

CIÊNCIA DE DADOS E MONITORAMENTO SOCIAL: UMA ANÁLISE DO PAINEL DE INTEGRIDADE PÚBLICA

Trabalho de Conclusão de Curso
apresentado como parte dos requisitos para
obtenção do grau de Especialista em Ciência de
Dados Aplicada a Políticas Públicas.

Aluno: Thiago Mendes Rodrigues

Orientador: Prof. Me. Francisco Eduardo
Holanda Bessa

Brasília – DF

Outubro/2022

CIÊNCIA DE DADOS E MONITORAMENTO SOCIAL: UMA ANÁLISE DO PAINEL DE INTEGRIDADE PÚBLICA

Autor: Thiago Mendes Rodrigues
Ministério da Economia

Ciência de Dados; Integridade Pública; Monitoramento Social

Resumo

O artigo analisa a possibilidade de monitoramento social dos programas de integridade do Poder Executivo Federal a partir do tratamento de dados abertos governamentais disponibilizados pela CGU no Painel de Integridade Pública, com apoio de outras bases de dados. O mérito da abordagem é aplicar conceitos da ciência de dados nesse tema, bem como compreender incentivos ao monitoramento social sobre a Integridade Pública, enquanto política pública. As plataformas utilizadas foram o GitHub, Google Collab, Big Query e Data Studio. As principais conclusões são de que a iniciativa precursora da CGU converge com recomendações da OCDE sobre integridade pública, gestão de riscos e ciência de dados; que isso pode significar um incentivo ao monitoramento social e replicação sistemática pelos integrantes do Sistema de Integridade Pública do Poder Executivo Federal – SIPEF, como abertura estruturada, organizada e espontânea dos dados; e que há janela para aprimorar a qualidade e indicadores dos dados, bem como a interoperabilidade entre as bases de dados.

1. INTRODUÇÃO

A Integridade Pública refere-se ao alinhamento consistente e à adesão a valores, princípios e normas éticas comuns para sustentar e priorizar o interesse público sobre os interesses privados (OCDE, 2018), e tem ganhado relevância no debate público nos últimos anos no Brasil, a partir da definição da integridade como princípio da governança pública¹; e da instituição do Sistema de Integridade Pública do Poder Executivo Federal – Sipef².

O contexto do esforço brasileiro mais recente de estruturação e promoção da Integridade Pública decorre da progressiva adoção da vertente teórica de combate à corrupção com foco na prevenção, baseada no conceito denominado neste artigo de “virtude cíclica”. As outras duas vertentes vigentes – e que estruturaram historicamente as políticas públicas nesse tema no Brasil – têm foco na repressão, baseadas no conceito weberiano de patrimonialismo, predominante nos anos 90, e no conceito de *rent-seeking*, advindo da teoria da escolha pública, hoje hegemônico (VIOL, 2021, p. 123).

O *rent-seeking* pode ser entendido como privilégio especial³ (transferência de renda econômica a pessoas ou grupos de interesse), que resulta diretamente em ineficiências, com danos indiretos piores ainda: a maior parte produtiva da sociedade arca com atividades que podem causar externalidades negativas duradouras ao bem-estar econômico e social. Essa causa explica mais a estagnação dessas sociedades do que o custo direto dos privilégios especiais (TULLOCK et al., 2005, p. 61).

Paralelamente, destaca-se que a corrupção pode ser um fenômeno induzido (e esperado) em contextos de burocracia excessiva, descrédito social e impunidade, à medida que fornece meios ágeis para concretização de negócios e favorece aqueles que a utilizam como forma de burlar as exigências normativas (LIMA *et al.*, 2019, p. 65). Esse fenômeno pode ser agravado num contexto de exaustão estatal da burocracia de controles, cuja boa intenção em busca por eficácia incorre na

¹ Ver Decreto nº 9.203, de 22 nov. 2017. Esta norma criou ainda o Comitê Interministerial de Governança (CIG); definiu a criação dos comitês internos de governança; reforçou o papel da gestão de riscos; e definiu a instituição dos programas de integridade.

² Ver Decreto nº 10.756, de 27 jul. 2021.

³ Análogo às “janelas de corrupção”, segundo Braga (2022), que cita como exemplos: benefícios fiscais em favor de pessoas ou setores, com possíveis benefícios indevidos a agentes públicos; edição de normas regulatórias enviesadas; restrição de concorrência pelo afastamento arbitrário de relevantes agentes do mercado; aplicação de multas e sanções passíveis de negociações escusas (suborno, troca de favores) a fim do perdão das penalidades.

repressão como tradição, com foco na criação de mais regras (controles), normas mais rígidas, e fiscalização de conformidade mais minuciosa. Esse contexto viciado pode agravar o risco de majoração de custos de transação e de incentivos adversos a “soluções de contorno” (suborno, captura regulatória, inobservância deliberada das normas).

Com viés preventivo, por sua vez, a virtude cíclica pode ser entendida como a abordagem voltada à aplicação estruturada de instrumentos, recursos e iniciativas (sensibilização, mobilização, incentivos, monitoramento) com foco em promover a Integridade Pública institucionalmente e junto à sociedade, de forma a alcançar – com a devida aferição – uma melhoria da percepção da “cultura de integridade”, que pode ser sintetizada com base nos elementos da Figura 1 abaixo:

Figura 1 – Elementos da Cultura de Integridade Pública



Fonte: OCDE (2018, p .5).

A virtude cíclica, outrossim, reflete em parte a tipificação moral, contextualização e contínua formação de “capital social”, um atributo da sociedade que envolve elementos como as redes horizontais de cooperação mútua, engajamento cívico, cooperação espontânea, confiança interpessoal, normas e sistemas, cuja sinergia contribui para aumentar a eficiência da sociedade, facilitando as ações coordenadas (PUTNAM, 1997).

A consistência do capital social favorece a solidez da democracia e a eficiência econômica social, de forma que a cooperação geral, enquanto equilíbrio condicional, depende da existência de laços de confiança mútua, que reforcem os mecanismos de cooperação entre os habitantes e favoreçam o desempenho das instituições políticas; esse mesmo desempenho institucional eficiente atua positivamente sobre o contexto social, reduzindo a incerteza e reforçando ainda mais o nível de confiança e cooperação no interior da população (REIS, 2003, p. 38).

Esclarece-se que o “equilíbrio condicional” deriva de analogia entre os incentivos a um contexto social de cooperação espontânea e uma solução do dilema dos prisioneiros, da teoria dos jogos. Nesse caso, na ocorrência de reiterada e infinita conduta “*tit-for-tat*”, é esperado que os “jogadores” sejam incentivados a cooperar por medo de retaliação, já que seus oponentes podem se comportar como eles no futuro. Porém, quando se parte da situação do “egoísmo universal⁴”, o equilíbrio de cooperação geral não será espontâneo, exceto sob condições específicas de súbita instabilidade interna, baixa interação entre grupos divergentes, e capacidade – e disposição – da minoria cooperativa em retaliar a maioria não-cooperativa da população (ZAGARE, 1984, p. 58-62 *in* REIS, 2003, p. 37).

A cooperação espontânea dependerá sempre de uma taxa de desconto suficientemente baixa na preferência temporal dos atores, de modo a permitir que eles renunciem à possibilidade de um ganho imediato que seria propiciado pelo abandono da estratégia cooperativa comum, a fim de evitar um equilíbrio ineficiente no futuro (REIS, 2003, p. 38). A não-cooperação incondicional geral, porém, uma vez alcançada, é um equilíbrio mais estável, no sentido de incentivos e incerteza de curto prazo. Daí a importância do reforço de percepção de contexto de confiança mútua e de estabilidade das “regras do jogo”, a fim de viabilizar a conduta geral de cooperação espontânea (solução de equilíbrio desejável), menos pelo medo de retaliações, e mais por meio de incentivos e sanções previsíveis.

Assim, é esperada uma “mudança de cultura” consistente a partir de aspectos como:

- a) percepção de garantia de fruição dos benefícios esperados (imediatos ou não) decorrentes da cooperação mútua;
- b) redução de custos de observância de normas, a partir de regras claras e simples, e de estabilidade nas “regras do jogo”;
- c) redução de incerteza no contexto social, a partir do aumento de previsibilidade da conduta alheia, alcançado com a aplicação consistente e impessoal de sanções e incentivos; e
- d) redução do medo e da necessidade de uso repressivo de força pelo Estado.

Por analogia, esses aspectos dão consistência à efetividade da virtude cíclica esperada decorrente da “cultura de integridade” e da Integridade Pública, enquanto política pública de esforço preventivo, de forma a promover o amadurecimento social a partir da indução de contexto

⁴ Situação na Teoria dos Jogos em que todos os jogadores adotam a estratégia inicial de não cooperar.

de confiança institucional e de honestidade.

É importante que a Integridade Pública promova também a contribuição ao debate sobre como definir, padronizar, aferir e monitorar empiricamente conceitos como a “cultura de integridade” e seus elementos, bem como a efetividade do esforço de prevenção junto à sociedade. É essencial (e desafiador) evidenciar os benefícios imediatos e futuros (diretos ou não) do esforço de prevenção – não tão perceptíveis como os da repressão –, de forma que a Integridade Pública não seja vista como retórica nem mera agenda de conformidade de ritos, ou, no limite da contradição, um constrangimento social por simbolizar um modismo⁵ ou por reforçar a opacidade da “cultura de controles” da burocracia.

Destaca-se criticamente que a própria definição de “capital social” a partir de enumeração vaga de atributos não chega ser conceitual, e peca, nesse sentido, por admitir em si fenômenos diversos de natureza distinta (REIS, 2003, p. 44). Esse é outro alerta válido à Integridade Pública, enquanto política pública em estruturação: é essencial a capacidade de dar plena compreensão de seu escopo e sentido, e de aferir sua efetividade nas dimensões em que se pretende intervir, considerando que a sociedade é a beneficiária última das ações de integridade pública.

Outro elemento essencial ao conceito de virtude cíclica aplicada pelo Poder Público é o levantamento sistemático de riscos à integridade⁶, como insumos ao levantamento, direcionamento e priorização da articulação de ações de prevenção a serem adotadas. Esse papel central da gestão de riscos ajuda a explicar também a crescente relevância dada ao Modelo de Três Linhas (IIA, 2020) pelas organizações do setor público, enquanto estrutura que explicita a interlocução, alinhamento e interdependência entre diversos níveis funcionais organizacionais no que se refere à sinergia entre governança e gestão de riscos, com a consequente proposta/revisão de controles internos mitigadores de tais riscos.

A sinergia entre gestão de riscos e estratégia organizacional é ressaltada, por sua vez, pela OCDE, a fim de salvaguardar a integridade nas organizações do setor público, a partir de recomendações de compromissos como: a utilização de dados e indicadores em conjunto com o

⁵ Em analogia a um “mar de soluções”, Braga (2020) defende a importância de que o burocrata de ponta (gestor na ponta das políticas públicas) seja entendido e seduzido no processo de implementação de iniciativas que garantam efetividade à integridade como realidade palpável, sob risco de dissociação entre modismos de nichos da burocracia (com métricas, prêmios e certificados) e as limitações e problemas complexos da realidade do burocrata de ponta, como carências de toda ordem, improvisos, pressão da sociedade e de órgãos de controle.

⁶ Risco à integridade é a possibilidade de ocorrência de evento de corrupção, fraude, irregularidade ou desvio ético ou de conduta que venha a impactar o cumprimento dos objetivos institucionais (Decreto nº 10.756/2021).

levantamento de riscos à integridade, a fim de desenvolver uma estratégia com objetivos e prioridades delineados (OCDE, 2018, p. 5); e a garantia de uma abordagem estratégica que inclua a avaliação dos riscos à integridade, capaz de monitorar inclusive as deficiências de controle (*ibid.*, p. 12). Assim, gestão de riscos, governança e monitoramento sistemático com base em dados e evidências são elementos que contribuem à consistência da Integridade Pública.

Considerada sua vertente teórica, frisa-se que a Integridade Pública constitui política pública, enquanto conjunto de programas ou ações governamentais integradas e articuladas à provisão serviços à sociedade (BRASIL, 2018, p. 13), notadamente a prevenção, detecção, punição e remediação de práticas de corrupção e fraude, de irregularidades e de outros desvios éticos e de conduta.

A sociedade se beneficia da integridade de forma indireta, já que é possível, por exemplo, que a provisão de bens e serviços (educação, segurança, investimentos em saúde, construção de estradas, compras e contratações) seja efetivada pelo poder público em ritos e processos de trabalho que não sejam íntegros. Por outro lado, essas entregas serão precárias, pois carregarão a “mancha” de eventuais vícios, como sobrepreço, desvios de finalidade, opacidade; e uma ineficiência inerente decorrente de ritos que padecem de integridade em seus processos. Por isso, enquanto política pública, a Integridade Pública objetivamente também traduz valores sociais de reação à percepção de permissividade cultural ao “rouba, mas faz”⁷.

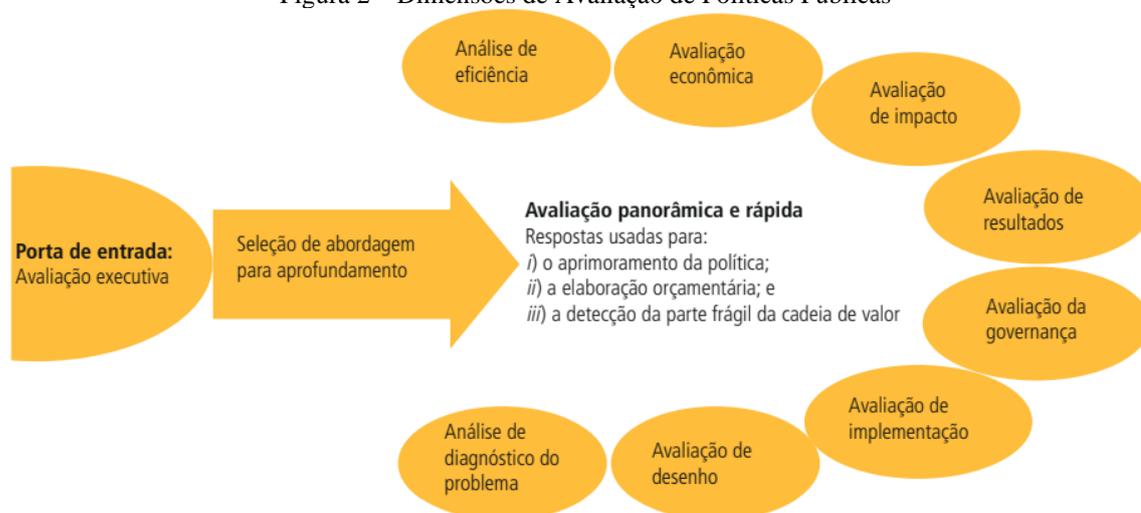
Nesse sentido, e para além da orientação burocrática e monitoramento normativo pela CGU, o incentivo ao monitoramento social dos programas de integridade tem base em iniciativas como o Painel de Integridade Pública, com apoio da ciência de dados.

O monitoramento social (aqui entendido como monitoramento do aparelho estatal pela sociedade), em sentido amplo, tem previsão no Decreto nº 9.203/2017, que define que é diretriz da governança pública manter o processo decisório orientado pelas evidências e pelo apoio à participação da sociedade.

Cabe esclarecer que não se confunde, porém, o monitoramento social com a avaliação da política pública. Conforme se depreende da Figura 2, a avaliação é multidimensional e robusta metodologicamente quanto aos impactos diretos e indiretos – esperados ou adversos – de uma política pública, bem como quanto à interação e percepção da população beneficiária.

⁷ O termo “rouba, mas faz” é parte do léxico político brasileiro e teve origem em críticas à gestão de Adhemar Costa Barros nos anos 1950 (COTTA, 2008, p. 7).

Figura 2 – Dimensões de Avaliação de Políticas Públicas



Fonte: Brasil (2018b, p. 30).

Já o monitoramento se restringe aos indicadores diretamente relacionados à sua execução (BRASIL, 2018b, p. 241), de forma que a participação da sociedade pode permitir maior conhecimento sobre a política e geração de informação gerencial qualitativa (*ibid.*, p. 166), bem como fomentar demanda por maior qualidade dos dados fornecidos pelo governo eletrônico, apontado como instrumento de grande potencial para elevar a eficiência governamental no Brasil, e capaz de estreitar a relação entre sociedade e Estado (*ibid.*, p. 167).

Nesse sentido, recentemente foi disponibilizada pela CGU a funcionalidade de acesso às bases de dados referentes aos programas de integridade de diversas organizações do setor público monitoradas, no âmbito de atualização do Painel de Integridade Pública (BRASIL, 2022a). Outra iniciativa de destaque foi a pesquisa realizada de 23 jun. a 31 ago. 2021 pela CGU para aferição da percepção sobre Integridade Pública junto a servidores públicos dos órgãos integrantes do Sipef. A conjugação inicial dessas duas bases de dados permitiu tratamento mais adequado de informações para elaboração do presente artigo.

Um aspecto que merece aprofundamento acadêmico, por outro lado, é que apesar da corrupção ser recorrentemente considerada um problema nacional grave, e existirem diversas iniciativas no Brasil para seu combate e controle, não se tem alcançado maior engajamento do cidadão comum, nem melhora sensível de percepção quanto à efetividade de seu enfrentamento (ODILLA, 2022).

Esse fenômeno pode ter diversas causas, que podem se constituir em objeto de estudo que não faz parte do escopo deste artigo. Porém, isso não reduz a importância de compreensão adequada desse fenômeno, nem da sistematização de iniciativas de estruturação, monitoramento e

disponibilização de dados no esforço de sensibilização institucional preventiva e engajamento social no tema.

Assim, consideradas essas abordagens conceituais até agora delineadas, este artigo busca responder: com base em dados e evidências, e a partir dos dados disponibilizados pela CGU sobre os programas de integridade, quais melhorias podem ser incorporadas ao Painel de Integridade Pública para possibilitar o aprimoramento do monitoramento e do controle social do Sipef?

Os objetivos específicos a fim de responder ao problema de pesquisa podem ser sintetizados em: identificar a necessidade de eventuais aprimoramentos na qualidade da disponibilização dos dados, em dimensões⁸ como completude e consistência, a fim de permitir maior inteligibilidade e monitoramento social por parte da sociedade; verificar a possibilidade de eventuais acréscimos de informações e outras métricas no portal; identificar em que medida os dados disponibilizados pela CGU podem possibilitar o aprendizado coletivo pelas unidades que conduzem os programas de integridade nos diferentes órgãos e entidades abrangidos pela atuação da CGU; e examinar a interoperabilidade com outras bases de dados disponíveis.

2. DESENVOLVIMENTO

2.1. REFERENCIAL TEÓRICO

Considerada a explicação introdutória sobre a relação entre monitoramento social e Integridade Pública, é possível avançar agora em questões que envolvem o uso e tratamento de dados atinentes a esses temas.

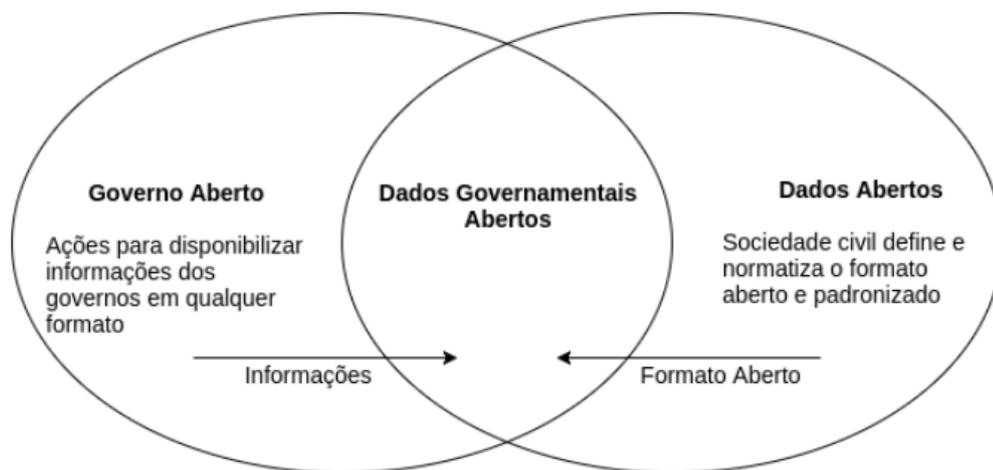
Os dados abertos governamentais (DAG) em geral devem atender a oito princípios: devem ser completos; primários; atuais; acessíveis; processáveis por máquina; de acesso não-

⁸ Van Nederpelt e Black (2020) listaram 65 dimensões para Qualidade de Dados, por exemplo.

discriminatório; de formato não-proprietário; e de licença livre (BRASIL, 2022b). Há outras listas mais minuciosas de princípios aplicáveis, dependendo da organização que os elenca.

Nesse sentido, frisa-se que é insuficiente que as bases de DAG existam se sua disponibilidade, bem como seu nível de acesso, impuserem constrangimento à verificação e tratamento (BRASIL, 2018a, p. 58). A Figura 3 ajuda a entender a relação entre o Governo Aberto, Dados Abertos e DAG:

Figura 3 – Relação entre Governo Aberto, Dados Abertos e Dados Abertos Governamentais



Fonte: Cardoso (2019, p. 47).

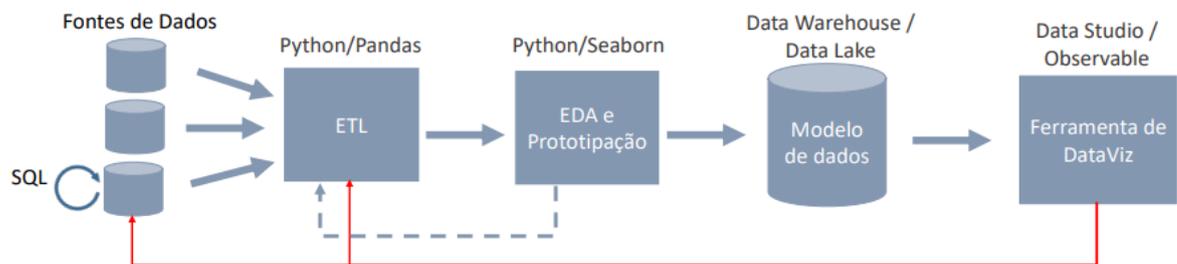
Em síntese, a explicação é que no universo de informações geradas ou mantidas pelo setor público (ISP), o Governo Aberto disponibiliza seus dados no formato que a sociedade (entidades diversas) padroniza, de forma que essa interseção permite a existência de Dados Abertos Governamentais que atendam aos princípios e requisitos definidos de forma padronizada.

No contexto brasileiro, destaca-se que o volume de DAG sobre Integridade Pública ainda é incipiente, de forma que chama a atenção a iniciativa precursora da CGU em abrir as bases de dados sobre o monitoramento dos programas de integridade no Painel de Integridade Pública (BRASIL, 2022a). Tal inovação converge com as sugestões de procedimentos para criação de plataforma que congregue bases de dados relacionadas à integridade pública (OCDE, 2022, p. 28), e de desenvolvimento de capacidades de letramento em dados e tecnologia da informação (*ibid.*, p. 2), o que pode ajudar a superar o desafio de aumentar o engajamento da sociedade no exercício de monitoramento social do sistema de prevenção da corrupção, em sentido amplo.

Esse contexto da Integridade Pública se constitui em campo fértil à aplicação da Ciência de Dados, compreendido como campo da ciência que transforma dados em informações, a partir de dimensões como: fundamentos teóricos (compreensão de um fenômeno, pergunta de pesquisa, monitoramento de política pública); métodos quantitativos (matemática, estatística, probabilidade); e métodos computacionais (linguagem de programação, algoritmos, bancos de dados) (PENG; SILVA; NAGATA; 2021, p. 2).

A partir da Figura 4, é possível entender um exemplo de fluxo de processos em Ciência de Dados:

Figura 4 – Fluxo de Processos em Método de Trabalho de Ciência da Dados



Fonte: Pereira (2021, p. 7).

Onde:

- SQL (*Structured Query Language*) é uma linguagem de consulta estruturada voltada a bancos de dados relacionais;
- ETL (*Extract Transform Load*) é a rotina de tratamento de dados, de forma que:
 - Extração (*Extract*) se refere a procedimentos como definição e seleção das bases de dados mesmo que em diversos formatos (CSV, ODS, XLS, XML, JSON), extração dos dados a serem utilizados, e aferição de sua compatibilidade para uso;
 - Transformação (*Transform*) se refere a procedimentos de tratamento e refino, como limpeza, junção e criação de informação a partir dos dados brutos; e
 - Carga (*Load*) se refere a subir os dados transformados a um armazém de dados (*Data Warehouse – DW*), por exemplo.
- EDA (*Exploratory Data Analysis*) se refere à análise exploratória de dados, normalmente a partir de inferência estatística, testes de hipóteses e uso de ferramentas de estatística e matemática computacionais; e

- DataViz é um conceito que se refere à comunicação e representação visual de dados.

Enquanto método de trabalho, recomenda-se que todo tratamento necessário (limpeza, correção, adição, redução, criação) de bases de dados seja realizado nos processos de SQL e ETL, de forma a mitigar riscos de inconsistências e ruídos de interoperabilidade entre os ambientes e ferramentas de dados.

2.2. MÉTODO

Foram utilizadas três fontes de dados com vistas ao exercício simulado de monitoramento social:

- Painel de Integridade Pública (BRASIL, 2022c), na posição de 4 mar. 2022;
- Resultados da Pesquisa sobre Integridade Pública (BRASIL, 2021); e
- Relação da alocação de servidores por órgão, a partir de dados do Sistema de Pessoal Civil da Administração Federal – Sipec (BRASIL, 2022d).

A seguir listam-se os passos do método de análise, sua aplicação e as respectivas dificuldades encontradas:

- I. *download* dos dados abertos do Painel de Integridade Pública em planilhas formato CSV (*Comma Separated Values* ou Valores Separados por Vírgulas), a fim de examinar a qualidade de sua disponibilidade e complexidade de tratamento;
- II. *upload* dos dados brutos estruturados ao **GitHub**, plataforma gratuita que, entre outras aplicações, permite criar repositórios remotos em nuvem para tratamento de arquivos e versionamento de projetos;
- III. tratamento dos dados no **Google Collab**, ambiente de programação em linguagem Python;
 - no tratamento de planilhas do Painel de Integridade Pública (12 ao todo), foram constatadas inconsistências, e providenciada a correção da tabela 1.2 (linhas: 51-79; 83; 150-153; 158; 164; 170; e 180-185), pois o formato CSV não permite nomes compostos sem o uso de aspas; e a mudança do delimitador (de “,” para “;”). Destaca-se que esse tratamento é feito no processo de ETL, e não diretamente nas bases de dados brutos.

- verificou-se ainda que, apesar de constar nos gráficos do painel, as tabelas não traziam os dados agregados por categoria (natureza institucional) dos 186 órgãos monitorados, de forma que se fez necessário buscar outra base de dados, não-estruturada, para complementar a informação pretendida.
- IV. acesso, definição e estruturação dos dados dos resultados do relatório final da pesquisa sobre Integridade Pública (2ª fonte de dados) na plataforma GitHub;
- a Figura 5 destaca a capa do relatório, com 25.688 respostas recebidas;
 - pelo formato em PDF, foi necessário transcrever os dados do relatório no repositório;
 - foi necessário comunicar ao Comitê Gestor do Programa de Integridade do Ministério da Economia (Prevenir) o acesso para uso acadêmico dos dados do resultado da pesquisa, pois não são DAG.

Figura 5 – Capa do Relatório Final da Pesquisa sobre Percepção da Integridade Pública



Fonte: BRASIL (2021, p. 1).

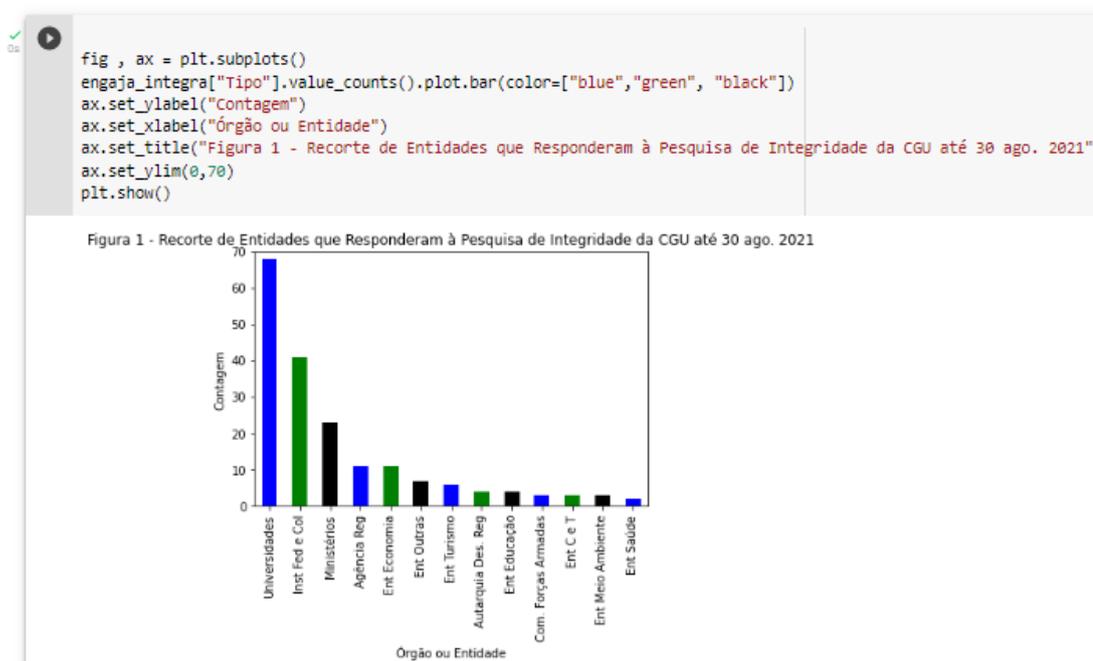
- V. Inserção dos dados sobre alocação de servidores por órgão (3ª fonte de dados), a fim de verificar a interoperabilidade com outra base de dados já existente, bem como testar uma nova métrica: o percentual de respondentes da pesquisa por cada um dos 186 órgãos e entidades listados;

- as chaves primárias (identificadores únicos de registros) das três bases de dados coincidiram na forma de siglas dos órgãos e entidades, o que permitiu boa interoperabilidade entre as fontes de dados;
- foram identificadas diversas inconsistências na apresentação do percentual de respondentes da pesquisa em relação ao corpo funcional, devido a distorções ou ausência de dados.

VI. Aplicação de métodos de análise descritiva e qualitativa dos dados;

Um aspecto importante sobre o tratamento conjunto de várias fontes de dados foi a possibilidade de recuperação do aspecto da agregação, de forma a mensurar representação dos órgãos de acordo com sua categoria (natureza institucional). A Figura 6 exemplifica um elemento de análise descritiva (por natureza de órgão ou entidade) das 25.688 respostas recebidas de 186 órgãos e entidades monitorados divididos em 13 categorias⁹:

Figura 6 – Exemplo de Análise Descritiva: Distribuição de Respostas à Pesquisa de Percepção sobre Integridade Pública, por Categoria



Fonte: elaborado pelo autor (2022).

⁹ Universidades; Institutos Federais e Colégios; Ministérios; Agências Reguladoras; Entidades – Ministério da Economia; Outras Entidades; Entidades – Turismo; Autarquias de Desenvolvimento Regional; Entidades – Educação; Comandos das Forças Armadas; Entidades – Ciência e Tecnologia; Entidades – Meio Ambiente; e Entidades – Saúde.

A Figura 7, por sua vez, traz exemplo de como uma análise quantitativa relativamente simples e rápida pode requerer diversas fontes de dados simultaneamente.

No caso, o interesse simulado de monitoramento social foi saber, por exemplo, dentre os órgãos e entidades do Sipef (escopo) que responderam à questão 7 do Painel de Integridade Pública¹⁰ (objeto), quantas, quais e que categorias de entidades (recorte de categoria) retornaram respostas à pesquisa de percepção da integridade, numa escala relativa restrita à faixa de 5% a 15% de seu quadro de servidores (intervalo de interesse), dado disponível no Sipef.

Destaca-se que, por enquanto, não foi identificado acesso aberto aos microdados das respostas à pesquisa de percepção da integridade, o que limita sensivelmente o estudo de relação entre os dados do painel e os da pesquisa. Ao final desse artigo, a abertura é uma recomendação de aprimoramento, inclusive.

O uso das várias bases de dados se fez necessário tanto para estruturar o mais adequada e coerentemente o escopo exemplificativo, como também evidenciar como a produção de informação pode amadurecer o quão mais dados e indicadores estiverem disponíveis em rotinas futuras, a exemplo do referido horizonte de perspectiva de abertura de microdados.

Figura 7 – Análise quantitativa: entidades sob monitoramento no Painel de Integridade Pública que responderam à Questão 7 do Painel, e à pesquisa de percepção de integridade, numa escala relativa de 5% a 15% de seu quadro funcional

Tipo	Nome	Rel
Ent C e T	CNEN	7.9
Ent C e T	CNPQ	13.92
Ent Economia	CVM	6.24
Ent Educação	CAPE S	10.63
Ent Meio Ambiente	IBAMA	7.41
Ent Meio Ambiente	ICMBIO	13.63
Ent Meio Ambiente	JBRJ	11.76
Ent Saúde	FIOCRUZ	8.73
Ent Saúde	FUNASA	12.08
Ent Turismo	IBRAM	8.56

Fonte: elaborado pelo autor (2022)

- VII. Integração da plataforma do Google Collab ao **Big Query** (armazém de dados), de forma a exportar os dados tratados ao **Data Studio**. Destaca-se que todas as ferramentas utilizadas no exercício de monitoramento social são gratuitas.

¹⁰ Sobre rotina de realização (sim ou não) de levantamento de riscos à integridade.

2.3. RESULTADOS

A partir do exercício de monitoramento social realizado, foi possível verificar a interoperabilidade entre ambientes, bases de dados e ferramentas, com apoio da Ciência de Dados, envolvendo ¹¹: o Painel de Integridade Pública (monitoramento dos programas de integridade) com apoio de outras fontes de dados; GitHub; Google Drive; Google Collab (ambiente de programação em Python); Google Big Query (DW); e Google Data Studio (*DataViz* - visualização de dados).

Foram identificadas relativamente poucas e reversíveis inconsistências, de forma geral, nas dimensões das bases de dados da fonte de dados principal (Painel de Integridade Pública), destacando-se que a chave primária (sigla dos órgãos e entidades) foi a mesma encontrada nas bases de dados complementares necessárias (alocação de servidores e Pesquisa de Integridade).

Foi identificada também a janela de aprimoramentos relativos a esse movimento inicial de disponibilidade de dados abertos governamentais sobre Integridade Pública, sendo possíveis aprimoramentos tais como:

- maior convergência da categorização dos órgãos e entidades nas diversas bases de dados, com possível uso de chaves estrangeiras;
- maior clareza na eventual e esperada atualização das categorias, notadamente as composições de “entidades”, de forma a permitir monitoramento social comparativo ao longo do tempo;
- maior integração com outras bases de dados abertos governamentais disponíveis, de forma a permitir a elaboração de indicadores de forma mais intuitiva, como a relação entre respostas recebidas em eventuais pesquisas qualitativas e o quantitativo de servidores, ou então dados das atividades de ouvidoria e corregedoria de cada órgão ou entidade;
- aumento de disponibilidade de dados estruturados sobre Integridade Pública, notadamente por parte das unidades de gestão da integridade (UGI);
- abertura estruturada de microdados dos resultados das Pesquisas de Percepção da Integridade, de forma a explicitar não só o volume absoluto de respostas, mas também as nuances de como os órgãos e entidades responderam às questões específicas de

¹¹ Ver: <https://www.gov.br/cgu/pt-br/assuntos/etica-e-integridade/dados-do-painel-integridade-publica>; <https://github.com/>; <https://console.cloud.google.com/bigquery>; e <https://datastudio.google.com/>. Acesso em: 14 set. 2022.

percepção¹²; e

- possibilidade de definição de escopo de monitoramento mais qualitativo (em escala contínua) de elementos dos programas de integridade, sem prejuízo da aferição binária de conformidade e controles, de forma a permitir, em conjunto com as outras hipóteses de melhoria, o uso de ferramentas de inferência, predição e *machine learning* com mais consistência.

Por fim, destaca-se que a CGU deu início em 5 set. 2022 à segunda edição da Pesquisa de Percepção sobre Integridade Pública a todos os órgãos e entidades do Sipef, no âmbito da avaliação dos programas de integridade. É uma oportunidade à produção de mais insumos para aplicação sistemática de conceitos e ferramentas da ciência de dados, a partir da reflexão e eventual absorção das sugestões elencadas nesse artigo.

3. CONSIDERAÇÕES FINAIS

A ciência de dados é um campo multidisciplinar do conhecimento com ferramentas cada vez mais ao alcance da sociedade, de forma que é oportuna a precursora iniciativa da CGU em abrir as bases de dados de monitoramento institucional no Painel de Integridade Pública.

Essa iniciativa de criar um instrumento de exercício do monitoramento social converge com a importante diretriz no setor público brasileiro voltada à participação social em políticas públicas. Ao mesmo tempo, esse monitoramento social não exige tanto rigor técnico (ao contrário das rotinas de avaliação e auditoria), e se constitui em um importante canal de aproximação entre a sociedade e as iniciativas do Estado relativas à Integridade Pública, seja para compreender ou criticar as estratégias do setor público.

A virtude cíclica da Integridade Pública é fortalecida, assim, pois ao mesmo tempo em que o corpo institucional monitora com base em dados e evidências a estruturação de elementos da “cultura de integridade”, e promove a prevenção a riscos de corrupção em sentido amplo, também oferece insumos e instrumentos ao monitoramento social.

¹² A CGU divulga um ranking de órgãos e entidades de acordo com o volume absoluto de respostas recebidas.

Nesse sentido, e em proveito da referida e relativamente recente disponibilização de dados abertos governamentais pela CGU sobre o monitoramento institucional dos programas de integridade do Sipef, esse artigo focou a dimensão do monitoramento social, com propostas de medidas de melhoria e ampliação do escopo dos dados.

As principais conclusões são de que a iniciativa precursora da CGU converge com recomendações recentes da OCDE sobre integridade pública, gestão de riscos e ciência de dados; que a abertura de dados pode resultar em incentivo ao monitoramento social e à replicação sistemática pelos integrantes do Sipef (abertura estruturada, organizada e espontânea dos dados); e que há hipóteses consistentes para o aprimoramento da qualidade, disponibilidade e indicadores dos dados, bem como a interoperabilidade entre as bases de dados.

Por fim, e com bases nas conclusões e resultados alcançados, sugere-se como tema para futuros trabalhos a pesquisa e elaboração de indicadores de monitoramento (institucional e social) que permitam aferir com mais acurácia, sensibilidade e tempestividade as nuances entre os diferentes programas de integridade do Sipef. Nesse sentido, reforça-se a sugestão apontada para abertura de microdados das Pesquisas de Percepção realizadas.

Destaca-se também a estrutura de Rodrigues (2022, p. 8), que apresentou a aplicação-piloto de um conjunto de 33 indicadores divididos em 3 dimensões ponderadas em 9 capacidades elencadas dos programas de integridade. É uma tendência natural e esperada a busca por mais maturidade, à medida em que o monitoramento social da Integridade Pública se consolida, e o exercício de disponibilização de novos dados e indicadores estruturados converge no mesmo sentido.

Referências bibliográficas

BRAGA, Marcus V. A. Navegando em um mar de soluções – dilemas da implementação da integridade na ponta. **Estadão**. Gestão, Política e Sociedade. Brasil, 5 jul. 2020. 2020. Disponível em: < <https://politica.estadao.com.br/blogs/gestao-politica-e-sociedade/navegando-em-um-mar-de-solucoes-dilemas-da-implementacao-da-integridade-na-ponta/> >. Acesso em 14 out. 2022.

BRAGA, Marcus V. A. As janelas da corrupção. **Jornal i**. Opinião i. Portugal, 14 set. 2022. 2022. Disponível em: <https://ionline.sapo.pt/artigo/780908/as-janelas-da-corrupcao?seccao=Opiniao_i>. Acesso em: 14 out. 2022.

BRASIL. Casa Civil da Presidência da República. **Avaliação de políticas públicas: guia prático de análise *ex ante***. v. 1. Brasília: Ipea – Instituto de Pesquisa Econômica Aplicada, 2018a. 194 p.

BRASIL. Casa Civil da Presidência da República. **Avaliação de políticas públicas: guia prático de análise *ex post***. v. 2. Brasília: Ipea – Instituto de Pesquisa Econômica Aplicada, 2018b. 318 p.

BRASIL. Pesquisa com Servidores Públicos Federais sobre Integridade Pública – Relatório Final. CGU. 1º set. 2021. 9 p. Brasília, DF, 2021.

BRASIL. **CGU publica Painel de Integridade Pública com novos gráficos e funcionalidades**. 2022a. Disponível em: < <https://www.gov.br/cgu/pt-br/assuntos/noticias/2022/03/cgu-publica-painel-de-integridade-publica-com-novos-graficos-e-funcionalidades> >. Acesso em: 29 mai. 2022.

BRASIL. O que são dados abertos?. Portal Brasileiro de Dados Abertos. 2022b. Disponível em: < <https://dados.gov.br/pagina/dados-abertos> >. Acesso em: 14 out. 2022.

BRASIL. Dados do Painel Integridade Pública. Controladoria-Geral da União. 2022c. Disponível em: < <https://www.gov.br/cgu/pt-br/assuntos/etica-e-integridade/dados-do-painel-integridade-publica> >; base 4 mar. 2022. Acesso em: 14 out. 2022.

BRASIL. Pessoal: Alocação de Servidores. Portal Brasileiro de Dados Abertos. 2022d. Disponível em: < <https://dados.gov.br/dataset/raio-x-da-administracao-publica-federal/resource/904e6ac0-60da-40f5-ad7d-62a7dd568438> > ; acesso em 14 out. 2022

CARDOSO, Paulo H.. **Ciência de dados aplicada a dados governamentais abertos sob a ótica da Ciência da Informação**. 2019. Dissertação (Mestrado). Programa de Pós-Graduação em Ciência da Informação. Universidade Federal do Rio de Janeiro – UFRJ, 2019.

COTTA, Luiza C. V. **Adhemar de Barros (1901-1969): a origem do “rouba, mas faz”**. 2008. Tese de Doutorado. Universidade de São Paulo.

LIMA, Lauro V. A.; RUFINO, Maria. A.; MACHADO, Márcia. R. **Criando dificuldades para vender facilidades: corrupção, burocracia e crescimento corporativo no Mercosul**. Revista de Contabilidade e Organizações, 13, 2019. 64-74. Disponível em: < <https://www.revistas.usp.br/rco/article/view/158324/157722> >. Acesso em: 14 out. 2022.

OCDE. **Recomendação do Conselho da OCDE sobre Integridade Pública**. 2018. Disponível em: < <https://www.oecd.org/gov/ethics/integrity-recommendation-brazilian-portuguese.pdf> >. Acesso em: 14 out. 2022.

OCDE. **Modernizando a avaliação dos riscos para a integridade no Brasil: Rumo a uma abordagem comportamental e orientada por dados**. OECD Publishing. Paris, 2022, 36 p. Disponível em: <https://doi.org/10.1787/61d7fc60-pt> . Acesso em: 14 out. 2022.

ODILLA, Fernanda. Apesar de ter muitas iniciativas de combate à corrupção, Brasil tem dificuldade de envolver a população no processo. **Interesse Nacional**, 19 mai. 2022. Disponível em: < <https://interessenacional.com.br/edicoes-posts/fernanda-odilla-apesar-de-ter-muitas-iniciativas-de-combate-a-corrupcao-brasil-tem-dificuldade-de-envolver-a-populacao-no-processo/> >. Acesso em 14 out. 2022.

PENG, Y.; SILVA, J. P. F.; NAGATA, M. H. **Um guia rápido sobre os conceitos fundamentais em ciência de dados: como começar, como fazer certo e com o que tomar cuidado**. 2021. Texto para Discussão. Laboratório de Aprendizado de Máquina em Finanças e Organizações – LAMFO, 2021. Disponível em: < <http://dx.doi.org/10.13140/RG.2.2.15272.29440> >. Acesso em: 14 out. 2022.

PEREIRA, Alex. Introdução a Ciência de Dados. 2021. Apresentação em Power Point. 52 p. Disponível em: <

<https://drive.google.com/file/d/16BG54hDZTeCHXvD4e9YYV74vvyoWsEPI/view?usp=sharing> >. Acesso em: 14 out. 2022.

RODRIGUES, Alisson R M. Avaliação de Programas de Integridade – Visão Sistêmica sobre o Sipef. In: Lançamento do Relatório OCDE – Modernizando a avaliação dos riscos para a integridade no Brasil, 2022, Brasília. CGU, 2022, 22 p. Disponível em: < <https://drive.google.com/file/d/1aTGgm90ndt3fSRq19tMHP9Mp6rsI0ujS/view?usp=sharing> >. Acesso em: 14 out. 2022.

TULLOCK, Gordon; SELDON, Arthur; BRADY, Gordon L. **Falhas de governo: uma introdução à teoria da escolha pública**. Rio de Janeiro: Instituto Liberal, 2005, 187 p.

VAN NEDERPEL, Peter; e BLACK, Andrew. **Dimensions of Data Quality (DDQ)**. Research Paper. DAMA NL Foundation, 2020, 113 p. Disponível em: < <https://www.dama-nl.org/wp-content/uploads/2020/09/DDQ-Dimensions-of-Data-Quality-Research-Paper-version-1.2-d.d.-3-Sept-2020.pdf> >. Acesso em: 14 out. 2022.