大数据和智能化双轮驱动赋能检察工作高质效发展

□冯涛

在数化万物、万物皆数的今天，数字检察是大势所趋。

最高人民检察院检察长应勇指出，数字检察是数字中国的重要组成部分，是数字中国在检察机关的具体体现，其根本是赋能检察机关法律监督，促进检察办案更加公正、检察管理更加科学、检察服务更加精准，推进检察工作现代化。这深刻阐释了什么是数字检察、数字检察的目标任务是什么。

作为构建“业务主导、数据整合、技术支撑、重在应用”的数字检察工作机制的重要一环，如何做好技术支撑，是检察信息技术部门要做的“必答题”。

从1991年起步至今，检察信息化已经走过32年发展历程，完成了网络化、信息化、初级智能化三个阶段的建设工作，为检察工作进入数字化发展阶段打下了坚实的基础。具体来看，在网络化阶段，修通了“路”，建成全国检察专线网“一张网”，实现了全国检察机关网络互联互通的目标；在信息化阶段，建造了“车”，建成四级检察机关统一使用的全国检察业务应用系统，案件办理全流程实现网上进行；在初级智能化阶段，探索了初级的“辅助驾驶”，全国检察机关研发运用了案卡回填、量刑计算、“三书”比对等60余款智能化辅助办案工具。而在数字化阶段，则是要构建“人车路协同”的“智能车联网”，建设更高级别的“辅助驾驶”，甚至个别场景下的“无人驾驶”，推动检察机关实现整体工作的数字化转型。

数字化原本是一个技术概念，指将文字、语音、图像、视频等各种复杂多变的信息转变为可识别、可传输的二进制代码数据，进而由计算机进行统一处理。随着互联网特别是移动互联网的快速发展，手机、PAD等数字设备在社会中得以普及应用，不断改变了人们的生产、生活、行为方式和价值观念，进而深层次推动社会变革，此时，数字化已逐步泛化为一个社会概念，数字社会、数字经济、数字政府应运而生，数字化不再聚焦技术本身，而转变为聚焦如何依托信息化手段和大数据、智能化等数字技术，实现业务工作的变革与重塑。

围绕一网运行、一网通办、一网赋能、一网运维“四个一网”建设目标，数字检察整体框架从技术上看可以概括为“12321”：“1”指夯实一个数字基础设施，“2”指强化大数据运用能力和智能化运用能力，“3”指提升数字办案、数字管理、数字服务三大系统数字化应用水平，“2”指创新检察数字化体系和安全保障体系，“1”指构建一个上下贯通、横向联通的检察机关与其他单位融合发展的数字检察生态。

提升三大系统数字化应用水平是数字检察的核心，是实现检察工作变革与重塑的有效载体，其关键在于检察机关大数据、智能化运用能力的提升，只有这样才能真正实现赋能。三大系统中，数字办案大系统分为上中下三层架构，其中，上层为管理分析子系统，包括流程监控、质量评查、统计报表、数据分析等；中层为案件办理子系统，包括共享协同、数字监督、信访、线索管理、流程办案等；下层为业务赋能子系统，包括知识服务、智能化辅助工具、音视频、数据服务等。数字管理大系统包括检察办公子系统、业绩考核子系统和档案管理子系统等，覆盖办公、事务办理全流程全场景，实现机关工作“能上网尽上网”，并拓展各种智能化应用。数字服务大系统包括诉讼服务、社会服务和检察公开等业务板块，部署在互联网面向人民群众提供“告（控告、举报）、办（事项办理，如律师申请互联网阅卷）、查（查询，如查询信访事项办理情况）、询（咨询，如检察官答疑）、公开（检察司法公开，如检察文书公开）、专区（专项保护，如未成年人保护）”等六大类功能服务。

需要特别注意的是，前期各地探索建设大数据法律监督模型提升了检察机关法律监督能力，深刻影响改变了检察履职方式，但这并不完全等同于数字检察的全部，而只是数字检察非常重要的核心内容，数字检察是三大系统的深化应用，赋能整体检察工作的变革与重塑。

大数据平台和智能化平台共同构建了数字检察的技术底座，是驱动检察工作高质效发展的双轮，也是实现技术支撑的有效载体。其中，大数据平台聚焦数据采集、数据治理、数据组织和数据服务，实现数据的“聚”“融”“通”“用”，满足三大系统复杂的各类数据使用需求，同时也为智能化建设奠定坚实数据基础。数据采集要实现检察机关内外部数据汇聚，这里要特别指出，数据汇聚应以检察机关内生数据为主，对于外部数据，除互联网开源数据外，其他数据应采用原则使用接口、分布式计算等方式使用数据，做到数据能用不存、可用不可见；数据治理要通过数据清洗、转化等提升数据质量；数据组织要按照业务需求完成各类业务数据库建设；数据服务要负责对数字监督模型运行和数据查询、数据统计、数据分析等各类数据应用提供数据。智能化平台采用感知层、数据层、计算层、表示层四层架构，感知层对语音、视频、图像、文书等数据进行解析，将非结构化数据结构化；数据层通过数据清洗、数据标注等方式，将粗数据转化为满足机器学习训练要求的数据；计算层根据检察业务实际需求构建和训练各类机器学习模型；表示层负责各项办案、管理、服务流程使用计算层模型。

可以说，数字检察建设得怎么样，要从业务上看整体工作质量、效率、效果有无大的提升，从技术上看检察工作中大数据和智能化运用的含量是否多。在以大数据和智能化推进数字检察建设过程中，要把握好三个方面的工作：**一是数据智能应用基础化。**将大数据、智能化技术从零散应用整合成数字化能力底座，各项建设紧扣大数据、智能化技术在检察业务具体场景中的应用展开。**二是业务应用模型化。**模型化是大数据、智能化运用的一大特色，通过将检察人员的工作经验和知识转化为模型，形成能力模型（功能）嵌入流程、使用数据，推动经验、知识、规则跨地域、跨层级、跨条线共享共用。**三是系统建设体系化。**要将检察机关各项应用系统整合为办案、管理、服务大系统，打破系统间的壁垒和数据孤岛，助力检察内部数据自由流动，实现全领域、全业务、全流程、全场景赋能。

数字检察建设号角已经吹响，检察技术信息部门要以“等不起”的紧迫感、“慢不得”的危机感、“坐不住”的责任感，主动答好技术支撑这张试卷，加快推进、实施数字检察战略，赋能新时代法律监督，更好地以检察工作现代化服务中国式现代化。

（作者为最高人民检察院检察技术信息研究中心副主任）

数字漫谈

找寻数字检察的源头活水

□刘哲

开展数字检察工作，要坚持“眼睛向内”，深度挖掘检察数据价值，充分激活和利用更多的“沉睡数据”。从这一点看，检察官要做好数字检察工作，必须要向内发力。

为什么一定要强调用足用好内部数据？因为用内部数据成本低、风险小且更方便。我们拥有全国统一的检察业务应用系统，同网同源便于互联互通，方便对内部办案数据进行加工、聚合、管理、应用，这与检察办案工作关联紧密，是我们练好内功的基础。但该系统当前主要功能是办案，要想将其打造成数字检察“天网”，还需对海量数据的存储、调阅、查询、分析进行结构化改造。对此，笔者有三点建议：

一是建立“案卡+文书”的高级查询功能。目前，检察业务应用系统只能对案卡数据进行查询，且筛选条件和显示内容不够灵活，根据实践所需，还应该确保所有的案卡都可以进行高级检索，可以自由组合检索关键词，还可支持较为高级的复杂组合。检察业务应用系统就像一个论文数据库，没有全文检索功能万万不可。因此，有必要将目前的高级查询功能升级为结合案卡和文书全文信息检索的功能配置，要向内发力，首当其冲需优先实现这一功能。检察业务应用系统已经积累了十余年的办案数据，如果将法律文书全文检索的高级查询功能建立起来，这十余年来积累的大部分办案数据都可以得到激活。

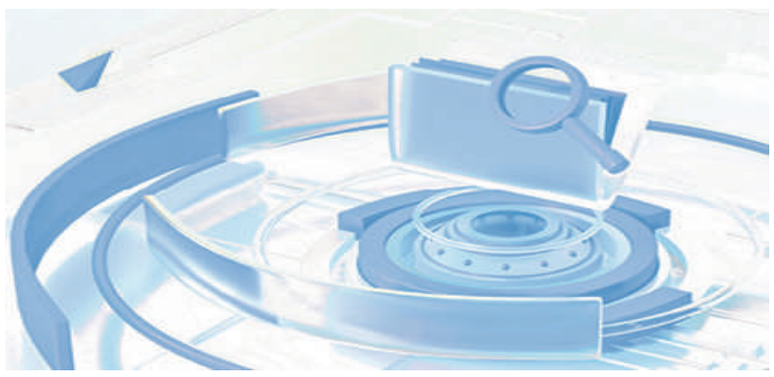
二是建立电子卷宗高级查询功能。只有文书的高级查询还远远不够，因为案卷中的很多信息，审查报告未必能够将其完全记载，不太可能将所有笔录都涵盖其中，很多书面证据更是不太可能包罗万象，比如对鉴定意见、审计报告、勘验检查笔录等一般只摘录其结论部分。虽然全文引用在个案中用不上，但在进行大数据分析时，每一个信息都至关重要，我们不仅要对审查报告进行全文检索，最好能够顺势对电子卷宗进行全文检索。实践中，我们也具备了对电子卷宗进行高级查询的技术基础和信息储备基础，实现电子卷宗高级查询功能是对“沉睡数据”的进一步激活。

三是留下“过手数据”，建立电子证据数据库。除了“沉睡数据”，我们还有不少“过手数据”没有留下来，这些数据就在随案移送的光盘里，比如讯问询问的同步录音录像、监控录像、手机和电脑的恢复数据和银行流水等电子表格，这些数据要么是无法打印，要么是打印起来数量过多，最终基本都以光盘的形式来存储运输。但目前的检察业务应用系统中只有电子卷宗，没有专门区域用于上传光盘数据，这就导致出现“案件走了数据也就跟着走了”的情形。

以手机的恢复数据为例，其中包括通讯录、通话记录、短信记录、聊天记录、交易记录、各种手机应用软件的信息，以及录音、录像等信息。这些数据合在一起就是一个人的数据档案，如果我们将这些单个的数据档案整合起来，就能够建立起涉案人员的广泛数据联系。又如银行流水信息和通讯记录信息，如果我们能够把大量的涉案人员信息联系起来，就可以建立起涉案人员资金往来和通讯交流的关系网。通过这些汇集电信诈骗、跨境赌博等某一类犯罪人员的电子数据，我们就可以摸排出此类行业的关系网络，从而更加容易发现哪些是漏网之鱼、哪些是关系节点。再如微信身份识别问题，只要我们串联起的人员关系网络足够大，就能够用自己的“过手数据”实现相当程度的身份识别。即使是看起来还不知如何用的监控录像和同步录音录像证据，也可以通过人脸识别技术进行扫描，看看本案的犯罪嫌疑人在其他案件的影像证据中是否出现过，从而较为容易地实现串并分析，收集更多的影像证据。

当把案卡、文书、卷宗、视听资料、电子证据等所有经办、经手的数据都整合起来之后，这些自有数据就可以极大地满足目前数字检察工作需要，比如进行广泛的身份识别、漏犯漏罪追查、案件串并分析、隐蔽犯罪关联等。不仅如此，这些数据还会不断得以累积，形成数字检察的源头活水，把这个源头活水利用起来，做大做强，数字检察就具有更加强大稳固的基础，能够真正实现行稳致远。

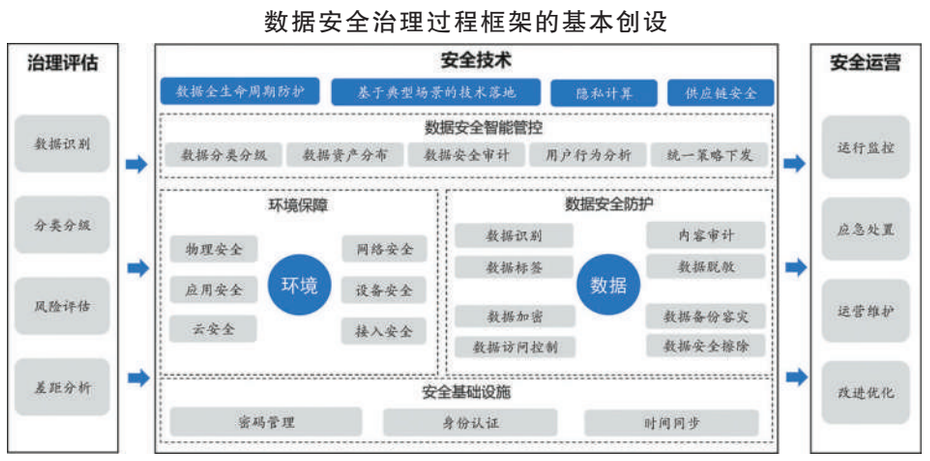
（作者为北京市人民检察院第一检察部副主任、三级高级检察官）



资料图片

数字视角

海量检察数据的安全性如何保障



□董岭　高彦恺

随着数字检察工作的深入开展，检察机关在大数据、人工智能等技术推动下，在开展主动监督、类案监督等方面取得显著成效，在此过程中积累的海量检察数据也得到进一步的扩展和融合。

数字检察扩展了检察数据收集的广度，“跨部门大数据办案”“行政执法和刑事司法衔接”“刑事数据调取”等跨网、跨部门业务协同场景打通了不同网络与单位之间的“数据孤岛”。数字检察推动了检察数据处理的深度融合，检察大数据法律监督模型的应用使得海量检察数据被挖掘出更多的业务价值。

高价值伴随高风险，在大量的应用场景背后，需重点关注的一个问题是如何保障海量检察数据的安全性？如何通过做好检察数据安全治理来构筑数字检察的基础底座？笔者从科技视角，寻求检察数据安全治理的可行性路径。

检察数据的安全治理应以全面评估为首要环节。技术人员应针对检察机关现有的业务应用进行调研、分析，确定业务间的关联关系、访问的关键路径、数据流向及演变过程，结合对基础安全管控措施的分析，明确各业务条线在推进数字检察工作中所面临的管理、技术等风险。

明确风险后，需要再进一步对检察数据进行分类分级。基于检察业务制定对应的数据分类分级实施规则，结合深度内容识别技术，包括关键字、正则表达式（指一个“规则字符串”，用其来表达一种过滤逻辑）、文件指纹、结构化数据指纹、智能分类等方式，对检察数据进行主动扫描，识别数据内容，最终形成检察数据的分类分级资产清单。此项清单是后期对不同类别、级别的检察数据进行针对性防护的依据。

技术能力落地是检察机关数据安全治理的核心环节。一方面，技术的落地要以检察业务场景为背景，明确不同业务场景的安全防护需求。另一方面，数据安全防护需要围绕数据全生命周期进行，包括采集、传输、处理、存储、销毁等环节。与此同时，防护手段要与分类分级结果相匹配，实现明确的、有侧重的防护。

各种典型业务场景与不同数据生命周期的交叉组合，显现出了不同的防护需求。技术人员可以从基于国密算法的敏感数据加密传输和存储、基于丰富脱敏算法的数据脱敏和基于不同检测方式的数据防泄漏等技术维度来匹配各种具体的防护需求。

另外，区块链和隐私计算技术（如同态加密、差分隐私、联邦学习）的发展，弥补了传统防护手段的不足，进一步增强了数据操作的溯源能力以及敏感数据的可用性。例如，基于隐私计算技术，通过构建统一隐私计算平台，不同检察机关通过输入加密后的检察数据，可以实现对此类数据的建模分析，在保证数据安全的同时推动数字检察技术发展。

随着检察大数据法律监督模型的提出，模型的训练也会涉及大量检察数据，因此，技术人员从模型的设计、研发、测试到后期应用等环节也应全面落实安全防护。例如，可采用安全需求分析手段，严格控制需求数据采集的范围；进行攻击面分析，完成系统架构攻击面收敛；对软件代码进行安全审计，完成问题代码修改完善；采用零信任技术，实现应用系统可信运维。

与此同时，检察数据安全治理应具备运行保障机制。技术人员可以通过构建可视化管控平台，对检察数据安全防护体系的运行情况进行常态化监控，对安全防护操作历史进行审计记录，对出现的异常行为进行监测告警、应急响应等，并定期总结运行情况并完成安全防护的改进优化。

（作者分别为九三学社河南省科技委员会副主任、信息安全专委会主任，天融信科技集团解决方案专家）

走进科技

数据管理的“智能好帮手”

随着全国检察业务应用系统2.0、统计系统2.0的上线运行，依托于案件流程、节点、文书、案卡等产生的结构化数据、非结构化数据量陡然增加，这也导致出现数据质量下滑和数据监管难度增加等问题。

在检察业务应用系统中，有300余种案件类别、4000余个文书模板、1000余张案卡、数万个案卡项目、300余张报表、上千万个数据点位。大量的案卡项目让检察官在填录时容易产生混淆，导致案卡填录质量参差不齐，这也给案件核查工作带来难度。

向科技要生产力成为不二选择。依托技术创新，构建一套好用、管用的案件管理数据质量监管系统，也成为实施数字检察战略的重要一环。完备的案件管理数据质量监管系统可以依据法律法规和办案规范设计校验规则，通过对不同案卡项目、不同流程、案卡与文书、案卡与报表、报表与报表的关联内容进行比对，自动发现案卡填录问题，解决案件管理顽疾。

在这样一套案件管理数据质量监管系统中，可设置完备的校验规则，匹配最新的案卡模型。在检测范围方面，可实现案件同流程、跨流程等多种案卡数据校验。就卡书比对而言，可对文书内容智能解析并筛查问题。实际应用过程中，不但案件管理部门工作人员可以批量核查案件，办案检察官也可进行自查，使用时，除了能够发现数据质量问题，还能够发现错捕、错诉、错判等办案不规范问题，以及数据注水、规避负向指标等问题。如何设置查询条件？对于检察官们普遍关注的问题，一套完善的案件管理数据质量监管系统应当可以根据时间、主题、部门等不同条件进行查询。在统计层面，可反查到个案，快速定位问题数据；在部署层面，更是简约好操作，通过一级部署即可实现三级使用的效果。

提升检察机关数据质量，为规范办案、业务管理、效率提升发挥积极作用，是创设案件管理数据质量监管系统的初心。随着校验规则的进一步增加、功能的进一步丰富、性能的进一步提高，案件管理数据质量监管系统便可深入挖掘检察机关内生数据价值、有效实现数据治理、发现不规范案件，为数字检察战略的发展发挥更大作用。（高航）

日光城里，数字检察『开花结果』

西藏自治区拉萨市检察院不断增强大数据战略思维，2022年成立拉萨市检察院数字检察监督指挥中心，并结合工作实际，在梳理以往办案经验的基础上，优化大数据已有模块，探索开发适合拉萨地域实际的监督新模块，实现监督效能新集成。自该指挥中心运行以来，共推送各类线索6000余条，运用“涉医保先行支付未追偿”模型筛查出涉医保类裁判不当案件线索9件，成功追回医保资金先行垫付3万余元；运用“廉租房违规申请”模型，准确发现违规申请人6户，向相关部门制发检察建议推动问题有效解决……一系列案件的成功办理，让数字检察监督效能得到有效彰显，以线索核查落地为抓手的法律监督效果正在逐渐显现。

（本报记者张超　通讯员江玟萱　蒋李/文图）