
O dado que sua empresa ainda não sabe que tem

Os quatro níveis em que a informação se organiza, se conecta e se torna estratégica — do arquivo interno ao dado público brasileiro

Marcos Cordioli

Resumo

A inteligência artificial não está apenas otimizando processos: está reorganizando a estrutura profunda da economia contemporânea em torno de um novo insumo estratégico, os dados. Essa reorganização acontece em quatro camadas interdependentes — o dado interno, o dado da cadeia produtiva, o dado de mercado e o dado público — cada uma com sua própria lógica, seus próprios atores e seus próprios desafios de governança. Entender essas camadas é o primeiro passo para transformar informação dispersa em inteligência aplicável.

Palavras-chave: Economia de dados. Data Lake. Cadeias produtivas. Dados públicos. Interoperabilidade. Governança de dados.

Abstract

Artificial intelligence is not merely optimizing processes: it is reorganizing the deep structure of the contemporary economy around a new strategic input, data. This reorganization takes place in four interdependent layers — internal data, supply chain data, market data and public data — each with its own logic, its own actors and its own governance challenges. Understanding these layers is the first step in turning scattered information into applicable intelligence.

Keywords: Data economy. Data Lake. Supply chains. Public data. Interoperability. Data governance.

1 | A pergunta que antecede qualquer arquitetura de dados

A inteligência artificial não está apenas otimizando processos existentes — está reorganizando a estrutura profunda da economia contemporânea em torno de um novo insumo estratégico: os dados. Essa reorganização, porém, não ocorre de forma linear. Ela se dá em camadas interdependentes, cada uma com sua própria lógica, seus próprios atores e seus próprios desafios de governança. Antes de investir em qualquer tecnologia, é necessário responder a uma pergunta mais elementar: em qual desses níveis os dados de uma organização efetivamente se encontram hoje. É essa pergunta que este texto tenta ordenar, a partir de quatro níveis distintos de problematização — ainda que essa separação seja, por natureza, provisória e analítica.

2 | O dado que a organização não sabe que tem

O primeiro nível é o Data Lake interno das corporações. Trata-se da capacidade da própria empresa reunir, integrar, estruturar e transformar em inteligência todos os dados produzidos por seus processos, operações, clientes, sensores, documentos e sistemas digitais — dados transacionais de ERPs e CRMs, dados operacionais de máquinas e IoT, dados comportamentais de clientes em plataformas digitais, registros administrativos, informações de cadeia de suprimentos, de recursos humanos, financeiras, de qualidade e, cada vez mais, de comunicação corporativa interna.

A maioria das organizações ainda opera com dados dispersos em múltiplos sistemas isolados, sem padronização de formatos, sem metadados consistentes e sem uma arquitetura de governança capaz de garantir a qualidade e a rastreabilidade das informações. O dado interno é, paradoxalmente, o mais próximo e o mais negligenciado. A problematização de partida é, portanto, anterior a qualquer integração externa: a organização ainda não sabe, na maioria dos casos, o que de fato sabe — nem sequer consegue inventariar com precisão onde seus dados estão, em que condição, com qual grau de confiabilidade.

O dado interno é, paradoxalmente, o mais próximo e o mais negligenciado: a maioria das organizações ainda não sabe o que já sabe.

Tecnologias e soluções para esse problema existem e estão sendo mobilizadas com grande velocidade. O desafio já não é predominantemente técnico — é de consciência organizacional e de escolha adequada. Num mercado que oferece muito e entrega resultados variáveis, muitas corporações, no afogadilho das mudanças tecnológicas, contratam a primeira solução que aparece, gerando mais camadas de inconsistência do que de inteligência. Saber o que se precisa antes de contratar o que se oferece constitui, em si, uma competência estratégica rara.

3 | O dado que a cadeia produtiva ainda não compartilha

O segundo nível envolve os Data Lakes integrados às cadeias produtivas. Aqui entram fornecedores, produtores integrados, cooperados, operadores logísticos e todos aqueles que participam diretamente do sistema econômico da empresa. Na agroindústria, por exemplo, isso significa integrar dados de propriedades rurais, produção agrícola, rastreabilidade sanitária, logística de transporte e processamento industrial em um único fluxo coerente de informações.

O desafio aqui é de outra ordem: não se trata apenas de tecnologia, mas de negociação de interesses, de acordos sobre propriedade do dado e de construção de confiança entre atores com assimetrias significativas de poder econômico. A definição de quem detém o dado da cadeia, de quem captura o valor gerado por ele e de quem responde pelos riscos de sua exposição ainda não encontrou modelos consolidados — e essa lacuna tem implicações diretas sobre a soberania de dados dos agentes econômicos menores da cadeia.

4 | **O dado que vale mais para quem não o produziu**

O terceiro nível envolve as relações de distribuição e circulação econômica. São empresas que não necessariamente integram o sistema produtivo principal, mas que mantêm relações estratégicas de mercado — redes varejistas, distribuidores, marketplaces e canais de comercialização. Neste caso, os dados permitem compreender circulação de produtos, comportamento de consumo, dinâmica territorial de vendas e padrões econômicos que raramente são visíveis a partir de qualquer ponto isolado da cadeia.

É aqui que começa a emergir algo qualitativamente novo: a possibilidade de que o valor do dado ultrapasse o interesse original de quem o produziu. Um produtor rural que rastreia sua produção pode estar, sem perceber, gerando informações extraordinariamente valiosas para uma rede de supermercados que nunca pisou em sua propriedade. Esse deslocamento entre quem produz e quem captura valor é, como se verá adiante, uma das características mais distintivas da economia de dados.

5 | **O dado público: o mais estratégico e o mais fragmentado**

O quarto nível talvez seja o mais complexo — e o mais decisivo: os dados públicos. Hoje, essas informações estão dispersas entre governos federal, estaduais e municipais, autarquias, universidades, concessionárias, institutos de pesquisa, agências reguladoras e empresas privadas de monitoramento territorial. Essa dispersão é real e seus custos são elevados — mas sua reversão é possível por meio de metodologias explícitas e ações concretas.

Existe fragmentação horizontal entre esferas de governo: dados de receitas federais, estaduais e municipais contêm massas volumosas de informações sobre empresas — registros de movimentação fiscal que poderiam ser cruzados para compreender dinâmicas econômicas locais com precisão considerável. Hoje, essas bases não conversam entre si, mesmo habitando o mesmo território. Existe fragmentação por domínio setorial: o mesmo fenômeno é registrado de formas radicalmente diferentes por autoridades distintas, criando lacunas que impedem a compreensão causal. Um acidente de trânsito com vítima, por exemplo, é registrado pela polícia como ocorrência com lesão corporal. Dias depois, a mesma vítima falece no hospital por complicações das lesões; no sistema de saúde, esse óbito entra como estatística de mortalidade hospitalar, mas na estatística oficial de mortalidade por acidente de trânsito da polícia o caso não aparece, porque o registro ocorreu no momento do acidente, não em sua evolução temporal. O resultado é que ninguém compreende o verdadeiro impacto do acidente sobre a mortalidade: a polícia não vê seus casos "leves" transformarem-se em mortes, a saúde não consegue rastrear a origem do óbito, e o judiciário opera com uma terceira classificação, baseada em culpa e responsabilidade, ortogonal às duas anteriores. Essas três visões do mesmo fenômeno, capturadas em um Data Lake integrado, revelariam rotas de violência e seus impactos reais sobre a população.

Existe ainda fragmentação por tipo de pessoa jurídica: dados de pessoas físicas — identidade, educação, saúde, previdência — residem em sistemas distintos dos dados de empresas. Um indivíduo que é simultaneamente trabalhador registrado,

microempreendedor e beneficiário de programa social existe como três entidades separadas em bases distintas. E existe fragmentação por concessionária: dados de iluminação pública, saneamento, energia e rodovias — informações críticas para a gestão territorial — residem em empresas que operam sob concessão. Uma concessionária de saneamento sabe onde há ruas sem cobertura adequada, em quais bairros há maior incidência de vazamentos, qual é a correlação entre cobertura e indicadores de saúde. Quando esses dados são organizados em um Data Lake público, transformam-se em inteligência territorial compartilhável.

Há, por fim, uma categoria pouco discutida: os dados que ninguém produz naturalmente, mas cuja produção é crítica. Levantamentos socioeconômicos de precisão, inventários de infraestrutura, mapeamentos de cobertura vegetal e risco ambiental, caracterizações de mobilidade urbana — esses dados são frequentemente contratados por governos junto a empresas especializadas em sensoriamento remoto, pesquisa social ou consultoria. O desafio é que, após produzidos, permanecem como ativos isolados, raramente integrados em Data Lakes acessíveis para políticas futuras. Construir repositórios públicos permanentes, com metodologias de atualização incremental e acesso estruturado, permite amortizar seus custos de produção ao longo de múltiplas aplicações e ciclos de política pública.

O ponto central é este: a fragmentação hoje existente não constitui uma condição permanente. É um estado que pode ser progressivamente revertido por meio de metodologias explícitas de convergência — o que requer construir um grande Data Lake de dados públicos, agregando as esferas federal, estaduais e municipais, além de outras instituições e concessões, e requer também equacionar o desafio do sigilo. Muitos desses dados envolvem graus de privacidade econômica, pessoal ou institucional que não podem ser ignorados. A solução não é publicizar dados individualizados, mas a dimensão estatística e estrutural daquilo que os dados revelam. A técnica da agregação estatística — mantendo privado aquilo que é singular e tornando público aquilo que é padrão — permite que a inteligência territorial seja compartilhada sem violar direitos de privacidade. O Consórcio de Dados público, estruturado com essa disciplina, torna-se não apenas uma aspiração tecnológica, mas uma infraestrutura de inteligência pública capaz de sustentar decisões informadas em todos os níveis de governo.

6 | O problema transversal: dois andares em construção simultânea

O desafio transversal a esses quatro níveis é a interoperabilidade — e é nele que reside, talvez, a tensão mais produtiva de todo esse debate. É útil identificar dois andares de um mesmo edifício em construção simultânea. O primeiro andar é o da organização interna dos dados: cada corporação, cada instituição pública, cada ator da cadeia produtiva ainda está, em sua grande maioria, resolvendo seus próprios problemas de qualidade, padronização, metadados e governança interna. É um trabalho de base, indispensável e urgente. O segundo andar é o da interoperabilidade entre Data Lakes: a capacidade de conectar ecossistemas distintos, de fazer com que dados produzidos em contextos diferentes dialoguem de forma estruturada e de construir inteligência coletiva a partir dessas conexões.

O ponto crítico é este: o segundo andar precisa ser projetado enquanto o primeiro ainda está sendo construído. Não é possível esperar que todas as organizações tenham seus Data Lakes internos plenamente organizados para então começar a pensar em integração — pois, quando isso ocorrer, os padrões de incompatibilidade já estarão consolidados e serão muito mais difíceis e custosos de superar. Construir o primeiro e o segundo andares em paralelo, com uma arquitetura mínima de interoperabilidade desde o início, é a condição para que a conexão entre esses quatro níveis seja possível. Isso exige não apenas soluções técnicas, mas protocolos de confiança, padrões de certificação, mecanismos de governança compartilhada e marcos jurídicos capazes de mediar interesses de atores profundamente distintos. A LGPD oferece um ponto de partida regulatório, mas está longe de ser suficiente para estruturar ecossistemas territoriais de inteligência que envolvam, simultaneamente, dados corporativos, dados de cadeias produtivas e dados públicos.

7 | Considerações finais

Nenhuma organização ocupa apenas um desses quatro níveis. Toda empresa é, ao mesmo tempo, produtora de dado interno, elo de uma cadeia produtiva, agente de um mercado de circulação e beneficiária — ou vítima — da fragmentação dos dados públicos que a cercam. Mapear em qual desses níveis seus dados hoje se encontram, e o que falta para movê-los ao nível seguinte, é o primeiro passo estratégico antes de qualquer investimento em tecnologia.

A pergunta que decide o futuro competitivo de uma organização não é mais quantos dados ela tem, mas em quantos desses quatro níveis ela já sabe operar.

Essa mudança de pergunta — de quantidade para capacidade de operação — é o que separa organizações que acumulam dados como arquivo morto daquelas que os transformam, de fato, em inteligência aplicável. É também, como se argumenta em texto complementar sobre a arquitetura rizomática desses ecossistemas, a condição prévia para que qualquer organização possa participar, com posição definida, da economia de dados que já está em formação.

Declaração sobre o uso de inteligência artificial

A inteligência artificial generativa foi utilizada como ferramenta de apoio à organização das informações fornecidas para este artigo. O uso ocorreu como auxílio ao trabalho intelectual, sem substituir a análise dos documentos e das fontes. A seleção e a interpretação das informações, assim como a revisão e a responsabilidade pela versão final do texto, cabem ao autor. Ao final, a IA foi utilizada para a correção ortográfica e semântica do texto e também para gerar o resumo e as palavras-chave.

Sobre o autor

Marcos Cordioli é consultor, empresário, pesquisador e palestrante brasileiro, mestre em Educação pela PUC-SP, com atuação nas áreas de educação, transformação digital, inovação, cidades inteligentes, economia de dados e inteligência artificial.

Foi Secretário de Inovação, Aceleração Econômica, Turismo e Comunicação de Maringá e Secretário de Cultura de Curitiba, além de Presidente da Fundação Cultural de

Curitiba. Desde 2015, atua na modelagem de ecossistemas de inovação e startups, aproximando tecnologia, conhecimento, negócios e transformação institucional.

É cofounder da Apolus.ai, startup dedicada a transformar conhecimentos dispersos em ecossistemas de inteligência artificial, ampliando a capacidade de empresas e instituições de organizar, compartilhar e utilizar conhecimento como inteligência coletiva. É também vice-presidente da ACIM-Maringá.

Na área de tecnologia, pesquisa os processos econômicos e tecnológicos associados à economia de dados, estudando dados, conhecimento e inteligência artificial como elementos estruturantes da nova economia. Dedicar-se também ao estudo da transformação digital de cidades, com atenção ao uso sistêmico de IoT, dados, inteligência artificial, conectividade e plataformas digitais na construção de cidades inteligentes.

No campo da inteligência artificial, desenvolve duas frentes complementares. Na educação, conduz o projeto "Tecnologias de IA na Sala de Aula", dedicado às possibilidades de utilização da inteligência artificial nos processos de ensino e aprendizagem, na formação de educadores e na gestão educacional. Nas empresas e instituições, atua na área de educação corporativa, pesquisando a gestão de conteúdos e conhecimentos como base para a construção de inteligência coletiva e para o desenvolvimento de novas soluções apoiadas por IA.

Marcos Cordioli também possui trajetória acadêmica na área de educação, tendo atuado como professor em cursos de graduação, especialização e mestrado, além de pesquisador, palestrante e consultor. Participou de eventos, pesquisas e projetos educacionais em todos os estados brasileiros e em diferentes países.

Contato: marcos.cordioli@gmail.com | (41) 99682-9123 | LinkedIn: [linkedin.com/in/marcos-cordioli](https://www.linkedin.com/in/marcos-cordioli) | Instagram: [@marcoscordioli](https://www.instagram.com/marcoscordioli) | Linktree: linktr.ee/marcos.cordioli