O BANCO NACIONAL DE DADOS DO CNMP NA ERA DA COMPLEXIDADE INVESTIGATIVA:

FUNDAMENTOS CONSTITUCIONAIS, ESTATÍSTICOS E ALGORÍTMICOS DA RASTREABILIDADE PENAL

The CNMP National Data Repository in the Age of Investigative Complexity: Constitutional, Statistical, and Algorithmic Foundations of Penal Traceability

Fábio Medina Osório

Sócio administrador do Medina Osório Advogados. Doutor em Direito Administrativo pela Universidade Complutense de Madri (Espanha). Mestre em Direito Público pela Faculdade de Direito da Universidade Federal do Rio Grande do Sul (UFRGS). Ex-Promotor de Justiça no Rio Grande do Sul. Ex-Secretário Adjunto de Justiça e Segurança Pública do Estado do Rio Grande do Sul. Ex-Ministro-Chefe da Advocacia-Geral da União. Presidente da Comissão Especial de Direito Administrativo Sancionador do Conselho Federal da OAB. Conselheiro do MDA – Movimento de Defesa da Advocacia. Presidente do Instituto Internacional de Estudos de Direito do Estado (IIEDE).

Este artigo expressa a opinião acadêmica do autor e não de qualquer instituição da qual faz parte ou já integrou.

Resumo

Este ensaio examina a necessidade institucional de um Banco Nacional de Dados sob governança do Conselho Nacional do Ministério Público (CNMP), concebido como infraestrutura de rastreabilidade, coerência e autocrítica na persecução penal. Sustenta-se que a titularidade privativa da ação penal pública, o controle externo da atividade policial e o poder investigatório do Ministério Público, quando interpretados no contexto da Era Digital, pressupõem condições materiais de inteligibilidade que não se realizam sem bases estruturadas, padronização semântica e auditabilidade sistêmica. Na investigação orientada por dados, a eficiência e a integridade do sistema de justiça penal dependem de metodologia uniforme de coleta, normalização, resolução de identidade e registro de trilhas decisórias e de acesso. O artigo demonstra — a partir de evidências comparadas extraídas da literatura especializada nacional e internacional — que o cenário brasileiro atual é marcado por fragmentação severa: vinte e sete sistemas distintos de estatísticas criminais, ausência de padrão semântico nacional, recusa de estados em compartilhar microdados e episódios documentados de vulnerabilidade dos bancos de dados a atores criminosos. Propõe-se, assim, um modelo de "unidade informacional" do Ministério Público, que não substitui

bancos nacionais do Executivo (nem pretende absorver bancos estaduais), mas organiza o núcleo de dados do MP e estabelece interoperabilidade governada com sistemas externos, conforme padrões de qualidade, segurança e governança algorítmica. A proposta é contextualizada à luz da Lei Geral de Proteção de Dados (LGPD — Lei 13.709/2018), da Resolução CNMP 318/2025 (BDP/MP), da Portaria MJSP 1.123/2026 (Sinic) e dos marcos normativos internacionais relevantes.

Palavras-chave: Banco nacional de dados — CNMP — Ministério Público — Controle externo — Estatística — Inteligência artificial — Auditabilidade — Rastreabilidade — LGPD — Proteção de dados.

Abstract

This essay discusses the institutional need for a National Data Repository governed by Brazil's National Council of the Public Prosecutor's Office (CNMP), conceived as infrastructure for traceability, coherence, and institutional self-critique in criminal prosecution. It argues that the Prosecutor's exclusive authority to bring public criminal actions, its external oversight of police activity, and its investigative powers, when interpreted in the Digital Age, require material conditions of intelligibility that cannot be achieved without structured databases, semantic standardization, and systemic auditability. In data-driven investigations, efficiency and integrity depend on uniform methodologies of collection, normalization, entity resolution, and robust trails for access and decision-making. Drawing on comparative evidence from national and international specialized literature, the paper demonstrates that the current Brazilian landscape is characterized by severe fragmentation — twenty-seven distinct criminal statistics systems, absence of national semantic standards, states refusing to share microdata, and documented episodes of database vulnerability to criminal actors. The paper proposes an informational unity model for the Public Prosecutor's Office, which does not replace Executive-branch national databases nor absorb state police databases, but organizes the Prosecutor's own core data and enables governed interoperability with external systems under quality, security, and AI governance standards. The proposal is contextualized in light of Brazil's General Data Protection Law (LGPD — Law 13,709/2018), CNMP Resolution 318/2025 (BDP/MP), Ministry of Justice Ordinance 1,123/2026 (Sinic), and relevant international normative frameworks.

Keywords: National data repository — CNMP — Public Prosecutor — External oversight — Statistics — Artificial intelligence — Auditability — Traceability — Data protection — LGPD.

Sumário

---

1. Introdução

A Constituição de 1988 consagrou um conjunto de garantias clássicas — publicidade, transparência, fundamentação, devido processo legal, contraditório, ampla defesa — que, historicamente, foram lidas como exigências orientadas ao ato decisório final: a sentença, o acórdão, o ato administrativo sancionador. A Era Digital deslocou o centro de gravidade desse debate. Hoje, a restrição concreta de direitos, no âmbito penal, frequentemente se materializa antes do julgamento: nas escolhas investigativas, nos critérios de priorização de alvos, na construção de narrativas probatórias, nos registros de inteligência, na seleção do que é buscado e do que é ignorado. Em outras palavras: a decisão, no mundo contemporâneo, é composta por uma cadeia de microdecisões, muitas vezes invisíveis, cuja legitimidade depende de rastreabilidade.

Esse cenário exige reconhecer uma premissa metodológica: não se compreende o que não se consegue reconstruir. A publicidade formal de atos e a transparência clássica já não bastam quando a persecução penal se torna dependente de bases massivas, buscas estruturadas e correlações algorítmicas. Se as informações são fragmentadas, se os registros são semanticamente incompatíveis entre Estados, se não há trilhas de auditoria, a própria

racionalidade do sistema perde densidade: a investigação pode produzir resultado, mas não produz inteligibilidade; pode gerar ação penal, mas fragiliza a capacidade de revisão crítica; pode condenar, mas dissolve a legitimidade do percurso.[1]

O diagnóstico empírico confirma essa premissa com contundência. Pesquisa abrangente sobre a situação das tecnologias de segurança pública nas Unidades da Federação brasileiras revelou que doze estados sequer utilizam tecnologias disruptivas e outros nove não responderam a pedidos de informação durante a pesquisa.[2] O Anuário Estatístico de Segurança Pública 2023-2024, elaborado conjuntamente pelo Ipea e pela Secretaria Nacional de Segurança Pública (Senasp/MJSP), é ainda mais preciso: o Brasil conta com vinte e sete sistemas distintos de estatísticas criminais apenas entre as polícias civis, e o país "segue sem ter um sistema de informações de segurança pública estruturado, com dados confiáveis".[3] O vácuo regulatório é igualmente documentado: a Lei Geral de Proteção de Dados (LGPD — Lei nº 13.709/2018) prevê, em seu art. 4º, exceção para atividades de segurança pública e persecução penal, mas essa exceção, na ausência de lei específica que a discipline, transforma-se em zona de opacidade, dificultando o controle e a transparência sobre como os dados são tratados pelos órgãos estatais.[4]

---

[1]TSUNODA, Denise Fukumi; CÂNDIDO, Ana Clara; GUIMARÃES, André José Ribeiro. Tecnologias disruptivas em segurança pública: uma análise situacional brasileira. Revista Tecnologia e Sociedade, v. 20, n. 61, p. 317-333, jul./set. 2024. DOI: 10.3895/rts.v20n61.18408. Os autores constatam que "é fundamental estabelecer bases de dados unificadas, padronizar os processos de coleta e registro de informações em todas as unidades federativas" para que seja possível "realizar pesquisas e análises de forma adequada", identificando que doze estados brasileiros sequer utilizam tecnologias disruptivas e outros nove não forneceram informações durante a pesquisa.

[2]IPEA; SENASP/MJSP. Anuário Estatístico de Segurança Pública 2023-2024. Brasília: Ipea, 2025. DOI: https://dx.doi.org/10.38116/ri-anuario-estatistico-2023-2024. O documento explicita que "o Brasil segue sem ter um sistema de informações de segurança pública estruturado, com dados confiáveis", descrevendo a existência de "27 sistemas distintos de estatísticas criminais, considerando apenas as polícias civis". O Anuário documenta que a recusa de alguns estados em divulgar microdados, sob a justificativa de proteção da LGPD, é um obstáculo grave à integração nacional.

[3]AMBROSIO, Gleiner Pedroso Ferreira; BARBOSA, André Luis Jardini. O paradigma da implantação da inteligência artificial na segurança pública brasileira: regulação versus eficiência. Revista de Estudos Jurídicos da UNESP, v. 28, n. 48, 2024. Os autores ressaltam que a LGPD "possui uma exceção em seu artigo 4º, determinando que a lei não se aplica ao tratamento de dados realizados para fins exclusivos de segurança pública, defesa nacional ou atividades de investigação e repressão penal", alertando que essa exceção "cria um vácuo regulatório, dificultando o controle e a transparência sobre como esses dados são geridos pelos órgãos estatais". O estudo também registra que o Índice Global de Crime Organizado (2023) coloca o Brasil em posição alarmante (22º lugar geral e 8º em mercados criminosos), com baixa resiliência institucional.

[4]AMBROSIO; BARBOSA, op. cit. O texto narra que, "em 2023, uma investigação da Polícia Federal revelou que o PCC (Primeiro Comando da Capital) conseguiu acessar o sistema de câmeras do Detecta", utilizando o banco de dados estatal "para monitorar uma viatura descaracterizada da Polícia Civil, levantando dados como chassi e proprietário, em meio a um plano de assassinato contra o Senador Sérgio Moro". O episódio demonstra que a ausência de controles técnicos e regulatórios adequados pode transformar bancos de dados estatais em instrumentos operacionais do crime organizado.

Nesse contexto, importa observar o papel do Conselho Nacional do Ministério Público (CNMP) que, segundo sua própria definição oficial, executa a fiscalização administrativa, financeira e disciplinar do Ministério Público no Brasil e de seus membros, respeitando a autonomia da instituição. O órgão, criado em 30 de dezembro de 2004 pela Emenda Constitucional nº 45, teve sua instalação concluída em 21 de junho de 2005, com sede em Brasília-DF. Formado por 14 membros que representam setores diversos da sociedade, o CNMP tem como objetivo imprimir uma visão nacional ao MP, o que é decorrência do princípio constitucional da unidade institucional.

Ao Conselho cabe orientar e fiscalizar todos os ramos do MP brasileiro: o Ministério Público da União (MPU), composto pelo Ministério Público Federal (MPF), Ministério Público Militar (MPM), Ministério Público do Trabalho (MPT) e do Distrito Federal e Territórios (MPDFT); e o Ministério Público dos Estados (MPE).

Presidido pelo procurador-geral da República, o Conselho é composto por quatro integrantes do MPU, três membros do MPE, dois juízes indicados um pelo Supremo Tribunal Federal e outro pelo Superior Tribunal de Justiça, dois advogados indicados pelo Conselho Federal da Ordem dos Advogados do Brasil, e dois cidadãos de notável saber jurídico e reputação ilibada, indicados um pela Câmara dos Deputados e outro pelo Senado Federal.

Antes da posse no CNMP, os nomes apresentados são apreciados pela Comissão de Constituição e Justiça e de Cidadania (CCJ) do Senado Federal, depois vão ao Plenário do Senado e seguem para a sanção do presidente da República.

Pautado pelo controle e pela transparência administrativa do MP e de seus membros, o CNMP é uma entidade aberta ao controle social e às entidades brasileiras, que podem encaminhar reclamações contra membros ou órgãos do MP, inclusive contra seus serviços auxiliares.

Tais princípios devem ser interpretados em harmonia com os princípios da eficiência, impessoalidade, legalidade, devido processo legal, economicidade, moralidade administrativa, interdição à arbitrariedade dos poderes públicos e direito à compreensão acerca dos conteúdos das decisões tomadas pelas autoridades públicas.[5]

---

[5]BRASIL. Ministério da Justiça e Segurança Pública. Governo do Brasil oficializa novo sistema e protocolo para fortalecer coleta, gestão e uso de informações criminais no país. Portal Gov.br, 06 jan. 2026 (atualizado em 24 jan. 2026). O documento esclarece que o Sinic "passará a ser a fonte única para a emissão da Certidão Nacional Criminal e da Folha de Antecedentes Criminais", substituindo progressivamente os sistemas fragmentados de "tribunais, polícias civis e institutos de identificação das Unidades da Federação". A portaria determina que a adesão ao Protocolo Nacional de Reconhecimento de Pessoas será critério técnico para priorizar "o repasse de recursos do Fundo Nacional de Segurança Pública".

A implementação da unidade institucional do Ministério Público, na esfera criminal e no combate à criminalidade violenta e organizada, envolve o controle nacional do exercício do poder investigatório da instituição e o controle externo das polícias de modo integrado e harmônico, através de um planejamento estratégico e articulado nacionalmente.

É nesse ponto que a arquitetura constitucional do Ministério Público ganha centralidade. O CNMP deve velar pela unidade institucional na gestão da inteligência do Ministério Público brasileiro e sobretudo essa gestão deverá surtir um primeiro grande impacto na segurança pública e nas investigações criminais em todo o território nacional. O Ministério Público não é apenas o titular privativo da ação penal pública (art. 129, I, da Constituição Federal), mas também exerce o controle externo da atividade policial (art. 129, VII), além de deter poderes de requisição e, no horizonte jurisprudencial consolidado pelo Supremo Tribunal Federal (RE 593.727, Tema 184), poderes investigatórios compatíveis com a Constituição, desde que sob garantias. O tripé acusação–controle–investigação coloca o MP numa posição inevitável: ele é destinatário e fiscal do produto investigativo. Todavia, destinatário e controlador só podem operar em ambiente de dados se dispuserem de infraestrutura adequada. A ausência dessa infraestrutura produz uma contradição essencial: o MP carrega responsabilidades constitucionais crescentes, mas herda um universo informacional disperso, heterogêneo e frequentemente opaco.

Daí decorre a hipótese deste ensaio: o controle externo, embora não seja hierárquico, possui natureza conformativa no mundo digital. Ele conforma o mínimo de registrabilidade, auditabilidade e padronização semântica exigível para que a atividade policial seja controlável, comparável e corrigível, e para que a titularidade da ação penal se exerça com coerência nacional. Nessa perspectiva, o Banco Nacional de Dados do CNMP surge como infraestrutura do próprio Ministério Público: núcleo de memória institucional, base de padronização e ponte de interoperabilidade governada com sistemas externos.

É essencial, contudo, delimitar o objeto para evitar equívocos. O Banco Nacional do CNMP não pretende substituir bancos nacionais do Poder Executivo. O Ministério da Justiça e Segurança Pública (MJSP) instituiu o Sistema Nacional de Informações Criminais (Sinic), pela Portaria nº 1.123/2026, como base oficial de consolidação e disponibilização de informações criminais. O ambiente legislativo recente — assentado na Lei do SUSP (Lei nº 13.675/2018) — igualmente desenha bancos nacionais temáticos no enfrentamento ao crime organizado, com lógica federativa de interoperabilidade. O Banco do CNMP tem vocação própria: organizar o núcleo de dados do Ministério Público e permitir interoperabilidade controlada, auditável e finalística — sem absorção indiscriminada de bancos policiais estaduais.

2. Controle externo, ação penal e poder investigatório: o tripé constitucional da rastreabilidade

O controle externo não é comando administrativo. Essa afirmação, embora correta, costuma ser mal utilizada: como se a ausência de hierarquia implicasse ausência de potência institucional. Na Era Digital, ocorre precisamente o contrário. Quando a atividade policial se materializa em sistemas, registros e cadeias informacionais, o controle externo precisa incidir sobre aquilo que torna a atividade verificável: registros mínimos, integridade dos metadados, rastreabilidade de alterações, preservação de versões, padronização mínima de remessa e capacidade de reconstruir decisões investigativas.

A titularidade privativa da ação penal pública impõe ao MP a responsabilidade por organizar a acusação com base em prova compreensível e criticável. Isso exige, de maneira crescente, que os atos investigativos cheguem ao MP acompanhados de metadados essenciais e de trilhas que permitam aferição posterior. Cada peça de um inquérito policial encaminhado ao Ministério Público carrega, na era digital, metadados implícitos — carimbos temporais, identificadores de terminais, logs de acesso, histórico de alterações — que, quando preservados, permitem aferir a integridade probatória, e, quando suprimidos ou corrompidos, inviabilizam o controle.

A investigação, por sua vez, não pode ser concebida como "caixa-preta" administrativa: ela é o campo onde direitos fundamentais se tensionam cotidianamente. O controle externo ganha densidade quando se converte em exigência de auditabilidade, e essa auditabilidade, em ambiente informacional, é sempre um fenômeno de padrão. Sistemas de análise forense que aplicam modelos de linguagem de grande escala (Large Language Models — LLMs) a evidências extraídas de dispositivos móveis — como o framework desenvolvido pela Agência Nacional de Polícia da Coreia do Sul — demonstram que a estruturação mínima de metadados é condição de validade epistêmica: sem identificação precisa de remetente, destinatário, carimbo temporal e contexto de conversação, a evidência digital perde a cadeia de custódia que a torna utilizável na persecução.[6]

[6]MJSP/CNMP. Resolução CNMP nº 318, de 28 de outubro de 2025 (BDP/MP); MJSP. Portaria nº 1.123, de 5 de janeiro de 2026 (Sinic). A articulação entre esses dois instrumentos normativos é o núcleo da proposta de interoperabilidade governada sustentada neste artigo: o CNMP governa os dados processuais do MP, enquanto o Sinic consolida o histórico criminal nos órgãos do Executivo. A interoperabilidade entre essas bases — sob protocolos técnicos acordados e com trilhas de auditoria — é condição de inteligibilidade sistêmica.

A dimensão interorganizacional desse desafio é igualmente relevante. No Brasil, as competências investigativas são distribuídas entre polícias civis, polícia federal, polícia militar (em alguns estados), e o próprio Ministério Público — com atribuições partilhadas que historicamente geram distorções na produção e no compartilhamento de inteligência.[7] A ausência de um modelo estruturado de inteligência orientada por dados — como o Intelligence-Led Policing (ILP) praticado no Reino Unido (National Intelligence Model) e adotado como diretriz pelo Subsistema de Inteligência de Segurança Pública (SISP) no Brasil — resulta em patrulhamento baseado em intuição, taxas de resolução de casos historicamente baixas e incapacidade de detecção de redes criminosas com atuação interestadual.[8]

3. A investigação como fenômeno informacional: quando eficiência depende de linguagem e método

A ineficiência investigativa no Brasil não se explica apenas por escassez de recursos humanos ou tecnológicos. Ela se explica, em parte significativa, pela ausência de linguagem comum entre bases de dados. A investigação orientada por dados depende de encontrar relações entre registros dispersos — pessoas, endereços, veículos, armas, vínculos societários, comunicações, georreferências. Quando cada Estado registra em idioma próprio, o sistema nacional não enxerga redes — enxerga fragmentos.

---

[7] IPEA; SENASP/MJSP, op. cit. O Anuário registra que o Sinesp VDE passou a coletar 28 indicadores padronizados a partir de 2023 e que apenas 11 Unidades da Federação utilizam o Sinesp PPE (Procedimentos Policiais Eletrônicos), com a maioria dos estados usando sistemas próprios e exportando planilhas ou ferramentas de Business Intelligence. A obra documenta que alguns estados recusam enviar microdados alegando barreiras ligadas à LGPD "de forma inadequada", comprometendo a validade estatística do sistema nacional. PYTLOWANCIV, Diogo Fernando Sampaio. Intelligence-Led Policing e sua Possibilidade de Implementação no Brasil. Revista Brasileira de Ciências Policiais, v. 15, n. 1, p. 103-123, jan./abr. 2024. ISSN Eletrônico 2318-6917. O autor aponta que o Brasil possui "polícias com atribuições partilhadas (polícia ostensiva e polícia judiciária separadas)", gerando distorções na aplicação da inteligência, e que o sucesso do modelo de Intelligence-Led Policing "exige uma maior integração institucional, correlação de compartilhamento de informações e proximidade entre diferentes agências".

[8] MJSP/CG-RIBPG. XXII Relatório da Rede Integrada de Bancos de Perfis Genéticos (RIBPG): Dados estatísticos e resultados — Nov/2024 a Mai/2025. Brasília: MJSP, maio 2025. O relatório descreve que a RIBPG adota "uma arquitetura em rede (federada): existem 23 Bancos de Perfis Genéticos locais (BPGs), geridos pelas unidades de perícia estaduais, distrital e da Polícia Federal, que são conectados e processados centralmente pelo BNPG". O banco já acumula "mais de 254 mil perfis genéticos", com taxa de coincidência (hit rate) de 7,08%, e realiza "compartilhamento internacional de perfis genéticos através da INTERPOL", tendo o Brasil enviado "mais de 32.900 perfis de vestígios de crimes e mais de 11.100 perfis de restos mortais para a base global" até maio de 2025. O modelo da RIBPG demonstra que a arquitetura federada — com padrões técnicos rígidos, governança central e interoperabilidade internacional — é compatível com o federalismo brasileiro e pode ser replicada em outras esferas.

Esse fenômeno produz um paradoxo: digitaliza-se a investigação, mas preserva-se a lógica analógica do registro. As consequências são previsíveis: a busca não funciona, a correlação é precária, os homônimos proliferam, a duplicidade se instala, e a análise estatística perde validade. A qualidade do dado deixa de ser detalhe técnico e passa a ser requisito de eficiência e de legitimidade.

O Anuário Estatístico de Segurança Pública 2023-2024 documenta esse paradoxo com precisão: a maioria dos estados usa sistemas próprios de coleta (como o RAI em Goiás, o SROP em Mato Grosso e o Millenium no Distrito Federal) e depois exporta planilhas ou emprega ferramentas de Business Intelligence para repassar dados estatísticos ao governo federal via Sinesp VDE. Alguns estados recusam enviar microdados alegando barreiras ligadas à LGPD "de forma inadequada", comprometendo a validade estatística e, por extensão, a racionalidade das políticas públicas fundadas nesses dados.

A pesquisa internacional sobre tecnologias disruptivas em segurança pública oferece contraste instrutivo. A plataforma SafetySmart, operada pela SoundThinking, Inc. nos Estados Unidos, processa mais de 1,3 bilhão de registros estruturados e não estruturados de múltiplas jurisdições por meio de um motor de busca federada, o CrimeTracer, que permite "acessar e cruzar informações cruciais de várias agências de TI através de cidades, condados, estados e de todo o país". O módulo CaseBuilder estrutura digitalmente todas as informações do caso em formato unificado, eliminando processos manuais e sistemas isolados. A comparação não é uma recomendação de privatização da inteligência criminal — modelo que suscita sérias objeções de soberania informacional e controle democrático, como se discute mais adiante —, mas uma demonstração de que a padronização semântica e a busca federada são tecnicamente viáveis e operacionalmente transformadoras.

No nível da microanálise probatória, pesquisas recentes demonstram que a estruturação de metadados de mensagens extraídas de smartphones — com campos padronizados de remetente, destinatário, carimbo temporal, identificador de sala de bate-papo e tipo de mensagem — permite que modelos de linguagem (como GPT, em suas versões mais avançadas, e Claude, em suas versões mais avançadas) automatizem a leitura, a compreensão de contexto e a extração de evidências criminais ocultas, reduzindo drasticamente o tempo necessário para análise de volumes massivos de dados em prazos processuais rígidos. O estudo italiano sobre Grafos de Conhecimento e NLP aplicados à análise de mensagens de investigações reais de fraude e corrupção aponta na mesma direção: a estruturação de metadados (lista de participantes, horários, remetentes, anexos, entidades identificadas por NER — Reconhecimento de Entidades Nomeadas) é pré-

condição para que investigadores extraiam insights sem ler manualmente todo o material apreendido.

A "IA contestável" — conceito proposto por pesquisadores alemães da Universidade Federal da Bundeswehr de Munique — vai além: propõe que sistemas de análise de inteligência criminal sejam não apenas explicáveis, mas contestáveis, permitindo ao investigador humano questionar, corrigir e refinar os outputs algorítmicos por meio de modelagem semântica e supervisão humana estruturada.[9] Esses desenvolvimentos convergem para uma mesma conclusão: a qualidade do dado de entrada — sua completude, padronização semântica, rastreabilidade e integridade — determina o teto de qualidade da análise de saída, seja ela feita por humanos ou por algoritmos.

4. Banco nacional do CNMP e bancos nacionais do Executivo: distinções necessárias

A Resolução CNMP nº 318/2025 institui a Base de Dados Processuais do Ministério Público (BDP/MP) e estabelece regras de tratamento, governança e utilização. Trata-se do núcleo institucional do Banco Nacional do CNMP: dados processuais e extrajudiciais do MP, organizados sob padrão nacional e governança própria. A base justifica-se pela natureza constitucional do MP como titular da ação penal e fiscal do ordenamento jurídico: sem memória institucional estruturada, o exercício dessas funções é sistematicamente dependente de informações produzidas por terceiros — o que compromete tanto a independência funcional quanto a qualidade da persecução.

---

[9]POZZI, Riccardo; BARBERA, Valentina; PRINCIPE, Renzo Alva; GIARDINI, Davide; PALMONARI, Matteo. Combining Knowledge Graphs and NLP to Analyze Instant Messaging Data in Criminal Investigations. In: Proceedings of WISE 2024 (Web Information Systems Engineering). Springer, 2024. DOI: https://doi.org/10.1007/978-981-96-0567-5_30. O artigo descreve um pipeline de análise de mensagens extraídas de smartphones apreendidos que integra Grafos de Conhecimento (armazenados no Neo4j) e modelos de NLP, com metadados extraídos por um parser que identifica "lista de participantes, números de telefone, horários de início e fim, remetente e anexos". Os autores demonstram que o enriquecimento semântico por meio do pipeline NEEL (Named Entity Recognition and Linking) é essencial para que promotores e policiais possam pesquisar e extrair insights sem ler manualmente todo o material. KIM, Kyung-Jong; LEE, Chan-Hwi; BAE, So-Eun; CHOI, Ju-Hyun; KANG, Wook. Digital forensics in law enforcement: A case study of LLM-driven evidence analysis. Forensic Science International: Digital Investigation, v. 54, art. 301939, 2025. DOI: https://doi.org/10.1016/j.fsidi.2025.301939. O estudo demonstra que o banco de dados estruturado gerado a partir de um celular "contém até 31 colunas detalhadas, incluindo metadados fundamentais como: aplicativo de origem, tipo de mensagem, conteúdo, ID único da sala de bate-papo, nome e número de telefone do remetente e do destinatário, e o carimbo de tempo", e que, antes de alimentar os algoritmos de investigação, esses dados sofrem "anonimização (nomes são mascarados por Reconhecimento de Entidades Nomeadas — NER, e números de telefone são randomizados) para evitar vazamento de dados sensíveis e violação de direitos constitucionais".

O Sinic, por sua vez, foi instituído pelo MJSP, pela Portaria nº 1.123/2026, como base oficial de consolidação e disponibilização de informações criminais — indiciamentos, denúncias e condenações — com vocação de se tornar a "fonte única" para emissão da Certidão Nacional Criminal e da Folha de Antecedentes Criminais, substituindo progressivamente os sistemas fragmentados de tribunais, polícias civis e institutos de identificação das Unidades da Federação.[10] O ecossistema do SUSP (Lei nº 13.675/2018) fornece a moldura legal para integração nacional de dados de segurança pública, com o Sinesp como sistema de referência para as estatísticas policiais.[11]

Assim, o Banco do CNMP deve ser desenhado como: (i) base nacional do MP (núcleo), compreendendo os dados processuais e extrajudiciais produzidos pelo Ministério Público em todas as esferas; (ii) interoperabilidade governada com bases federais e estaduais (ponte), mediante protocolos técnicos, acordos de compartilhamento e trilhas de auditoria; e (iii) camada analítica e estatística (inteligência institucional), que permita ao CNMP exercer sua função de planejamento, avaliação e controle sobre a persecução penal. A legitimidade do projeto depende justamente dessa distinção: não duplicar, não absorver indiscriminadamente, mas integrar com governança.

A distinção entre controlador e operador, nos termos da LGPD (Lei nº 13.709/2018), é aqui essencial. O CNMP, como controlador público dos dados da BDP/MP, define as finalidades e os meios de tratamento; as polícias e demais órgãos que alimentam o sistema operam como fontes; e eventuais empresas de tecnologia contratadas para desenvolver conectores estaduais e normalizar dados atuam como operadoras técnicas, subordinadas às instruções

[10]SOUNDTHINKING, INC. Form 10-K: Annual Report Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934 (Fiscal Year Ended December 31, 2024). United States Securities and Exchange Commission, 2025. Commission File Number 001-38107; Nasdaq: SSTI. O relatório descreve o CrimeTracer como capaz de processar "mais de 1,3 bilhão de dados estruturados e não estruturados de múltiplas jurisdições", operando por meio de "busca federada de campos estruturados" e cruzando dados locais com "bilhões de registros de dados públicos" via integração com a plataforma Thomson Reuters CLEAR. O sistema demonstra "o Poder da Rede" (The Power of the Network), permitindo "acessar informações cruciais não apenas dos sistemas de TI de uma agência específica, mas cruzando fronteiras locais, municipais, estaduais e nacionais", com vínculo a bases federais como NIBIN e NCIC. O relatório identifica como lacunas críticas da segurança pública atual "o sub-relato de crimes violentos, o patrulhamento baseado na intuição (gut-based) e as baixíssimas taxas de resolução de casos, que atingiram o pior nível em 40 anos (menos de 50% para homicídios)".

[11]UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024 (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689. ELI: http://data.europa.eu/eli/reg/2024/1689/oj. O AI Act classifica sistemas de IA voltados à aplicação da lei como de alto risco, exigindo gestão de riscos, transparência, supervisão humana e registro em base de dados da Comissão Europeia. O regulamento proíbe a identificação biométrica remota em tempo real em espaços públicos como regra geral, admitindo exceções apenas mediante autorização judicial para ameaças terroristas ou crimes graves organizados (tráfico de seres humanos, terrorismo, crimes ambientais organizados, sabotagem, pertencimento a organização criminosa). O AI Act determina que os sistemas integrados aos frameworks de interoperabilidade da UE (Schengen Information System, Eurodac, ECRIS-TCN, Visa Information System) devem estar em conformidade até o final de 2030.

do controlador e sujeitas a auditorias periódicas. Essa arquitetura de responsabilidade é condição de conformidade com o art. 23 da LGPD, que impõe ao Poder Público o dever de publicar suas regras de tratamento e de indicar o encarregado (DPO — Data Protection Officer).

O Sinic incorpora, como diretriz normativa expressa, registros de pessoas condenadas por integrar organizações ou facções criminosas — o que densifica a inteligência criminal contra o crime organizado de âmbito nacional. A experiência da Rede Integrada de Bancos de Perfis Genéticos (RIBPG), que já acumula mais de 254 mil perfis genéticos em arquitetura federada (23 bancos estaduais conectados ao Banco Nacional de Perfis Genéticos — BNPG), demonstra que essa interoperabilidade é tecnicamente viável e institucionalmente sustentável.[12] O modelo da RIBPG — com padrão técnico definido em Manual de Procedimentos Operacionais, software padronizado (CODIS) e conectividade à base da INTERPOL — oferece um template para o Banco do CNMP: arquitetura federada, padrão técnico centralizado, governança pública e auditabilidade externa.

5. Transparência métrica na persecução penal: estatística como escuta institucional

Na persecução penal, a estatística não deve ser reduzida a relatórios anuais ou contagem de ocorrências. Na Era da complexidade, estatística é a forma científica de escuta institucional: identifica padrões, revela anomalias, detecta desigualdades e permite autocrítica. Essa função só é possível com dados comparáveis e com qualidade aferível.

O Anuário Estatístico de Segurança Pública 2023-2024 documenta que o crime organizado (como PCC e CV) atua fortemente em regiões de fronteira (Norte e Centro-Oeste), utilizando rotas transnacionais para escoamento de drogas e armas, mas "não existe, atualmente, uma iniciativa federal ou banco de dados único e consolidado que integre as diversas instituições (Senappen, CNJ, Polícia Federal, Coaf, Abin) para um diagnóstico abrangente do crime organizado". A ausência de uma base integrada força os pesquisadores

[12]NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg, MD: NIST, 2023. (NIST.AI.100-1). DOI: https://doi.org/10.6028/NIST.AI.100-1. O framework parte da "premissa de que riscos emergem da interação entre componentes técnicos e fatores sociais e institucionais, exigindo documentação, controle e gestão contínua" por meio das funções Govern, Map, Measure e Manage. O NIST AI RMF enfatiza a importância de "limpar dados, documentar metadados e adotar tecnologias de aprimoramento de privacidade ao treinar sistemas automatizados" e alerta que a "coleta enviesada ou perda de contexto original dos dados pode tornar a IA não confiável". O documento exige "rastreabilidade de dados (data traceability)" como capacidade de "rastrear e auditar internamente os conjuntos de dados (datasets) usados pela IA e seus metadados essenciais".

a construírem proxies — marcadores indiretos — utilizando as bases existentes (Sinesp), como a razão entre homicídios consumados e tentados, apreensões de armas de grosso calibre e taxas de óbitos intencionais dentro do sistema prisional.

Ganham relevância, aqui, os frameworks internacionais de qualidade estatística. O Data Quality Assessment Framework (DQAF) do FMI e os Princípios Fundamentais das Estatísticas Oficiais das Nações Unidas (Resolução ONU 68/261) consagram integridade, confiabilidade, confidencialidade e uso responsável como condições de confiança pública.[13] Esses princípios têm implicação direta para o Banco do CNMP: (i) o Princípio 6 da ONU determina que dados individuais coletados por agências estatísticas devem ser "estritamente confidenciais e usados exclusivamente para fins estatísticos", o que impõe uma separação estrutural entre a camada analítica agregada do banco e os dados processuais individuais, com controles de acesso diferenciados; (ii) o Princípio 8 prescreve que "a coordenação entre agências estatísticas dentro dos países é essencial para alcançar consistência e eficiência no sistema estatístico", justificando o papel do CNMP como coordenador nacional da estatística do Ministério Público; e (iii) o Princípio 9 defende a padronização internacional de conceitos e classificações, orientando as escolhas de schema e de ontologia jurídica para o sistema.

A pesquisa sobre policiamento preditivo no Brasil revela que os estados e municípios têm adotado uma "autorregulamentação" na aplicação de algoritmos, "sujeitando a segurança pública a falhas metodológicas, discricionariedade governamental, vazamento de dados e viés discriminatório".[14] Essa autorregulamentação fragmentária compromete não apenas a eficiência investigativa, mas a legitimidade das estatísticas produzidas: quando o algoritmo de um estado é alimentado com dados que "retratam a seletividade própria do sistema de segurança pública e de justiça criminal", as inferências estatísticas amplificam o viés em

[13]UNESCO. Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO, 2021. Código SHS/BIO/REC-AIETHICS/2021. A Recomendação estabelece que "dados relativos a infrações, processos criminais e condenações, e medidas de segurança relacionadas" são dados sensíveis cuja divulgação "pode causar danos excepcionais aos indivíduos", exigindo "segurança total para dados pessoais e sensíveis". O documento veda expressamente o uso de IA para "pontuação social ou vigilância em massa" e determina que, quando Estados adquiram sistemas de IA para aplicação da lei e sistemas judiciários, "devem ser criados mecanismos independentes para monitorar o impacto social e econômico de tais sistemas". A Recomendação exige que os "conjuntos de dados usados para treinar sistemas de IA sejam de alta qualidade e não reforcem preconceitos, desigualdades ou discriminação".

[14]MAORO, Falk; GEIERHOS, Michaela. Contestable AI for criminal intelligence analysis: improving decision-making through semantic modeling and human oversight. Frontiers in Artificial Intelligence, v. 8, art. 1602998, jul. 2025. DOI: 10.3389/frai.2025.1602998. Os autores propõem um modelo de "IA contestável" para análise de inteligência criminal que integra "modelagem semântica e supervisão humana", exigindo que "os modelos sejam auditáveis, justos e livres de vieses humanos". O estudo demonstra como a extração de entidades por NLP e NER pode transformar textos livres de relatórios policiais em metadados estruturados (formato JSON), superando o problema dos "relatórios narrativos em texto livre preenchidos por policiais, que são barulhentos, cheios de erros gramaticais e difíceis de minerar".

vez de corrigi-lo. O Tribunal de Contas do Estado de São Paulo, ao auditar o sistema Detecta, constatou "conflitos entre sistemas operacionais, falta de infraestrutura e treinamento", o que ilustra que a ausência de governança estruturada afeta tanto a validade operacional quanto a confiabilidade estatística dos dados.

Um episódio documentado ilustra, com dramaticidade, o risco da ausência de governança: em 2023, uma investigação da Polícia Federal revelou que o PCC (Primeiro Comando da Capital) conseguiu acessar o sistema de câmeras do Detecta, utilizando o banco de dados estatal para monitorar uma viatura descaracterizada da Polícia Civil em meio a um plano de assassinato.[15] O episódio demonstra que bancos de dados de segurança pública sem controles adequados de acesso, autenticação, monitoramento de anomalias e gestão de vulnerabilidades podem ser instrumentalizados pela própria criminalidade organizada — convertendo-se de ferramenta de proteção em vetor de ameaça.

6. Padronização algorítmica e auditabilidade: da busca ao grafo

A padronização algorítmica não significa impor um software único aos Estados. Significa impor propriedades mínimas: (i) schema canônico versionado — estrutura de dados com campos e tipos definidos, controlada por versão para garantir retrocompatibilidade; (ii) dicionário semântico — vocabulário controlado de termos jurídicos e criminológicos que assegura que o mesmo fenômeno seja descrito da mesma forma em todos os sistemas; (iii) regras de resolução de identidade — algoritmos que identificam se dois registros referem-se ao mesmo indivíduo, entidade ou evento, eliminando duplicidades e homônimos; (iv) logs imutáveis — registros de acesso e operação que não podem ser alterados retroativamente, essenciais à cadeia de custódia digital; (v) trilhas de transformação (data lineage) — rastreamento de todas as transformações sofridas pelo dado desde a coleta até o uso analítico; e (vi) métricas de qualidade por fonte e por UF — indicadores mensuráveis de completude, consistência, precisão e atualidade.

A governança de riscos de IA, como enfatiza o NIST AI RMF 1.0 (Artificial Intelligence Risk Management Framework), parte da premissa de que riscos emergem da interação entre

[15]KÜÇÜK, Dilek; CAN, Fazli. Computational Law: Datasets, Benchmarks, and Ontologies. arXiv, 2025. Preprint 2503.04305v2. O artigo apresenta um survey abrangente de datasets e ontologias para processamento de linguagem natural no domínio jurídico, discutindo o benchmark FEDLEGAL como arquitetura na qual "modelos de aprendizado de máquina são treinados em bases de dados distribuídas (que contêm documentos jurídicos sensíveis) sem que esses dados locais precisem ser centralizados em um único servidor, mitigando problemas de privacidade na previsão de causas e sentenças legais". O Federated Learning oferece um modelo de treinamento distribuído compatível com o federalismo e com a proteção de dados sensíveis.

componentes técnicos e fatores sociais e institucionais, exigindo documentação, controle e gestão contínua.[16] O framework organiza a governança de sistemas de IA em quatro funções — Govern (governar), Map (mapear), Measure (medir) e Manage (gerenciar) —, aplicáveis diretamente ao ciclo de vida dos algoritmos utilizados na busca, correlação e análise de dados criminais.

O AI Act europeu (Regulamento (UE) 2024/1689) consagra obrigações de gestão de riscos, transparência e governança especialmente relevantes quando sistemas impactam direitos fundamentais e atividades de enforcement.[17] O regulamento classifica sistemas de IA voltados à aplicação da lei como de alto risco, exigindo avaliação de impacto sobre direitos fundamentais antes da implantação, supervisão humana estruturada, testes de precisão e avaliação de disparidades demográficas. Embora seja norma de direito europeu, o AI Act funciona como parâmetro de referência para a governança de sistemas similares no Brasil, especialmente diante da ausência de legislação específica para IA aplicada à segurança pública.

O Relatório Final do Departamento de Justiça dos Estados Unidos sobre IA e Justiça Criminal (2024) — elaborado em cumprimento à Seção 7.1(b) da Executive Order 14.110 (revogada em 20 de janeiro de 2025 pelo Presidente Trump, sem prejuízo dos documentos produzidos em sua vigência) — aponta que ferramentas de IA usadas para "identificar suspeitos criminais, prever crimes, aplicar técnicas de forense digital, monitorar redes sociais ou rastrear localização física de indivíduos" devem ser submetidas a avaliações de impacto (AI Impact Assessments) e práticas estruturadas de gerenciamento de risco, com procedimentos para auditar os dados de entrada e evitar ciclos de feedback

[16]UNITED STATES DEPARTMENT OF JUSTICE. Artificial Intelligence and Criminal Justice: Final Report. Washington, D.C.: U.S. DOJ, 3 dez. 2024. Elaborado em cumprimento à Seção 7.1(b) da Executive Order 14110 (revogada em 20 de janeiro de 2025). O relatório identifica como aplicações de IA de alto impacto na justiça criminal "identificar suspeitos criminais, prever crimes, aplicar técnicas de forense digital, monitorar redes sociais ou rastrear localização física de indivíduos", exigindo para esses sistemas "avaliações de impacto da IA (AI Impact Assessments) e práticas de gerenciamento de risco". O DOJ reconhece que "a coleta de dados criminais é historicamente falha, exigindo procedimentos estruturados para auditar os dados de entrada, evitar ciclos de feedback discriminatórios e estruturar bases de dados limpas e representativas". O relatório também detalha que modelos de previsão criminal integram "metadados fora do escopo de aplicação da lei, como dados de saúde pública (CDC), elevação do terreno, zoneamento, clima e proximidade de transporte público".

[17]GROSSI, Alexandre Viezzer. A aplicação da Inteligência Artificial na segurança pública brasileira: o caso de São Paulo e a análise do PL nº 2338/2023. Revista Políticas Públicas & Cidades, v. 14, n. 4, 2025. ISSN: 2359-1552. DOI: https://doi.org/10.23900/2359-1552v14n4-72-2025. O autor observa que "os estados e municípios vêm adotando uma autorregulamentação na aplicação de algoritmos", sujeitando a segurança pública a "falhas metodológicas, discricionariedade governamental, vazamento de dados e viés discriminatório". O texto defende que "antes de se buscar a eficiência irrestrita, é indispensável a regulação nacional prévia (inspirada em normativas como a LGPD brasileira e o AI Act europeu) para garantir a legitimidade do uso tecnológico no território nacional".

discriminatórios.[18] O Memorando M-25-21 da Casa Branca (OMB, 2025) reforça essa orientação, determinando que Diretores de IA e Diretores de Dados coordenem critérios de interoperabilidade entre agências e invistam em "ativos de dados de qualidade, infraestrutura de tecnologia e governança na coleta, curadoria e preparo da informação".[19]

Esses marcos ajudam a dar densidade contemporânea ao argumento central: bancos de dados e algoritmos não são apenas ferramentas; são infraestruturas de poder que exigem auditabilidade. A Recomendação da UNESCO sobre a Ética da Inteligência Artificial (2021) é explícita ao proibir o uso de IA para "pontuação social ou vigilância em massa" e ao exigir que sistemas implantados por Estados para aplicação da lei se submetam a mecanismos independentes de supervisão, garantindo que os dados de treinamento "não reforcem preconceitos, desigualdades ou discriminação".[20]

No plano técnico, a arquitetura de grafo de conhecimento (Knowledge Graph) — como proposta no sistema baseado em Neo4j para análise de mensagens de investigações criminais — oferece uma alternativa ao modelo relacional clássico para representar a complexidade das redes investigadas: em vez de tabelas, o grafo representa entidades (pessoas, organizações, locais) e suas relações (comunicou-se com, transferiu dinheiro para, apareceu no mesmo local que) com enriquecimento semântico por NER (Named Entity Recognition) e transcrição automática de áudios. O benchmark FEDLEGAL, discutido na literatura de Direito Computacional, propõe o Federated Learning como

[18]INTERPOL. Rules on the Processing of Data (RPD). Lyon: INTERPOL, 2019. O documento estabelece que "o sucesso das investigações policiais internacionais depende intrinsecamente da disponibilidade de dados globais atualizados" e que "todos os dados compartilhados obedecem a padrões internacionais rígidos, possuindo fundamentação legal e recursos de segurança embutidos". A INTERPOL gerencia bancos especializados (Dados Nominais com histórico criminal, fotos e impressões digitais; Perfis de DNA; material de exploração sexual infantil; Documentos de Viagem Roubados ou Perdidos; Veículos e Obras de Arte Roubados; armamentos rastreados via iARMS e Rede de Informações Balísticas) acessíveis pelo sistema I-24/7. A arquitetura da INTERPOL demonstra que "oficiais da linha de frente (como guardas de fronteira) podem submeter uma consulta de forma simultânea tanto ao banco de dados nacional quanto ao banco de dados da INTERPOL, obtendo os cruzamentos de informações de ambos em questão de segundos".

[19]VOUGHT, Russell T. M-25-21: Accelerating Federal Use of AI through Innovation, Governance, and Public Trust. Washington, D.C.: Executive Office of the President, Office of Management and Budget, 3 abr. 2025. O memorando estimula "o compartilhamento de dados, modelos algorítmicos e códigos-fonte entre as agências do Governo Federal" e recomenda que "Diretores de IA (Chief AI Officers) e Diretores de Dados (Chief Data Officers) coordenem ativamente critérios de interoperabilidade de dados entre agências governamentais". O documento encoraja a "padronização de formatos de dados e interoperabilidade em todo o governo federal para facilitar a adoção e a integração algorítmica da IA".

[20]PYTLOWANCIV, op. cit. O autor descreve que nos EUA, após os atentados de 11 de setembro, foi criado o National Criminal Intelligence Sharing Plan, que "estabeleceu diretrizes para compartilhamento de informações, padrões de infraestrutura e a criação de centros de fusão (fusion centers) para fortalecer o compartilhamento de conhecimentos entre agências". No Brasil, o autor cita o Subsistema de Inteligência de Segurança Pública (SISP) e a Política Nacional de Inteligência de Segurança Pública (Pnisp), enfatizando que a principal atuação do Intelligence-Led Policing deve ser direcionada à mitigação de ameaças como organizações criminosas e grupos extremistas.

arquitetura alternativa para treinar modelos de IA em dados jurídicos sensíveis sem centralizar fisicamente os dados — preservando a privacidade das bases distribuídas e, ao mesmo tempo, permitindo aprendizado coletivo.[21]

7. Infraestrutura pública de dados, convênios interinstitucionais e soberania informacional

A viabilização do Banco do CNMP, em escala federativa, pressupõe convênios e protocolos de interoperabilidade com os sistemas estaduais, com o Sinic do MJSP, com o Sinesp e com bases temáticas como a RIBPG. Esses instrumentos devem definir: (a) os tipos de dados objeto de compartilhamento (categoria, finalidade e sensibilidade); (b) as bases jurídicas aplicáveis em cada caso (art. 7º, III ou VI; art. 11, II, f; e art. 23 da LGPD, conforme o dado seja ou não de natureza sensível); (c) as salvaguardas técnicas exigíveis (criptografia em repouso e em trânsito, controle de acesso baseado em função, autenticação multifator, logs de acesso imutáveis); (d) as responsabilidades de cada parte (controlador, cocontrolador ou operador); e (e) os mecanismos de auditoria e prestação de contas.

A empresa eventualmente contratada pelo CNMP deve atuar como operadora técnica, implementando conectores estaduais e normalizando dados para o padrão nacional, sob governança do controlador público. Essa relação deve ser regida por contrato de tratamento de dados (art. 39 da LGPD), com cláusulas de auditorias periódicas, vedação de uso dos dados para finalidades alheias ao contrato, notificação obrigatória em caso de incidente de segurança (art. 48 da LGPD) e encerramento seguro do tratamento ao término do contrato. O modelo não é de privatização da inteligência criminal — que suscita graves objeções de soberania informacional e accountability democrático —, mas de terceirização técnica com responsabilidade pública preservada.

A experiência internacional fornece parâmetros relevantes. O Tribal Law and Order Act norte-americano de 2010 demonstra que a integração de bancos de dados entre entes de diferentes esferas pode ser viabilizada por mecanismos de "acesso gradual", condicionados

[21]NATIONS UNIES. Résolution 68/261: Principes fondamentaux de la statistique officielle. A/RES/68/261, 29 jan. 2014. O Princípio 6 determina que "dados individuais coletados por agências estatísticas (sejam referentes a pessoas físicas ou jurídicas) devem ser estritamente confidenciais e usados exclusivamente para fins estatísticos", impondo uma separação estrutural entre dados estatísticos oficiais e dados para investigação criminal. O Princípio 8 estabelece que "a coordenação entre agências estatísticas dentro dos países é essencial para alcançar consistência e eficiência no sistema estatístico". O Princípio 9 defende a "padronização internacional de conceitos, classificações e métodos para assegurar a consistência dos sistemas". Esses princípios fornecem o lastro normativo multilateral para as exigências de qualidade, integridade e confidencialidade aplicáveis ao componente estatístico do Banco do CNMP.

ao cumprimento de requisitos técnicos e jurídicos.[22] O modelo dos fusion centers americanos — centros onde agências criminais de níveis local, estadual e federal integram e compartilham inteligência — oferece referência para a articulação entre o Banco do CNMP e os centros de inteligência das polícias estaduais e federal.

Na esfera global, a arquitetura da INTERPOL demonstra que bancos de dados de inteligência criminal com alcance transnacional são viáveis sob governança rigorosa: todos os dados compartilhados pelos países-membros "obedecem a padrões internacionais rígidos, possuindo fundamentação legal e recursos de segurança embutidos", com acesso estruturado pelo sistema seguro I-24/7 e capacidade de consulta simultânea às bases nacionais e à base central, em tempo real.[23] Isso reforça que a soberania informacional não é incompatível com a interoperabilidade — desde que o acesso seja controlado, a finalidade seja definida e o dado permaneça sob governança de autoridade pública.

O objetivo do Banco do CNMP não é "copiar tudo", mas criar uma ponte auditável que permita busca e correlação em escala nacional, com preservação de sigilos funcionais e conformidade com a LGPD. Dados de inteligência criminal, comunicações protegidas por sigilo profissional, dados de saúde mental de réus, informações de vítimas de crimes sexuais e dados de testemunhas protegidas requerem tratamento diferenciado, com controles de acesso mais restritivos e finalidades mais estreitamente definidas.

8. Proteção de dados pessoais e salvaguardas na persecução penal

A articulação entre segurança pública e proteção de dados pessoais é um dos nós mais complexos do ordenamento jurídico brasileiro contemporâneo. A LGPD (Lei nº 13.709/2018), em seu art. 4º, III, exclui de seu âmbito de aplicação o tratamento de dados para fins exclusivos de segurança pública, defesa nacional, segurança do Estado e atividades de investigação e repressão de infrações penais — excluindo tais operações de tratamento da incidência geral da lei e remetendo-as a lei específica a ser editada.

Essa exceção, contudo, não equivale a ausência de proteção. Dois argumentos convergentes sustentam essa afirmação. Primeiro, o argumento constitucional: os direitos fundamentais à privacidade (art. 5º, X), à proteção de dados (art. 5º, LXXIX, com a EC nº 115/2022) e ao

[22]Confira-se, a propósito, artigo que escrevi sobre o tema: OSÓRIO, Fábio Medina. O direito à compreensão na Era da complexidade tecnológica: fundamentos constitucionais, estatísticos e algorítmicos da transparência decisória. Revista dos Tribunais, v. 1077/2025, jul. 2025. DTR\2025\7689.

[23]OSÓRIO, Fábio Medina. O direito à compreensão na Era da complexidade tecnológica: fundamentos constitucionais, estatísticos e algorítmicos da transparência decisória. Revista dos Tribunais, v. 1077/2025, jul. 2025. DTR\2025\7689.

devido processo legal (art. 5º, LIV) constituem limites intransponíveis mesmo para a persecução penal, independentemente de lei ordinária. Segundo, o argumento sistêmico: a ausência de lei específica não cria um vácuo normativo absoluto, pois incidem sobre a matéria: (i) o Código de Processo Penal (CPP), que disciplina a produção probatória e a integridade das cadeias de custódia; (ii) as resoluções do CNMP sobre manuseio de dados e sigilo funcional; (iii) a Convenção 108+ do Conselho da Europa, da qual o Brasil não é parte, mas que funciona como parâmetro interpretativo para a proteção de dados em contextos de aplicação da lei; e (iv) as diretrizes da UNESCO sobre ética na IA, que impõem salvaguardas específicas para dados relativos a infrações, processos criminais e condenações.

Para fins de aplicação ao Banco do CNMP, os princípios de proteção de dados operam do seguinte modo. O princípio da finalidade determina que cada tipo de dado só pode ser tratado para a finalidade que justificou sua coleta — dado coletado para fins de identificação criminal não pode ser reutilizado para fins de perfilamento comportamental ou vigilância contínua. O princípio da necessidade impõe que o banco colete apenas o mínimo de dados indispensável para as finalidades definidas, vedando a coleta especulativa ou o armazenamento de dados desnecessários. O princípio da adequação exige que o meio de tratamento seja proporcional à finalidade perseguida. O princípio da transparência impõe a publicação das regras de tratamento e a designação de encarregado de dados (DPO). O princípio da segurança exige a adoção de medidas técnicas e administrativas para proteger os dados contra acessos não autorizados, destruição, perda e alteração. O princípio da responsabilização e prestação de contas (accountability) impõe ao controlador a obrigação de demonstrar conformidade e de responder por danos causados em decorrência do tratamento.

Esses princípios impõem, na prática, um conjunto de salvaguardas operacionais para o Banco do CNMP: (a) mapeamento de categorias de dados e avaliação de sensibilidade (dados relativos a infrações, origem racial, saúde, orientação sexual e vida privada recebem proteção reforçada); (b) controle de acesso baseado em função (RBAC — Role-Based Access Control), com perfis diferenciados para consultores de inteligência, promotores de justiça, administradores de sistema e auditores; (c) logs de acesso imutáveis, auditados periodicamente por órgão externo; (d) anonimização ou pseudonimização dos dados para finalidades estatísticas e analíticas, preservando os dados identificados apenas para finalidades processuais específicas; (e) avaliação de impacto à proteção de dados (DPIA — Data Protection Impact Assessment) antes da implantação de novos módulos analíticos, especialmente os que utilizem IA; e (f) plano de resposta a incidentes, com notificação ao

CNMP, à Autoridade Nacional de Proteção de Dados (ANPD) e, quando pertinente, ao titular dos dados afetados.

Um risco específico merece atenção destacada: o viés algorítmico. Quando os dados que alimentam um sistema de IA criminal foram coletados em contexto de policiamento seletivo — com sobrerrepresentação de determinados grupos populacionais nos registros de suspeitos, infrações e condenações —, os algoritmos treinados nessa base reproduzem e amplificam a discriminação estrutural, violando os princípios da igualdade (art. 5º, I, da CF) e da não discriminação. A mitigação exige: (i) auditoria de viés (bias audit) sobre os dados de entrada e sobre os outputs do sistema; (ii) testes de disparidade demográfica nos resultados analíticos; (iii) documentação das escolhas de design e das limitações do modelo; e (iv) supervisão humana obrigatória sobre decisões que impactem direitos individuais.

9. Conclusão: uma nova arquitetura de controle externo

O controle externo, no mundo digital, não se esgota em inspeções e recomendações. Ele se realiza como exigência de rastreabilidade. A rastreabilidade, por sua vez, depende de linguagem comum, metodologia de coleta, qualidade do dado e auditabilidade de acessos e transformações. Sem essas condições, o controle externo permanece retórico — uma garantia formal que não alcança o campo onde as decisões são efetivamente tomadas: na cadeia de microdecisões investigativas que antecedem o ato acusatório.

A construção do Banco Nacional de Dados do CNMP, assentada na BDP/MP e articulada a bases nacionais do Executivo — em especial o Sinic — por interoperabilidade governada, representa uma arquitetura institucional capaz de aumentar eficiência investigativa, fortalecer direitos fundamentais e permitir autocrítica do sistema de justiça penal. Essa arquitetura é necessária, mas não suficiente: ela precisa ser acompanhada de lei específica para o tratamento de dados na persecução penal (a ser aprovada na forma exigida pelo art. 4º, §1º, da LGPD), de uma Autoridade Nacional de Proteção de Dados (ANPD) fortalecida em sua capacidade de fiscalizar o Poder Público, e de uma cultura institucional de governança de dados que ainda precisa ser construída nas organizações de segurança pública brasileiras.

A literatura internacional converge, com variações de ênfase, em torno de cinco lições fundamentais para a construção desse tipo de infraestrutura: (i) a fragmentação de dados é o principal obstáculo à inteligência criminal eficaz, e a padronização semântica é condição

prévia à integração; (ii) a centralização de dados sem governança adequada cria riscos de abuso, discriminação e instrumentalização por atores criminosos; (iii) a supervisão humana é insubstituível — algoritmos identificam padrões, mas não exercem julgamento; (iv) a accountability externa (auditoria, supervisão parlamentar, controle jurisdicional) é condição de legitimidade; e (v) a interoperabilidade federada — arquitetura em que os dados residem nos órgãos de origem e são acessados por consulta controlada, como no modelo da RIBPG e da INTERPOL — é mais compatível com o federalismo brasileiro e com os princípios de proteção de dados do que a centralização física irrestrita.

Ao Ministério Público, como titular da ação penal e guardião do Estado Democrático de Direito (art. 127, caput, da CF), cabe liderar esse processo — não porque o Banco Nacional seja de sua propriedade exclusiva, mas porque nenhum outro ator institucional reúne, com a mesma abrangência constitucional, o mandato de acusar, controlar e investigar. A unidade informacional do MP não é centralismo; é o pressuposto epistêmico de uma persecução penal que aspira à coerência, à equidade e à possibilidade de ser corrigida.

---

10. Referências bibliográficas

AMBROSIO, Gleiner Pedroso Ferreira; BARBOSA, André Luis Jardini. O paradigma da implantação da inteligência artificial na segurança pública brasileira: regulação versus eficiência. Revista de Estudos Jurídicos da UNESP, v. 28, n. 48, 2024.

ARRIETA, Alejandro Barredo et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, v. 58, p. 82-115, 2020.

CHMIELINSKI, Kasia et al. The CLeAR Documentation Framework for AI Transparency: recommendations for practitioners and context for policymakers. Cambridge, MA: Shorenstein Center/HKS, 2024.

GROSSI, Alexandre Viezzer. A aplicação da Inteligência Artificial na segurança pública brasileira: o caso de São Paulo e a análise do PL nº 2338/2023. Revista Políticas Públicas & Cidades, v. 14, n. 4, 2025. ISSN: 2359-1552. DOI: https://doi.org/10.23900/2359-1552v14n4-72-2025.

INTERPOL. Rules on the Processing of Data (RPD). Lyon: INTERPOL, 2019.

INTERNATIONAL MONETARY FUND (IMF). Data Quality Assessment Framework (DQAF). Washington, D.C.: IMF, 2012.

IPEA; SENASP/MJSP. Anuário Estatístico de Segurança Pública 2023-2024. Brasília: Ipea, 2025. DOI: https://dx.doi.org/10.38116/ri-anuario-estatistico-2023-2024.

KERDVIBULVECH, Chutisant. Big Data and AI-driven evidence analysis: a global perspective on citation trends, accessibility, and future research in legal applications. Journal of Big Data, v. 11, n. 180, 2024.

KIM, Kyung-Jong; LEE, Chan-Hwi; BAE, So-Eun; CHOI, Ju-Hyun; KANG, Wook. Digital forensics in law enforcement: A case study of LLM-driven evidence analysis. Forensic Science International: Digital Investigation, v. 54, art. 301939, 2025. DOI: https://doi.org/10.1016/j.fsidi.2025.301939.

KÜÇÜK, Dilek; CAN, Fazli. Computational law: datasets, benchmarks, and ontologies. arXiv, 2025. Preprint 2503.04305v2.

MAORO, Falk; GEIERHOS, Michaela. Contestable AI for criminal intelligence analysis: improving decision-making through semantic modeling and human oversight. Frontiers in Artificial Intelligence, v. 8, art. 1602998, 2025. DOI: 10.3389/frai.2025.1602998.

MJSP/CG-RIBPG. XXII Relatório da Rede Integrada de Bancos de Perfis Genéticos (RIBPG): Dados estatísticos e resultados — Nov/2024 a Mai/2025. Brasília: MJSP, maio 2025.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg, MD: NIST, 2023. (NIST.AI.100-1). DOI: https://doi.org/10.6028/NIST.AI.100-1.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). Recommendation of the Council on Artificial Intelligence. Paris: OECD, 2019 (revisada em 2024).

OSÓRIO, Fábio Medina. O direito à compreensão na Era da complexidade tecnológica: fundamentos constitucionais, estatísticos e algorítmicos da transparência decisória. Revista dos Tribunais, v. 1077/2025, jul. 2025. DTR\2025\7689.

PADIU, Bogdan; IACOB, Radu; REBEDEA, Traian; DASCALU, Mihai. To what extent have LLMs reshaped the legal domain so far? A scoping literature review. Information, v. 15, n. 11, 2024.

POZZI, Riccardo; BARBERA, Valentina; PRINCIPE, Renzo Alva; GIARDINI, Davide; PALMONARI, Matteo. Combining Knowledge Graphs and NLP to Analyze Instant

Messaging Data in Criminal Investigations. In: Proceedings of WISE 2024. Springer, 2024. DOI: https://doi.org/10.1007/978-981-96-0567-5_30.

PYTLOWANCIV, Diogo Fernando Sampaio. Intelligence-Led Policing e sua Possibilidade de Implementação no Brasil. Revista Brasileira de Ciências Policiais, v. 15, n. 1, p. 103-123, jan./abr. 2024. ISSN: 2318-6917.

RIGANO, Christopher. Using Artificial Intelligence to Address Criminal Justice Needs. NIJ Journal, n. 280. Washington, D.C.: National Institute of Justice, jan. 2019. NCJ 252038.

SOUNDTHINKING, INC. Form 10-K: Annual Report Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934 (Fiscal Year Ended December 31, 2024). U.S. Securities and Exchange Commission, 2025. Commission File Number 001-38107; Nasdaq: SSTI.

TSUNODA, Denise Fukumi; CÂNDIDO, Ana Clara; GUIMARÃES, André José Ribeiro. Tecnologias disruptivas em segurança pública: uma análise situacional brasileira. Revista Tecnologia e Sociedade, v. 20, n. 61, p. 317-333, jul./set. 2024. DOI: 10.3895/rts.v20n61.18408.

UNESCO. Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO, 2021. Código SHS/BIO/REC-AIETHICS/2021.

UNITED NATIONS. Fundamental Principles of Official Statistics. Resolution 68/261. New York: United Nations Statistics Division, 2014. A/RES/68/261.

UNITED STATES DEPARTMENT OF JUSTICE. Artificial Intelligence and Criminal Justice: Final Report. Washington, D.C.: U.S. DOJ, 3 dez. 2024.

VOUGHT, Russell T. M-25-21: Accelerating Federal Use of AI through Innovation, Governance, and Public Trust. Washington, D.C.: Executive Office of the President, Office of Management and Budget, 3 abr. 2025.

11. Referências legislativas

BRASIL. Constituição da República Federativa do Brasil de 1988.

BRASIL. Emenda Constitucional nº 115, de 10 de fevereiro de 2022. Inclui a proteção de dados pessoais entre os direitos e garantias fundamentais (art. 5º, LXXIX, CF).

BRASIL. Lei nº 13.675, de 11 de junho de 2018. Institui o Sistema Único de Segurança Pública (SUSP).

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD).

BRASIL. Supremo Tribunal Federal. RE 593.727 (Tema 184). Poderes investigatórios do Ministério Público. Brasília: STF.

CNMP. Resolução nº 318, de 28 de outubro de 2025. Base de Dados Processuais do Ministério Público (BDP/MP).

MJSP. Portaria nº 1.122, de 5 de janeiro de 2026. Protocolo Nacional de Reconhecimento de Pessoas em Procedimentos Criminais.

MJSP. Portaria nº 1.123, de 5 de janeiro de 2026. Sistema Nacional de Informações Criminais (Sinic).

UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024 (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689. ELI: http://data.europa.eu/eli/reg/2024/1689/oj.

UNITED STATES OF AMERICA. Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Federal Register, 30 out. 2023. Revogada pelo Presidente Trump em 20 de janeiro de 2025.

UNITED STATES OF AMERICA. Tribal Law and Order Act of 2010. Pub. L. 111-211, 124 Stat. 2258.