



如何让大数据资源池发挥显著功效

业务数据资源化 大家谈

杨程鹏

2013年以来,全国四级检察机关上线运行检察业务应用系统,实现了检察业务数据化。经过多年沉淀和扩容,检察系统已经积累了大量检察业务数据,形成了可进一步利用的数据资源池。如何激活这“一池春水”助推检察工作高质量发展?作为检察业务工作中枢和业务数据管理专责部门的案件管理部门责无旁贷。笔者从宁夏回族自治区检察机关业务数据资源化实践入手,浅谈对业务数据资源化的理解和把握。

业务数据为什么要实现资源化

“数化万物,万物皆数”已是当今数字时代的重要特征,每一个组织和每一个个体都是一个“实时数据库”,每时每刻都在产生各类数据。如果这些数据能够被有效收集、汇总、加工、利用,其产生的生产价值、社会价值将是巨大的。这也是国家层面将数据与土地、劳动力、资本、技术一起并列在生产要素的重要原因。为有效应对数字时代的新机遇、新挑战,国家已经从立法、制度机制等方面进行了顶层设计。检察机关自检察业务应用系统上线以来,已实现个案线上办理,沉淀了大量个案数据,但这些个案数据汇总形成的大数据资源池并

没有得到充分利用,也没有发挥应有的价值,主要原因在于内部数据耦合性强、缺乏外部数据支撑、数据应用创新不够。为解决以上三个问题,需要对业务数据进行资源化处理。

什么是业务数据资源化

检察业务数据来自于检察业务工作,最终又要服务于检察业务工作,即业务生产数据、数据反哺业务。要实现这一目的,需将业务数据改造成为可以被进一步利用的生产要素,这个过程称之为业务数据资源化。业务数据资源化主要包括数据结构重构、数据多源重组和数据业务重塑三个方面。

一是数据结构重构。检察业务应用系统是以案件为中心进行业务数据设计的,案件数据的耦合性强、关联度高、逻辑性强,这种设计模式对于单个案件的办理是合适的。但在大数据应用中,需要以涉案人员为中心进行业务数据结构设计,现有数据结构便不能很好地满足实际需要,需要对现有数据结构解构后再重构,才能进一步使用。具体是以涉案人员的唯一编码为主标识,建立涉案人标识、涉案人标签(属性)和涉案人案件三个模块。为适应外部数据重组需要,建立涉案人模块,包括主标识(身份证号)及其他关联标识(手机号、社交账号等);为进一步记录涉案人信息,建立涉案人标签(属性)模块,包括涉案人的个人基本信息(年龄、性别、职业、学历等)及其他相关标签(是否是未成年人等);为关联涉案人案件信息,建立涉案人

案件模块,包括案件基本信息、案件办理信息等。

二是数据多源重组。大数据应用的价值在于通过多源数据之间的比对碰撞分析,发现规律或者异常,但由于研发单位不同,多源数据在数据标准、数据结构、存储方式等方面是不统一的。为此,需对多源数据进行重新组合。具体可建立多源数据元数据库(元数据库本质上即为构造数据库的基础信息,以此为基础可建立同样结构的数据库)、多源数据资源目录数据库、多源数据原始数据库三个数据库。为实现多源数据库逻辑结构识别,需要建立多源数据元数据库;为验证多源数据库的完整性、一致性,需要建立多源数据资源目录数据库;为保存原始多源数据,需建立多源数据原始数据库。

三是数据业务重塑。最高检提出要大力实施数字检察战略,破解检察机关法律监督职能作用发挥还不够充分这一核心问题。大数据法律监督工作是数字检察的一个重要突破口,就是要充分利用内外部数据实现对检察机关法律监督工作的系统性重塑,实现从个案办理到类案监督再到系统治理的新型业务模式。一线办案检察官要破除过去个案办理的思维定式,要敢于从个案办理中发现大数据法律监督规则,善于总结提炼形成大数据法律监督模型,精准发现类案监督线索,将被动式监督变为主动式监督,实现监督理念、手段、方式、成效的系统性重塑。

业务数据资源化如何展开

以宁夏为例,今年以来,宁夏回

族自治区检察机关大力推动数字检察战略实施,出台《宁夏检察机关大数据赋能新时代法律监督实施方案》,明确推进数字检察战略实施的路线图、时间表;修订《宁夏回族自治区人民检察院对各市级人民检察院检察业务工作考评办法(2023年度)》,将数字检察工作列为考评范围并加大考评分值;举办全区检察机关大数据法律监督模型竞赛。其中,银川市检察院建设的大数据法律监督平台,优化重组检察业务结构化数据,识别并整合公安、法院的非结构化数据(不能直接用有关系型数据库存储并处理的),创新法律监督方式,通过将公安、检察院和法院制发的刑事案件起诉书意见书、起诉书、判决书进行“三书”比对,发现监督线索439份、成案36件。具体做法是对公安、检察院、法院的纸质起诉书意见书、起诉书、判决书进行OCR识别、按照预先设定的数据标准自动存储,建立刑事案件“三书”数据库。在此基础上,建立刑事案件“三书”数据库之间的关联关系,并与检察业务应用系统中结构化数据建立关联关系,进而建立大数据法律监督模型,开展类案监督和系统治理工作。

下一步,宁夏回族自治区检察院将聚焦建设集数据治理、数据返回、模型管理、线索管理于一体的大数据法律监督管理平台,不断提升业务数据治理能力,完善数据安全保障体系,助推全区检察机关法律监督工作实现系统性重塑。

(作者为宁夏回族自治区人民检察院案件管理办公室副主任)

房文龙 郭树合

法律监督模型是开展大数据法律监督工作的重要工具,检察机关依托法律监督模型开展法律监督工作,可有效提高法律监督工作质效。法律监督模型的建立和应用需要大量的数据支持,数据的规范性、有效性和安全性对于法律监督模型的构建至关重要。

数据信息的规范性

一是建立数据采集和管理机制,明确数据的来源和采集方式。数据采集的方式通常包括人工采集和自动采集,人工采集如现场调查、进行庭审记录等;自动采集如自动从检察机关内部数据、互联网数据等处抓取。可采用多种渠道获取数据,并对数据进行交叉验证和比对,确保数据的来源可靠。

二是制定数据的标准和格式规范,确保数据的一致性和可比性。在法律监督模型中,文本、图像等不同的数据类型和指标需要采用不同的标准和格式规范。检察机关可通过建立数据标准化机制,对数据进行分类整理,建立数据仓库和数据分析平台,方便数据的查询和使用,还可以采用数据验证技术保障数据标准化的实施。

三是对数据进行严格的审核和校验,避免出现数据错误、遗漏的情况。除了数据验证技术外,还可以采用数据质量评估方法,帮助识别和解决不良数据问题,确保数据的可靠性和有效性,常用的数据质量评估方法包括描述性统计分析、异常值检测、数据完整性检查等。

数据信息的有效性

一是科学建立数据仓库。数据仓库是一个集成、稳定的数据集,用于支持办案、决策和管理。在建设数据仓库之前,需要明确业务需求,确定需要哪些数据以及如何使用这些数据,并同步考量数据的时效性和稳定性等因素。在确定好技术方案开始实施数据仓库的建设时,要完成数据的采集、清洗、转换和加载等工作。数据清洗方面,原始数据中可能存在重复记录、缺失值、异常值等问题,需要通过数据清洗的方式进行处理。数据清洗的方法包括删除重复记录、填充缺失值、修复异常值等。

二是科学设置数据类型和指标。根据不同的需求选择合适的数据类型和指标,确保数据的针对性和实用性。在法律监督模型中,不同的应用场景需要采用不同的数据类型和指标。例如,在刑事案件中,需要采用犯罪记录、行为轨迹等数据来分析犯罪嫌疑人行为模式;在民事案件中,需要采用财产关系、合同协议等数据来分析案件的事实和法律依据。

三是科学设置算法规则。不同的算法模型适用于不同的监督对象和监督内容。因此,在科学设置算法规则时,需要根据具体的监督对象和监督内容选择合适的数据模型,并考虑算法模型的复杂度和可解释性,以确保算法的可靠性和稳定性。在科学设置算法规则时,需要建立高质量的数据集和标注规则。数据集应该涵盖各种类型的法律文书,并尽可能地覆盖各种情况和细节。标注规则应当清晰明确,避免歧义和误判。同时,还需定期对数据集和标注规则进行更新和优化,以确保算法的准确性和可靠性。

数据信息的安全性

一是确定不同用户的访问角色和权限。可以设置管理员、审核员、查看员等不同的角色,根据角色的不同设置相应的权限。管理员可以对所有数据进行操作和管理,审核员可以对审核任务进行管理和审核,查看员只能查看已经审核通过的数据。同时,还需要考虑角色之间的相互制约关系,以避免越权操作和滥用权限。为了满足不同用户的需求,需要对权限进行合理细分,可根据不同的数据对象、数据类型、数据范围等因素设置不同的权限。例如,可以设置对某些数据对象的编辑权限或只读权限,以保护数据的机密性和完整性。同时,还需要考虑权限的承接性,即当一个用户的角色发生变化时,其相应的权限也需实现自动更新和调整。采用加密技术和身份认证等手段,防止数据被非法获取或篡改。在法律监督模型中,可采用多种身份认证技术,如口令、指纹、面部识别等。其中,生物特征识别技术具有较高的安全性和便捷性,被广泛应用于法律监督模型中。

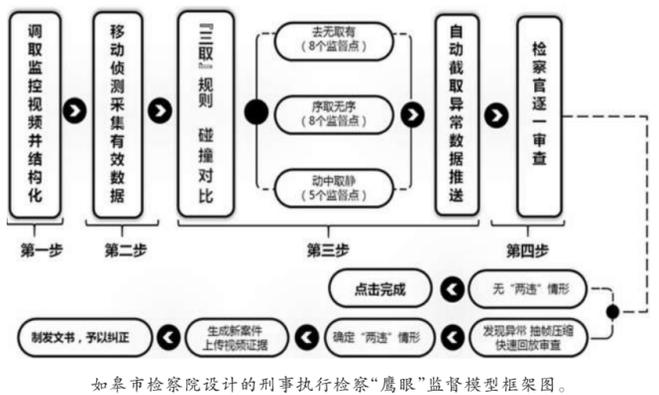
二是完善数据备份和恢复机制。在模型构建过程中,数据备份和恢复机制可以在数据丢失或损坏时快速恢复,确保法律监督模型的正常运行。因此,在构建法律监督模型时,可以设置每天、每周或每月等不同的备份频率,以满足不同场景下的数据备份需求,并通过磁盘镜像、数据库备份、文件还原等技术实现数据备份或丢失数据的恢复。

三是定期进行安全审计和漏洞扫描。通过建立审计日志和监控机制,记录用户的操作行为和访问记录,定期进行分析和评估,及时发现和修复潜在的安全问题。一旦发现异常行为或者安全事件,应立即采取相应的措施进行处理和防范。同时,还需建立监控机制,对系统进行实时监控和预警,以保证系统的稳定性和可靠性。

(作者单位:山东省东营市河口区人民检察院、山东省人民检察院)

数据信息应具备的三个特性

“鹰眼”确保每一秒监控都在眼前



如皋市检察院设计的刑事执行检察“鹰眼”监督模型框架图。

葛林卫 朱志军 戴莉莉

监管场所重大事件、事故或隐患中,往往隐藏着监管人员违法违规甚至是犯罪的行为,对监控视频进行审查是检察机关开展监所检察监督最有效的渠道和方式。但监管场所每天都会产生海量视频数据,对此,检察机关难以及时、全面审查,往往会导致监管场所违法违规行为及司法工作人员职务犯罪发现难、查处难,严重制约着新时期刑事执行检察工作的高质量发展。

为维护监管场所秩序稳定,精准发现违法违规行为,在江苏省检察院和南通市检察院的指导帮助下,如皋市检察院研发了刑事执行检察“鹰眼”监督模型(以下简称“鹰眼”)并投入使用。“鹰眼”依托视频结构化、AI识别和抽帧压缩(即以100倍

速将监控视频资料抽帧压缩,保证不丢失一秒监控视频画面,并将审查效率提升100倍)三项关键技术,利用数据碰撞实现AI智能和人工审查相结合,有效解决了监管场所普遍存在的监控录像多、检察人员少、数据利用难等问题,打造了全过程可监督、全节点可查寻、全链条可溯源的刑事执行检察智慧新模式。

【模型构建】

数据汇集:数据来自监管场所监控视频,包含管谈话室、讯问室、过道、监室、户外活动等各项功能。该监控视频由“鹰眼”自动调取备份,数据来源直接、可靠,实现监管场所监控行为的全覆盖,有效确保数据的真实性、全面性。

数据清洗:“鹰眼”通过AI多算法融合协同技术,对海量数据进行转换处理,将非结构化数据按照人员身

份(在押人员、民警、医生、工勤人员、陌生人等)、出现时间、位置(监室等)、事件(人员聚集、倒地、按报警器)进行标签化处理,将其解构为结构化数据,建立数据仓。在此基础上,利用AI智能算法将无人、无车辆的视频智能过滤,保留“有人有车”画面,后按照“三取”规则(去无取有、序取无序、动中取静),自动截取“应当有人”等正常画面,留取“不应当有人”等异常情境画面;自动截去有序活动镜头,留取“三人以上聚集”等异常行为画面;自动截去正常动态镜头,留取“人员长时间不动”等异常静止画面。通过对数据自动精准筛选、对比耦合、抓取识别,浓缩提炼有价值的视频。

数据模型:依据最高检对“两违”(监管场所或监管人员违法违规行为)工作的具体要求和看守所管理规定,结合一线刑事执行检察工作实际,该院科学设置21项运算子规则,对可能发生的监管人员虐待被监管人、违规约束被监管人、被监管人救治不及时等违法违规情形。“鹰眼”比照21项子运算规则自动识别并截留异常监控录像,即时推送给办案人员逐一审查。一旦发现违法违规线索,办案人员即利用该院自主研发的抽帧压缩技术,在保证不丢失一秒监控视频画面的同时,以100倍的倍速对抓取的异常画面进行抽帧压缩,压缩后的待审查视频存储大小减少至百分之百,人力审查时间缩短至百分之一,实现了人工智能要时间、要空间的目标。经审查确认系违法违规行为后,办案人员通过制发纠正违法通知书或检察建议的方式予以纠正,“鹰

眼”截留的监控录像将被作为办案证据上传附件,确保有据可依。

【模型应用实效】

“鹰眼”借用大数据的力量,颠覆了“大海捞针”式的原有监督模式和“一张嘴、一支笔、一张纸”式的传统巡回检察监督模式,直接把证据和线索摆到检察官面前,极大提高了刑事执行检察监督质效。自“鹰眼”运行以来,累计发出预警593次,准确率达91.2%,依托该模型,检察官精准发现日常监管和在押人员权益保障等方面存在的“两违”线索189条,据此发出的78份检察建议或纠正违法通知书均被采纳。

今年4月,江苏省检察院将“鹰眼”向全省推广应用。目前,该系统已被拓展运用到监狱检察、巡回检察和指定居所监视居住检察等多项刑事执行检察业务和各类专项检察中。

此外,如皋市看守所主动与如皋市检察院会签文件,共同构建跨部门大数据协同办案规范化制度体系,取得较好成效。当地看守所“两违”行为大幅下降,并借鉴“鹰眼”自主研发执法动态AI系统,相关工作经验在全省监管系统推广。经过一年多的运行,“鹰眼”一方面凸显了其应用场景的广阔性和实用性,有效提升了刑事执行检察工作质效,另一方面倒逼看守所执法规范化水平不断提高,真正实现了双赢多赢共赢。

(作者单位:江苏省南通市人民检察院、江苏省如皋市人民检察院)

近看模型

检察长眼中的数字检察

黑龙江省佳木斯市人民检察院 检察长 王媛媛

数字检察历经智慧检务、智慧检察等发展阶段应时而生、乘势而上、聚势而强,已经成为关乎检察工作现代化发展历史进程的系统性、革命性工程。黑龙江省佳木斯市检察机关主动拥抱、积极回应数字革命浪潮,全面加强数字意识和数字思维,及时将大数据方法、数据模型应用与检察工作相结合,以大数据赋能高质量办案为切入点、发力点,不断实现大数据在服务大局、检察为民工作中的赋能功效。

“眼睛向内”提质效,数据赋能法律监督更有刚性。近年来,我院注重从内部入手、从主责主业抓起,以数字检察推动“四大检察”全面发展。聚焦刑事检察监督线索发现难点问题,建立2021年以来生效刑事判决数据库,探索开展审判监督与侦查

做实数据赋能法律监督需“身体力行”

监督线索碰撞识别工作,及时发现抗诉案件线索6件,成功办理抗诉案件3件;发现漏罪漏犯线索5件,成功追诉漏犯5人。聚焦民事检察“不专”“不会”等问题,着力拓展监督的广度和深度,建立民事公告送达裁判文书数据库,完善“大数据初步筛查+办案人员介入审查”工作模式,据此办理民事公告送达违法监督案件349件。聚焦行政检察“不敢”“不力”等问题,以大力推进检察机关内部刑事检察与行政检察反衔接工作为突破口,建立数字检察领导小组办公室、侦查监督与协作配合办公室、案件管理办公室一体化运行机制,对近年来行政法机关移送公安机关的刑事案件及已办结492件刑民不衔接案件开展数据比对,梳理发现刑民衔接案件线索40件,目前已制发行政违法行为监督检察建议2件。

“视野向外”促发展,数据赋能服务大局更有准度。我院牢牢把握围绕中心、服务大局的首要政治任务,以数字检察夯基固本、稳预期、利长远的作用。我们牢记习近平总书记“粮食安全”是“国之大者”,耕地是粮食生产的命根子”重要指示精神,依托数据赋能,构建起黑土地刑事检察与公益诉讼检察综合保护格局。我院对近年来办理的110件非法占用农用地案件进行大数据分析解构,发现公益诉讼线索48件,提起黑土地领域刑事附带民事公益诉讼11件,督促清理、复垦非法占用土地800余亩,追缴涉农补贴资金12万余元。坚决保护国有财产安全,通过开展数据赋能环境保护征收公益诉讼检察专项监督工作,与环保部门、税务部门共同建立涉税排污单位信息库,对1241家排污单位开展数据比对工作,排查出应缴未缴环境保护排

污单位48家,据此制发诉前检察建议9份,督促相关行政部门追缴税款、滞纳金80余万元。

“脚步向下”办实事,数据赋能检察为民更有实效。我院坚持以人民为中心、让人民满意为各项工作的出发点、落脚点,以数字检察助力做实人民群众可感受、能体验、得实惠的检察为民工作。持续推进佳木斯市检察机关“东极暖阳”强制报告法律监督平台试运行工作,以强制报告制度落实为切入点,相继引入附条件不起诉考察帮教、社区矫正法律监督、未成年人检察公益诉讼线索发现等功能,以点带面推动未成年人检察综合履职。

同时,加强与涉及其他“五大保护”的相关部门沟通协作,打破数据壁垒,加强共建互促,在“掌上佳木斯”App开设“东极暖阳”未成年人保护板

块,为促进“六大保护”协同发力贡献检察智慧。守护人民群众的“养老钱”



佳木斯市检察院数字检察办公室组织相关人员开展数据分析研判。