立法、执法、司法、守法等各方面都有了更高要求。法律法规数据库为全社会学法用法提供了一个有益平台,要推动社会各方面共建、知晓、运用法律法规数据库,推动全社会形成尊法学法守法用法的法治氛围,推进法律服务向需要覆盖的一切领域依法、合理、有序扩展。

作者单位: 安庆市司法局

### 参考文献

[1] 孟兵 李杰臣编著:《Python 爬虫、数据分析与可视化——从入门到精通》,机械工业出版社 2021 年 5 月第一版,第 298 页。

[2] 张明新 李冠叶:《大数据时代我国法律服务行业的新发展》,《江苏师范大学学报(哲学社会科学版)》,第 46 卷第 6 期,第 96 页。

[3] 中国注册会计师协会组织编写：《公司战略与风险管理》，中国财政经济出版社 2022 年 2 月出版, 第 169 页。

[4] 张配豪：《传统律所的“无讼”式转型》，《人民周刊》 2017 年第 14 期, 第 41 页。

[5] 王晨：《建设好运用好国家法律法规数据库为人民群众提供权威便利公共法律服务》，2021 年 2 月 24 日在国家法律法规数据库开通仪式上的讲话。