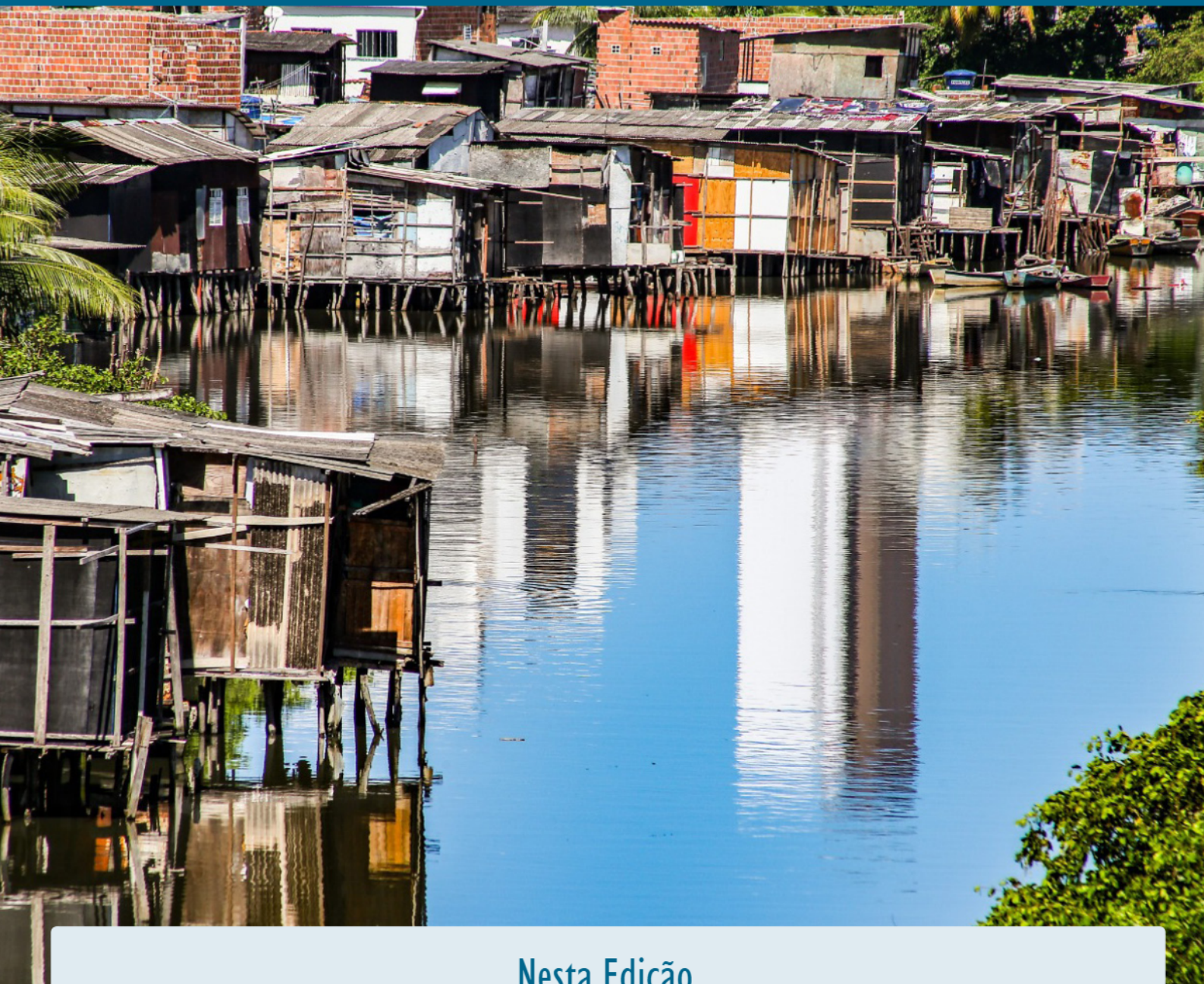


e-ISSN 2595-668X

Revista da CGU

Brasília • Volume 14 • Nº 26 • Jul-Dez 2022



Nesta Edição

Dossiê Especial “Ciência de Dados na Administração Pública: Desafios e Oportunidades”

Revista da CGU

Brasília • Volume 14 • Nº 26 • Jul-Dez 2022

WAGNER DE CAMPOS ROSÁRIO

Ministro da Controladoria-Geral da União

JOSÉ MARCELO CASTRO DE CARVALHO

Secretário-Executivo da Controladoria-Geral da União

ANTÔNIO CARLOS BEZERRA LEONEL

Secretário Federal de Controle Interno

GILBERTO WALLER JÚNIOR

Corregedor-Geral da União

MARCOS GERHARDT LINDENMAYER

Ouvidor-Geral da União (substituto)

ROBERTO CÉSAR DE OLIVEIRA VIEGAS

Secretário de Transparência e Prevenção da Corrupção

JOÃO CARLOS FIGUEIREDO CARDOSO

Secretário de Combate à Corrupção

Revista da CGU / Controladoria-Geral da União

<http://revista.cgu.gov.br/>

e-ISSN 2595-668X

Administração Pública

As opiniões expressas nos artigos aqui publicados são de inteira responsabilidade de seus autores e não expressam, necessariamente, as da Revista da CGU.

Diagramação: Assessoria de Comunicação Social • Ascom/CGU

Foto da Capa • Wesley Marcelo D'Almeida dos Santos

CONTROLADORIA-GERAL DA UNIÃO

SIG (Setor de Indústrias Gráficas), Quadra 02, Lotes 530 a 560

Edifício Sohestre - Brasília - DF / CEP: 70610-420 • Brasília/DF

Telefone: (61)2020-6826

E-mail: revista@cgu.gov.br

www.gov.br/cgu

REVISTA DA CGU

Periódico científico voltado à difusão e promoção do conhecimento e da pesquisa nos seguintes temas: *accountability* e políticas anticorrupção; controle interno e auditoria pública; governança e riscos; correição e responsabilização administrativa; transparência e controle social; integridade e ética pública; ouvidoria e participação social; prevenção e combate à corrupção; e metodologia de pesquisa aplicada aos referidos temas.

EQUIPE EDITORIAL

Comissão Editorial

Editor-Chefe M. Sc. Daniel Matos Caldeira • Universidade de Lisboa, Portugal

Editora Dra. Danusa da Matta Duarte Fattori • Controladoria-Geral da União (CGU), Brasil

Editora M.Sc. Flávia Lemos Sampaio Xavier • Controladoria-Geral da União (CGU), Brasil

Editor M. Sc. Flavio Rezende Dematté • Instituto Brasiliense de Direito Público (IDP), Brasil

Editor M. Sc. Bruno Dantas Faria Affonso • Universidade Federal Fluminense (UFF), Brasil

Editores Especiais • Dossiê Especial “Ciência de Dados na Administração Pública: Desafios e Oportunidades”

Dra. Solange Oliveira Rezende • Universidade de São Paulo (USP)

Dr. Ricardo Marcacini • Universidade de São Paulo (USP)

Dr. Alex Lopes Pereira • Escola Nacional de Administração Pública (Enap)

Editores Especiais

M.Sc. George Moura Colares • Controladoria-Geral da União (CGU)

Dr. Marcus Vinicius de Azevedo Braga • Universidade Federal do Rio de Janeiro (UFRJ)

M. Sc. Djalma Peçanha Gomes • Universidade Federal do Mato Grosso do Sul (UFMS), Brasil

Conselho Editorial

Dr. Andityas Soares de Moura Costa Matos • Universidade Federal de Minas Gerais (UFMG)

Dr. Carlos Frederico Pereira da Silva Gama • Universidade Federal de Tocantins (UFT)

Dra. Cecília Olivieri • Universidade de São Paulo (USP)

Dr. César Augusto Tibúrcio Silva • Universidade de Brasília (UnB)

Dr. Dawisson Elvécio Belém Lopes • Universidade Federal Minas Gerais (UFMG)

Dr. Frederico Lustosa da Costa • Universidade Federal Fluminense (UFF)

Dr. Ivan Ricardo Gartner • Universidade de Brasília (UnB)

Dra. Janaína Lima Penalva da Silva • Universidade de Brasília (UnB)
Dr. José Célio Silveira Andrade • Universidade Federal Bahia (UFBA)
Dr. José Matias-Pereira • Universidade de Brasília (UnB)
Dr. Leonardo Secchi • Universidade do Estado de Santa Catarina (UDESC)
Dr. Luciano da Ros • Universidade Federal de Santa Catarina (UFSC)
Dr. Marco Cepik • Universidade Federal do Rio Grande do Sul (UFRGS)
Dr. Matthew M. Taylor • American University
Dra. Paula Veiga • Universidade de Coimbra
Dr. Robert Gregory Michener • Escola Brasileira de Administração Pública e de Empresas da Fundação Getúlio Vargas (FGV-EBAPE)
Dr. Rubens Pinto Lyra • Universidade Federal da Paraíba (UFPB)
Dra. Sandra Isabel Rasteiro Firmino • Universidade de Lisboa
Dr. Saulo de Oliveira Pinto Coelho • Universidade Federal de Goiás (UFG)

Avaliadores

Esta edição contou com o trabalho dos seguintes pareceristas:

Bruno Cunha	Gustavo de Queiroz Chaves
Bruno Fagali	Herbert Kimura
Cedric Pin	Joacy Victor Araújo
Claudio Henrique Fontenelle Santos	Lidiane Marques
Claudiomar Matias Rolim Filho	Lincoln Augusto Santana Telhado
Éder Gualberto	Marcio Denys Pessanha Goncalves
Edgard Augusto de Oliveira	Mario Larangeira Jr.
Fabio Lopes Soares	Moisés Dutra
Frederico José Lustosa da Costa	Nilo Cruz Neto
Gelson Silva Junquilha	Pedro Masson Sesconetto Souza
Giovanni Pacelli Carvalho Lustosa da Costa	Rafael Dalla Rosa
Gladston Luiz da Silva	Robson Zuccolotto

ÍNDICE

EDITORIAL • Ciência de Dados na Administração Pública: Desafios e Oportunidades • 164

DOSSIÊ ESPECIAL

INTERVIEW • TAKA ARIGA - Chief Data Scientist Director, Innovation Lab, US Government Accountability Office • 169

Classificação de processos judiciais segundo Objetivos de Desenvolvimento Sustentável da Agenda ONU 2030 • 172

Classification of lawsuits according to the Sustainable Development Goals of the UN 2030 Agenda

Clasificación de demandas según los Objetivos de Desarrollo Sostenible de la Agenda 2030 de la ONU

Luiz Fux, Pedro Felipe de Oliveira Santos, Aline Carlos Dourado Braga, Pamella Sada Dias Edokawa e Júlio Luz Sisson de Castro

Entorno legal y adopción de *blockchain* como herramienta para prevenir la corrupción en contrataciones públicas: reflexiones sobre iniciativas europeas y los marcos normativos brasileños • 190

Entorno legal e adoção de blockchain como ferramenta para prevenir a corrupção em contratações públicas: reflexões sobre iniciativas europeias e os marcos normativos brasileiros.

Legal surroundings and adoption of blockchain as a tool to prevent corruption in public contracts: reflections on European initiatives and the Brazilian regulatory framework.

Samir Sallen Silva Santos e Lucio Alves Angelo Junior

Como somos vistos? Análise da imagem organizacional pública utilizando ciência de dados • 203

How are we seen? public organizational image analysis using data science

¿cómo somos vistos? análisis de imagen organizacional pública utilizando ciencia de los datos

Carolina Coelho da Silveira, Carla Bonato Marcolin e Carlos Henrique Rodrigues

Estimando a Arrecadação da Dívida Ativa da União com Machine Learning: Uma análise baseada nos dados de arrecadação do período de 2015 a 2021 • 222

Estimating the Union's Active Debt Revenues with Machine Learning: An analysis based on tax revenues data from 2015 to 2021.

Estimación de los ingresos de la deuda activa de la Unión con Machine Learning: un análisis basado en datos de ingresos fiscales de 2015 a 2021.

Rubens Quaresma Santos

Ciência de Dados como Ferramenta de Apoio à Tomada de Decisão: Classificação Hierárquica Descendente de Pedidos de Acesso à Informação na Prefeitura de São Paulo • 238

La Ciencia de Datos como Herramienta de Apoyo a la Toma de Decisiones: Clasificación Jerárquica Descendente de Solicitudes de Acceso a la Información en la Alcaldía de São Paulo

Data Science as a Tool to Support Decision-Making: Descending Hierarchical Classification of Access to Information Requests in the Municipality of São Paulo

Claudio Henrique Fontenelle Santos e Ana Lúcia da Silva Romão

Relato Técnico • A regulação baseada em dados: a era digital na agência reguladora de São Paulo • 261

Data-based regulation: the digital era in the public services regulatory agency in the state of São Paulo

Regulación basada en datos: la era digital en la agencia reguladora de los servicios públicos del estado de São Paulo

Itamar Aparecido de Oliveira

Relato Técnico • Proposta de framework de internacionalização no contexto da evolução da plataforma Lattes • 270

José Francisco Salm Junior, Paulo Henrique de Assis Sant'ana e Geraldo Sorte

Relato Técnico • Ciência de Dados no Controle Social em Educação • 285

Data Science in Social Control in Education

Ciencia de Datos en Control Social en Educación

Glleddson Fryttys Menezes Leite e Hugo Leonardo Tosto Cuoco

Relato Técnico • Alice: Desafios, resultados e perspectivas da ferramenta de auditoria contínua de compras públicas governamentais com uso de inteligência artificial • 296

Alice: Challenges, results and perspectives of the continuous auditing tool for government procurement using artificial intelligence

Alice: Desafíos, resultados y perspectivas de la herramienta de auditoría continua para compras gubernamentales utilizando inteligencia artificial

Andre Luiz Monteiro da Rocha, Matheus Scatolino de Rezende e Tiago Chaves Oliveira

ARTIGOS

Contribuições das denúncias e representações autuadas em 2020 para o controle externo realizado pelo Tribunal de Contas do estado de Santa Catarina • 310

Contributions of complaints and representations filed in 2020 to the external control performed by the Court of Auditors of the state of Santa Catarina

Contribuciones de quejas y denuncias presentadas en 2020 al control externo realizado por el Tribunal de Cuentas del estado de Santa Catarina

Fernanda Camila De Carli e Fabiano Maury Raupp

Dados abertos: uma análise da Empresa Brasileira de Serviços Hospitalares e de sua rede de hospitais universitários federais • 324

Open data: an analysis of the Empresa Brasileira de Serviços Hospitalares and its network of federal university hospitals

Datos abiertos: un análisis de la Empresa Brasileira de Servicios Hospitalares y su red de hospitales universitarios federales

Gabriela de Oliveira Silva, Renata Mendes de Araujo e Fernando Filgueiras

A trajetória mexicana na consolidação do direito de acesso à informação pública: histórico e desafios • 342

The Mexican path in consolidating the right of access to public information: historic and challenges

La trayectoria mexicana en la consolidación del derecho de acceso a la información pública: historia y retos

Luma Poletti Dutra

EDITORIAL • Ciência de Dados na Administração Pública: Desafios e Oportunidades

<https://doi.org/10.36428/revistadacgu.v14i26.617>

A Ciência de Dados é uma área de pesquisa interdisciplinar para o estudo e análise de dados que está recebendo cada vez mais atenção por popularizar o uso de métodos científicos para extração de conhecimento para tomada de decisão. Matemática, Estatística, Ciência da Computação, Engenharia de Software, Aprendizado de Máquina, Análise de Dados, Mineração de Dados e Textos, Gestão de Projetos e *Big Data* são exemplos de áreas exploradas por um cientista de dados.

A administração pública é um campo que não só detém inúmeras bases de dados, mas que precisa cada vez mais desenvolver competências para transformá-los em evidências que fundamentem a melhor tomada de decisão. A título de exemplificação, só em relação aos dados abertos acessíveis por meio do Portal [Dados.gov.br](https://dados.gov.br), são disponibilizados 13.726 conjuntos de dados que abrangem 218 organizações públicas classificados em 20 grandes temas, que vão da defesa e segurança à indústria, passando pelo meio-ambiente, educação e saúde.

Para fins de contextualização dos desafios e oportunidades da Ciência de Dados na administração pública, a Revista da CGU emoldura a discussão a partir de dois grandes panoramas: a transição digital e o desenvolvimento social, propondo uma reflexão a partir do contraste do tema em pauta com a fotografia que ilustra a capa desta edição.

A foto vem do Recife e é de autoria de Wesley Marcelo D'Almeida dos Santos, sendo intitulada “Reflexo da desigualdade”. Esta foto obteve a segunda colocação no III Concurso Nacional de Fotografia da Ouvidoria-Geral da União.

A cidade do Recife notabiliza-se por ser a sede do Porto Digital, um dos principais parques tecnológicos e ambiente de inovação do Brasil, resultado de uma ação coordenada entre governo, academia e empresa para requalificação urbana de uma região antes degradada do seu centro histórico da cidade transformando-a em um polo estratégico em soluções tecnológicas aliando ainda a inclusão digital da comunidade em seu entorno.

Nesse particular, o conhecimento dos dados, principal insumo dessa transição digital que avança em passos largos para uma nova economia, intermedeia uma nova relação entre os cidadãos e o governo no provimento de políticas públicas.

O governo brasileiro empreendeu significativos esforços se preparando para essa realidade e recentemente foi reconhecido como segundo líder em governo digital no mundo pelo relatório *GovTech Maturity Index 2022* do Banco Mundial.

Posto isto, esta edição convida seus leitores à reflexão sobre em que medida esse esforço da agenda pública em direção ao governo digital pode contribuir e está efetivamente contribuindo para uma melhor prestação de serviços públicos e superação das desigualdades sociais.

Para trazer uma visão de um órgão de controle com destacada experiência avaliativa, o presente Dossiê Especial “Ciência de Dados na Administração Pública: Desafios e Oportunidades” é aberto com uma entrevista de Taka Ariga, cientista-chefe de Ciência de Dados e Diretor do Laboratório de Inovação da Agência Governamental de *Accountability* do Governo dos Estados Unidos – GAO. Taka Ariga chamou atenção para questões que precisam

ser observadas sobre decisões de políticas públicas baseadas em inteligência artificial, a importância do relacionamento com a academia e aplicações práticas da inteligência em políticas públicas, incluindo a perspectiva da previsão e mitigação de corrupção.

Em sequência à entrevista, o presente dossiê apresenta nove trabalhos inéditos, meticulosamente selecionados pela mais concorrida chamada de trabalhos da história da Revista da CGU. São cinco artigos científicos e quatro relatos técnicos, os quais são apresentados a seguir.

O artigo "*Classificação de processos judiciais segundo Objetivos de Desenvolvimento Sustentável da Agenda ONU 2030*", de Luiz Fux, Pedro Felipe Santos, Aline Braga, Pamella Edokawa e Júlio Luz de Castro, apresenta a ferramenta RAFA 2030, que viabiliza a integração entre a Agenda 2030 da ONU e o STF via aprendizagem de máquina e análise de dados fazendo a classificação de ações por ODS da Agenda 2030.

O artigo seguinte "*Entorno legal y adopción de blockchain como herramienta para prevenir la corrupción en contrataciones públicas: reflexiones sobre iniciativas europeas y los marcos normativos brasileños*", de Samir Santos e Lucio Alves Angelo Junior, apresenta os principais desafios e oportunidades para a adoção, pelo governo brasileiro, de blockchain como ferramenta para prevenção a corrupção em contratações públicas.

Já o artigo "*Como somos vistos? Análise da imagem organizacional pública utilizando ciência de dados*", de Carolina Silveira, Carla Marcolin e Carlos Henrique Rodrigues, apresenta por meio de uma técnica inovadora uma análise da imagem organizacional da ANEEL veiculada na imprensa. O artigo apresenta resultados interessantes, inclusive uma disparidade recorrente entre o conteúdo da manchete e o conteúdo da reportagem em si.

O autor Rubens Quaresma Santos propõe no artigo "*Estimando a Arrecadação da Dívida Ativa da União com Machine Learning: uma análise baseada nos dados de arrecadação do período de 2015 a 2021*" a aplicação de algoritmos de aprendizado de máquina para projetar a arrecadação futura em substituição da técnica da suavização exponencial hoje empregada pela Procuradoria-Geral da Fazenda Nacional. Os modelos propostos mostraram-se ancorados em parâmetros mais amplos e robustos que o cálculos atuais, viabilizando uma decisão mais bem amparada.

Em relação ao artigo intitulado "*Acesso à informação no Brasil e ciência de dados: classificação hierárquica descendente em pedidos realizados à Prefeitura de São Paulo de 2012 a 2019*", os autores Cláudio Henrique Fontenelle Santos e Ana Lúcia Romão apresentam técnicas de análise inteligente de dados para organizar e compreender os pedidos de acesso à informação pelos cidadãos, permitindo um melhor entendimento das demandas da sociedade nesse assunto.

O Relato Técnico "*A Regulação baseada em dados: a era digital na agência reguladora de São Paulo*" de Itamar Aparecido de Oliveira apresenta a experiência da Diretoria Técnica de Saneamento Básico da Agência de Serviços Públicos do Estado de São Paulo (Arsp) em seu trajeto de modernização das atividades regulatórias por meio de estratégias de análise de dados. Os resultados indicaram que a adoção de metodologias regulatórias baseadas em análise de dados contribuíram para melhores análises regulatórias e decisões mais refinadas, fortalecendo a segurança no ambiente regulado.

Por meio do Relato Técnico "*Proposta de framework de internacionalização no contexto da evolução da Plataforma Lattes*", os autores José Francisco Salm Junior, Paulo Henrique de Assis Santana e Geraldo Sorte apresentam a proposta de intervenção denominada de *framework* CV-PTBR que objetiva estabelecer um protocolo de trabalho conjunto baseado em diretrizes entre as equipes do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e a Fundação para a Ciência e a Tecnologia – FCT de Portugal para assimilação das tecnologias e seus elementos para o uso nos projetos de aprimoramento das rotinas de importação de dados na Plataforma Lattes e a evolução das integrações de dados na Plataforma CIÊNCVIAVITAE de Portugal.

Em sequência, o Relato Técnico "*Ciência de Dados no Controle Social em Educação*", de Gilleddson Frytts Menezes Leite e Hugo Leonardo Tosto Cuoco, discute o uso de ferramenta de *business intelligence* para análise de dados do Sistema de Informações sobre Orçamentos Públicos em Educação (SIOPE) e do Censo Escolar, visando

a análise inteligente das despesas em educação e facilitando a visualização de dados e comparação entre diferentes entes da federação.

O dossiê ciência de dados é encerrado com o Relato Técnico "Alice: Desafios, resultados e perspectivas da ferramenta de auditoria contínua de compras públicas governamentais com uso de inteligência artificial", de Andre Luiz Monteiro da Rocha, Matheus Scatolino de Rezende e Tiago Chaves Oliveira, o qual apresenta o histórico da construção e os principais resultados gerados pelo Analisador de Licitações, Contratos e Editais (Alice), o sistema de identificação de indícios de irregularidades em compras públicas.

Finalizando esta edição, apresentamos três artigos científicos recepcionados pelo fluxo contínuo de submissões à Revista da CGU.

O artigo “A trajetória mexicana na consolidação do direito de acesso à informação pública: histórico e desafios”, de Luma Poletti Dutra, discute o pioneirismo do México na regulação do direito do acesso à informação na América Latina por meio de uma pesquisa qualitativa que envolveu entrevistas realizadas com pesquisadores e ativistas mexicanos.

Os autores Gabriela de Oliveira Silva, Renata Mendes de Araujo e Fernando Filgueiras no artigo “Dados abertos: uma análise da Empresa Brasileira de Serviços Hospitalares e de sua rede de hospitais universitários federais” analisaram o grau de alinhamento da Empresa Brasileira de Serviços Hospitalares (Ebserh) e de sua rede de hospitais universitários federais aos princípios da Política Nacional de Dados Abertos (PNDA).

No artigo “Contribuições das denúncias e representações para o controle externo realizado pelo Tribunal de Contas do Estado de Santa Catarina”, os autores Fernanda Camila De Carli e Fabiano Maury Raupp discutiram a participação social a partir do nível de interação das espécies de controle incidentes pelo Tribunal de Contas de Santa Catarina sobre a gestão pública a partir da lente teórica do Novo Serviço Público e da *accountability* diagonal.

Por fim, agradecemos a atividade voluntária de revisão pelos pareceristas que contribuíram para esta edição por meio de avaliações construtivas para o aperfeiçoamento dos trabalhos. A todos eles, cujos nomes encontram-se divulgados ao final desta edição, nosso agradecimento.

Desejamos a todos uma boa leitura!



Daniel Matos Caldeira

Universidade de Lisboa, Portugal

matoscaldeira@gmail.com

<https://orcid.org/0000-0001-9823-8356>

Doutorando em Administração Pública pela Universidade de Lisboa. Auditor Federal de Finanças e Controle da Controladoria-Geral da União.



Dr. Alex Lopes Pereira (Enap)

<https://orcid.org/0000-0002-6405-429X>

Doutor em Engenharia Eletrônica e de Computação pelo Instituto Tecnológico de Aeronáutica – ITA. É integrante da carreira de Analista de Planejamento e Orçamento. Atualmente, trabalha no Ministério da Economia, professor de Ciência de Dados e empreendedor.



Dr. Ricardo Marcacini (ICMC/USP)
Universidade de São Paulo (USP), Brasil
<https://orcid.org/0000-0002-2309-3487>

Ricardo Marcacini tem doutorado em Ciência da Computação e Matemática Computacional pelo ICMC/USP, e atualmente é docente do Departamento de Ciências de Computação (SCC) no ICMC/USP. É orientador do programa de mestrado e doutorado de Ciência de Computação e Matemática Computacional do ICMC/USP, na qual também ministra disciplinas do tema de Aprendizado de Máquina e Inteligência Artificial. Atua no Laboratório de Inteligência Computacional (LABIC), com pesquisas nas áreas de Inteligência Artificial, Mineração de Textos, Mineração de Opiniões e Sensoriamento da Web.



Dra. Solange Oliveira Rezende
Universidade de São Paulo (USP), Brasil
<https://orcid.org/0000-0002-5233-7639>

Solange Rezende é professora no Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo (USP), desde 1991. Tem experiência na área de Ciência da Computação, com ênfase em Inteligência Artificial, atuando principalmente nos temas relacionados com Mineração de Dados e Textos, Análise de Sentimentos e Sistemas de Recomendação. O foco das pesquisas realizadas é como extrair Conhecimento de Dados e Informação. Possui graduação em Licenciatura em Ciências Habilitação em Matemática pela Universidade Federal de Uberlândia (UFU) (1986), mestrado em Ciências de Computação e Matemática Computacional pela Universidade de São Paulo (1990) e doutorado em Engenharia Mecânica - São Carlos pela Universidade de São Paulo (1993). Realizou pós-doutorado na Universidade de Minnesota, USA (1995-1996). É orientadora no Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional; no Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria e Vice Coordenadora do MBA em Inteligência Artificial e Big Data do ICMC/USP. Atua como pesquisadora no Laboratório de Inteligência Computacional (labic.icmc.usp.br) do Departamento de Ciências de Computação e no C4AI. Informações específicas sobre publicações e orientações podem ser acessadas em: <http://lattes.cnpq.br/8526960535874806>.

Dossiê Especial

INTERVIEW • TAKA ARIGA - Chief Data Scientist Director, Innovation Lab, US Government Accountability Office

<https://doi.org/10.36428/revistadacgu.v14i26.613>

1. Public policy decisions based on Artificial Intelligence have been gaining ground. This scenario offers several risks, especially those about the decision quality. Considering the context for pressuring public authorities' commitment to reinforce policy-based decisions in evidence, what is the GAO doing about this scenario?

[Taka] While the use of artificial intelligence is rapidly transforming wide-ranging public sector missions, implementations of AI solutions by governmental entities necessarily need to be held to a higher standard of accountability. This is because AI-based decisions can have outsized and opaque impacts on the society. Certainly there are complex technical and data science challenges that need to be appropriately addressed in order to deploy effective AI systems. They include challenges with data quality, machine learning performance tuning, operational integration, and cybersecurity resiliency. However, governmental entities often have to grapple with other non-technical concerns such as regulatory compliance with privacy and civil rights statutes, availability of a workforce with adequate digital skills to interpret outputs that are based on probabilities, and inflexibility of acquisition processes, just to name a few. Bottom line, implementing AI solutions within the public sector context is complicated. Over the past several years, there have been a concerted push towards implementing responsible AI—meaning that outputs should be explainable, policy decisions made based on AI results should be transparent, and use of AI should not promulgate disparate impacts. These are certainly worthwhile ideals towards accountability. That being said, these efforts towards responsible AI often are expressions of qualitative aspirations that lack sufficient details to implement. This is why the Government Accountability Office published first-of-its-kind AI Accountability Framework to help

inform governmental agencies on specific practices that can verifiably demonstrate accountability. These practices cover the spectrum of AI lifecycle and span across four pillars—governance, data, performance, and monitoring—at both organizational and system component levels. GAO auditors and others within the oversight community are using this framework as a blueprint to conduct evidence-based evaluations of AI implementation across governmental entities. In short, independent verifications is key towards making sure that aspirations of responsible AI manifest into reality. Because AI is becoming ubiquitous, GAO expects governmental agencies will continue to confront complex implementation challenges. Our goal is to make sure that oversight capacity continually evolve alongside at the speed of advances in machine learning, data engineering, and other related disciplines. Link to the GAO AI Accountability Framework is at: <https://www.gao.gov/products/gao-21-519sp>.

2. Has the GAO made evaluations about specific AI algorithms applied by public agencies or third parties in policy process? What were the challenges and the results?

[Taka] While GAO has the technical capacity to evaluate individual machine learning models and related artifacts, our oversight work typically do not opine at the algorithmic level of details. GAO's goal is not to zoom into and unpack complexities of individual ingredients but rather assessing an AI program holistically for effectiveness, efficiency, economics, equity, and ethics. To that end, GAO's growing body of oversight, insight, and foresight work related to AI can be found on our website at GAO.gov. In addition, GAO has done work highlighting systemic challenges such as the need for more digital

talents within public sector workforce, implementation of zero trust cybersecurity controls, promoting an innovation economy, and acquisition difficulties. Since the publication of GAO's AI Accountability Framework in summer of 2021, GAO auditors and others within the public sector oversight community have relied this blueprint to conduct evidence-based evaluation of AI programs across governmental entities. The Framework itself is purposefully grounded in the language of the Yellow Book (a.k.a., Generally Accepted Government Auditing Standards) and the Green Book (a.k.a., Standards for Internal Control in the Federal Government) which affords a baseline familiarity of key accountability concepts for the auditors.

3. How does the GAO relate to the academia? Is there an internal policy for the development of the skills and knowledge of its staff about data science and artificial intelligence? Is there any department that monitors and incorporates scientific knowledge into GAO procedures?

[Taka] Academic institutions are invaluable sources of experimentations that help to continually push boundaries of AI capabilities. The challenge is figuring out how to bridge capabilities based on research and theories with applied implementations that need to happen in the real (imperfect) world. GAO's Innovation Lab staff routinely review research papers on the latest advances in data science as we build our own AI solutions. Auto-summarization of texts and data-sparse machine learning models are two exciting recent examples that come to mind. In addition, GAO makes a significant level of training resources available to our staff, including online skill development, subscriptions to technical journals, attending conferences, and participating in knowledge exchange discussions with external stakeholders. As we experiment with new machine learning techniques, we carefully evaluate performance potential alongside weighing our ability to train/tune them for GAO use cases. We also frequently dialogue with academic institutions, think tanks, advocacy groups, and international stakeholders on evolving policies and accountability issues related to AI and other emerging technologies. These collaborations help us mitigate blind spots, anticipate changing landscapes, and strengthen our oversight capacity.

4. What is the role of artificial intelligence in GAO's routine? If possible, go through the following aspects in your answer:

- the use of artificial intelligence in the evaluation of public policies.
- AI solution typologies for your business: internal development or through acquisitions? Are you following some framework for this?
- use of artificial intelligence in inspections.
- human role in guiding artificial intelligence on GAO

[Taka] As a congressional watchdog, it is important for GAO be a hands-on practitioner of AI so that we can rely on our own experience to independently conduct evidence-based evaluations. After all, that is our core belief of trust but verify. In practice, this means that our growing teams of data scientists focus on custom design, development, and deployment of our own AI solutions for use in audit contexts. This approach allows GAO to explore and combine different machine learning and modeling techniques while enabling a complete control over parameter tuning, workflow designs, operational integration, and compliance towards federal information assurance requirements. This level of flexibility is typically not afforded by commercial AI products, which in turn hinders our ability to meet specific audit quality standards such as conducting detailed code reviews and drafting technical records of analyses. Moreover, commercially available AI products are typically trained using commercially available or open-source data, which often do not perform effectively under characteristic of governmental data. It is important to note that GAO's AI ethos is that humans must always be in the loop and never out of the loop. In other words, we develop AI capabilities only as augmentations of audit procedures and never to supplant professional judgements. GAO is concurrently moving forward with agency-wide data governance and data literacy efforts to make sure our multi-disciplinary workforce can effectively govern, operate, scale, and interpret AI capabilities.

5. How do you see the advances of artificial intelligence for corruption prediction and mitigation?

[Taka] While the use of AI to combat corruption and fraud risks is an active area of research within the data science community, significant technical and design challenges remain for appropriate use in oversight con-

texts. Generalized analyses are extraordinarily difficult because details related to corruption and fraudulent activities are typically sparse, and often associated with particularized circumstances. It is also unfortunate that this is a market space with much hyped promises. I do see tremendous potential in the use deep learning techniques and graph-based databases to surface anomalies that may indicate corrupt or fraudulent practices. However, while deep learning may produce compelling results, the outputs are nearly impossible to linearly explain, which in turn, poses roadblocks toward substantiating findings within audit contexts. GAO's Innovation Lab is actively exploring prototype use of graph databases and looking at how we might unlock a new class of machine learning capabilities that can be appropriately integrated within audit procedures.

6. Is there any government innovation based on artificial intelligence that you wish to point out for us?

7. How is the Data Science (DS) process, including Analytics and AI tools, at the GAO?

If possible, go through the following aspects in your answer:

- data literacy for the audit workforce.
- data governance.
- availability of DS projects to internal and external stakeholders? (management panels, academic events, ...).
- Examples or ideas of Data Science, Analytics Tools and AI projects that add more value to the Government? (Continuous Auditing, predictive models, pattern identification, ...)

Examples of using new technologies (Blockchain, Network-Graphs, Cloud,) to improve audit processes or public policies.

[Taka] The Comptroller General established the Innovation Lab in 2019 to spearhead hands-on exploration of emerging capabilities such as AI, blockchain, extended reality, and cloud services. The Lab's unique role is aimed for us to better articulate accountability implications, build oversight capacity, and to adopt our own use cases. We built a next-generation compute platform—dubbed the Analytics Foundry—for use by our growing teams of data scientists, statisticians, modelers, and technologists in support of the full innovation lifecycle from ideation, iteration, incubation, and through implementation. Here are select examples of our work:

- Establishing the first non-crypto prototype of a distributed ledger (a.k.a blockchain) that spans across multiple agency/entity operating boundaries to support grants management use case.
- Designed, trained, and tuned our own natural language processing models to enhance clarity of report writing, including automated detection of compliance with evolving writing guidance/standards.
- Developing a corpus of precedence papers to support appropriate use of specific machine learning techniques within oversight contexts and in compliance with audit standards.
- Deployed a first-of-its-kind interactive simulation tool that illustrates complex tradeoff decisions on use of digital identity verifications to potentially reduce improper payments.
- Exploring use of extended reality as a risk reduction tool for onsite work in high-risk operating environments.
- Implementing data literacy and data governance efforts specifically aligned to the audit tradecraft and standards.

Taka is the first Chief Data Scientist appointed by the Comptroller General of the United States for the Government Accountability Office. He also leads GAO's Innovation Lab in driving problem-centric experiments across oversight, insight, and foresight work through data science and emerging technologies. As a member of the federal Senior Executive Service, Taka is responsible for working with GAO stakeholders to adopt prospective views on impacts of emerging capabilities such as AI, cloud computing, blockchains, RPA, extended reality, and IoT.

Taka is a seasoned data science executive with over 22 years of experience helping private and public sector organizations make sense of hidden correlations, behaviors, relationships, patterns, and anomalies. He is passionate about fostering a data-informed culture, using data science as a catalyst to address complex regulatory, risk, operational, and business intelligence challenges.

Classificação de processos judiciais segundo Objetivos de Desenvolvimento Sustentável da Agenda ONU 2030¹

Classification of lawsuits according to the Sustainable Development Goals of the UN 2030 Agenda

Clasificación de demandas según los Objetivos de Desarrollo Sostenible de la Agenda 2030 de la ONU

Luiz Fux², Pedro Felipe de Oliveira Santos³, Aline Carlos Dourado Braga⁴, Pamella Sada Dias Edokawa⁵ e Júlio Luz Sisson de Castro⁶

<https://doi.org/10.36428/revistadacgu.v14i26.548>

Resumo: O Supremo Tribunal Federal (STF), a partir de novembro de 2020, classifica alguns de seus processos em Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030 da ONU. O objetivo desta classificação é integrar efetivamente a Agenda no dia a dia do tribunal. Neste contexto, uma ferramenta para apoio tecnológico à classificação tem imenso potencial para automatizar as tarefas manuais e repetitivas de ler as peças e registrar as etiquetas. A iniciativa RAFA 2030 surgiu com o objetivo de ajudar os servidores a melhor classificar os processos. Este artigo tem os objetivos de apresentar a integração entre a Agenda 2030 e a rotina da corte e a própria ferramenta tecnológica RAFA 2030, em seus aspectos técnicos de desenvolvimento. Atualmente, as principais entregas deste projeto consistem em ferramentas gráficas para processamento de linguagem natural (*co-occurrence graphs*, nuvem de palavras), algoritmos de aprendizagem de máquina, redes neurais, busca por contexto e contagem de palavras-chave, além de outras ferramentas disponíveis em R (*Shiny*) e Python (*Keras*, *Tensorflow* e *Pytorch*). Os resultados iniciais sugerem imenso potencial para aplicações de NLP e aprendizagem de máquina na classificação de documentos jurídicos em temas da Agenda 2030.

Palavras-chave: Agenda ONU 2030. Processamento de Linguagem Natural. Aprendizagem de Máquina. Redes Neurais.

Abstract: Since 2020, the Brazilian Federal Supreme Court has classified its cases according to the sustainable development goals (SDGs) of the the United Nations 2030 Agenda. In this context, a tool for technological support to classification has immense potential to automate the manual and repetitive tasks of reading the text and registering the labels. The RAFA 2030 initiative came up with the goal of helping to classify cases. This article aims to present an integration between Agenda 2030, the work routine in STF and technical aspects of development

1. Artigo submetido em 18/07/2022 e aceito em 01/12/2022.

2. Ministro do Supremo Tribunal Federal, jurista e professor universitário.

3. Desembargador Federal do Tribunal Regional Federal da 6ª Região

4. Analista Judiciário(a) do Supremo Tribunal Federal.

5. Analista Judiciário(a) do Supremo Tribunal Federal.

6. Analista Judiciário(a) do Supremo Tribunal Federal.

about RAFA 2030. Currently, the main results of this project consist of graphical tools for NLP (co-occurrence graphs, tool cloud), machine learning keys, neural networks, context search and keyword counting, in addition to other tools available in R. (Shiny) and Python (Keras, Tensorflow and Pytorch). Initial results indicate great potential for applications of NLP documents and machine learning of legal documents in Agenda 2030 themes.

Keywords: UN 2030 Agenda. Natural Language Processing. Machine Learning. Neural Network.

Resumen: El Supremo Tribunal Federal (STF), a partir de noviembre de 2020, clasifica algunos de sus procesos en los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 de la ONU. El objetivo de esta clasificación es integrar la Agenda en el día a día del tribunal. En este contexto, una herramienta de apoyo tecnológico a la clasificación tiene un potencial inmenso para automatizar las tareas manuales y repetitivas de lectura de piezas y registro de etiquetas. La iniciativa RAFA 2030 surgió con el objetivo de ayudar a los funcionarios judiciales a clasificar mejor los procesos. Este artículo tiene como objetivo presentar la integración entre la Agenda 2030 y la rutina del STF y la propia herramienta tecnológica RAFA 2030, en sus aspectos técnicos de desarrollo. Actualmente, los principales productos intermedios de este proyecto consisten en herramientas gráficas para PNL (grafos de coocurrencia, nube de palabras), algoritmos de aprendizaje automático, redes neuronales, búsqueda de contexto y conteo de palabras clave, además de otras herramientas disponibles en R (Shiny) y Python (Keras, Tensorflow y Pytorch). Los resultados iniciales sugieren un inmenso potencial para las aplicaciones de PNL y aprendizaje automático en la clasificación de documentos legales en temas de la Agenda 2030.

Palabras clave: Agenda 2030 de la ONU. Procesamiento del Lenguaje Natural. Aprendizaje Automático.

1. INTRODUÇÃO

O Supremo Tribunal Federal (STF), Corte Constitucional e instância máxima do Poder Judiciário brasileiro produz uma quantidade expressiva de dados organizados em forma de texto, provenientes das petições recursais e das ações originárias que recebe, além das próprias decisões judiciais que profere. As ações que tramitam no STF são classificadas, a fim de permitir a indicação de assuntos e de outras etiquetas específicas de prioridade, como COVID-19 e parte maior de 65 anos, mas não havia, até setembro de 2020, qualquer catalogação relacionada com a Agenda 2030 da Organização das Nações Unidas (ONU).

No evento (*webinar*⁷) o Ministro Luiz Fux anunciou o início do trabalho de integração entre a Agenda 2030 e o STF, com o estabelecimento de diversas ações, das quais se destaca a classificação manual de ações de controle concentrado e de casos com reconhecida repercussão geral pelos objetivos de desenvolvimento sustentável (ODS) da Agenda 2030 da ONU.

A finalidade deste artigo é apresentar o projeto de integração entre a Agenda 2030 da ONU e o STF e a ferramenta RAFA 2030, desenvolvida para apoiar, via

aprendizagem de máquina e análise de dados, a classificação de ações por ODS da Agenda 2030. A questão que se quer responder é: Ferramentas tecnológicas podem ajudar na classificação de processos de acordo com ODS da Agenda 2030? Assume-se aqui que a aplicação de ferramentas tecnológicas no dia a dia dos tribunais se baseia em dois pontos centrais: o primeiro relacionado à melhoria dos fluxos de trabalho das cortes, que envolve, entre outras ações, a automação de tarefas manuais e repetitivas realizadas por seus servidores; e o segundo referente à combinação das inteligências humana e artificial para a melhoria da prestação jurisdicional.

A aplicação de aprendizagem de máquina e ferramentas centradas em inteligência artificial nos tribunais recebeu impulsos importantes recentemente, conforme FGV (2021), muito embora ainda existam discussões éticas sobre soluções com este perfil. Ainda que o tempo decorrido entre as primeiras aplicações – como a ferramenta Victor (STF e UnB) iniciada em 2017 – e os dias atuais seja relativamente curto, é possível perceber que vários tribunais desenvolveram projetos de inteligência artificial, conforme apresentado em FGV (2021) e CNJ (2022).

7. *Webinar* disponível para acesso no portal eletrônico <https://portal.stf.jus.br/hotsites/webinar-cortes/>. Acesso em: 01 mai. 2022.

O panorama atual pode ser conferido a partir de pesquisa realizada pelo Conselho Nacional de Justiça (CNJ), que lançou, em junho de 2022, um painel⁸ com informações sobre os projetos de inteligência artificial no Poder Judiciário. Mesmo sem estudar profundamente cada uma das iniciativas, é possível afirmar que várias tarefas envolvem processamento de linguagem natural (NLP), como classificação e agrupamento de textos.

Sobre a iniciativa objeto deste artigo, é importante registrar que atualmente a ferramenta RAFA 2030 se encontra em fase de testes, etapa realizada em conjunto por várias secretarias do STF, e os resultados iniciais mostram imenso potencial para utilização em fluxos diários de trabalho. Conforme será apresentado a seguir, as principais entregas da iniciativa envolvem ferramentas gráficas para processamento de textos e apoio à tomada de decisão, algoritmos de redes neurais para classificação automática de peças em ODS da Agenda 2030 da ONU, busca de palavras e leis por contexto e contagem de palavras de interesse para cada ODS, definidas a priori por especialistas do tribunal.

2. ATIVIDADE DE CLASSIFICAÇÃO DE PROCESSOS POR ODS

Esta seção se dedica à atividade de classificar processos a partir de objetivos de desenvolvimento sustentável, que é uma tarefa bastante complexa e desafiadora. Desde 2021, tal tarefa é realizada por servidores destacados de quatro unidades do tribunal⁹ e envolve: a) o estudo da Agenda 2030 em seus objetivos e metas; b) a avaliação da correlação entre o conteúdo da Agenda 2030 e as questões submetidas a julgamento no tribunal e as teses firmadas pela Corte; e c) o registro do resultado dessa atividade em sistemas informatizados do tribunal. Cumpre informar que tal tarefa é realizada em diversos momentos do fluxo processual do STF, a saber: antes da distribuição de ações de controle concentrado, quando do ingresso do processo na pauta dirigida do Plenário, com a inclusão de temas pelo reconhecimento da repercussão geral no Plenário Virtual, ao haver divulgação da questão no Informativo STF e ao ser publicado acórdão na página de pesquisa de jurisprudência do tribunal¹⁰.

A entrega inicial, datada de outubro de 2020, já apresentava a quantidade de 348 etiquetas de ODS em 177 processos avaliados, como mostra a Figura 1.

FIGURA 1 - PAINEL STF E AGENDA ONU 2030 – OUTUBRO/2020



Fonte: Hotsite STF (2022).

8. Painel: Projetos com Inteligência Artificial no Poder Judiciário. 2022. Disponível em: https://paineisanalytics.cnj.jus.br/single/?appid=29d-710f7-8d8f-47be-8af8-a0152545b771&sheet=b8267e5a-1f1f-41a7-90ff-d7a2f4ed34ea&lang=pt-BR&opt=ctxmenu_currsel/. Acesso em: 11 mai. 2022.

9. Integraram a força tarefa para classificação de processos, em outubro de 2020, no STF: a Secretaria de Gestão de Precedentes (SPR), a Secretaria de Altos Estudos, Pesquisas e Gestão da Informação (SAE) e a Assessoria do Plenário (APL); a partir de 2021, também passou a realizar essa atividade a Secretaria Judiciária (SEJ).

10. Os dados decorrentes dessa identificação podem ser visualizados pelos usuários em diversos pontos do portal do Supremo, como, por exemplo, ao consultar processos no acompanhamento processual informatizado ou ao visitar a página de pesquisa de jurisprudência, ao ler informativo STF, ao acompanhar o calendário da pauta do plenário, ao estudar o boletim repercussão geral em pauta e até mesmo ao examinar o novíssimo programa de transparência do Tribunal, denominado Corte Aberta.

Importa dizer que o escopo do projeto de classificação por ODS foi delimitado a partir de recortes estratégicos, não abrangendo todas as ações em tramitação na Corte. Envolve a avaliação de ações de controle concentrado (Ação Direta de Inconstitucionalidade-ADI, Ação Declaratória de Constitucionalidade-ADC, Ação Direta de Inconstitucionalidade por Omissão-ADO e Ação de Descumprimento de Preceito Fundamental-ADPF), além dos casos em que o Tribunal reconhece a repercussão geral da questão constitucional objeto do recurso extraordinário, nos termos do art. 1.035 do Código de Processo Civil¹¹. Verificada a aderência de tal peça com um ou mais ODS da Agenda 2030 há o

registro das etiquetas. Dessa análise, pode-se, ainda, concluir que o processo não é correlacionado a ODS algum¹².

Essa nova atividade classificatória bem como a própria institucionalização da Agenda no STF foram objeto de normativo próprio: a Resolução STF n. 710, de 20 de novembro de 2020¹³. Em junho de 2022, pouco menos de dois anos do início desse trabalho, o STF já conta com mais de 3.300 etiquetas de ODS em processos monitorados (informações disponíveis em painel do hotsite¹⁴ da Agenda 2030 no STF, acessível no portal do Tribunal), cuja imagem segue abaixo:

FIGURA 2 - PAINEL STF E AGENDA ONU 2030 – JUNHO/2022

Processos Monitorados: 2108			Ocorrências de ODS: 3336		
Indicação de processos com aderência à Agenda 2030 no STF					
ODS1	ODS2	ODS3	ODS4	ODS5	ODS6
Erradicação da Pobreza	Fome Zero e Agricultura Sustent.	Saúde e Bem-Estar	Educação de Qualidade	Igualdade de Gênero	Água Potável e Saneamento
62	42	324	122	36	39
ODS7	ODS8	ODS9	ODS10	ODS11	ODS12
Energia Acessível e Limpa	Trabalho Decente e Cresc. Econ.	Indústria, Inovação e Infra	Redução das Desigualdades	Cidades e Comunidades Sust.	Consumo e Produção Respons.
53	441	104	325	92	66
ODS13	ODS14	ODS15	ODS16	ODS17	
Ação Contra Mud. Global Clima	Vida na Água	Vida Terrestre	Paz, Justiça e Instituições Eficazes	Parcerias e Meios de Implem.	
13	14	84	1.300	219	

Fonte: STF (2022).

11. Art. 1.035. O Supremo Tribunal Federal, em decisão irrecorrível, não conhecerá do recurso extraordinário quando a questão constitucional nele versada não tiver repercussão geral, nos termos deste artigo.

[...] 5º Reconhecida a repercussão geral, o relator no Supremo Tribunal Federal determinará a suspensão do processamento de todos os processos pendentes, individuais ou coletivos, que versem sobre a questão e tramitem no território nacional. (CPC, acessível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm, acesso em 14 de julho de 2022).

12. No caso da classificação efetuada pela Coordenadoria de Jurisprudência, foram desenvolvidos critérios de pesquisa para cada uma das 169 metas, sendo utilizada a ferramenta da página de pesquisa de jurisprudência para auxílio na correlação dos acórdãos publicados pelos 17 ODS da Agenda 2030. A equipe já realizou a análise e classificação por ODS dos acórdãos publicados desde 2016, nos casos de ações de controle concentrado e nos casos em que reconhecida a repercussão geral.

13. A Resolução STF n. 710/2020, destacou grupo de trabalho coordenado pelo Secretário-Geral do STF para atuar em nas diversas ações ali elencadas e eleitas para serem executadas no biênio de 2020-2022 com o fito de integração da Agenda 2030 da ONU ao STF.

14. Hotsite disponível para acesso em <https://portal.stf.jus.br/hotsites/agenda-2030/>, com painel gerencial acessível conforme STF (2022).

É possível observar que alguns ODS têm maior número de etiquetas, como o ODS 16 - Paz, Justiça e Instituições Eficazes, cujo tema é fortemente relacionado com a atuação de cortes constitucionais, a exemplo do STF. Vale registrar que, com a classificação de processos estabelecida, foi aberta a possibilidade de priorização dos julgamentos de ações que impactem positivamente os objetivos e metas da Agenda 2030, tendo, já no primeiro semestre de 2021, 76% dos processos incluídos na pauta do Plenário com uma ou mais etiquetas de ODS da Agenda 2030, conforme informado pelo Ministro Luiz Fux no evento Suprema Corte e Diálogos sobre a Agenda 2030. A produção de dados sobre a classificação dos processos em ODS é importante para aferir a efetiva contribuição da Corte com a pauta da Agenda 2030 (*accountability*).

A alteração dos critérios de identificação dos processos submetidos ao STF e a consequente alteração na seleção de matérias a serem julgadas representam um avanço na internacionalização da Corte com equiparação a outros tribunais constitucionais de todo o mundo. Destaca-se, neste ponto, o caráter inédito e pioneiro da iniciativa de classificação dos processos por ODS em uma Corte Suprema.

A agregação da Agenda 2030 da ONU ao STF impactou, também, a própria jurisprudência da Corte, como são exemplos o julgamento da ADI 2.096, Rel. Min. Celso de Mello, em 13/10/2020 e o reconhecimento da repercussão geral no Tema 1.194 (ARE 1352872) apresentado pelo Ministro Luiz Fux.

2.1. Desenvolvimento inicial da ferramenta RAFA 2030

A decisão de desenvolver uma ferramenta tecnológica para apoio à atividade de classificação de processos em ODS da Agenda 2030 foi apresentada no início do projeto e da classificação manual, em setembro de 2020. A associação entre a classificação manual e a ferramenta tecnológica tem o objetivo de tornar a própria classificação mais aderente ao que julga o STF, usando como referência apenas as questões jurídicas em debate ou as teses firmadas pela Corte. Optou-se, assim, por não incluir na iniciativa assuntos já vinculados aos processos ou outros metadados disponíveis em sistemas informatizados.

A ideia inicial era desenvolver um simples contador de palavras cujos significados sugerissem etiquetas de ODS. A iniciativa RAFA começou, portanto, entregando o número de ocorrências de tais palavras

em cada peça jurídica de interesse. Ao longo do tempo, no entanto, a equipe técnica de desenvolvimento da ferramenta percebeu que apenas contar a ocorrência de palavras não era a solução adequada, uma vez que não é razoável ignorar o contexto das palavras descon siderando a posição destas em relação às palavras vizinhas (LIU; KUSNER; BLUNSOM, 2020).

Para ilustrar, anota-se que um texto pode trazer muitas ocorrências da palavra “segurança” sem ter relação ao ODS 16 - Paz, Justiça e Instituições Eficazes. A expressão segurança alimentar, por exemplo, guarda relação com o ODS 2 – Fome zero e Agricultura Sustentável e não com o ODS 16 já citado. A primeira decisão técnica no sentido de aumentar as funcionalidades da iniciativa foi a de ajustar gráficos sobre a disposição das palavras no texto, como nuvens de palavras e grafos de coocorrência, e os apresentar para os analistas.

2.2. Workshop - Grupo de Trabalho Agenda 2030

A primeira versão da iniciativa foi apresentada aos servidores do tribunal no Workshop do Grupo de Trabalho da Agenda 2030 no STF, evento realizado para tratar de questões relativas à classificação de processos e aos requisitos mínimos da iniciativa RAFA 2030.

Entre os dias 29 e 30 de abril de 2021, 60 servidores do tribunal se reuniram para: a) compartilhar experiências e lições aprendidas com a atividade de classificação manual, b) estabelecer critérios comuns para classificação de processos em ODS da Agenda 2030 e c) formular os requisitos básicos para a ferramenta tecnológica de apoio à classificação.

O evento foi constituído de apresentações sobre as estratégias utilizadas para classificação pelas equipes. Contou também com a apresentação inicial da ferramenta RAFA 2030, além de trazer painéis com a participação de assessoras de dois gabinetes do tribunal, que relataram a relevância da iniciativa de classificação.

Houve, ainda, uma parte prática, em que foram promovidas oficinas para classificar experimentalmente alguns processos, cujos acórdãos e petições foram enviados previamente. Tais oficinas contaram com a apresentação do que a RAFA 2030 podia fazer naquele ponto do desenvolvimento, como a contagem de palavras-chave e recursos visuais para ajudar na indicação de ODS da Agenda 2030.

A experiência foi considerada bastante positiva por dois motivos. Primeiro porque a primeira versão ferramenta foi bem avaliada no apoio à atividade de classificação. E segundo, porque os desenvolvedores puderam verificar, na prática, como era feita a classificação, bem como quais eram as dificuldades enfrentadas pelas equipes nessa tarefa. Essa proximidade com a classificação manual tornou mais fácil a identificação de pontos relevantes para o desenho definitivo da ferramenta RAFA 2030.

Após o evento, a equipe do projeto decidiu aumentar as funcionalidades da iniciativa. Alguns servidores procuraram espontaneamente o projeto para sugerir mudanças, ao mesmo tempo que os próprios gestores do projeto estudaram os melhores caminhos para aprimorar a ferramenta.

Como já haviam processos etiquetados em número suficiente para ajustar algoritmos mais sofisticados do que a contagem de palavras-chave, a iniciativa RAFA 2030 organicamente caminhou para utilização de processamento de linguagem natural (NLP) em tarefas de classificação, área da aprendizagem de máquina que experimentou enormes avanços recentemente (HOVY, 2021). Por outro lado, ficou claro para a equipe de desenvolvimento que os servidores do tribunal também desejavam uma interface gráfica associada à ferramenta. Em razão disso, a iniciativa RAFA 2030 passou a se preocupar com a experiência do usuário e com a apresentação visual dos resultados.

Por fim, a proximidade entre os servidores participantes do evento permitiu que parcerias internas fossem criadas com objetivo de testar a ferramenta e promover curadoria nos resultados, etapa fundamental para avaliação do desenvolvimento matemático e de programação realizado. O artigo segue com uma apresentação detalhada sobre as melhorias implementadas na ferramenta, produto das avaliações compartilhadas no workshop e de eventos externos com cunho interinstitucional, tais como Sextas Inteligentes¹⁵ e IX SPES - Seminário de planejamento estratégico sustentável do poder judiciário 2022¹⁶.

2.3. Aspectos gerais da ferramenta RAFA 2030

O primeiro protótipo funcional da ferramenta RAFA 2030 (Redes Artificiais Focadas na Agenda 2030 da ONU) foi lançado em evento promovido pelo STF, no 16 de maio de 2022¹⁷. Neste encontro, especialistas em ciência de dados e em direito registraram a importância da iniciativa, bem como seu enorme potencial para ajudar em tarefas de classificação processual nos objetivos de desenvolvimento sustentável da Agenda ONU 2030.

A ferramenta foi desenhada e construída por servidores da Secretaria de Gestão Estratégica e da Secretaria de Gestão de Precedentes do tribunal, sob coordenação da Secretaria Geral da Presidência do STF, com a participação da Secretaria Judiciária (curadoria) e, ainda, com o apoio da Secretaria de Tecnologia da Informação (disponibilização das peças).

A aplicação se concentra atualmente em duas tarefas específicas: a) classificação automática de grandes conjuntos (lotes) de processos judiciais via aprendizagem de máquina e *deep learning*; e b) apoio visual e estatístico para classificação individual de processos judiciais. A primeira tarefa consiste em utilizar algoritmos de redes neurais para verificar se textos de acórdãos e de petições iniciais apresentam aderência aos ODS 16 (paz, justiça e instituições eficazes), 10 (redução das desigualdades), 8 (trabalho decente e crescimento econômico) e 3 (saúde e bem-estar).

O fluxo é bastante simples: a área jurídica do tribunal indica conjuntos de peças e a RAFA 2030 utiliza textos etiquetados manualmente para treinar e etiquetar novas entradas. A entrega, nesta frente da iniciativa, são as classificações que resultam dos algoritmos ajustados.

A segunda tarefa é centrada no desenvolvimento de um aplicativo para apoiar a decisão de classificar novos processos de acordo com ODS da Agenda 2030. Este aplicativo conta com diversos gráficos utilizados em processamento de linguagem natural (NLP), além de contagem de palavras-chave e busca do contexto. A entrega, neste caso, é a interface que possibilitará aos servidores classificar mais processos jurídicos em menos tempo e com maior qualidade.

15. Evento realizado em 11 de março de 2022, denominado Sextas Inteligentes, pela plataforma Zoom, com a participação de integrantes dos Núcleos de Gerenciamento de Precedentes de todo o país.

16. O resultado de outros testes, efetuados em no final de 2021, foram compartilhados em notícia veiculada no portal do STF, acessível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=481995&ori=1>.

17. Está no canal do YouTube do STF o vídeo do evento de lançamento da RAFA 2030, acessível em <https://www.youtube.com/watch?v=Fhl0snROqfU>, acesso em 8 de junho de 2022.

Nenhuma das tarefas, no entanto, busca substituir os especialistas. A iniciativa RAFA pressupõe integração entre as inteligências artificial e humana, e por este motivo as sugestões de ODS feitas pelas redes neurais ou pelas ferramentas gráficas do aplicativo necessariamente serão revisadas por analistas do tribunal quando a ferramenta estiver em produção.

Sobre o conjunto de palavras-chave utilizado na ferramenta, importa registrar que ele foi elaborado pela equipe técnica da área judicial (Coordenadoria de Jurisprudência - COJU) e o objetivo foi reunir palavras que indicam forte relação da peça jurídica em tela com os objetivos da Agenda 2030, usando suas metas e indicadores. Uma alta ocorrência da palavra “feminicídio”, por exemplo, pode indicar que o processo em análise tem forte relação com o ODS 5 – Igualdade de Gênero.

A metodologia utilizada para construir a iniciativa RAFA 2030 em sua atual composição será melhor explicada no capítulo a seguir, que se ocupa de relacionar aspectos técnicos de programação, base de dados e processamento de linguagem natural (NLP) com o problema de classificar processos judiciais segundo ODS da Agenda 2030.

3. METODOLOGIA

A iniciativa RAFA foi desenvolvida em linguagem R e *Python*, prioritariamente. Isso significa que, com exceção de alguns poucos passos intermediários, tudo aquilo que será apresentado nesta seção envolve apoio computacional de alguma linguagem de programação (R ou *Python*, neste caso). O corpo técnico da iniciativa escolheu usar R e *Python* pois são as linguagens mais populares para ciência de dados, mas os registros deste artigo podem ser reproduzidos em outras linguagens de programação com algum esforço de código e pesquisa de pacotes/bibliotecas. Outras opções populares são: Scala, para utilização do Ecossistema *Hadoop*, Julia, Go e C++.

Dado que *Python* e R são linguagens com grandes comunidades de usuários, não é difícil encontrar informações sobre processamento de linguagem natural na

internet. Reproduzir códigos de fontes não oficiais, no entanto, pode ser complicado e até limitar o potencial de desenvolvimento. Neste contexto, um possível caminho envolve pesquisar algoritmos e estratégias para resolver a tarefa de interesse em boas plataformas (*Towards Data Science* e *Towards AI do Medium*, *Kaggle* e *MachineHack*)¹⁸ e partir para análise da documentação oficial dos frameworks/pacotes utilizados, como *Tensorflow*, *Keras* e *Pytorch* em *Python* ou *Quanteda* e *Text* em R. Evidentemente, tais pesquisas não excluem a necessidade de formação. Apesar de todos os membros técnicos da Secretaria de Gestão Estratégica envolvidos na iniciativa possuírem graduação/pós-graduação em estatística, houve necessidade de complementar a formação superior com cursos sobre aprendizagem de máquina, *deep learning* e inteligência artificial.

O primeiro desafio metodológico para aplicações de aprendizagem de máquina costuma ser a obtenção dos dados. Na maioria das vezes, limpar e manipular dados em texto é bem mais complexo do que a manipulação exigida por dados em formato numérico ou categórico e nem sempre os textos estão disponíveis publicamente para uso. Na iniciativa RAFA são utilizados dados do próprio tribunal, o que certamente torna esta etapa mais fácil. Explicações mais detalhadas sobre a base de dados serão dadas na seção subsequente.

3.1. Base de dados

Os principais dados de entrada da RAFA são textos de petições iniciais e de acórdãos de processos do STF. São arquivos em formato .PDF que podem ser nato digitais ou imagens. Imagens são típicas em processos digitalizados e exigem um passo de reconhecimento ótico de caracteres (OCR). Maiores informações sobre tarefas de OCR estão disponíveis em Ooms (2022). As peças em PDF estão disponíveis no site do STF. Basta pesquisar pelo processo de interesse na homepage, como mostra a figura a seguir:

18. Disponíveis em: *Towards Data Science* (<https://towardsdatascience.com/>); *Towards AI* (<https://towardsai.net/>); *Medium* (<https://medium.com/>); *Kaggle* (<https://www.kaggle.com/>); *MachineHack* (<https://machinehack.com/>), com acesso em: 04 de mai. 2022.

FIGURA 3 – CONSULTA PROCESSUAL NO PORTAL STF



Fonte: Homepage STF¹⁹.

O exemplo em tela mostra uma pesquisa referente ao processo ADPF-779. Clicando em pesquisar, abre-se a seguinte página:

FIGURA 4 – CONSULTA PROCESSUAL ADPF-779

 The image shows the detailed page for process ADPF 779 on the STF portal. The page has a blue header with the text 'ADPF 779' and 'PROCESSO ELETRÔNICO PÚBLICO'. Below the header, there are buttons for 'Dje', 'Jurisprudência', 'Peças', 'Push', and a printer icon. The main content area is white and contains the following information:

- NÚMERO ÚNICO: 0112261-18.2020.1.00.0000
- ARGUIÇÃO DE DESCUMPRIMENTO DE PRECEITO FUNDAMENTAL
- Origem: DF - DISTRITO FEDERAL
- Relator: MIN. DIAS TOFFOLI
- Relator do último incidente: MIN. DIAS TOFFOLI (ADPF-MC-Ref)

 Below this information, there is a table with two columns:

REQTE.(S)	PARTIDO DEMOCRATICO TRABALHISTA
ADV.(A/S)	PAULO ROBERTO IOTTI VECCHIATTI (242668/SP) E OUTRO(A/S)
INTDO.(A/S)	PRESIDENTE DA REPÚBLICA
PROC.(A/S)(ES)	ADVOGADO-GERAL DA UNIÃO

 At the bottom of the page, there is a section titled 'AGENDA 2030 DA ONU:' with three icons representing the Sustainable Development Goals:

- 5 IGUALDADE DE GÊNERO (Gender Equality)
- 10 REDUÇÃO DAS DESIGUALDADES (Reduced Inequalities)
- 16 PAZ, JUSTIÇA E INSTITUIÇÕES EFICAZES (Peace, Justice and Strong Institutions)

Fonte: Homepage STF.

19. Homepage STF disponível em: <https://portal.stf.jus.br>. Acesso em: 10 jul. 2022.

Clicando-se no botão peças, os documentos estarão dispostos à esquerda e podem ser baixados facilmente, desde que não estejam em segredo de justiça. Para verificar quais processos foram classificados em ODS da Agenda 2030 basta entrar no *hotsite* da Agenda 2030 no STF e baixar um arquivo em .xlsx com os processos e os metadados de classificação. Combinando tal arquivo com a busca de peças no *site* do STF, é possível obter as peças de onde os textos podem ser extraídos, bem como as classificações manuais em ODS da Agenda 2030.

Com os textos devidamente extraídos dos processos de interesse, é possível começar a etapa de processamento de linguagem natural, que será apresentada na próxima seção. Boas referências em NLP são os livros dos autores Martin (2009), Vajjala *et al* (2020) e Hvitfeldt e Silge (2021), este último exclusivo para uso em linguagem R.

3.2. Processamento de Linguagem Natural

As ciências sociais receberam bem a análise de texto e o processamento de linguagem natural (NLP) em seu conjunto de técnicas, como mostram as referências Nay (2021), Jockers (2014) e Kulas (1988). Em resumo, as técnicas de NLP tem foco em processar o texto para, através de treinamento supervisionado ou não, realizar tarefas de predição, além de análise visual e descritiva. A tarefa mais conhecida de processamento de linguagem, talvez, seja a análise de sentimentos, que se concentra em, dado um conjunto grande de textos classificados e dicionários produzidos por especialistas, adivinhar se um extrato de texto é positivo, negativo ou neutro (LIU, 2020). Variações desta abordagem usam dicionários mais sofisticados e objetivam identificar sentimentos como alegria ou raiva²⁰. Na prática, as atividades de predição buscam, a partir de conjuntos de treinamento devidamente etiquetados, prever se um novo texto pertence ou não à uma determinada categoria. Atividades de análise descritiva e visual se concentram em buscar relações entre as palavras e extratos dos textos, sem fazer previsões ou tentar adiantar comportamentos em textos novos.

A iniciativa RAFA conta com a parte de predição, especificamente na sugestão de classificação para lotes de processos e com a parte de análise descritiva e visual, uma vez que o aplicativo apresenta contagens de palavras e inúmeros gráficos. A etapa de classificação

é, seguramente, mais pesada em termos de processamento e exige maiores conhecimentos técnicos. As funcionalidades utilizadas unicamente no aplicativo usam o texto bruto, isto é, sem o passo adicional de limpeza e manipulação, característico de atividades de estatística e programação.

As peças em PDF são processadas de forma a extrair apenas os textos importantes para tarefa de classificação. As principais etapas de pré-processamento (limpeza) são:

- Extração de textos em PDFs com imagens (OCR) e nato digitais;
- Remoção de palavras de parada (*stopwords*);
- Remoção de caracteres especiais, tais como #, @ e &;
- Remoção de espaços em branco desnecessários;
- Formatação das palavras em minúsculo.

Este processamento é feito para tornar os textos mais densos e melhorar a performance dos algoritmos de aprendizagem de máquina. A natureza dos documentos alvo de limpeza influenciam os resultados, dado que as próprias *stopwords* podem mudar de área para área. Textos jurídicos, por exemplo, possuem palavras de parada diferentes de textos literários, dado que os autores de literatura usam recursos de estilo não necessários em textos técnicos. Desta forma, é importante frisar que cada problema de NLP exige uma limpeza de texto específica.

A etapa seguinte transforma os textos em vetores de números e é conhecida pelo nome de *embedding* (GUPTA *et al.*, 2020). Máquinas não processam textos nativamente, sendo necessário um passo de codificação, para transformar coleções de palavras em conjuntos numéricos. Existem diversas formas de *embedding*, sendo as mais simples baseadas em dicionário (conjunto de todas as palavras de um texto) e as mais sofisticadas baseadas em contexto. Na construção dos modelos para predição de ODS, foram utilizadas as seguintes tecnologias de *embedding*: *one-hot encoding*, *bag of words* (BoW), *tf-idf* e *bert*. Maiores informações sobre tais métodos podem ser encontradas em Meijer, Truong e Karimi (2021) e David e Renjith (2021).

Com os textos limpos e convertidos em vetores de números, algoritmos podem ser ajustados para prever categorias, neste caso, etiquetas de ODS. Muitas

20. Detalhes em: <https://medium.com/towards-data-science/using-sentiment-analysis-to-explore-emotions-within-text-ac48e3e93999>. Acesso em: 12 mai. 2022.

abordagens são possíveis, mas as principais envolvem aprendizagem supervisionada e algoritmos de *machine learning* (aprendizagem de máquina), tais como SVM, *Naive Bayes*, *XGBoost* e *CatBoost*, além de redes neurais profundas, características de *deep learning*. Também é possível utilizar variações de abordagens não supervisionadas, tais como similaridade e agrupamento de textos. As redes neurais foram escolhidas pois apresentaram boa performance em diversas métricas, conforme seções subsequentes mostram.

3.3. Testes – Atividade de Classificação Automática

A primeira estratégia utilizada para classificação foi *multilabel*, o que significa que cada processo avaliado podia ter todas as etiquetas (17 ODS) ou nenhuma delas. Considerando cada combinação possível como um rótulo individual, seriam aproximadamente 131.071 etiquetas possíveis para cerca de 2 mil processos analisados, ou seja, existiriam mais categorias do que dados, um problema recorrente para modelos de previsão que envolvem treinamento e aprendizagem. Uma alternativa elegante para contornar este problema é a utilização de modelos com probabilidade de pertencimento. Modelos deste tipo, após o treinamento, devolvem não uma coleção de etiquetas, mas a probabilidade de cada etiqueta ser utilizada no registro que se quer prever a classificação.

Para tal modelo, uma possibilidade é utilizar o *framework StarSpace* do Facebook, cuja implementação em R pode ser encontrada no pacote *ruimtehol*. O *StarSpace* é, fundamentalmente, uma máquina de gerar *embeddings*. No caso de textos, bons *embeddings* possuem vetores parecidos para palavras de significado próximo em linguagem natural, considerando o contexto de utilização. A implementação em R do *StarSpace* inclui algoritmos de similaridade e redes neurais, que, no caso de problemas *multilabel*, geram saídas com probabilidade de pertencimento. Apesar do *StarSpace* produzir *embeddings* robustos, classificar com um único algoritmo todos os ODS se mostrou uma tarefa muito pesada para o método, principalmente nos objetivos mais gerais, como no caso do décimo sexto, relacionado à Paz, Justiça e Instituições Eficazes.

A consequência natural da utilização do *framework StarSpace* é um estudo sobre similaridade de processos judiciais. Algumas iniciativas do judiciário utilizam similaridade para agrupar processos com diferentes propósitos, a exemplo da Berna (TJGO, 2020) e Athos (STJ, 2020).

No início, a iniciativa RAFA se ocupou de empregar similaridade com objetivo de classificação posterior. Na prática, calculava a similaridade entre um novo processo candidato a receber etiquetas de ODS e as peças jurídicas da base já etiquetada, com objetivo de avaliar a similaridade média entre a peça nova e aquelas classificadas em cada um dos ODS. O ponto de corte para decidir se o novo processo receberia ou não uma determinada etiqueta era dado pela similaridade média entre os processos classificados na própria base. A continuidade desta abordagem se tornou difícil pela heterogeneidade dos textos (petições em especial) e pela falta de textos etiquetados para a maioria dos ODS. Para contornar a situação, o problema *multilabel* foi quebrado em problemas binários.

Essa estratégia permitiu tratar cada ODS de forma individual, separando aqueles com número suficiente de processos daqueles com registros insuficientes para aprendizagem de máquina. Também permitiu ajustar um modelo para cada ODS, escolhendo os algoritmos candidatos em uma coleção maior de modelos. Como consequência de haver mais opções de modelos, surgiram também mais opções de ensemble, abordagem que define o rótulo (etiqueta) por votação simples ou ponderada (GANAIÉ, 2021).

Alguns ODS possuem poucos processos etiquetados, o que naturalmente impede que máquinas aprendam a classificá-los usando poucos exemplos. Se existem, por exemplo, 2 mil processos classificados e apenas 3 etiquetados em um determinado ODS, a máquina pode classificar nenhuma peça em tal ODS e ainda assim performar bem em termos gerais. Por este motivo, foram selecionados apenas os ODS 16, 8, 10 e 3, respectivamente, os objetivos com maiores números de entradas. Considerando cada um destes 4 ODS como um problema único de aprendizagem, a iniciativa RAFA alcançou resultados satisfatórios usando redes neurais recorrentes (SHERSTINSKY, 2020).

O quadro resumo a seguir compara as abordagens utilizadas:

QUADRO 1 – SÍNTESE DAS ABORDAGENS TESTADAS

ABORDAGEM	OBJETIVO	RESULTADOS
Classificação <i>Multilabel</i> .	Etiquetar processos em múltiplos ODS da Agenda 2030.	Baixa performance em ODS gerais, tais como o 16.
Similaridade.	Calcular a similaridade entre novas entradas e processos já etiquetados.	Baixa performance em ODS com poucas entradas, tais como o 13 e o 14.
Classificação Binária.	Etiquetar processos em um único ODS da Agenda 2030.	Performance boa nos 4 ODS avaliados – 16, 8, 10 e 3.

Fonte: *Elaboração própria.*

Tendo justificado a escolha das redes neurais, resultados pormenorizados podem ser apresentados, como as métricas *F-score*, *recall* e precisão. O próximo capítulo demonstra os resultados obtidos com as redes e as principais notas sobre a utilização do aplicativo para apoio à classificação.

4. RESULTADOS E DISCUSSÕES

Os resultados obtidos no processamento de grandes lotes de processos estão disponíveis num painel gerencial para uso interno. Tal painel mostra a evolução da iniciativa nos testes realizados semanalmente, dado que ajustes são feitos nos algoritmos quando os especialistas fazem curadoria dos resultados anteriores. O fluxo é o seguinte: os algoritmos são executados em sua configuração e parâmetros atuais. Os resultados são enviados para especialistas do tribunal, que devolvem notas sobre as classificações feitas pela iniciativa RAFA 2030. As análises de tais notas motivam mudanças nos algoritmos e o ciclo de melhorias se repete, alternando curadoria e reajuste de algoritmos. Este movimento motivou, por exemplo,

a mudança na precedência entre os tipos de peças utilizados. No começo, a iniciativa RAFA utilizava preferencialmente acórdãos, usando petições apenas para processos sem informação útil mais recente. Ocorre que, apesar dos acórdãos serem mais fáceis de manipular, a classificação manual é feita, hoje, basicamente em petições. Essa troca de precedência (acórdão para petições) gerou aumento significativo da acurácia para as redes neurais dos ODS 16, 8, 10 e 3. Também há de se destacar que, em alguns momentos, as divergências observadas entre a marcação dos servidores e aquelas sugeridas pela RAFA geraram boas discussões sobre o processo de classificação em si, uma vez que a máquina enxerga padrões muitas vezes não perceptíveis rapidamente por humanos.

4.1 Performance Processamento em Lote

A performance das redes neurais recorrentes é avaliada através de métricas tradicionais de aprendizagem de máquina, tais como acurácia, sensibilidade, especificidade, precisão e F-score. O quadro 2 apresenta tais métricas e suas fórmulas:

QUADRO 2 – MÉTRICAS

MÉTRICA	FÓRMULA
Sensibilidade (<i>recall</i>)	$VP / (VP + FN)$
Especificidade	$VN / (FP + VN)$
Acurácia	$(VP + VN) / N$
Precisão	$VP / (VP + FP)$
F-Score	$2 \times (P \times S) / (P + S)$

Fonte: *Elaboração própria.*

No Quadro 2 as siglas e suas respectivas descrições correspondem a: verdadeiros positivos (VP), falsos negativos (FN), falsos positivos (FP), verdadeiros negativos (VN), precisão (P), sensibilidade (S) e total de elementos (N). Em outras palavras, VP ocorre quando o modelo prevê etiqueta e há etiqueta na realidade, isto é, os especialistas de fato etiquetaram aquele processo. VN ocorre quando não há etiqueta e o modelo não prevê etiqueta. FP ocorre quando o modelo prevê etiqueta e na verdade não há marcação. E, por fim, FN acontece quando o modelo não prevê etiqueta e os especialistas a identificam.

O Quadro 2 mostra que a acurácia é a métrica mais simples e avalia o percentual geral de acertos, entre marcações e não marcações. A sensibilidade (*recall*) avalia a capacidade do modelo classificar corretamente etiquetas de ODS. A especificidade, ao contrário, é indicada para avaliar não etiquetas. A precisão avalia a quantidade de verdadeiros positivos sobre a soma de todas as etiquetas (certas ou não). A métrica *F-Score* é uma média entre a precisão e a sensibilidade.

O Quadro 3 mostra a matriz de confusão, objeto que organiza as quantidades VP, VN, FP e FN:

QUADRO 3 – MATRIZ DE CONFUSÃO

MATRIZ DE CONFUSÃO		CLASSE PREDITA	
		POSITIVA	NEGATIVA
Classe original	Positiva	VP	FN
	Negativa	FP	VN

Fonte: Elaboração própria.

O Quadro 4 mostra as métricas observadas para a última execução das redes neurais recorrentes:

QUADRO 4 – MÉTRICAS RAFA 2030

	ACURÁCIA	SENSIBILIDADE	F-SCORE
ODS 16	77,5%	85,3%	69,9%
ODS 8	82,9%	84,1%	79,6%
ODS 10	91,9%	100%	30,8%
ODS 3	93,7%	83,3%	74,1%

Fonte: Elaboração própria.

O ODS 10 mostra o pior *F-Score* em razão da baixa precisão. Em síntese, neste ODS, a RAFA 2030 etiqueta muito mais processos do que deveria. Em contrapartida, as métricas de acurácia se mostram boas para todos os ODS e todas as sensibilidades estão acima de 80%. Estes resultados mostram que é possível utilizar aprendizagem de máquina e *deep learning* para sugerir classificações em ODS da Agenda 2030.

4.2. Utilização do aplicativo

A iniciativa RAFA também pretende apoiar analistas que buscam classificar um processo por vez. A execução em lote demanda esforço de código e computação, funcionando melhor para grandes conjuntos

de processos. Por este motivo, foi desenvolvido um aplicativo em *Shiny*, pacote da linguagem R. São disponibilizadas ferramentas gráficas para NLP, tais como gráficos de coocorrência e de palavras mais frequentes, bem como nuvens de palavras e bigramas, além de pesquisas de palavras-chave e busca por contexto de palavras específicas.

O aplicativo fica disponível em um *link* similar a um endereço *web* e pode ser utilizado em qualquer navegador (*Google Chrome*, *Mozilla*, etc). Nenhum pré-requisito técnico ou de hardware é necessário, pois o aplicativo foi criado para ser útil, funcional e intuitivo aos servidores da área jurídica. O aplicativo pode ser utilizado até mesmo em celular, basta haver um navegador *web* e acesso à internet. A primeira aba do apli-

cativo tem informações sobre o seu uso e a segunda contém informações sobre a Agenda 2030, com *links* oficiais para a ONU e para o *hotsite* do STF. O usuário encontra um campo para *upload* do processo na terceira aba (chamada RAFA 2030). Depois que o aplicativo lê a peça em PDF, aparecem sugestões de clas-

sificações obtidas aprendizagem de máquina (ODS 16,10,8 e 3) e gráficos como a nuvem de palavras. Ainda nesta aba, o usuário pode baixar o texto limpo para utilização em algoritmos próprios. A figura a seguir mostra a aba RAFA 2030 do aplicativo:

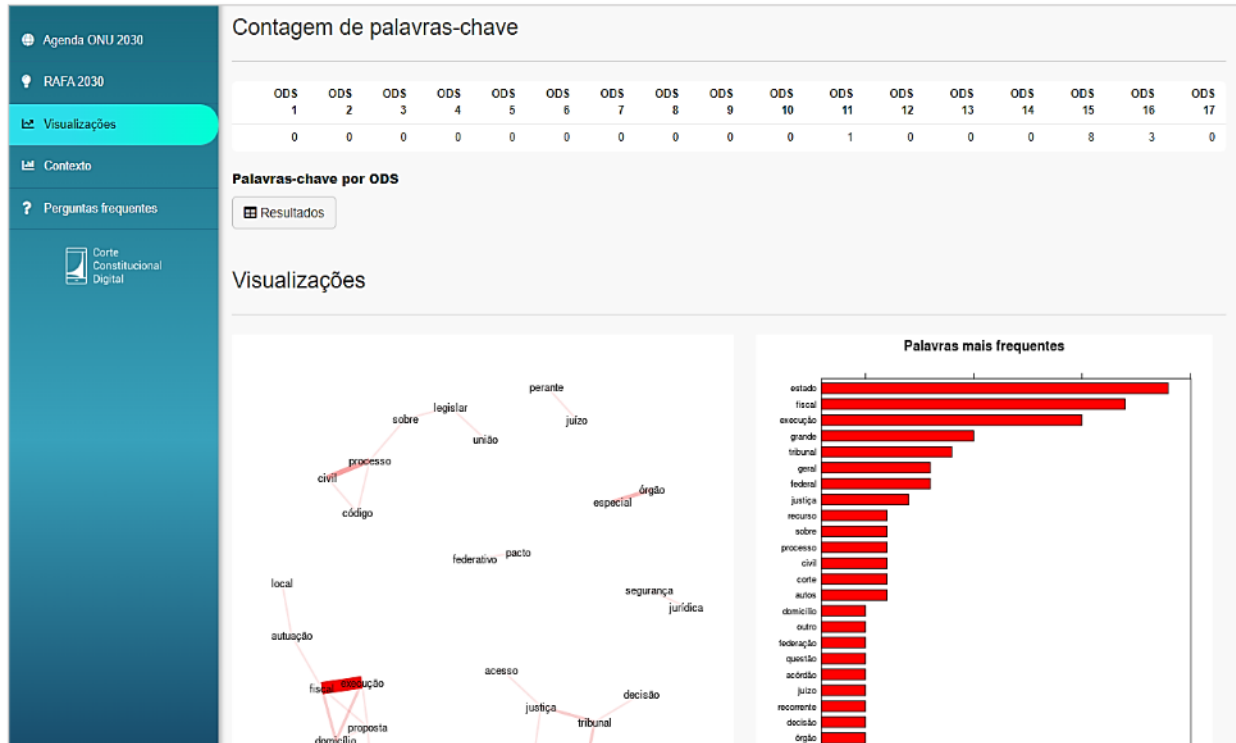
FIGURA 5 – ABA RAFA 2030 DO APLICATIVO



Fonte: Elaboração própria.

As demais abas mostram gráficos mais avançados, contagem e exibição das palavras-chave localizadas e um buscador de contexto, como mostram as figuras 6 e 7.

FIGURA 6 – ABA VISUALIZAÇÕES DO APLICATIVO



Fonte: Elaboração própria.

FIGURA 7 – ABA CONTEXTO DO APLICATIVO



Fonte: Elaboração própria.

Com esses recursos é possível mapear palavras importantes do documento e analisar o contexto de cada uma delas. Um exemplo clássico é dado pela palavra “Rio”, que pode ser curso de água ou palavra inicial de nomes de estados, como Rio de Janeiro e Rio Grande do Norte. Uma simples busca de contexto pode ajudar a não classificar um dado processo no ODS 14 – Vida na Água, por exemplo. Com o uso do aplicativo, analistas podem fazer um mapa geral do documento, de forma automatizada e em poucos segundos após o *upload* de uma peça jurídica. Espera-se que a ferramenta possa ajudar a diminuir o tempo para análise e classificação dos documentos, de forma a permitir que servidores sejam alocados em atividades menos manuais e repetitivas.

Alguns módulos ainda podem ser incluídos e o aplicativo encontra-se em fase de testes por usuários do próprio tribunal. O aumento de processos para treinamento tende a trazer mudanças nas redes neurais, contudo não deve impactar no processo de classificação, já que em última análise, esta avaliação será feita por um servidor. Até a estabilização do algoritmo espera-se que o aplicativo seja de grande utilidade na padronização das classificações, o que gera um efeito cíclico de melhoria na performance dos algoritmos.

5. CONCLUSÃO

O principal objetivo do artigo foi apresentar a ferramenta tecnológica RAFA 2030, desenvolvida dentro de projeto estratégico de institucionalização da Agenda 2030 no STF para apoiar a classificação de processos em objetivos de desenvolvimento sustentável. O trabalho não esgota todas as possibilidades de melhoria da iniciativa, dado que tanto o projeto quanto a ferramenta de apoio à classificação estão em constante avaliação e evolução.

As contribuições deste artigo envolvem a apresentação de algumas etapas para desenvolvimento do aplicativo, bem como para o ajuste de redes neurais. Extratos de código com exemplos estão disponíveis em repositório no *GitHub*²¹.

Possibilidades futuras envolvem o ajuste de algoritmos que combinam texto e metadados, a exemplo do *CatBoost*²² e também ajustes de *ensemble*, isto é, com a reunião de vários algoritmos de aprendizagem de máquina em um esquema de votação para melhor classificar processos em ODS. Sobre o aplicativo, módulos com buscas mais sofisticadas podem ser incluídos, como por exemplo, *queries* de distância entre pares de palavras em extratos de texto. Também podem ser adicionadas abas sobre a similaridade, de forma a apresentar para o analista exemplos de textos etiquetados parecidos com o de interesse. As principais soluções testadas e integradas à iniciativa são combinações de Inteligência Artificial propriamente dita com mecanismos de automação.

REFERÊNCIAS

CNJ - Conselho Nacional de Justiça. Programa Justiça 4.0. 2022. Disponível em: <https://www.cnj.jus.br/tecnologia-da-informacao-e-comunicacao/justica-4-0/>. Acesso em: 10 mai. 2022.

_____. Projetos com Inteligência Artificial no Poder Judiciário. 2022. Disponível em: https://paineisanalytics.cnj.jus.br/single/?appid=29d710f7-8d8f-47be-8af8-a9152545b771&sheet=b8267e5a-1f1f-41a7-90ff-d7a2f4ed34ea&lang=pt-BR&opt=ctxmenu_currsel/. Acesso em: 11 mai. 2022.

DAVID, M. S.; RENJITH, S. Comparison of word embeddings in text classification based on RNN and CNN. In: *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2021. p. 012029. Disponível em: <https://doi.org/10.1088/1757-899x/1187/1/012029>. Acesso em: 12 mai. 2022.

FGV - Fundação Getúlio Vargas. Portal FGV. 2021. Artificial Intelligence in the Judiciary: the most complete research on the subject. Disponível em: <https://portal.fgv.br/en/news/artificial-intelligence-judiciary-most-complete->

21. Disponível em: <https://github.com/agenda2030rafa/rafa2030>.

22. Detalhes em: <https://catboost.ai/en/docs/>. Acesso em: 13 jul. 2022.

[-research-subject](#). Acesso em: 04 de mai. 2022.

GANAIÉ, M. A. et al. Ensemble *deep learning*: A review. **arXiv preprint arXiv:2104.02395**. 2021. Disponível em: <https://doi.org/10.48550/arXiv.2104.02395>. Acesso em: 12 mai. 2022.

GUPTA, S. et al. Task-optimized word embeddings for text classification representations. **Frontiers in Applied Mathematics and Statistics**, v. 5, p. 67, 2020. Disponível em: <https://doi.org/10.3389/fams.2019.00067>. Acesso em: 12 mai. 2022.

HOSSIN, M.; SULAIMAN, M. N. A review on evaluation metrics for data classification evaluations. **International journal of data mining & knowledge management process**, v. 5, n. 2, p. 1, 2015. Disponível em: <http://dx.doi.org/10.5121/ijdkp.2015.5201>. Acesso em: 06 jun. 2022.

HOVY, D. **Text Analysis in Python for Social Scientists: Prediction and Classification**. Cambridge University Press, 2021. Disponível em: <https://doi.org/10.48550/arXiv.2104.02395>. Acesso em: 06 jun. 2022.

HVITFELDT, E.; SILGE, J. **Supervised machine learning for text analysis in R**. Chapman and Hall/CRC, 2021.

JOCKERS, M. L. **Text analysis with R for students of literature**. Cham: Springer, 2014.

KULAS, Jack. Philosophy and Natural-Language Processing. In: Kulas, J., Fetzer, J.H., Rankin, T.L. (eds) *Philosophy, Language, and Artificial Intelligence*. Studies in Cognitive Systems, vol 2. Springer, Dordrecht. 1988. Disponível em: https://doi.org/10.1007/978-94-009-2727-8_1. Acesso em: 07 abr. 2022.

LIU, Bing. **Sentiment analysis: Mining opinions, sentiments, and emotions**. Cambridge university press, 2020.

LIU, Q.; KUSNER, M. J.; BLUNSOM, P. A survey on contextual embeddings. **arXiv preprint arXiv:2003.07278**, 2020. Disponível em: <https://doi.org/10.48550/arXiv.2003.07278>. Acesso em: 15 abr. 2022.

MARTIN, J. H. **Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition**. Pearson/Prentice Hall, 2009.

MEIJER, H. J.; TRUONG, J.; KARIMI, R. Document embedding for scientific articles: Efficacy of word embeddings vs TFIDF., 2021. Disponível em: <https://doi.org/10.48550/arXiv.2107.05151>. Acesso em: 12 mai. 2022.

NAY, J. Natural Language Processing for Legal Texts. In D. Katz, R. Dolin, & M. Bommarito (Eds.), **Legal Informatics** (pp. 99-113). Cambridge: Cambridge University Press. 2021. Disponível em: <https://doi.org/10.1017/9781316529683.011>. Acesso em: 07 abr. 2022.

OOMS, J. 2022. Tesseract: Open Source OCR Engine. Disponível em: <https://CRAN.R-project.org/package=tesseract>. Acesso em: 02 fev. 2022.

PEIXOTO, F. H. Projeto Victor: relato do desenvolvimento da inteligência artificial na repercussão geral do Supremo Tribunal Federal. **Revista Brasileira de Inteligência Artificial e Direito-RBIAD**, v. 1, n. 1, p. 1-22, 2020.

PYTHON SOFTWARE FOUNDATION. *Python* Language, 2022. Disponível em: <https://www.Python.org/>. Acesso em: 06 jun. 2022.

R CORE TEAM. R: **A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2022. Disponível em: <https://www.R-project.org/>. Acesso em: 06 jun. 2022.

SHERSTINSKY, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. **Physica D: Nonlinear Phenomena**, v. 404, p. 132306, 2020. Disponível em: <https://doi.org/10.48550/arXiv.1808.03314>. Acesso em: 06 jun. 2022.

STF - Supremo Tribunal Federal. Hotsite Agenda 2030 STF. 2022. Disponível em: <http://portal.stf.jus.br/hotsites/agenda-2030/>. Acesso em: 11 abr. 2022.

_____. Painel: STF e Agenda 2030. 2022. Metadados de processos. Disponível em: <https://transparencia.stf.jus.br/single/?appid=bc086310-9ce5-4b55-9a8a-e420a907ccca&sheet=db127d33-d4c1-47e0-8bab-3079214719c6&opt=c->

[txmenu](#). Acesso em: 11 abr. 2022.

STJ - Superior Tribunal de Justiça. Revolução tecnológica e desafios da pandemia marcaram gestão do ministro Noronha na presidência do STJ. 2020. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/23082020-Revolucao-tecnologica-e-desafios-da-pandemia-marcaram-gestao-do-ministro-Noronha-na-presidencia-do-STJ.aspx>. Acesso em: 14 mai. 2022.

TJGO - Tribunal de Justiça do Estado de Goiás. Lançamento Oficial da ferramenta Berna - Sistema de Inteligência Artificial. Youtube, 26 mai. 2020. Disponível em: <https://www.youtube.com/watch?v=MY8OG7UGdhM>. Acesso em: 14 mai. 2022.

VAJJALA, S. et al. **Practical natural language processing: a comprehensive guide to building real-world NLP systems**. O'Reilly Media, 2020.



Luiz Fux

ORCID: <https://orcid.org/0000-0003-1996-7572>

Ministro do Supremo Tribunal Federal, jurista e professor universitário.

E-mail: gabineteluzifux@stf.jus.br

Foi ministro do Tribunal Superior Eleitoral de 2014 a 2018 e ministro do Superior Tribunal de Justiça de 2001 a 2011, promotor de justiça do Ministério Público do Estado do Rio de Janeiro de 1979 a 1982 e juiz de direito fluminense de 1982 até 1997, quando foi promovido a desembargador do Tribunal de Justiça do Rio de Janeiro. É bacharel (1976) e doutor (2009) em direito pela Faculdade de Direito da Universidade do Estado do Rio de Janeiro (UERJ). Desde 1995 é professor titular da UERJ, além de ter lecionado na Escola de Magistratura do Estado do Rio de Janeiro e na Pontifícia Universidade Católica do Rio de Janeiro. É membro da Academia Brasileira de Letras Jurídicas e da Academia Brasileira de Filosofia. Presidiu a comissão de juristas que elaborou o anteprojeto do Código de Processo Civil, em vigor desde 2016.



Pedro Felipe de Oliveira Santos

ORCID: <https://orcid.org/0000-0003-4600-3760>

Desembargador Federal do Tribunal Regional Federal da 6ª Região.

E-mail: pedro.felipe@trf1.jus.br

Vice-Diretor da Escola de Magistratura Federal da Sexta Região. Magistrado de Enlace no Brasil para a Convenção de Haia de 1980. Atuou como Secretário Geral da Presidência do Supremo Tribunal Federal, Juiz Instrutor e Juiz Auxiliar do Supremo Tribunal Federal. Atuou como Juiz Auxiliar do Conselho Nacional de Justiça. Atuou como Defensor Público Federal. Doutorando em Direito pela Universidade de Oxford. Mestre em Direito pela Universidade de Harvard. Graduado em Direito pela Universidade de Brasília. Professor de Direito Constitucional.



Aline Carlos Dourado Braga

ORCID: <https://orcid.org/0000-0002-1713-3833>

Analista Judiciária do Supremo Tribunal Federal.

E-mail: alinecd@stf.jus.br

É graduada em Direito pelo Uniceub-Centro de Ensino Unificado de Brasília, pós-graduada em Direito Público pela Universidade Cândido Mendes e responsável pela interlocução para a execução das ações do Projeto Agenda 2030 no STF. Secretária de Gestão de Precedentes do STF.



Pamella Sada Dias Edokawa

ORCID: <https://orcid.org/0000-0001-7387-8533>

Analista Judiciária do Supremo Tribunal Federal.

E-mail: pamella.edokawa@stf.jus.br

Coordenadora do Escritório de Gestão da Estratégia do Supremo Tribunal Federal. Graduada em Estatística pela UnB e especialista em regulação financeira. É responsável pelo núcleo de Análise Estatística e pela gerência de Gestão da Estratégia. É analista no STF desde 2016.



Júlio Luz Sisson de Castro

ORCID: <https://orcid.org/0000-0002-7755-850X>

Analista Judiciário do Supremo Tribunal Federal.

E-mail: juliols@stf.jus.br

Analista judiciário desde 2005, atualmente ocupa o cargo de Supervisor do Núcleo de Gerenciamento de Precedentes do STF, tendo trabalho na 12ª Vara Federal de Brasília e no Superior Tribunal de Justiça. É pós-graduado em Processo Civil pelo UniCeub-Centro de Ensino Unificado de Brasília. Atua no projeto Agenda 2030 no STF.

Entorno legal y adopción de *blockchain* como herramienta para prevenir la corrupción en contrataciones públicas: reflexiones sobre iniciativas europeas y los marcos normativos brasileños¹

Entorno legal e adoção de blockchain como ferramenta para prevenir a corrupção em contratações públicas: reflexões sobre iniciativas europeias e os marcos normativos brasileiros.

Legal surroundings and adoption of blockchain as a tool to prevent corruption in public contracts: reflections on European initiatives and the Brazilian regulatory framework.

Samir Sallen Silva Santos e Lucio Alves Angelo Junior

<https://doi.org/10.36428/revistadacgu.v14i26.528>

Resumen: La contratación pública representa parte relevante del gasto de los países. En Brasil, puede alcanzar alrededor de un 12% de su Producto Nacional Bruto (PNB). Según la Convención de las Naciones Unidas contra la Corrupción, los contratos públicos son instrumentos de gran importancia para lograr intereses colectivos, pero muy vulnerables a actos de corrupción y fraude causadores de daños al desarrollo económico. Recientemente han surgido las Tecnologías de Registros Distribuidos (TRD, o DLTs, por sus siglas en inglés), dotadas de características que pueden aportar beneficios significativos al sector público. Entre ellas, la garantía de inmutabilidad e integridad de registros múltiples y la desintermediación generada por un mecanismo de consenso entre ellos. Su uso permite la automatización de procedimientos a través de los denominados *smart contracts* en procesos de licitación pública, reduciendo posibles brechas a hechos ilícitos. Así las cosas, este trabajo examina algunas de las iniciativas europeas sobre dicha tecnología, el entorno jurídico brasileño y, al final, propone una reflexión sobre los primeros pasos hacia su implementación y las interacciones necesarias a una efectiva coordinación institucional para el uso de la *blockchain* por la administración pública en Brasil.

Palabras clave: contrataciones públicas, *blockchain*, corrupción, entorno legal, iniciativas.

Resumo: A contratação pública representa parte relevante do gasto dos países. No Brasil, pode alcançar ao redor de 12% do Produto Interno Bruto (PIB). Segundo a Convenção das Nações Unidas contra a Corrupção, os contratos públicos são instrumento de importância para atingir interesses coletivos, mas altamente vulneráveis a atos de corrupção e fraude, causadores de prejuízos ao desenvolvimento econômico. Recentemente, surgiram as Tecnologias de Registros Distribuídos (ou DLTs, segundo a sigla em inglês), dotadas de características que podem aportar benefícios significantes ao setor público. Entre elas, a garantia de imutabilidade e integridade de múltiplos registros e a desintermediação gerada por um mecanismo de consenso entre estes. Seu uso permite a automatização de procedimentos através dos denominados *smart contracts* em processos de licitação pública, reduzindo possíveis brechas a atos ilícitos. Assim, este trabalho examina algumas das iniciativas europeias sobre

1. Artigo submetido em 08/07/2022 e aceito em 01/12/2022.

essa tecnologia, o entorno jurídico brasileiro e, ao final, propõe uma reflexão sobre os primeiros passos para sua implementação e as interações necessárias a uma efetiva coordenação institucional para o uso da tecnologia *blockchain* pela administração pública no Brasil.

Palavras-chave: contratações públicas, *blockchain*, corrupção, entorno legal, iniciativas.

Abstract: Public procurement represents a relevant amount of public expenses in every country. In Brazil, they might reach around 12% of its Gross Domestic Product (GDP). According to the United Nations Convention against Corruption, public contracts are important tools to accomplish collective interests, although highly vulnerable to corrupt acts and fraud, which harm economic development. Recently, Distributed Ledger Technologies (DLTs) have emerged, endowed with features capable of introducing significant benefits to the public sector. Among them the guarantee of immutability and integrity of multiple records and the disintermediation generated by a consensus mechanism. Its usage allows for automatizing processes through the denominated “*smart contracts*” in public tender procedures, reducing possible breaches to illicit acts. That being said, this paper examines a few European initiatives on such technology, the Brazilian legal framework, and, in the end, proposes a reflection on the first steps towards its implementation and the interactions necessary for effective institutional coordination in the application of *blockchain* by the public administration in Brazil.

Keywords: public procurement, *blockchain*, corruption, legal framework, initiatives.

1. INTRODUCCIÓN

Últimamente, el uso de las tecnologías de registro distribuido (TRD), entre las cuales la *blockchain* (o cadena de bloques) es la más conocida, ha sido probado por algunos países en diversas áreas y discutido dentro de la Unión Europea como una herramienta que quizá represente la fase más transformadora de la era de la *internet*, con potencial de mermar la corrupción y el fraude, así como crear confianza para el intercambio de valores.

Su celebrado potencial disruptivo se produce principalmente por las características de descentralización de registros, trazabilidad, confianza y transparencia. Todo esto sumado a una de sus aplicaciones específicas, los *smart contracts*, que son capaces de automatizar procesos y prescindir de gran parte de la intervención humana, tiene grandes posibilidades de contribuir con la lucha contra la corrupción, especialmente en el ámbito de la contratación pública.

Sin embargo, antes de su implementación, un mapeo sobre iniciativas internacionales potencialmente útiles y una descripción del entorno jurídico brasileño parecen inolvidables. Por tanto, el presente artículo ha utilizado el método deductivo a través de una investigación bibliográfica, con datos e informaciones secundarias colectadas en libros, artículos y en línea. La propuesta es examinar el contexto europeo de estudios e implementación de la *blockchain*, lo que,

a pesar de sugerir una posible limitación a este contexto geográfico, permitió analizar brevemente el actual cuadro legal brasileño relacionado y reflexionar sobre las instituciones a que les tocaría participar y coordinar iniciativas a ese respecto en Brasil.

2. BLOCKCHAIN EN EL SECTOR PÚBLICO – INICIATIVAS EUROPEAS

Uno de los grandes retos para el despliegue de la tecnología *blockchain* es, sin duda, superar las fronteras entre países, conectándolos entre sí para que de hecho su potencial sea realmente explotado. Respecto a su implantación en el continente europeo, Muñoz Carmona (2018) señala que:

A pesar de los esfuerzos que ello supone, resulta indispensable dotar a la cadena de bloques de un carácter transfronterizo por cuanto ello supondrá el fomento de la participación de posibles interesados que sean ciudadanos de otros estados miembros comunitarios en procedimientos administrativos incoados por Administraciones públicas nacionales, aprovechando así las bondades de la tecnología *blockchain* como herramienta para dotar de transparencia a los referidos procedimientos y conseguir la simplificación de la tramitación de los expedientes, fomentando el respeto a los principios que constituyen

la clave de bóveda del funcionamiento de la Unión Europea, como son la libre circulación de personas y la libertad de establecimiento. (Muñoz Carmona, 2018. p. 49)

Por ello, cabe subrayar la importancia de tenerse el comando del desarrollo de estas iniciativas en manos del gobierno sin olvidar de ponerlas en la pauta de discusión y legitimación de los parlamentos. En Europa, por ejemplo, los principales esfuerzos e iniciativas en este sentido son del Parlamento Europeo y de órganos creados y gestionados por el bloque de países como forma de garantizar que estos sistemas sean interoperables y que se puedan discutir y decidir cuestiones sensibles como regulación y gobernanza de la red, protección de datos personales y admisibilidad de registros como fuente de prueba judicial.

En Brasil ya hay iniciativas, aunque quizá aisladas, de uso fructífero de *blockchain* en el ámbito de la administración pública. Pero de alguna manera le parece faltar, como se verá adelante, una coordinación nacional y la existencia de un fórum de discusiones en que el gobierno, parlamento e instituciones de representación civil estén presentes.

A modo de ejemplo, algunas iniciativas actuales en Europa pueden ser valiosas para los objetivos de este trabajo.

2.1. Informe del Parlamento Europeo (2017)

En febrero de 2017, el Servicio de Estudios del Parlamento Europeo elaboró un informe titulado “Cómo puede cambiar nuestra vida la tecnología de la cadena de bloques” para servir de referencia a los diputados y miembros del Parlamento Europeo, como forma de estimular el debate y el interés por esta tecnología. Este estudio analiza ocho ámbitos en los que la cadena de bloques tiene una posible repercusión sustancial. En cada uno de ellos se explica cómo se podría desarrollar la tecnología, los posibles efectos que podría tener este desarrollo y cuáles son las posibles cuestiones que se debe anticipar en materia de políticas.²

2.2. Observatorio y foro de la cadena de bloques (2018)

En febrero de 2018, la Comisión Europea lanzó el Observatorio y Foro *Blockchain* de la Unión Europea con el objetivo de acelerar la innovación y el desarrollo de la cadena de bloques dentro de la UE, con un claro propósito de hacer consolidar la posición de Europa como líder mundial en esta nueva tecnología.

Su trabajo está destinado a supervisar las iniciativas de *blockchain* en Europa, producir una fuente completa de conocimientos sobre la temática, crear un foro atractivo y transparente para compartir información y opiniones y hacer recomendaciones sobre el papel que puede desempeñar la UE en su uso.³

2.3. Resolución sobre las tecnologías de registros distribuidos del Parlamento Europeo (2018)

En octubre de 2018 el Parlamento Europeo aprobó la Resolución sobre las tecnologías de registros distribuidos y las cadenas de bloques para “fomentar la confianza con la desintermediación”. Entre las ventajas que se ha destacado sobre las TRD, están la posibilidad de proporcionar un marco de transparencia, reducir la corrupción, detectar la evasión fiscal, facilitar las políticas de prevención del blanqueo de capitales, detectar la malversación de activos, garantizar la integridad de los datos y la capacidad de proporcionar exámenes de auditoría a prueba de manipulaciones. Además de estas, se enumera una larga lista de ventajas respecto a la mejora de eficiencia de los costes de transacción con la eliminación de intermediarios, mecanismos de encriptación y control que pueden mejorar el grado de confianza en las transacciones y la posibilidad de que los ciudadanos controlen sus propios datos, decidiendo con quién compartirlos. Además, deja claro que las TRD fomentan la seudonimización de usuarios y no su anonimización.

Al mismo tiempo, dicha Resolución solicita a la Comisión Europea que evalúe y desarrolle un marco jurídico europeo para solventar posibles problemas jurisdiccionales que puedan plantearse en caso de fraude o delitos en el intercambio de TRD.⁴ Así, cuando el Parlamento Europeo hace esa solicitud a la Comi-

2. (Boucher, Nascimento e Kritikos, 2017) Para saber más, acceda http://publications.europa.eu/resource/cellar/9964fbfd-6141-11e7-8dci-01aa75ed71a1.0005_01/DOC_1, última consulta en 09/05/2022

3. Para más informaciones acceder a <https://www.eublockchainforum.eu/>, última consulta en 09/05/2022

4. Véase la Resolución en la íntegra en https://www.europarl.europa.eu/doceo/document/TA-8-2018-0373_ES.html, última consulta en 09/05/2022

sión Europea lo que ambiciona no es que se regulen nuevos fenómenos, sino que se eliminen barreras a su desarrollo (Cordero Valdavidia, 2019).

2.4. Asociación Europea de Blockchain (EBP) e Infraestructura Europea de Servicios de Blockchain (EBSI) (2018)

En abril de 2018, 27 Estados miembros (entre ellos España), Liechtenstein y Noruega firmaron una declaración para crear la Asociación Europea de Blockchain (en inglés *European Blockchain Partnership* - EBP).⁵ El objetivo del grupo es desarrollar una estrategia de la UE para construir una infraestructura de blockchain y mejorar los servicios públicos, el EBSI (sigla del inglés *European Blockchain Services Infrastructure*).⁶

La red EBSI deberá hacer realidad el intercambio de datos entre países de una manera sencilla y supondrá un mejor acceso a los servicios transeuropeos. EBSI ha sido diseñada bajo cinco principios: pública y permissionada, escalable, abierta, sostenible e interoperable.⁷

En relación con las Administraciones públicas, el EBSI las permite protegerse contra el fraude, aumentar la confianza y la seguridad y hacer que la verificación de la autenticidad de los datos sea sencilla y más eficiente.

2.5. Dictamen del Comité Económico y Social Europeo (CESE) sobre «La tecnología de cadena de bloques y el mercado único de la UE: ¿hacia dónde vamos?» (Dictamen de iniciativa, 2020)

Más recientemente, el 13 de febrero de 2020, se publicó en el Diario Oficial de la Unión Europea un dictamen (aprobado en plenaria por el CESE en octubre de 2019) en el que se pone de manifiesto el poder transformador de las TRD en diversos sectores de la sociedad y se destaca la confianza, la transparencia, la democracia y la seguridad como valores que permite esta tecnología. Añade que la propuesta de valor social

de blockchain está relacionada con la identidad auto soberana, la inmutabilidad y el concepto de ausencia de intermediario.

Además, aclara la diferenciación entre los beneficios que la TRD puede generar en los ámbitos público y privado y los riesgos relacionados con las criptomonedas (que también hacen uso de blockchain). Por lo tanto, el Dictamen sugiere que el tema de las criptomonedas debe tratarse por separado debido al riesgo de ser utilizado para el lavado de dinero y la evasión fiscal.

Además, este documento insta a la Comisión Europea a poner en marcha una iniciativa global de blockchain, y a definir un enfoque y una visión comunes de la UE basados en los ODS (Objetivos de Desarrollo Sostenible)⁸, reforzando la asociación con el Observatorio y el Foro de la Tecnología Blockchain. Para ello, pide que se aúnen los esfuerzos de los representantes de las instituciones de la UE, en particular el CESE y el Comité de las Regiones, la industria, los consumidores, la sociedad civil, los Estados miembros y el mundo académico, entre otros, para la creación de una plataforma de partes interesadas en la cadena de bloques de la UE.

Sin embargo, el dictamen deja claro que garantizar la seguridad jurídica y la protección de la intimidad son retos que deben superarse y, al final, advierte que es necesario actuar de inmediato si la UE no quiere perder la oportunidad única que tiene de mantener su liderazgo en materia de TRD en el mercado mundial.

3. ALGUNOS MARCOS NORMATIVOS ACTUALES PARA APLICACIÓN DE BLOCKCHAIN EN CONTRATACIONES PÚBLICAS EN BRASIL

Dentro de la administración pública, especialmente el Gobierno Federal brasileño, que suele tener un grado más avanzado de modernización en comparación con otros entes políticos (Estados, Distrito Federal y Municipios), se puede decir que hay un ambiente favorable para la implementación de proyectos

5. https://administracionelectronica.gob.es/pae/Home/pae_Actualidad/pae_Noticias/Anio2022/Abril/Noticia-2022-04-08-EBSI-la-infraestructura-europea-de-blockchain-en-marcha.html, última consulta en 09/05/2022

6. <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/ebsi>

7. Op. Cit. nota 4

8. "En 2015, la ONU aprobó la Agenda 2030 sobre el Desarrollo Sostenible, una oportunidad para que los países y sus sociedades emprendan un nuevo camino con el que mejorar la vida de todos, sin dejar a nadie atrás. La Agenda cuenta con 17 Objetivos de Desarrollo Sostenible, que incluyen desde la eliminación de la pobreza hasta el combate al cambio climático, la educación, la igualdad de la mujer, la defensa del medio ambiente o el diseño de nuestras ciudades." <https://www.un.org/sustainabledevelopment/es/>, última visita en 19/05/2022

que promueven el uso de nuevas tecnologías, especialmente en relación con la Política Nacional de Gobierno Abierto, establecida por el Decreto nº 10.160/2019.⁹

Según información facilitada por la Contraloría General de la Unión de Brasil (CGU), gobierno abierto se describe como una nueva visión de la Administración Pública con el fin de promover acciones y proyectos basados en los principios de transparencia, rendición de cuentas, participación ciudadana; y tecnología e innovación.

El primero se refiere al mantenimiento de información abierta, comprensible, de libre acceso y oportuna sobre las actividades gubernamentales, según estándares básicos de datos abiertos, mientras el segundo se relaciona a normas y mecanismos establecidos para que los actores públicos justifiquen sus acciones y cómo aceptan o deben aceptar las responsabilidades que se les asignan.

Referente a la participación ciudadana, se puede decir que las acciones dirigidas para su promoción deben buscar que la sociedad debata, colabore y haga propuestas que contribuyan a un gobierno más eficaz y receptivo. Y, de hecho, su integración a un contexto de tecnología e innovación permitirá que se pongan en práctica de forma efectiva todos los demás principios, por lo que el gobierno reconoce la importancia de facilitar a los ciudadanos el acceso a las nuevas tecnologías.¹⁰

3.1. Rasgos de La Política Nacional de Gobierno Abierto

Vistos los lineamientos establecidos por el Decreto, se puede afirmar que todos están relacionados con los beneficios que pueden surgir de la implementación de la tecnología *blockchain* en la contratación pública. Según su artículo 2º:

Art. 2º Son directrices de la Política Nacional de Gobierno Abierto:

I - mayor **disponibilidad de información sobre las actividades del gobierno**, incluidos datos sobre el gasto y el desempeño de las acciones y programas del gobierno federal;

II - fomentar la **participación social en los procesos de toma de decisiones**;

III - fomentar **el uso** de nuevas tecnologías que fomenten la innovación, **fortaleciendo la gobernanza pública** y aumentando **la transparencia y la participación social en la gestión y prestación de servicios públicos**;

IV - Aumento de los procesos de **transparencia**, acceso a **la información** y uso **de tecnologías que apoyen estos procesos**. (grifos de los autores de esta obra)¹¹

Ya su artículo 3º presenta un rol siete iniciativas/ acciones que deben ser instituidas y, así como ocurre con las directrices, la tecnología *blockchain* tiene el potencial de contribuir con todas. De hecho, transparencia y acceso a la información se beneficiarían de la lógica distribuida y descentralizada de la información, así como la gobernanza pública, en función del diseño de las redes, podría contar con mejores gestión, seguimiento y relación entre actores.

En añadidura, la trazabilidad de registros, más allá de potenciar la automatización de procesos, compras y auditoría es factor indisoluble de la prevención y lucha contra la corrupción, pues contribuye con la integridad y proporciona mejora de la eficiencia y de la prestación de servicios públicos, ya que la automatización de los procesos aportará mayores garantías de que, por ejemplo, las adquisiciones se realicen por valores adecuados.

Por lo tanto, se puede decir que en relación con la Política Nacional de Gobierno Abierto el camino está allanado para la implementación de las TRD como herramienta con gran potencial para contribuir al logro de los objetivos propuestos por el Decreto 10.160/2019.

9. BRASIL. Decreto nº 10.160, de 9 de diciembre de 2019. Establece la Política Nacional de Gobierno Abierto y la Comisión Interministerial de Gobierno Abierto. http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D10160.htm, última visita en 15/05/2022

10. Definiciones disponibles en <https://www.gov.br/cgu/pt-br/governo-aberto/governo-aberto-no-brasil/o-que-e-governo-aberto>, última visita en 24/06/2022

11. Original em português: BRASIL. Decreto nº 10.160, de 9 de dezembro de 2019. Institui a Política Nacional de Governo Aberto e o Comitê Interministerial de Governo Aberto. http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D10160.htm, última visita en 15/05/2022/

(Art. 2º São diretrizes da Política Nacional de Governo Aberto:

I - aumento da disponibilidade de informações sobre as atividades governamentais, incluídos os dados sobre os gastos e o desempenho das ações e dos programas do Governo federal;

II - fomento à participação social nos processos decisórios;

III - estímulo ao uso de novas tecnologias que fomentem a inovação, o fortalecimento da governança pública e o aumento da transparência e da participação social na gestão e na prestação de serviços públicos; e

IV - aumento dos processos de transparência, de acesso a informação e da utilização de tecnologias que subsidiem esses processos. (grifos do autor deste trabalho)

Además, Brasil es uno de los cofundadores de OGP (*Open Government Partnership*), una parceria internacional creada en 2011 con el objetivo de fomentar prácticas que contribuyan a la transparencia, la participación ciudadana, la rendición de cuentas y la innovación. El grupo cuenta actualmente con la participación de más de 77 países y 76 gobiernos subnacionales.¹²

Otra prueba de que Brasil ha participado activamente en los foros internacionales sobre innovación es el hecho de que también es miembro de la Red de Gobierno Abierto e Innovador de la OCDE en América Latina y el Caribe. Esta iniciativa ha sido creada en 2015 con el objetivo de promover diálogo e intercambio de experiencias sobre gobierno abierto, innovación en el sector público y gobierno digital. Brasil, junto con Colombia, copreside esta red, que trabaja para compartir las mejores prácticas y para afrontar los retos regionales en este ámbito.¹³

Sobre el grado de digitalización del sector público brasileño, antes de abordar su estado actual es importante señalar que hay dos formas de interpretar el gobierno electrónico:

La e-Administración o Administración Electrónica es la incorporación de las tecnologías de la información en la administración pública. Tiene dos vertientes. Desde un punto de vista intra-organizativo, implica transformar las oficinas tradicionales, convirtiendo los procesos en papel en procesos electrónicos, con el fin de crear una oficina sin papeles. Desde una perspectiva de las relaciones externas, implica habilitar la vía electrónica como un nuevo medio para la relación con la ciudadanía y empresas. Interesa resaltar que es una herramienta con un elevado potencial de mejora de la productividad y simplificación de los diferentes procesos del día a día. (Cordero Valdavidia, 2019, p. 19)

Así, en relación con el nivel de digitalización en el cual se encuentran los órganos que conforman el Estado brasileño, el país se encuentra en una fase sig-

nificativamente avanzada, por lo que este es más un factor positivo relacionado con el entorno ambiental para el uso de la tecnología *blockchain* en la contratación pública. Y hay varias razones que comprueban esta afirmación.

3.2. La Ley de Gobierno Digital

Recientemente, en agosto de 2021, se creó la Ley 14.129/2021 (Ley de Gobierno Digital)¹⁴, que define la adopción de soluciones digitales tanto para la gestión y tramitación de procesos administrativos internos (Arts. 5º a 13), como para la prestación de servicios públicos al ciudadano (Arts. 14 a 16).

Asimismo, la ley aborda la necesidad de interoperabilidad entre plataformas (Arts. 38 a 41), el tratamiento adecuado de datos personales (Art. 25), destaca que dichos datos deben ser fácilmente accesibles por los ciudadanos y no olvida el respeto a las garantías previstas en la Ley 13.709, de 14 de agosto de 2018 (Ley General de Protección de Datos Personales).

De hecho, la Ley de Gobierno Digital trae acertadamente transparencia y publicidad activa como reglas y el secreto como excepción (art. 29). También prevé divulgación obligatoria de facturas electrónicas relacionadas con compras públicas. Sobre este aspecto, vale destacar que, aunque la tradición latina considera la información como patrimonio del poder y de los que lo desempeñan, el artículo citado parece pretender poner el Brasil en pie de igualdad a la tradición de los países escandinavos, que priman por el principio de que todos los documentos son públicos a no ser que la ley se los declare secretos. (Arenilla Sáez, 2011, p.355)

De momento, para las aplicaciones de *blockchain* en la contratación pública, faltaría que esta ley permitiera el acceso y uso por parte de los organismos públicos que vigilan el fraude y la corrupción, como es el caso de la CGU, de todo el universo de facturas de las

12. Información disponible en: <https://www.opengovpartnership.org/es/our-members/>; y <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/o-que-e-a-iniciativa>, accesos en 17/05/2022

13. Disponible en <https://www.gov.br/cgu/pt-br/governo-aberto/noticias/2022/02/iv-encontro-da-rede-da-ocde-sobre-governo-aberto-e-inovador-na-america-latina-e-caribe>, último acceso en 17/05/2022

14. BRASIL. Ley nº 14.129, de 29 de marzo de 2021. Dispone sobre los principios, normas e instrumentos para el Gobierno Digital y para el aumento de la eficiencia pública(...) http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/L14129.htm, última visita en 15/05/2022

empresas que participan en las licitaciones públicas y no sólo de aquellas en las que la Administración figura como compradora.¹⁵

Por último, aunque no aborda específicamente el uso de TRD, la citada ley permite utilizar pruebas de concepto o implementar laboratorios de innovación y experimentación de nuevas tecnologías abiertas y libres, lo que permite la creación inmediata de un *sandbox*¹⁶ liderado por el gobierno federal para aplicar *blockchain* a la contratación pública (traducción libre al castellano):

(...) Art. 44. Las entidades públicas podrán instituir laboratorios de innovación, abiertos a la participación y colaboración de la sociedad para el desarrollo y experimentación de conceptos, herramientas y métodos innovadores para la gestión pública, la prestación de servicios públicos, el tratamiento de datos producidos por las autoridades públicas y la participación de los ciudadanos en el control de la administración pública.

(...)¹⁷

3.3. La Ley 14.133/2021 y los contratos electrónicos

De acuerdo con las disposiciones de la Ley 14.133/2021, la regla general para los contratos públicos es su forma preferentemente electrónica, pudiendo la

Administración determinar que los licitadores realicen también sus actos en mismo formato como condición de su validez y eficacia. In litteris:

Art. 12. En el proceso de licitación, se observará lo siguiente: (...)

VI - los actos serán preferentemente digitales, de forma que permitan ser producidos, comunicados, almacenados y validados por medios electrónicos. (...)

Art. 17 (...)

§ 2º Las licitaciones públicas se realizarán preferentemente de forma electrónica, permitiéndose el uso de la asistencia física, siempre que se justifique, levantando acta de la sesión pública y grabando en audio y vídeo. (...)

§ 4º En los procedimientos realizados por medios electrónicos, la Administración podrá determinar, como condición de validez y eficacia, que los licitadores realicen sus actos en formato electrónico.¹⁸

Por lo que se observa, este aspecto es otro factor positivo para la implementación de *blockchain* en la contratación pública. Sin embargo, es importante fijarse que actualmente los procesos de licitación ya son por regla general electrónicos en Brasil, por lo menos a nivel federal. Además, hay un nivel razonablemente satisfactorio de transparencia en los portales en línea de contratación pública. El sistema Compras.

15. El artículo 23 de la Ley de licitación pública establece que el valor estimado de la contratación debe ser compatible con los practicados en el mercado y con los que figuren en las bases de datos públicas que contengan información sobre los precios presentados en las administraciones públicas o en las licitaciones públicas (si existe integración). Sin embargo, más adelante, en el § 1º del mismo art. 23, la ley establece que la estimación también debe tener en cuenta un estudio que se realizará sobre la base de datos nacional de facturas electrónicas, en forma de reglamento. De todos modos, aún no existe regulación y, por tanto, no está claro qué facturas puede utilizar la Administración, es decir, si solo aquellas en que es compradora o todo el universo de facturas a que tiene acceso por su poder y competencia recaudatoria. Aquí la opinión es que, para este caso, la normativa a editar debería incluir la posibilidad de que la Administración pueda acceder a todo el universo de facturas, anonimizando y suprimiendo algunos datos de la factura para garantizar el derecho a la intimidad en los negocios

16. "Un *Sandbox Regulatorio* se identifica como un banco de pruebas en el que la autoridad supervisora permitirá probar proyectos innovadores de base tecnológica en el ámbito financiero, con usuarios reales, bajo el control y supervisión de la autoridad competente. (...) En términos de regulación y supervisión, supone una de las *mayores innovaciones realizadas hasta la fecha por parte de las autoridades supervisoras*, ya que el *objetivo último del Sandbox será apoyar el desarrollo de proyectos innovadores bajo un marco legal y seguro*, que por la novedad de su modelo negocio, no ven cabida en el marco regulatorio actual." Definición extraída de <https://sandboxspain.com/> en 15/05/2022

17. Original em português: BRASIL. DECRETO Nº 14.129, DE 29 DE MARÇO DE 2021. Dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública(...) http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/L14129.htm, última visita en 15/05/2022
(...) Art. 44. Os entes públicos poderão instituir laboratórios de inovação, abertos à participação e à colaboração da sociedade para o desenvolvimento e a experimentação de conceitos, de ferramentas e de métodos inovadores para a gestão pública, a prestação de serviços públicos, o tratamento de dados produzidos pelo poder público e a participação do cidadão no controle da administração pública

18. Original em português: BRASIL. Decreto nº 14.129, de 29 de março de 2021. Dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública(...) http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/L14129.htm, última visita en 15/05/2022

Art. 12. No processo licitatório, observar-se-á o seguinte: (...)

VI - os atos serão preferencialmente digitais, de forma a permitir que sejam produzidos, comunicados, armazenados e validados por meio eletrônico. (...)

Art. 17 (...)

§ 2º As licitações serão realizadas preferencialmente sob a forma eletrônica, admitida a utilização da forma presencial, desde que motivada, devendo a sessão pública ser registrada em ata e gravada em áudio e vídeo. (...)

§ 4º Nos procedimentos realizados por meio eletrônico, a Administração poderá determinar, como condição de validade e eficácia, que os licitantes pratiquem seus atos em formato eletrônico. (...)

gov.br, antes llamado Comprasnet, por ejemplo, ofrece a los ciudadanos la posibilidad de visualizar todas las etapas del proceso y las interacciones llevadas a cabo tanto por los licitadores como por el contratista, este último representado por la figura del subastador.

Lo que va a permitir el uso de *blockchain* en la contratación pública es la mejora de la digitalización, la transparencia y la integridad de los procesos de licitación y, por lo que se ve, la normativa actual está redactada satisfactoriamente para su aplicación, aunque hay que revisarla en algún punto, lo que es absolutamente normal y esperable cuando se trata de una nueva tecnología. Por lo tanto, podemos afirmar que en Brasil los vientos normativos soplan a favor de la aplicación de las TRD en la contratación pública.

4. LA DECISIÓN N° 1.613/2020 DEL TRIBUNAL DE CUENTAS (TCU) Y LA RED BLOCKCHAIN BRASIL (RBB). ¿A QUIÉN CABE LA COORDINACIÓN DE LA INFRAESTRUCTURA NACIONAL?

Según la guía de conceptos y usos potenciales de *blockchain* en el sector público¹⁹, concebida en el seno de la Estrategia Nacional de Lucha contra la Corrupción y el Lavado de Activos (ENCCLA, sigla en portugués), es posible afirmar que, a pesar de haber órganos y entidades que han incorporado la tecnología a algunas de sus rutinas, todavía no parece existir uso consolidado y selectivo a gran escala alineado a los objetos de dicha estrategia.

Además, tampoco se detecta una participación masiva del gobierno brasileño en proyectos internacionales destinados a integrar o promover la infraestructura transfronteriza de *blockchain*, aunque se debe reconocer las discusiones del Grupo de Trabajo creado en la estructura del Comité Central para la Gobernanza de Datos, destinado a tratar de propuestas relacionadas a la tecnología presentada en este artículo²⁰, y la creación de la propia RBB, mencionada en el título de este apartado. Es decir, lo que parece existir son iniciativas aisladas, intentos de algunos órganos o empresas, pero no se puede olvidar que todavía es necesaria la creación de una red nacional adecuada-

mente coordinada y que esté preparada para cambiar datos, informaciones y conocimientos con otras redes mundiales.

No obstante, es relevante señalar que en el año 2020 el Tribunal de Cuentas de la Unión (TCU - Brasil) ha realizado un extenso trabajo para identificar áreas de aplicación de las TRD en el sector público. Este trabajo pretendió evaluar los principales riesgos y factores críticos de éxito, así como los retos para el control gubernamental en aquel momento.

A lo largo del proceso, el ministro (ponente) Aroldo Cedraz, comentó que “la característica descentralizadora de las tecnologías *blockchain* y TRD podría acelerar la transformación digital del Estado, ya que la posibilidad de realizar transacciones autenticadas sin necesidad de una autoridad central facilita la implementación de servicios públicos digitales orientados al ciudadano”.²¹

Mediante la Decisión n° 1.613/2020 - PLENARIO los ministros allí reunidos decidieron entre otros puntos (traducción libre al español):

9.2. remitir el Informe a la Contraloría General de la Unión (CGU) y a la Asociación de Integrantes de los Tribunales de Cuentas de Brasil (Atricon) para que evalúen la conveniencia y oportunidad de adoptar medidas en sus áreas de actuación, teniendo en cuenta los aspectos de la nueva tecnología que pueden afectar a las actividades de fiscalización.²²

Además, dicha Decisión determinó que:

(...) a la Secretaría Especial de Desburocratización, Gestión y Gobierno Digital (SEDGG) del Ministerio de Economía, a la Secretaría de Coordinación y Gobierno de Empresas Estatales del Ministerio de Economía, al Consejo Nacional de Justicia y al Consejo Nacional del Ministerio Público que, si identifican organismos y entidades bajo su supervisión que consideren el uso de la tecnología *Blockchain/ DLT*.²³

19. Disponible en <http://enccla.camara.leg.br/acoefs/arquivos/resultados-enccla-2020/blockchain-no-setor-publico-guia-de-conceitos-e-usos-potenciais/view>, última consulta en 01/06/2022.

20. Conforme lo que dispone el Objetivo 8 del Anexo al Decreto n° 10.332/2020, que instituyó la Estrategia de Gobierno Digital en Brasil, respecto a *blockchain*

21. <https://portal.tcu.gov.br/imprensa/noticias/tcu-avalia-tecnologias-da-informacao-blockchain-e-livros-razao-distribuidos-para-o-setor-publico.htm>, última visita en 09/05/2022

22. <https://pesquisa.apps.tcu.gov.br/#/redireciona/acordao-completo/%22ACORDAO-COMPLETO-2406748%22>, última consulta 09/05/2022

23. *ibidem*

Sin embargo, en mayo de 2022 el propio TCU tomó la delantera firmando, en asociación con el BNDES (Banco Nacional de Desarrollo Económico y Social) un acuerdo de cooperación técnica para la creación de la Red *Blockchain* Brasil (RBB) y, más allá, nombraron a sí mismos patronos de la iniciativa, haciéndose dotados de poderes exclusivos (incluyendo el veto de propuestas). De esta manera, parecen haber dejado las demás organizaciones gubernamentales simplemente como posibles adherentes a la red. En otras palabras, todos aquellos organismos del gobierno federal a los que el propio TCU, a través de la Decisión nº 1613/2020, había asignado la tarea de adoptar medidas y considerar el uso de la tecnología *blockchain*.

Por lo tanto, es necesario reflexionar si los dos signatarios originales de dicho acuerdo deben poseer el rol de protagonistas y responsables por decidir sobre la infraestructura de lo que se pretende establecer como la principal red de *blockchain* utilizada por entes gubernamentales. De hecho, un análisis a las atribuciones que son propias a cada una las instituciones eventualmente involucradas en su uso debe ser un punto central de discusión. No por cuestiones de jerarquía, sino de coordinación adecuada.

Sin dudas la iniciativa es válida y útil. De otra mano, debido a la necesidad de implementar una red transversal que tiene como objetivo promover la confianza en el poder estatal de forma amplia, y porque involucra todo el funcionamiento de los participantes e intereses de la ciudadanía (como la protección de datos), la coordinación de esta iniciativa no debería excluir la participación activa del gobierno federal o mismo limitarles poderes en la red. En añadidura, es necesario señalar la necesidad de debate entre los miembros del parlamento y de manifestación de órganos como la Comisión Interministerial de Gobernanza (CIG), compuesta por el Ministro de Estado Jefe de Gabinete de la Presidencia de la República, el de la Economía y el de la Contraloría General de la Unión, de acuerdo con el Decreto nº 9.203, de 22 de noviembre de 2017.²⁴

Además, la CGU es responsable por tomar las medidas necesarias “para proteger el patrimonio público, el control interno, las auditorías públicas, las correcciones, la prevención y la lucha contra la corrupción, las actividades de ombudsman y el aumento de la transparencia de la gestión en la administración pública federal”²⁵ y, al mismo tiempo, coordinar la Comisión Interministerial de Gobierno Abierto, encargada de la “articulación intergubernamental necesaria para la ejecución de acciones conjuntas, el intercambio de experiencias, la transferencia de tecnología y el desarrollo de capacidades, en el ámbito de los planes de acción nacionales sobre gobierno abierto.”²⁶

Por su turno, un objetivo del acuerdo mencionado es cooperar para crear la Red *Blockchain* Brasil – RBB, y permitir la construcción de procesos públicos que se muestren fiables y transparentes, promoviendo la confianza de los ciudadanos en el poder público. Para ello, define que un número mínimo de socios públicos institucionalmente independientes deben formar parte de la red técnica, como forma de garantizar la total confianza en las operaciones ejecutadas. Sin embargo, el instrumento aún no define cuál sería este número mínimo de participantes, tampoco cuáles serían dichos socios. Es probable que, por precaución, se haya dejado esta decisión a la formulación del reglamento de la red.

El acuerdo también pone de manifiesto la preocupación de los patrocinadores por la gobernanza, aunque no se haya consultado a la CIG, al calificarla como elemento de máxima importancia para viabilizar la red, lo que parece adecuado a priori, teniéndose en cuenta que esta primera iniciativa pretende integrar a varios organismos y entidades públicas. Además, se define la gobernanza prevista como una estructura de toma de decisiones necesaria para mantener y hacer evolucionar la RBB.

Así, en relación con la estructura de la RBB, el acuerdo define que habrá tres categorías de participantes: los patrocinadores (TCU y BNDES); los asociados adherentes; y los adherentes.

24. BRASIL. Decreto nº 9.203, de 22 de noviembre de 2017. Prevé la política de gobierno de la administración pública federal directa, autárquica y fundacional. Diário Oficial da União, Brasília, 2017a. Disponible en: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/decreto/D9203.htm

25. BRASIL. Ley 13.844, de 18 de junio de 2019. Establece la organización básica de los órganos de la Presidencia de la República y de los Ministerios. Disponible en http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/L13844compilado.htm

26. BRASIL. Decreto nº 10.160, de 9 de diciembre de 2019. Establece la Política Nacional de Gobierno Abierto y la Comisión Interministerial de Gobierno Abierto. http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D10160.htm, última visita en 15/05/2022

TABLA 1 - LAS POTESTADES DE CADA TIPO DE PARTICIPANTE DE LA RBB

PARTICIPANTES PATROCINADORAS	PARTICIPANTES ASOCIADOS ADHERENTES	PARTICIPANTES ADHERENTES
Pueden: <ul style="list-style-type: none"> • presentar propuestas; • participar en las reuniones de gobernanza; • votar sobre las propuestas presentadas, incluido el voto de calidad; • vetar las propuestas presentadas. 	Pueden: <ul style="list-style-type: none"> • presentar propuestas; • participar en las reuniones de gobernanza; • votar sobre las propuestas presentadas. 	Pueden: <ul style="list-style-type: none"> • presentar propuestas; • participar en las reuniones de gobernanza.
Son, exclusivamente, participantes fundadores: <ul style="list-style-type: none"> • BNDES; • TCU. 	Deben ser entidades de la administración pública directa, indirecta y fundacional y empresas estatales de la Unión, de los Estados, del Distrito Federal y de los Municipios, órganos o entidades de los Poderes Legislativo y Judicial, o entidades privadas nacionales sin fines de lucro, cuya naturaleza sea considerada de interés público, según la evaluación de Gobernanza de la RBB.	Deben ser entidades de la administración pública directa, indirecta y fundacional y empresas estatales de la Unión, de los Estados, del Distrito Federal y de los Municipios, órganos o entidades de los Poderes Legislativo y Judicial, o entidades privadas nacionales sin fines de lucro, cuya naturaleza sea considerada de interés público, según la evaluación de Gobernanza de la RBB.
Son necesariamente validadores de la RBB.	Pueden ser validadores de la RBB.	no pueden ser validadores de la RBB
Pueden serlo: <ul style="list-style-type: none"> • Registradores; y • Usuarios de la RBB; 	También pueden serlo: <ul style="list-style-type: none"> • Registradores; y • Usuarios de RBB. 	Tienen la opción de ser: <ul style="list-style-type: none"> • Registradores; y • Usuarios de RBB.

Fuente: Producción propia. Datos accesibles en <https://github.com/RBB-Net/rbb/blob/master/documentos/ACT_TCU_BNDES_RBB.pdf>

Aunque la iniciativa contiene la peculiaridad de poseer un Tribunal de Cuentas y una empresa pública (es decir, entidad privada) como patrocinadores, no se olvide que su manejo, difusión y perfeccionamiento debe conllevar el desempeño irrestricto de funciones propias de otros órganos y entidades, pertenecientes a otros Poderes independientes y constitucionalmente instituidos. De esta manera, no parece haber razones (o misma posibilidad jurídica) que estos se queden supeitados, al menos en el plano del ejercicio de atribuciones, al arbitrio de otros, con facultades que no les son extensivas.

Sin embargo, se reconoce que el acuerdo, de manera correcta, notifica a todos los posibles participantes que deben comprometerse a “cumplir con las normas y regulaciones aplicables a los datos personales, así como las determinaciones de los organismos reguladores y supervisores sobre el tema”, debido a su implementación²⁷, “incluyendo la Ley N° 13.709 del 14 de agosto de 2018 - Ley de Protección de Datos Personales (LGPD).”

CONSIDERACIONES FINALES

A modo de conclusión, merece la pena señalar que el aislamiento interno o internacional no es la mejor estrategia, ya que se podría crear un sinfín de estructuras *blockchain*, casi seguramente de manera desordenada. La configuración de la cadena de bloques en el ámbito de los gobiernos debe ser planeada en detalles suficientes a la cooperación, visándose a acciones concertadas y coordinada sin sustraer las competencias asignadas por ley al gobierno federal. De hecho, el Estado debe establecer estándares para que las redes existentes en la circunscripción nacional sean compatibles entre sí y que necesariamente los sistemas y la gobernanza establecidas para uso dentro del país permitan integración a los sistemas y reglas adoptadas por la comunidad internacional. De nada serviría a la integridad de los actos administrativos tener cadenas de bloques que no se comunican unas con las otras porque los intereses y la necesidad de protección, incluso contra ilícitos de corrupción, son complejos y ul-

27. Disponible en https://github.com/RBBNet/rbb/blob/master/documentos/ACT_TCU_BNDES_RBB.pdf. Última consulta en 19/05/2022

trapasen las fronteras. Asimismo, las redes deben ser construidas sin muros y con vías que permitan el rastreo de cualquier transacción, sea donde sea.

Definitivamente, no parecen ser adecuadas prerrogativas exclusivas a un número limitado de organizaciones, lo que puede traer consecuencias potencialmente incompatibles con el ejercicio pleno de competencias ajenas, en la implementación de un sistema tan complejo y con múltiples (y todavía razonablemente desconocidas) implicaciones jurídicas, como las TRD. Por ser una cuestión de naturaleza técnica y jurídica, también implica derechos fundamentales, por lo que el gobierno brasileño necesita ganar las riendas que le son propias y empezar cuanto antes las negociaciones internacionales y discusiones internas sobre qué sistemas adoptar y bajo qué reglas, una vez que el Estado brasileño posee un entorno jurídico suficientemente favorable a la implantación de esta novedosa tecnología.

Como es importante que en el futuro todas las *blockchains* sean compatibles entre sí, parece que Brasil debe mantener conversaciones con todos los países (y bloques) sin excepción que puedan contribuir con iniciativas más avanzadas en términos de infraestructura, como el laboratorio LACChain²⁸ creado por el BID para América Latina y el Caribe y el EBSI en Europa.

Sin embargo, debido a que la Unión Europea ya cuenta con un grupo para el desarrollo de la infraestructura de servicios en *blockchain* y tiene vasta experiencia en relación con complejas regulaciones transfronterizas, valdría la pena que Brasil buscara establecer acuerdos tanto para estandarizar las regulaciones sobre el uso de la tecnología como para adoptar plataformas que operen en compatibilidad con las que se están desarrollando en regiones más avanzadas.

REFERENCIAS

BNDES/TCU. *Acordo de Cooperação nº D-121.2.0014.22 para a formação da Rede Blockchain Brasil (RBB)*. 2022. Disponible en <https://github.com/RBBNet/rbb/blob/master/documentos/ACT_TCU_BNDES_RBB.pdf>, acceso en 19/05/2022.

BOUCHER, Philip.; NASCIMENTO, Susana.; KRITIKOS, Mihalis. *Cómo puede cambiar nuestra vida la tecnología de la cadena de bloques: análisis en profundidad*: European Parliament. 2017.

BRASIL, CONTROLADORIA-GERAL DA UNIÃO. Evento: *IV Encontro da Rede da OCDE sobre Governo Aberto e Inovador na América Latina e Caribe*. Disponible en <<https://www.gov.br/cgu/pt-br/governo-aberto/noticias/2022/02/iv-encontro-da-rede-da-ocde-sobre-governo-aberto-e-inovador-na-america-latina-e-caribe>>, acceso en 17/05/2022.

BRASIL, CONTROLADORIA-GERAL DA UNIÃO. *O que é Governo Aberto*. Disponible en <<https://www.gov.br/cgu/pt-br/governo-aberto/governo-aberto-no-brasil/o-que-e-governo-aberto>>, acceso en 24/06/2022.

BRASIL. *Decreto nº 10.160, de 9 de diciembre de 2019*. Establece la Política Nacional de Gobierno Abierto y la Comisión Interministerial de Gobierno Abierto. Disponible en <http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D10160.htm>, acceso en 15/05/2022.

BRASIL. *Decreto nº 9.203, de 22 de noviembre de 2017*. Prevé la política de gobierno de la administración pública federal directa, autárquica y fundacional. Diário Oficial da União, Brasília, 2017a. Disponible en: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/decreto/D9203.htm>, acceso en 05/07/2022.

BRASIL. *Ley 13.844, de 18 de junio de 2019*. Establece la organización básica de los órganos de la Presidencia de la República y de los Ministerios. Disponible en <http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/L13844compilado.htm>, acceso en 05/07/2022.

28. "LACChain es una alianza global integrada por diferentes actores del entorno *blockchain*, liderada por el Laboratorio de Innovación del Grupo del Banco Interamericano de Desarrollo (BID Lab) para el desarrollo del ecosistema *blockchain* en América Latina y el Caribe." Para saber más acceda <https://www.lacchain.net/home>, última consulta en 16/05/2022

BRASIL. *Ley Nº 13.709 del 14 de agosto de 2018. Ley General de Protección de Datos Personales (LGPD)*. Disponible en <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm>, acceso en 05/07/2022.

BRASIL, TRIBUNAL DE CONTAS DA UNIÃO. *TCU avalia tecnologias da informação blockchain e livros-razão distribuídos para o setor público*. Disponible en <<https://portal.tcu.gov.br/imprensa/noticias/tcu-avalia-tecnologias-da-informacao-blockchain-e-livros-razao-distribuidos-para-o-setor-publico.htm>>, acceso en 09/05/2022.

BRASIL. TRIBUNAL DE CONTAS DA UNIÃO. *Acórdão nº1613/2020. Processo nº 031.044/2019-0. GRUPO I – CLASSE – Plenário. Relator: Aroldo Cedraz. Relatório de Levantamento. Processo AC-1613-23/20-P. Ata nº 23/2020 – Plenário. Brasília, DF, Sessão 24/6/2020 – Telepresencial*. Disponible en <<https://pesquisa.apps.tcu.gov.br/#/redireciona/acordao-completo/%22ACORDAO-COMPLETO-2406748%22>>, acceso en 09/05/2022.

CARMONA, Agustín. *Implicaciones jurídicas del uso de blockchain en la administración pública*. 2018.

EU Blockchain OBSERVATORY & FORUM. Disponible en <<https://www.eublockchainforum.eu/>>, acceso en 09/05/2022.

LACCHAIN. Página de inicio. Disponible en <<https://www.lacchain.net/home>>, acceso en 16/05/2022.

LEPORCHER, Yves.-Michel. *et al. Blockchain : de la teoría a la práctica, de la idea a la implementación*. Cornellà de Llobregat, Barcelona: Ediciones ENI, 2020.

NATARAJAN, Harish.; KRAUSE, Solvej.; GRADSTEIN, Helen. *Distributed ledger technology and blockchain*. 2017.

OPEN GOVERNEMENT PARTNERSHIP (OGP). *Página principal*. Disponible en <<https://www.opengovpartnership.org/es>>, acceso en 17/05/2022.

ORGANIZACIÓN DE LAS NACIONES UNIDAS. *Objetivos de Desarrollo Sostenible*. Disponible en <<https://www.un.org/sustainabledevelopment/es/>>, acceso en 19/05/2022.

PALAMENTO EUROPEU. *Tecnologías de registros distribuidos y las cadenas de bloques: fomentar la confianza con la desintermediación*. Resolución 2017/2772(RSP) de 3 de octubre de 2018. Disponible en <https://www.europarl.europa.eu/doceo/document/TA-8-2018-0373_ES.html>, acceso en 09/05/2022.

PORTAL DE LA ADMINISTRACIÓN ELECTRÓNICA. EBSI: *La infraestructura europea de blockchain en marcha*. Disponible en <https://administracionelectronica.gob.es/pae/Home/pae_Actualidad/pae_Noticias/Anio2022/Abril/Noticia-2022-04-08-EBSI-la-infraestructura-europea-de-blockchain-en-marcha.html>, acceso en 09/05/2022.

SÁEZ, Manuel. *Crisis y reforma de la administración pública*. Oleiros (La Coruña): Netbiblo, 2011.

SECO, Antonio. *Blockchain: Conceptos y aplicaciones potenciales en el área tributaria*. CIAT, 2017.

TAPSCOTT, Don; EUCHNER, Jim. *Blockchain and the Internet of Value: An Interview with Don Tapscott* Don Tapscott talks with Jim Euchner about *blockchain*, the Internet of value, and the next Internet revolution. *Research-Technology Management*, v. 62, n. 1, p. 12-19, 2019.

THE EUROPEAN Blockchain SERVICES INFRASTRUCTURE (EBSI). *Experience the future with the European Blockchain Services Infrastructure (EBSI)*. Disponible en <<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/ebsi>>, acceso en 05/07/2022.

VALDAVIDA, Magdalena. *Personas y Organizaciones Públicas. Revista Vasca de Gestión de Personas y Organizaciones Públicas*, Núm. 16 zk./2019. Págs. 16-34 or. 2019.



Samir Sallen Silva Santos

<https://orcid.org/0000-0002-2372-6563>

Universidad de Salamanca (USAL – Espanha)

samirsallen@gmail.com

Auditor Federal de Finanças e Controle da Controladoria-Geral da União, onde atua desde 2001. Possui formação em Direito e Administração, é advogado e Mestre em Direito com ênfase em Estratégias Anticorrupção e Políticas de Integridade pela Universidade de Salamanca - Espanha. Detém capacitação em Ações Investigativas e Inovadoras da CGU, além de conhecimentos em Auditoria baseada em riscos. Possui atuação/experiência em auditorias de combate a corrupção tanto investigativas (repressão), como também preventivas, além de ter ocupado cargo de chefia do Núcleo de Ações de Controle da Superintendência da CGU no Maranhão.



Lucio Alves Angelo Junior

<https://orcid.org/0000-0003-4268-3439>

Universidad de Salamanca (USAL – Espanha).

lucio.alves@usal.es

Advogado da União (AGU) desde 2013, é professor de Pós-graduação da Escola da Advocacia-Geral da União. Possui atuação/experiência junto aos Departamentos de Assuntos Internacionais e Patrimônio Público e Probidade (PGU/AGU), com ênfase em demandas relacionadas ao combate à corrupção, à recuperação de ativos e à cooperação internacional. Coordenou a Ação 08/2020, relativa ao uso de *blockchain*, na Estratégia Nacional de Combate à Corrupção e à Lavagem de Dinheiro (ENCCLA/MJSP) e foi representante titular da AGU (2021) no Gabinete de Gestão Integrada de referida Estratégia. Foi delegado do Brasil junto ao Escritório das Nações Unidas contra Drogas e Crime (UNODC), G20 e BRICS, em matérias anticorrupção (2017 em diante), tendo participado das negociações multilaterais para a Declaração Política da Assembleia-Geral da Nações Unidas contra a Corrupção (UNGASS 2021) e outros acordos internacionais.

Como somos vistos? Análise da imagem organizacional pública utilizando ciência de dados¹

How are we seen? public organizational image analysis using data science

¿cómo somos vistos? análisis de imagen organizacional pública utilizando ciencia de los datos

Carolina Coelho da Silveira, Carla Bonato Marcolin e Carlos Henrique Rodrigues

<https://doi.org/10.36428/revistadacgu.v14i26.559>

Resumo: Entende-se por imagem organizacional as percepções públicas em torno de uma organização. Neste sentido, os veículos de imprensa têm grande influência, contudo existem desafios relacionados ao formato e volume dos dados destas fontes. Assim, o objetivo deste artigo é utilizar técnicas de Ciência de Dados para analisar a imagem organizacional de uma organização pública através da imprensa nacional, com foco em portais de notícias e jornais. Para validação, foi utilizado o caso da Agência Nacional de Energia Elétrica (ANEEL). As etapas metodológicas consistiram em definição, coleta, preparação e análise dos dados através de técnicas de Processamento de Linguagem Natural. Os principais resultados reforçam os indícios de alta relação entre a imagem da ANEEL e a imagem do Governo, e que existe uma disparidade entre a imagem passada pela manchete e pela notícia na íntegra. Ainda, observou-se a utilização de estratégias de comunicação (rotulagem, *agenda-setting*, *linkage* e *framing*) por parte dos veículos de imprensa. Para estudos futuros, sugere-se a utilização de outras fontes de dados e também a validação a partir de outros casos.

Palavras-chave: Imagem Organizacional Pública; Notícias; Ciência de Dados, NLP.

Abstract: Organizational image is understood as the public perceptions around an organization. In this sense, press vehicles have a great influence on the organizational image, but there are challenges when analyzing the organizational image through these sources. Therefore, the objective of this article is to use Data Science techniques to analyze the organizational image of a public organization through the national press, focusing on news portals and newspapers. For validation, the case of the Agência Nacional de Energia Elétrica (ANEEL) was used. The methodological steps consisted of defining, collecting, preparing and analyzing data, and Natural Language Processing techniques were used. The main results reinforce the evidence of a strong relationship between the image of ANEEL and the image of the Government, and that there is a disparity between the image conveyed by the headline and the news in full. Also, the use of communication strategies (labeling, *agenda-setting*, *linkage* and *framing*) by the press vehicles was observed. For future studies, it's suggested the use of other data sources and also the validation from other cases.

Key-words: Public Organizational Image; News; Data Science, NLP.

Resumén: La imagen organizacional se entiende como las percepciones del público en torno a una organización. En este sentido, los vehículos de prensa tienen una gran influencia en la imagen organizacional, pero existen desafíos al analizar la imagen organizacional a través de estas fuentes. Por ello, el objetivo de este artículo es utilizar

1. Artigo submetido em 18/07/2022 e aceito em 25/11/2022.

técnicas de Ciência de los Datos para analizar la imagen organizacional de una organización pública a través de la prensa nacional, centrándonos en portales de noticias y periódicos. Para la validación se utilizó el caso de la Agência Nacional de Energia Elétrica (ANEEL). Los pasos metodológicos consistieron en definir, recopilar, preparar y analizar datos, y se utilizaron técnicas de Procesamiento del Lenguaje Natural. Los principales resultados refuerzan la evidencia de una fuerte relación entre la imagen de ANEEL y la imagen del Gobierno, y que existe una disparidad entre la imagen transmitida por el titular y la noticia en su totalidad. Asimismo, se observó el uso de estrategias de comunicación (etiquetado, *agenda-setting*, *linkage* y *framing*) por parte de los vehículos de prensa. Para futuros estudios, se sugiere el uso de otras fuentes de datos y también la validación de otros casos.

Palavras-clave: Imagen Organizacional Publica; Noticias; Ciência de los Datos; NLP.

1. INTRODUÇÃO

Uma organização é um sistema complexo de comunicações com um objetivo funcional e, durante sua existência e atuação, ela vai formando uma imagem junto aos seus diversos públicos em função de todo um conjunto de contatos realizados, e não somente pelas iniciativas planejadas de comunicação que ela toma para formar a imagem pretendida (Schuler, 2004). Essa imagem é chamada imagem organizacional, e pode ser descrita como uma compilação de opiniões e pontos de vista baseados em informações processadas de várias fontes ao longo do tempo, gerando uma imagem mental dos atributos da organização (Valerio & Pizzinatto, 2003; Zhang *et al.*, 2016).

Seguindo uma tendência já percebida na iniciativa privada, os órgãos públicos têm, cada vez mais, se preocupado com sua imagem organizacional. A imagem organizacional pública tem valor estratégico na medida em que interfere diretamente no relacionamento da organização com diferentes atores, bem como na sua legitimação e credibilidade perante a sociedade. No entanto, apesar de sua relevância, poucos estudos endereçam a imagem organizacional no contexto público, tanto nacional quanto internacionalmente (Da Silva & Gutierrez, 2013; De Sousa, Ardigo & Santos, 2021; Steppacher, 2012; Valerio & Pizzinatto, 2003).

Quando se trata da construção da imagem de uma organização, ressalta-se o papel dos veículos de imprensa. Estes veículos, ao publicarem uma notícia, atuam na associação e formação das crenças, ideias, sentimentos e impressões que uma pessoa ou grupo tem ou passa a ter sobre aquela organização (Otto *et al.*, 2022). Não obstante, a migração desses meios de comunicação para a internet aumenta o volume de informações em circulação, elevando seu poder de influência, uma vez que é possível acessar dados e in-

formações de qualquer lugar, a qualquer momento, expandindo a comunicação entre as pessoas e organizações (Bartoluzzio *et al.*, 2019).

Contudo, a popularização dos portais eletrônicos, o grande volume de informações geradas e o formato não-estruturado desses dados se apresentam como um desafio para as organizações que tentam se manter à par do que está sendo veiculado na mídia sobre elas. Existe uma dificuldade de resumo e análise através de esforço humano: não só altos custos seriam gerados, mas os resultados estariam expostos à subjetividade dos analistas (Marcolin *et al.*, 2021). Logo, são necessárias técnicas inovadoras para endereçar as complexidades que cercam a imagem de uma organização conforme veiculada pela imprensa. Nesse sentido, as técnicas de Ciência de Dados se apresentam como uma alternativa.

Considerando contexto, este artigo traz como pergunta de pesquisa: como aferir a imagem de uma organização pública através do que a imprensa veicula sobre ela? A fim de responder esta pergunta, o objetivo desse artigo é utilizar técnicas de Ciência de Dados para construir um instrumento capaz de analisar e oferecer insights sobre a imagem organizacional de uma instituição pública através da imprensa nacional, com foco em portais de notícias e jornais.

Para validação, foi utilizado o caso da Agência Nacional de Energia Elétrica - ANEEL. A ANEEL é uma agência reguladora, estabelecida como uma autarquia, vinculada ao Ministério das Minas e Energia, que tem como objetivo regular e fiscalizar a geração, transmissão e comercialização de energia elétrica, em conformidade com as políticas e diretrizes do governo federal. Apesar disso, a ANEEL tem autonomia decisória, financeira e gerencial, além de ter a competência técnica para normatizar questões do setor elétrico (Agência Nacional de Energia Elétrica, 2021).

Este trabalho está organizado da seguinte forma: na seção 2, apresenta-se o conceito de imagem organizacional e qual sua relevância no setor público, bem como discute-se o papel dos veículos de imprensa na formação dessa imagem. Na seção 3, apresenta-se a metodologia utilizada para se atingir o objetivo desta pesquisa. Os principais resultados são apresentados na seção 4. Na seção 5, são feitas as considerações finais, ressaltando as limitações e possibilidades de estudos futuros.

2. CONTEXTUALIZAÇÃO

Nesta seção é feita uma contextualização sobre o tema imagem organizacional e a influência da imprensa na sua formação. Esta seção inicia com uma caracterização da imagem organizacional (seção 2.1) e porquê é importante para o ente público gerenciá-la. Na seção 2.2 discute-se o papel da imprensa na formação dessa imagem, ressaltando as principais estratégias utilizadas por esses veículos.

2.1. A caracterização da imagem organizacional e a imagem organizacional pública

São três os aspectos que caracterizam a imagem organizacional. Primeiro, a imagem organizacional representa limitações da realidade na mente do público - ou seja, as pessoas formam uma ideia de uma determinada organização a partir do que consideram real sobre ela. Em segundo lugar, a formação de uma imagem organizacional para um determinado público decorre das mensagens recebidas e decodificadas de forma acumulada ao longo do tempo processadas a partir de traços psicológicos de cada pessoa. Por fim, a imagem organizacional representa as percepções individuais da realidade das pessoas - diferentes públicos podem ter diferentes impressões de uma única organização com base em suas origens econômicas, sociais e pessoais (Barich & Kotler, 1991; Bromley, 1993; Dowling, 1986).

Sendo o conceito de imagem organizacional ligado à área de marketing, Cezar (2019) destaca que se apoiar em suas teorias, conceitos e práticas enquanto instrumento para a melhoria de desempenho das organizações públicas é pertinente e necessário, porém ressalta que a maioria foi desenvolvido sob uma ótica mercadológica, e as organizações públicas não se orientam por uma lógica de lucratividade ou venda

como acontece no setor privado. Logo, é necessário delinear os pressupostos de marketing às lógicas das organizações públicas, considerando a forma de atuação dos órgãos, as leis a que estão expostos, as formas de governança e os serviços que oferecem, entre outros.

Mesmo com tais considerações, é fundamental para o ente público conhecer e gerenciar sua imagem organizacional, na medida em que ela interfere diretamente no relacionamento da organização com os seus diversos públicos (Valerio & Pizzinatto, 2003) e na legitimação em relação ao seu ambiente técnico-organizacional (Steppacher, 2012). Além disso, o alinhamento entre a imagem projetada pela organização e a percepção de seus públicos é de fundamental importância, uma vez que, quando este alinhamento não existe ou é deficiente, a credibilidade do discurso organizacional fica afetada (Da Silva & Gutierrez, 2013).

Entende-se, portanto, que a imagem real de uma organização não é necessariamente aquela cujos seus esforços se propõem a divulgar, sendo construída de forma diversa e quase que individual na mente do público. E, considerando que as organizações públicas devem se preocupar com diferentes questões ao exercer sua função de servir aos cidadãos, o estudo de imagem organizacional no setor pode trazer benefícios.

2.2. O papel da imprensa na formação da imagem organizacional

Em comunicação, entende-se por mídia os canais ou ferramentas usadas para armazenamento e transmissão de informações ou dados. Neste universo, existe a imprensa, que é um conjunto de veículos de comunicação que exercem o Jornalismo e outras funções de comunicação informativa (Martins, 2014).

O papel da imprensa, em sua essência, é produzir e transmitir informação, conteúdo de relevância e de interesse público; nesse contexto, a imprensa é vista como distribuidora social de mensagens, consideradas em sua maioria fidedignas pela população (Martins, 2014; Otto *et al.*, 2022). No entanto, a imprensa não apenas transmite um conjunto de estruturas interpretativas, mas também sistemas de relevância, reforçando valores, interesses, ênfases e convicções sobre o que é verdadeiro ou falso, o que vale a pena conhecer ou o que é negligenciável (Hjarvard, 2012).

Dada a força das suas posições e influência sobre a sociedade, Maia (2009) ressalta que a imprensa não é composta de meros veículos neutros, mas de organizações que controlam o acesso dos atores sociais aos

seus canais e regulam os fluxos de comunicação. Os profissionais da imprensa selecionam e editam eventos e discursos, enquadrando significados partir da própria lógica e de seus *modus operandi*. Consequentemente, quando a imprensa transmite uma notícia, ela começa a construir uma realidade social – no âmbito deste estudo, essa realidade social pode ser entendida como a imagem organizacional.

Nesse processo, a imprensa faz uso de algumas estratégias que impactam a imagem organizacional. A rotulagem, por exemplo, associa a organização a uma ação, fato ou tema (Otto *et al.*, 2022). Já na *agenda-setting*, há uma atenção e repetição da mídia sobre uma questão específica, concentrando a atenção do público em uma questão particular (Martins, 2014).

Ainda, há o *framing*, usado pela imprensa quando se deseja modelar, destacar ou enfatizar algo, dando enfoque utilizando certas expressões, palavras e adjetivos, permitindo que o receptor compreenda de forma recortada uma informação (Martins, 2014). Outra estratégia é o *linkage*, processo que acontece quando a imprensa vincula um acontecimento a outro e, consequentemente, um conjunto de tópicos torna-se associado a uma organização, fenômeno ou evento (Otto *et al.*, 2022).

Portanto, a mídia tem se constituído como fonte de definição e interpretação da realidade, e adquire grande importância enquanto meio pelo qual é possível oferecer uma visão particular das organizações. Como os órgãos públicos não são entidades isoladas dessa realidade, eles não podem ignorar a necessidade de conhecer e gerenciar a sua imagem organizacional, conforme veiculada pela imprensa (Da Fonseca, 2020).

3. PROCESSO METODOLÓGICO

3.1. Ciência de Dados e Processamento de Linguagem Natural

Entende-se *big data* como um conjunto de dados identificado por quatro características principais: volume, devido à grande quantidade de dados; velocidade, devido à rapidez com que os dados são gerados ou atualizados; variedade, devido aos dados estrutu-

rados e não-estruturados, provenientes de diversas fontes; e veracidade, devido às estruturas complexas dos grandes conjuntos de dados que os tornam ambíguos, imprecisos e inconsistentes (Tabesh, Mousavidin & Hasani, 2019).

Considerando tais aspectos, é necessária a utilização de ferramentas inovadoras para a coleta, armazenamento e análise desses dados (Tabesh, Mousavidin & Hasani, 2019). Logo, um dos objetivos da Ciência de Dados é considerar diferentes tipos de dados de forma conjunta, a fim de facilitar sua análise e apresentá-los como suporte na tomada de decisão organizacional (Chuprina, Postanogov & Kostareva, 2017).

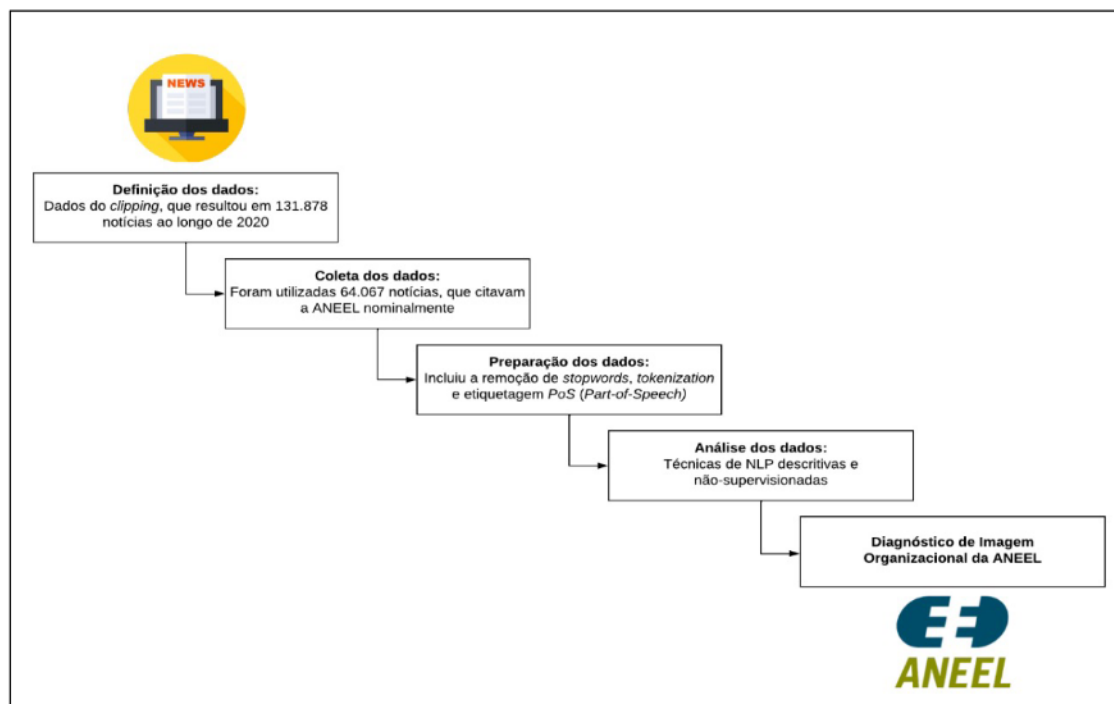
Para atingir os objetivos deste trabalho, foi realizada uma coleta de dados junto à ANEEL. A Agência contrata um serviço de *clipping*, que coleta notícias da imprensa especializada e não-especializada sobre a organização. As características e o volume dos dados do *clipping* de notícias aproximam-se do que Tabesh, Mousavidin & Hasani (2019) denominaram *big data*. Logo, esses conjuntos de dados são tão volumosos que técnicas metodológicas tradicionais de análise podem não ser adequadas para lidar com o volume e a multidimensionalidade.

3.2. As etapas metodológicas para a construção do instrumento

A partir da escolha das técnicas de análise mais recomendadas para o *clipping* de notícias, esta seção busca descrever as etapas metodológicas utilizadas para a construção do instrumento, cuja preocupação central foi permitir um processo de coleta, preparação, resumo, visualização e análise de dados (Fan & Gordon, 2014) produzidos pela imprensa nacional para extrair padrões e informações úteis acerca da imagem organizacional da ANEEL. Na prática, o instrumento criado para análise da imagem organizacional de notícias possui quatro etapas distintas: definição dos dados, coleta dos dados, preparação dos dados e análise dos dados (Stieglitz *et al.*, 2018). Os códigos na íntegra estão disponíveis na plataforma GitHub² dos autores.

2. Link para acesso: <https://github.com/carlamarcolin/topicmodellling/blob/master/codigos-aneel-imprensa.zip>.

FIGURA 1: ETAPAS METODOLÓGICAS PARA CRIAÇÃO DO INSTRUMENTO



Considerando a definição dos dados, o *clipping* de notícias contratado pela ANEEL mostrou-se a melhor opção, por possuir quantidade, qualidade e custos de obtenção mais adequados.

Para coleta, os dados utilizados foram as notícias do *clipping* para todo o ano de 2020. No total foram coletadas notícias de 3.232 veículos de imprensa diferentes, o que totalizou 131.878 notícias ao longo de 2020. Como o objetivo deste trabalho de pesquisa é desenvolver um instrumento diagnóstico para analisar a imagem organizacional a partir das publicações da imprensa nacional, foram filtradas apenas as notícias que citaram nominalmente a Agência ao longo de 2020 (cerca de 64.067 notícias, equivalente a 49%).

Na fase de preparação dos dados, utilizou-se a linguagem computacional Python, por ser a linguagem computacional mais utilizada para análise de dados em massa com características *big data*, que é o caso do *clipping* (Cielen & Meysman, 2016). Esta fase seguiu as melhores práticas indicadas em literatura (Altszyler *et al.*, 2017; Manning, Rhagavan & Schutze, 2010) considerando: (i) remoção de *stopwords*; (ii) *tokenization* ou tokenização e *n-grams*; e (iii) etiquetagem PoS (*Part-of-Speech*) ou “tagueamento”.

A remoção de *stopwords* consiste em remover palavras de alta frequência, mas baixo valor semântico, ou seja, expressões como “de”, “para”, “que”, entre

outras. A tokenização visa quebrar o texto em unidades mínimas, que podem ser palavras simples ou até mesmo compostas, bem como outras expressões como siglas. A tokenização auxilia a melhor performance computacional, permitindo a análise *n-grams*, que são uma sequência de *n*-caracteres. Por fim, a etiquetagem PoS consiste na identificação das classes gramaticais de cada palavra, que recebe uma etiqueta com sua respectiva classe, como verbo, substantivo e adjetivo. Também se realizou um tratamento no texto para a substituição de caracteres codificados incorretamente pelos caracteres corretos. Isso foi necessário pois o arquivo de texto das notícias não possuía padrão de codificação.

Por fim, na fase de análise dos dados, foram utilizadas técnicas de Processamento de Linguagem Natural (NLP) descritivas e não-supervisionadas. NLP é um campo dentro da Ciência de Dados formado por um subconjunto de técnicas computacionais baseadas em inteligência artificial, voltado para a compreensão da linguagem humana, sendo um método para extrair conhecimento a partir de um grande volume de textos e convertê-los de linguagem natural (falada ou escrita) em resultados utilizáveis (Aggarwal & Zhai, 2012; Maulud *et al.*, 2021). As técnicas descritivas tiveram como objetivo compreender o conjunto de dados e obter uma análise inicial da imagem organizacional.

Contudo, sendo o conceito de imagem organizacional complexo e multifacetado, optou-se por aprofundar essa análise através de uma técnica não-supervisionada, a modelagem de tópicos (*topic modeling*).

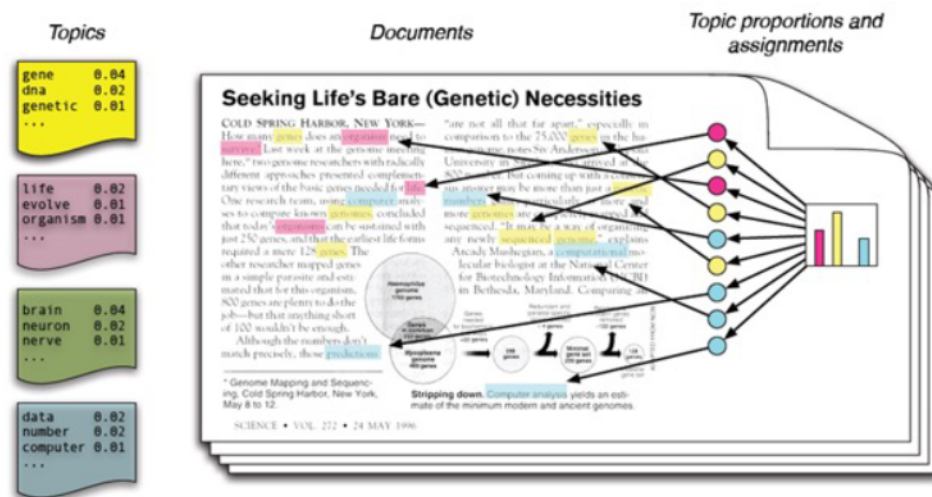
A modelagem de tópicos é adequada à tarefa e ao potencial de avanço para a pesquisa, fornecendo um caminho complementar aos métodos empíricos tradicionais (Schmiedel, Muller & Brocke, 2019; Langley & Klag, 2019). O objetivo desta técnica é uma representação multidimensional dos tópicos presentes em um conjunto de documentos, neste caso, notícias. Após o processo, cada documento é atribuído a cada tópico com diferentes pesos, especificando o grau de associação, revelando a estrutura latente do tópico formada por seu conjunto de palavras (Crain *et al.*, 2012). Embora se possa utilizar a modelagem de tópicos como técnica de agrupamento (Landauer, 2007), para este trabalho o objetivo é o processo de descoberta de tópicos, que representam as diferentes dimensões da imagem organizacional do ente público.

Das diferentes abordagens para modelagem de tópicos, foi escolhida a abordagem probabilística, que impõe um modelo probabilístico gerando a partir da estrutura semântica de um conjunto de documentos. Assumindo que os documentos são gerados de acordo com um determinado modelo probabilístico, descobrir sua estrutura semântica equivale a estimar os parâmetros do modelo, alguns deles representando um conjunto de tópicos (Stevens *et al.*, 2012). De certa forma,

como sugerido por Ashton *et al.* (2014), a abordagem probabilística, na modelagem de tópicos, expandiu os modelos de fatoração de matrizes incorporando um contexto de incerteza, permitindo a implementação de processos generativos. Neste trabalho, foi aplicado o LDA (*Latent Dirichlet allocation*), um modelo tradicional que assume que os documentos são misturas probabilísticas de tópicos latentes (Figura 2) e que os tópicos são distribuições de probabilidade sobre palavras (Blei, 2012). Além de ser um dos modelos mais comuns em uso em trabalhos de ciências sociais aplicadas (Eickhoff & Neuss, 2017), o LDA também tem a robustez necessária para lidar com dados qualitativos do discurso, resultando em um conjunto de tópicos mais coerente que representa a verdade, mais próximo do julgamento humano (Nikolenko *et al.*, 2017).

O modelo LDA é uma abordagem não supervisionada que requer três parâmetros: os hiperparâmetros de Dirichlet (α e β) e o número de tópicos (κ). A escolha de κ tem implicações significativas para os resultados gerados pelo modelo, uma vez que um número relativamente grande atribuído ao parâmetro pode dispersar a alocação de texto, e um número relativamente pequeno pode comprimir demais a alocação de texto, dificultando uma análise sólida. A escolha apropriada de κ é uma questão de pesquisa aberta (Lima Jr., 2020). Para este trabalho, foi utilizada uma abordagem iterativa, verificando a estabilidade do modelo considerando entre 3 e 10 tópicos, escolhendo $\kappa = 5$.

FIGURA 2: LATENT DIRICHLET ALLOCATION (LDA)



Fonte: Blei (2012).

FIGURA 4: PRINCIPAIS PALAVRAS ASSOCIADAS À ANEEL NAS NOTÍCIAS PUBLICADAS EM 2020



O destaque das palavras “empresa”, “setor”, “governo” e “pandemia” apresenta indícios de alta relação entre a imagem da ANEEL e a imagem do Governo. Em seguida, em uma segunda ordem de relevância, aparecem as palavras: sistema, luz, brasil, federal, presidente. Por fim, em um terceiro nível, as palavras: distribuidoras, valor, serviços, amapá, tarifa, geração, coronavírus, consumidor. Agrupando essas palavras, percebe-se que o nome da ANEEL aparece junto de temas relacionados à política brasileira, ao Estado, à temas sazonais do ano, aos consumidores e ao setor de energia elétrica

4.2. Análise descritiva

Nesta seção, buscou-se entender melhor a linguagem utilizada pela imprensa nacional nas notícias veiculadas sobre a ANEEL. Os bigramas e tri-

gramas estão relacionados com a utilização conjunta de palavras, e o intuito dessa técnica é definir quais são os termos e expressões mais utilizados pela imprensa para se referir à ANEEL. Já com a técnica da etiquetagem (*Part-of-Speech*), buscou-se identificar os verbos, adjetivos e substantivos associados à ANEEL com mais frequência.

A Figura 5 contrasta dois gráficos: o bigrama de todas as notícias que citam a ANEEL na íntegra (5A), e os bigramas apenas das manchetes das notícias que citam a ANEEL (5B). Com o intuito de ter uma visualização ainda mais contundente, o instrumento também mostra os trigramas (Figura 6). À exemplo dos bigramas, aqui também foram realizadas visualizações tanto dos dados das notícias na íntegra (6A) quanto das manchetes (6B).

FIGURA 5: BIGRAMAS DAS NOTÍCIAS NA ÍNTEGRA E DAS MANCHETES DAS NOTÍCIAS

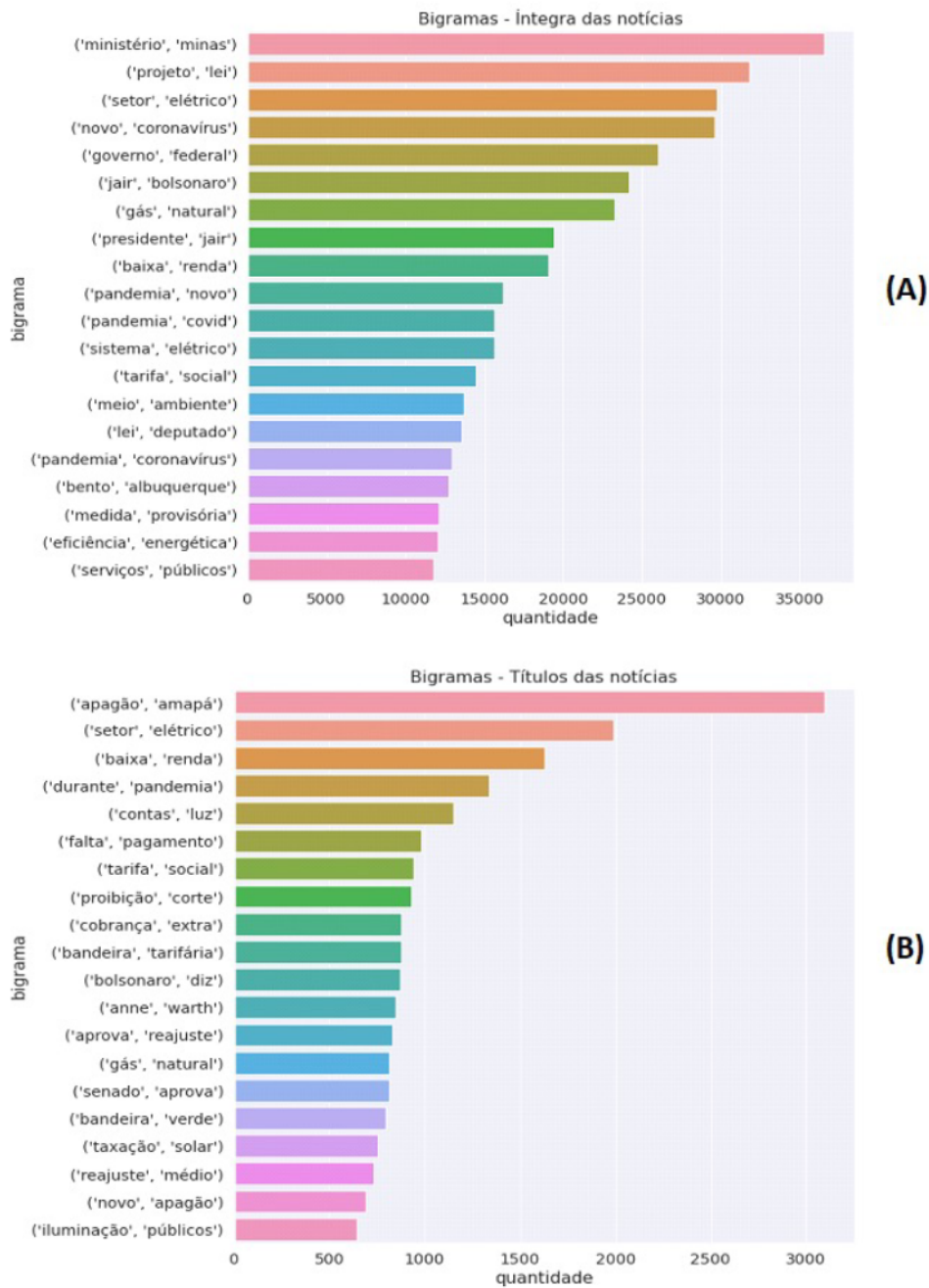
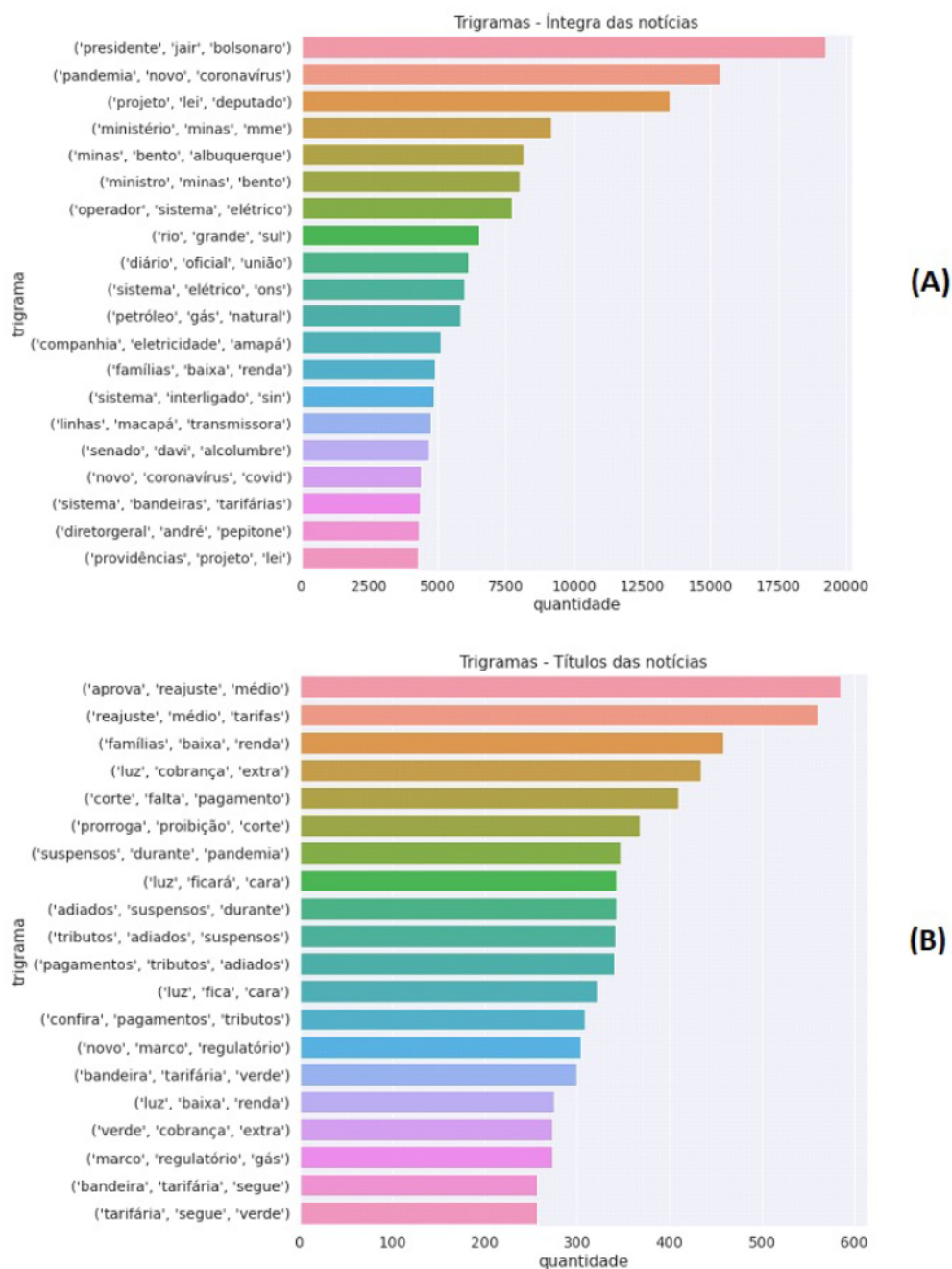


FIGURA 06: TRIGRAMAS DAS NOTÍCIAS NA ÍNTEGRA E DAS MANCHETES DAS NOTÍCIAS



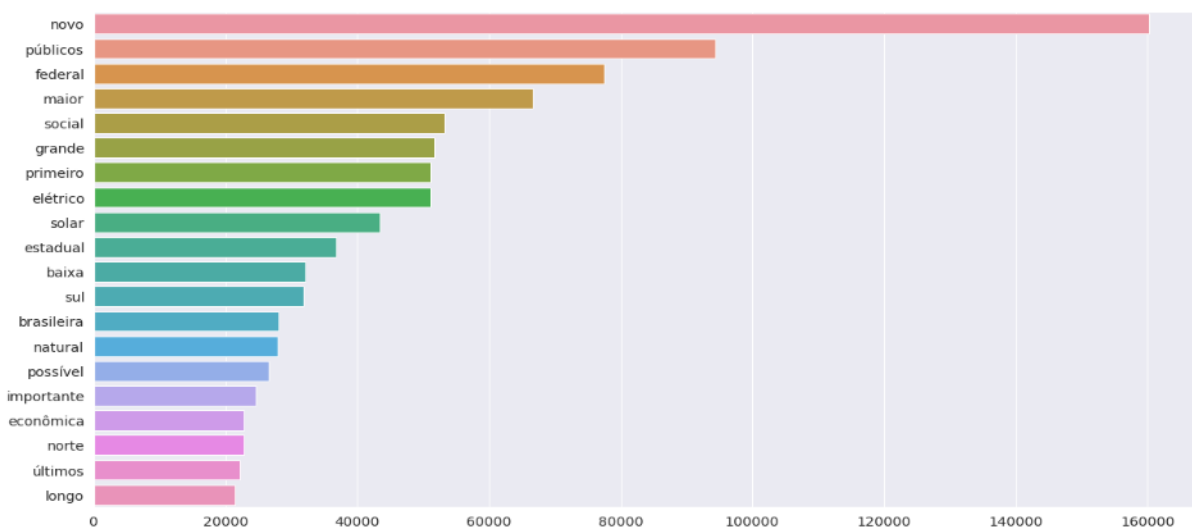
Observa-se que do bigrama referente às notícias na íntegra (Figura 5A), emergem palavras que fortalecem a evidência de associação da ANEEL ao contexto político, Estado, pandemia e setor elétrico. No entanto, considerando que parte do público que acessa as notícias não aprofunda a leitura, ficando somente com a informação evidenciada na manchete, para efeitos de comparação apresenta-se também um bigrama para os títulos das notícias (Figura 5B).

Nele, tem-se uma interpretação distinta das notícias na íntegra, sendo que as palavras mais frequentes são: apagão do Amapá, setor elétrico, baixa renda, durante a pandemia e conta de luz. Ou seja, apesar de a manchete ter o propósito editorial de síntese da notícia completa, percebe-se que as manchetes dão o entendimento menos de governo e Estado, e mais do setor elétrico.

Os trigramas das notícias na íntegra (Figura 6A) confirmam mais uma vez a forte associação da imagem da ANEEL com o governo e o Estado. Nos trigramas das manchetes, observa-se uma associação às questões ligadas ao setor elétrico em geral. Novamente, fica claro que o público que lê as manchetes pode ter uma associação distinta da imagem organizacional da ANEEL em relação ao público que lê as notícias na íntegra, uma vez que as manchetes não refletem o teor das notícias completas.

Outra análise realizada foi uma exploração das principais estruturas gramaticais presentes nas notícias: uso de adjetivos, substantivos e verbos na construção das mensagens. A Figura 7 apresenta os adjetivos mais comuns nas notícias. Destaca-se que nenhum dos principais adjetivos possuem um teor de significado negativo.

FIGURA 7: POS: ADJETIVOS MAIS FREQUENTES



A Figura 8 apresenta os substantivos que mais apareceram nas notícias. Percebe-se que são substantivos ligados a questões de política, pandemia e do setor elétrico. Ao associar os adjetivos mais utilizados com os substantivos mais proeminentes nos textos das notícias, percebe-se as junções de qualificação do governo, empresa e setor.

Por fim, a Figura 9 apresenta os verbos mais utilizados nos textos das notícias. São verbos que indicam ações dos substantivos mais utilizados. Percebe-se que são verbos que indicam ações, em sua maioria, de comunicar, extraídos de entrevistas ou falas.

FIGURA 8: POS: SUBSTANTIVOS MAIS FREQUENTES

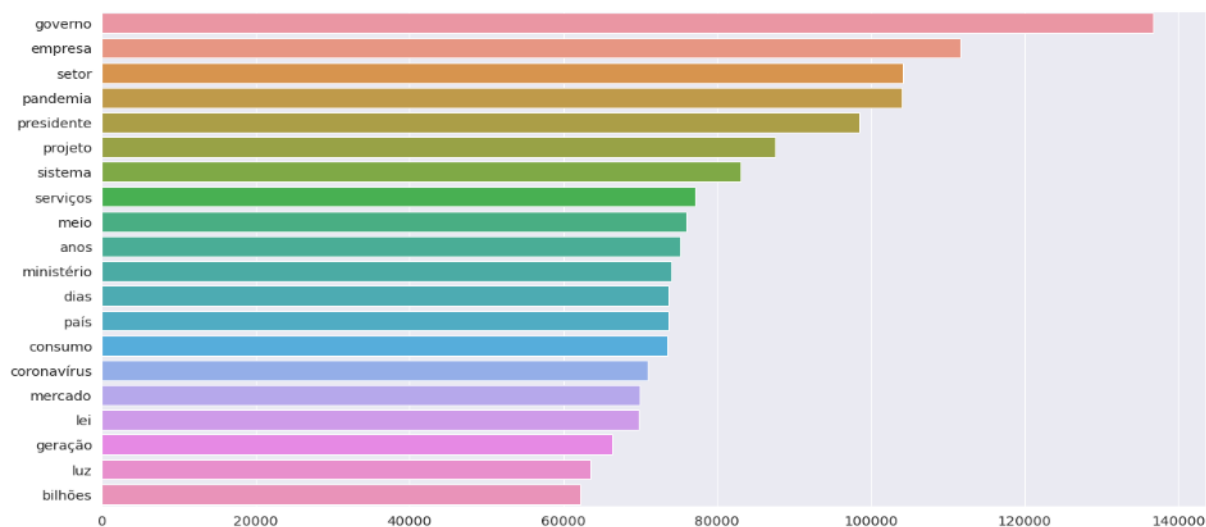
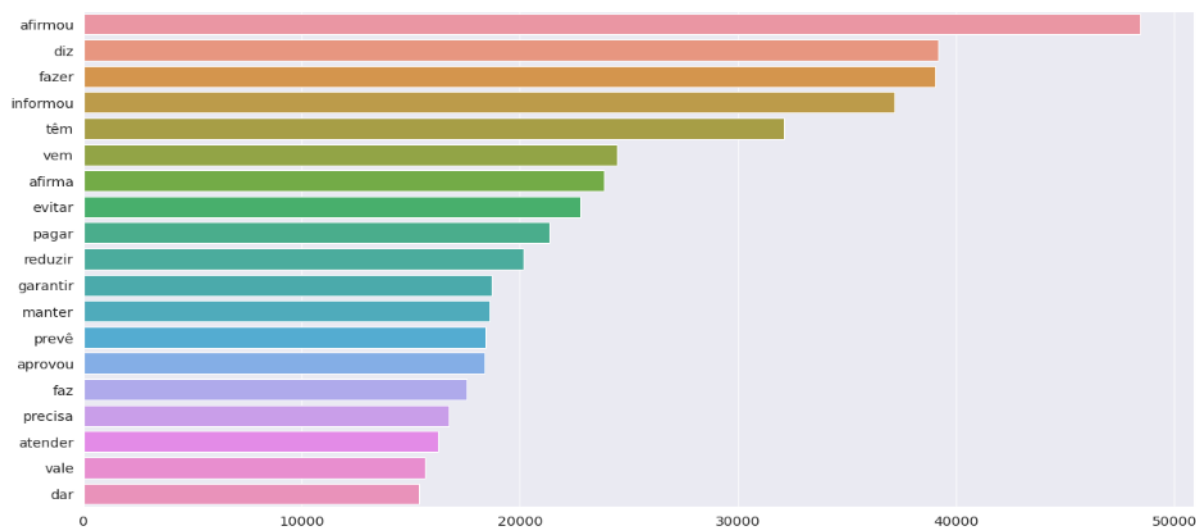


FIGURA 9: POS: VERBOS MAIS FREQUENTES

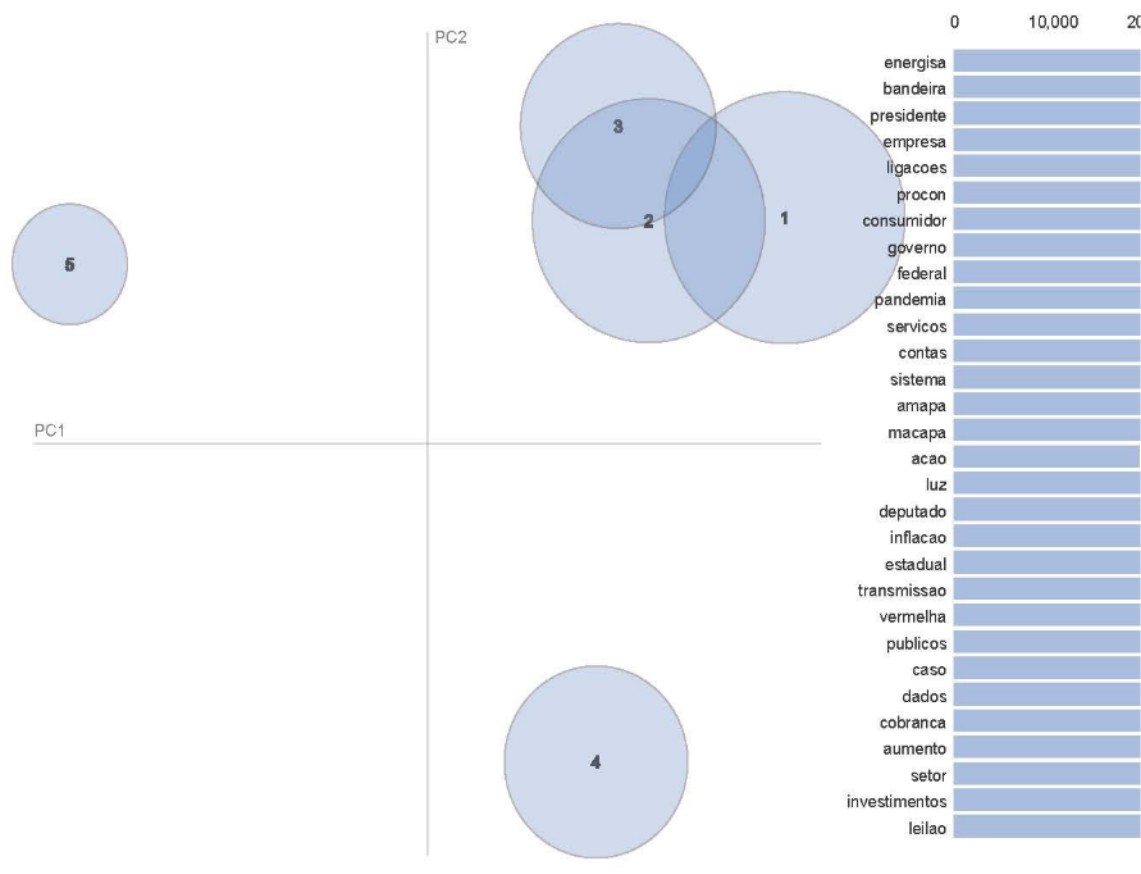


4.3. Modelagem de tópicos

Dados os resultados da seção anterior, para a análise de tópicos, as notícias na íntegra e as manchetes também foram analisadas separadamente. O intuito desta técnica é visualizar como os termos se agrupam em torno tópicos, representados pelos círculos, considerando também as diferentes estratégias utilizadas pela imprensa para veiculação de informações. As Figuras 10 e 11 apresentam o conjunto de todos os tópicos, respectivamente, para o texto das notícias na íntegra (10) e apenas para as manchetes (11).

O tamanho do círculo (tópico), mostra a sua relevância no conjunto dos dados analisados. Quanto maior for o círculo, maior a relevância daquele tópico abordado pela a imprensa nacional. Percebe-se também que os círculos se sobrepõem, ou seja, há alguns termos que são comuns a mais de um tópico. Quanto maior a área de sobreposição, maior a correlação entre os tópicos abordados. Outro aspecto a ser observado na visualização dos tópicos é a distância entre um círculo e outro. Quanto mais distantes forem, maior é a discricionariedade entre um tópico e outro, ou seja, os tópicos tratam de temáticas distintas.

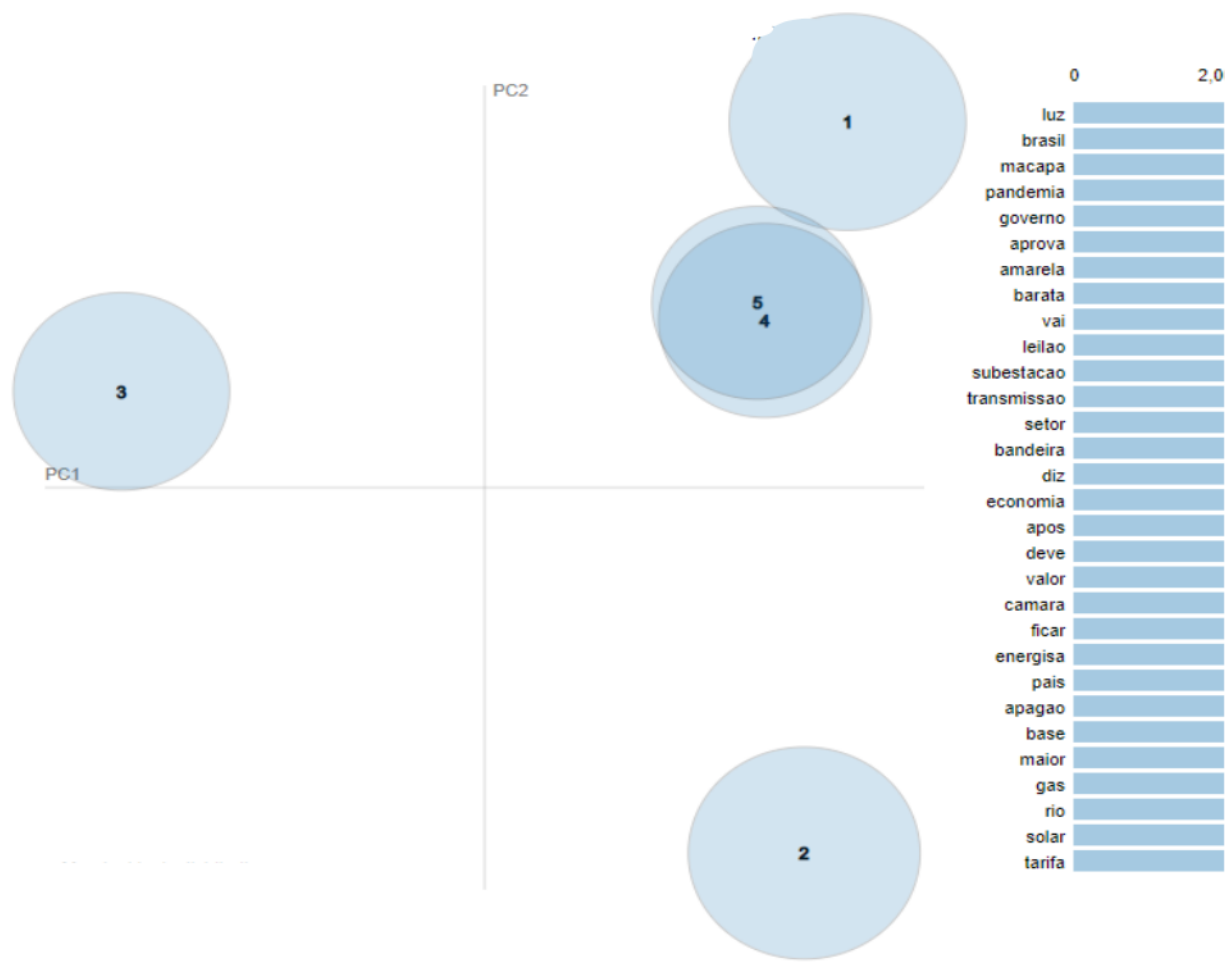
FIGURA 10: TÓPICOS DO TEXTO INTEGRAL DAS NOTÍCIAS QUE CITAM A ANEEL



As visualizações dos tópicos das notícias completas mostram cinco tópicos (Figura 10), em que os tópicos 1, 2 e 3 apresentam sobreposições de termos. Já os tópicos 4 e 5 estão mais distantes entre si e entre os demais. O tópico 1 apresenta uma forte associação com temas relacionados ao setor de energia elétrica: setor, transmissão, geração, empresa, Brasil, investimentos, mercado, país, usinas, elétrico. O tópico 2, apesar das sobreposições, apresenta uma associação forte com “impacto financeiro”: pandemia, luz, contas, valor, consumidor, inflação, cobrança, alta, aumento, redução, taxa, custo, tarifa, preços, IPCA, pagamento. O tópico 3, também com sobreposições com o 1 e principalmente com o tópico 2, apresenta forte asso-

ciação com o sistema bandeiras tarifárias: bandeira(s), sistema/programa, dados, vermelha, amarela, verde, água, reservatórios, economia, consumo, municípios, kwh. O tópico 4, bem distinto dos demais, têm forte relação com o governo, política e o Estado: governo, presidente, federal, Amapá, Macapá, deputado, ministério, lei, CPI, órgãos, ministro, Bolsonaro, apagão, justiça, Jair, Câmara. O tópico 5, também distinto dos demais, porém com pequena relevância se comparado com os demais (menor tamanho do círculo), agrupa termos que têm relação com jurídico e consumo: Energisa, ligações, Procon, ação, serviços, estadual, empresa, consumidor, reclamações, suspensão, fiscalização, fornecimento.

FIGURA 11: TÓPICOS DAS MANCHETES DAS NOTÍCIAS QUE CITAM A ANEEL



As visualizações dos tópicos das manchetes das notícias mostram também cinco tópicos (Figura 11), em que os tópicos 4 e 5 apresentam sobreposições de termos, estando bem próximos do tópico 1. Já os tópicos 2 e 3 estão mais distantes entre si e entre os demais. Não foi observada grande distinção entre o tamanho dos círculos.

O tópico 1 tem forte relação com o Estado e eventos específicos: Macapá, Energisa, país, apagão, pública, Covid, Norte, aumento, problema, Amazônia. O tópico 2, sem sobreposições e distante dos demais, apresenta uma forte associação com temas de governo, política e o Estado: pandemia, governo, economia, Câmara, Amapá, projeto, alta, Bolsonaro, Ministério, privatizações, petróleo. O tópico 3, também sem sobreposições e distante dos demais, apresenta relação com temas de economia: leilão, subestação, gás, inflação, Senado, IPCA, contas, recursos, comer-

cialização, sobe, ministro, demanda. O tópico 4, com sobreposição ao tópico 5 e bastante próximo ao tópico 1, apresenta temas voltados para o setor de energia elétrica: setor, investimentos, ambiental, elétrico, Procon, metas, leilões, usina, venda, geração, Cuiabá, obras, hidrelétricas, biodiesel, privatização. O tópico 5, apesar da sobreposição, apresenta forte associação com o sistema de bandeiras tarifárias: luz, Brasil, amarela, bandeira, barata, valor, tarifa/tarifária, crise, mercado, Brasília, preço, reajuste, corte.

5. DISCUSSÃO

Ao noticiar sobre a ANEEL, determinado veículo de imprensa pode ou não, de forma consciente ou inconsciente, refletir uma imagem organizacional na linguagem e na escolha das palavras, sentenças e temas

de suas notícias. Essas escolhas rotulam a realidade e, portanto, contribuem para formar a imagem organizacional da ANEEL.

Apesar da ampla quantidade de veículos de imprensa nacional que publicaram notícias mencionando a ANEEL em 2020, grande parte dessas notícias está concentrada em canais que não possuem uma linha editorial especializada no setor de energia elétrica. Observou-se que existem diferenças de teor entre as notícias veiculadas em canais especializados versus não especializados: veículos de imprensa especializados tendem a associar a imagem da ANEEL às questões mais técnicas do setor, como as atividades de geração, distribuição e transmissão de energia elétrica; enquanto veículos de imprensa não-especializados tendem a associar a imagem organizacional da ANEEL a questões políticas, de governo e do Estado. Esta associação da imprensa não-especializada merece atenção especial, pois representa um risco ao associar a imagem de políticos e do governo à imagem da ANEEL.

Outro resultado que chama atenção são as diferenças entre os conteúdos das notícias na íntegra e das manchetes. Ao comparar as Figuras 10 e 11, fica evidente que a distribuição dos tópicos e as principais palavras sofrem variação, ainda que, em tese, a mensagem passada deveria ter certa similaridade. Apesar de a manchete ter o propósito editorial de síntese da notícia completa, percebe-se que estas possuem mais termos sobre governo e Estado, e menos sobre o setor elétrico. Considerando que parte do público que acessa as notícias não aprofunda a leitura e que a imprensa utiliza de técnicas de linguagem para chamar a atenção do público nos títulos das notícias, o nome da ANEEL associado à estas manchetes pode criar na mente do público uma imagem mais associada à opinião pública sobre o governo (e seus governantes) em exercício, o que pode causar flutuações de percepções dependendo da aprovação (ou não) dos cidadãos.

Foram identificados também picos de veiculação de notícias quando o tema tem caráter de associação negativo à imagem da ANEEL, como: “aumento de tarifas de energia elétrica”, “apagão no Amapá”, “projeto de lei com taxa de energia solar”. Essa situação requer atenção, pois a repetição da mídia sobre uma questão específica é chamada de *agenda-setting*, e a repetição da mesma notícia por veículos de imprensa diferentes passa ao público uma ideia de importância relativa do assunto repetido em relação a outros as-

suntos e eventos. Isso concentra a atenção do público nessas questões particulares e fortalece a associação de temas negativos à imagem organizacional da ANEEL.

O agrupamento de termos proporcionado pela modelagem em tópicos (seção 4.3) também evidencia a utilização de estratégias de comunicação pelos veículos de imprensa (seção 2.2). Por exemplo, para os tópicos das notícias na íntegra (Figura 10): para termos evidenciados no tópico 1, não foi identificada nenhuma estratégia evidente, uma vez que ele apresenta termos informativos do setor elétrico. No entanto, para o tópico 2, que apresenta uma forte associação com economia e finanças, é possível identificar as estratégias de *agenda-setting*, *framing* (modelar a informação para que o público a compreenda de forma recortada) e *linkage* (vincular um acontecimento a outro). No tópico 3, cujo principal tema é o sistema de bandeiras tarifárias, também foi possível identificar o uso do *framing* e *agenda-setting*. No tópico 4, sobre governo, política e o Estado, é possível observar a estratégia do *linkage*. Por fim, no tópico 5, que agrupa termos que têm relação com jurídico e consumo, identificou-se o uso de rotulagem (associação de uma organização a uma ação, fato ou tema) e *framing*.

Para os tópicos associados às manchetes (Figura 11): para os termos evidenciados no tópico 1, que têm forte relação com o Estado e eventos específicos, foi possível identificar as estratégias de *agenda-setting* e *linkage*. No tópico 2, sobre governo, política e o Estado, também foi identificado o uso de *linkage*. No tópico 3, que apresenta relação com temas de economia e finanças, observou-se o uso do *agenda-setting*. No tópico 4, com temas voltados para o setor de energia elétrica, não foram identificadas estratégias evidentes. Por último, no tópico 5, cujo tema é o sistema de bandeiras tarifárias, foi identificado o uso do *framing*.

Entende-se que o uso dessas ferramentas, principalmente nos veículos não-especializados, tem como objetivo ajudar o público a entender uma realidade social – no âmbito deste estudo, essa realidade social está intrinsecamente relacionada à imagem organizacional. No entanto, seu uso indevido e não justificado favorece uma composição textual mais dramática do que realista. E, considerando que os jornalistas estão cada vez mais injetando seus pontos de vista nos seus trabalhos (Kuypers, 2002), pode acabar se tornando difícil para o leitor cuidadoso separar as informações verdadeiras daquelas que são tendenciosas.

6. CONCLUSÕES

O objetivo deste estudo foi utilizar técnicas de Ciência de Dados para construir um instrumento capaz de analisar e oferecer insights sobre a imagem organizacional de uma organização pública, conforme veiculado em portais de notícias e jornais. Para validação da proposta, foi utilizado o caso da Agência Nacional de Energia Elétrica – ANEEL.

Dentre os principais resultados, observou-se uma forte relação entre a imagem da ANEEL e a imagem do Governo. Ainda, evidenciou-se a existência de uma discrepância entre o conteúdo das manchetes e da íntegra das notícias publicadas sobre a ANEEL. Considerando que parte do público que acessa as notícias não aprofunda a leitura, a imagem organizacional da ANEEL transmitida pelas manchetes pode não estar compatível com aquela pretendida pela organização – que é balizada em equilíbrio, transparência e comprometimento com o interesse público (ANEEL, 2021). Também foi possível identificar que os veículos de imprensa utilizam estratégias de comunicação ao publicarem sobre a ANEEL, sendo as mais frequentes *agenda-setting*, *linkage* e *framing*.

Este estudo apresentou algumas limitações. A base de dados apresentou alguns problemas no que diz respeito a filtros e *tags* desajustados, o que oca-

sionou a presença de algumas notícias pouco relacionadas com o setor ou à Agência, devidamente filtradas. Além disso, os dados utilizados são de 2020, ano bastante atípico, marcado pelo desencadeamento de uma pandemia global e por episódios pontuais relacionados ao setor de energia, o que pode ter influenciado os resultados. Ainda, os portais de notícias e jornais não são as únicas fontes de informações que podem influenciar a imagem organizacional. Logo, para estudos futuros, sugere-se a utilização de outras fontes de dados e a validação a partir de outros casos.

Uma boa imagem organizacional, construída em tempo oportuno e alinhada com a estratégia de comunicação constituirá um referencial positivo para a organização em todos os aspectos. Por isso, para estudos futuros, também há a oportunidade de realizar o monitoramento do horizonte, no sentido de identificar, compilar e analisar os vários sinais de mudança (Hines *et al.*, 2018). A proposta de analisar a imagem organizacional a partir de notícias veiculadas em diversos canais dá abertura para que a organização, a partir destes mesmos dados, realize uma nova análise, mas agora com o intuito de coletar *insights* e identificar tendências disruptivas que, em algum momento, podem impulsionar mudanças e questões emergentes que impactam a organização.

REFERÊNCIAS

- Aggarwal, C. C. & Zhai, C. X. (2012). A survey of text classification algorithms. In *Mining Text Data* (pp. 163-222). Springer.
- Altszyler, E., Ribeiro, S., Sigman, M., & Slezak, D. F. (2017). The interpretation of dream meaning: Resolving ambiguity using Latent Semantic Analysis in a small corpus of text. *Consciousness and Cognition: An International Journal*, 56, 178–187.
- Agência Nacional de Energia Elétrica (2021). A ANEEL. <https://www.aneel.gov.br/a-aneel>.
- Ashton, T., Evangelopoulos, N. and Prybutok, V. (2014), Extending monitoring methods to textual data: a research agenda, *Quality and Quantity*, Vol. 22, No. 4, pp. 2277-2294.
- Barich, H. & Kotler, P. (1991). A framework for marketing image management. *Sloan Management Review*, 32 (2), 94–104.
- Bartoluzzio, A. I. S. S., Dos Anjos, L. C. M., Dos Santos, M. R. L. & Freire, R. S. (2019). Percepção dos Cidadãos sobre os Portais da Transparência do Governo Estadual e das Gestões Municipais em Pernambuco. *Revista da CGU*, 11 (18), 1115-1133.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.

- Bromley, D. B. (1993). Reputation, image and impression management. John Wiley & Sons.
- Cezar, L. C. (2019). Comunicação e marketing no setor público: diferentes abordagens para a realidade brasileira. Enap.
- Chuprina, S., Postanogov, I., & Kostareva, T. (2017). A way how to impart data science skills to computer science students exemplified by obda-systems development. *Procedia Computer Science*, 108, 2161- 2170.
- Cielen, D. & Meysman, A. (2016). Introducing data science: *big data*, machine learning, and more, using Python tools. Simon and Schuster.
- Crain, S. P., Zhou, K., Yang, S. H. & Zha, H. (2012). Dimensionality reduction and *topic modeling*: From latent semantic indexing to latent *dirichlet allocation* and beyond. In *Mining Text Data* (pp. 129-161). Springer.
- Da Silva, M. S. P. & Gutierrez, R. H. (2013). Avaliação da visão interna da imagem de autarquias federais: Estudo de caso da Comissão Nacional de Energia Nuclear. *Sustainable Business International Journal*, 30, 01-25.
- Da Fonseca, L. C. L. (2020). Uma análise da reputação das empresas do PIM na mídia. *Revista Eletrônica – Arte, Educação, Comunicação e Design*, 1 (3), 1-30.
- De Sousa, C. M., Ardigo, J. P. & Santos, A. R. (2021). A produção acadêmica nacional e internacional sobre imagem corporativa no setor público. *Gestão e Sociedade*, 15 (43), 4466-4496.
- Dowling, G. R. (1986). Managing your corporate image. *Industrial Marketing Management*, 15(2), 109–115.
- Eickhoff, M. & Neuss, N. (2017, June 5-10). Topic modelling methodology: its use in information systems and other managerial disciplines. *Proceedings of the 25th European Conference on Information Systems (ECIS)*, Guimarães, Portugal.
- Fan, W. & Gordon, M. D. (2014). The Power of Social Media Analytics. *Communications of the ACM*, 57 (6), 74-81.
- Hines, A., Bengston, D. N., Dockry, M. J. & Cowart, A. (2018). Setting Up a Horizon Scanning System: A U.S. Federal Agency Example. *World Futures Review*, 10(2), 136-151.
- Hjarvard, S. (2012). Mídiação: teorizando a mídia como agente de mudança social e cultural. *Matrizes*, 5 (2), p. 53-91.
- Kuyper, J. A. (2002). *Press Bias and Politics: How the Media Frame Controversial Issues*. Praeger.
- Landauer, T. K. (2007). LSA as a theory of meaning. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 3–34). Lawrence Erlbaum Associates Publishers.
- Langley, A. & Klag, M. (2019). Being where? Navigating the involvement paradox in qualitative research accounts. *Organizational Research Methods*, 22(2), 515-538.
- Lima Jr., A. V. D. (2020), Metric for selecting the number of topics in the LDA Model, Doctoral Dissertation, Federal University from Rio Grande do Sul, Porto Alegre, RS, Brazil.
- Maia, R. C. M. (2009). Mídia e diferentes dimensões da Accountability. *Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação*, 2.
- Manning, C. D., Raghavan, P. & Schütze H. (2010). *Introduction to information retrieval*. Cambridge University Press.
- Marcolin, C. B., Becker, J. L., Wild, F., Behr, A. & Schiavi, G. (2021). Listening to the voice of the guest: A framework to improve decision-making processes with text data. *International Journal of Hospitality Management*, 94, 1-12.
- Martins, S. A. (2014). Mídia e Opinião Pública: Estudo de Caso sobre o Mensalão nas Ópticas dos jornais Folha de São Paulo e o Estado de São Paulo. *Universitas: Arquitetura e Comunicação Social*, 11 (2), p. 47-58.

- Maulud, H. D., Zeebaree, S. R. M., Jacksi, K., Sadeeq, M. A. M. & Sharif, K. H. (2021). State of Art for Semantic Analysis of Natural Language Processing. *Qubahan Academic Journal*, 1(2), 21–28.
- Nikolenko, S. I., Koltcov, S., & Koltsova, O. (2017). Topic modelling for qualitative studies. *Journal of Information Science*, 43(1), 88-102.
- Otto, L. P., Thomas, G., Glogger, I. & De Vreese, C. H. (2022). Linking Media Content and Survey Data in a Dynamic and Digital Media Environment – Mobile Longitudinal *Linkage* Analysis. *Digital Journalism*, 10 (1), 200-215.
- Stevens, K., Kegelmeyer, P., Andrzejewski, D. and Buttler, D. (2012), Exploring topic coherence over many models and many topics, in Proceedings of the 2012 EMNLP-IJCNLP, Association for Computational Linguistics, Stroudsburg, PA, United States, pp. 952-961.
- Schmiedel, T., Müller, O. & Brocke, J. (2019). *Topic modeling* as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, 22(4), 941-968.
- Schuler, M. (2004). Management of the organizational image: a method for organizational image configuration. *Corporate Reputation Review*, 7 (1), 37-53.
- Steppacher, D. (2012). Pesquisa de imagem organizacional de um órgão público federal: a Escola de Administração da Universidade Federal do Rio Grande do Sul sob a ótica de seus alunos de graduação. 2012. [Especialização em Marketing, Universidade Federal do Rio Grande do Sul - UFRGS].
- Stieglitz, S., Mirbabaie, M., Ross, B. & Neuberger, C. (2018). Social media analytics: Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39, 156–168.
- Tabesh, P., Mousavidin, E. & Hasani, S. (2019). Implementing *big data* strategies: A managerial perspective. *Business Horizons*, 62(3), 347-358.
- Valerio, D. M. & Pizzinatto, N. K. (2003). Análise da Imagem Organizacional de Universidades por Meio da Matriz Familiaridade-Favorabilidade. *Revista de Administração Mackenzie*, 4 (1), 26-37.
- Zhang, J., Wu, B., Morrison, A. M., Tseng, C. & Chen, Y. C. (2016). How country image affects tourists' destination evaluations. *Journal of Hospitality & Tourism Research*, 42(6), 904–930.



Carolina Coelho da Silveira

Universidade Federal de Uberlândia

carolinasilveira@ufu.br

<https://orcid.org/0000-0001-6365-9871>

Mestranda no Programa de Pós-Graduação em Administração na Universidade Federal de Uberlândia (UFU), na área de Gestão Organizacional e Regionalidade. Graduada em Administração em 2022 pela UFU.



Carla Bonato Marcolin

Universidade Federal de Uberlândia

carla@ufu.br

<http://orcid.org/0000-0003-0260-5073>

Doutora em Administração (Linha Pesquisa Operacional) e Mestre em Administração (Linha Gestão de Sistemas e TI) pelo Programa de Pós-Graduação em Administração da UFRGS. Possui especialização em Finanças pela mesma instituição, e graduação pela Pontifícia Universidade Católica do Rio Grande do Sul (2010). Atualmente é professora da Faculdade de Gestão e Negócios (FAGEN) da Universidade Federal de Uberlândia (UFU) e realizou pós-doutoramento na Fundação Getúlio Vargas (EAESP/FGV). Trabalhou com análise e desenvolvimento de sistemas personalizados para PMEs e com gestão de contratos com órgãos públicos. Trabalha com R e possui familiaridade com SQL e estrutura de dados, bem como com técnicas de Processamento de Linguagem Natural (NLP) e de Aprendizado de Máquina (Machine Learning). Principais interesses de pesquisa incluem inteligência artificial, analytics, mineração de dados e principalmente mineração de texto.

**Carlos Henrique Rodrigues**

Agência Nacional de Energia Elétrica (ANEEL) e Faculdade SENAC-DF
caique@aneel.gov.br

Mestre em Administração (2012) pela Universidade de Brasília, Especialista em Gestão Empresarial (2003) e Graduado em Administração (2000), pela Universidade Federal de Uberlândia. Chefe Adjunto de Assessoria de Gestão na Agência Nacional de Energia Elétrica (ANEEL), atua na coordenação de atividades ligadas ao planejamento estratégico, gestão de processos, projetos e orçamento na Assessoria Institucional da Diretoria. Possui experiência profissional e na docência de ensino superior e pós-graduação, e em atividades de instrutoria na área de Administração e Gestão de Pessoas, atuando em temas como: cargos e salários, avaliação de desempenho, cultura organizacional, carreira profissional, qualidade de vida, treinamento, desenvolvimento e educação corporativa, satisfação, qualidade de vida no trabalho e desenvolvimento.

Estimando a Arrecadação da Dívida Ativa da União com Machine Learning: Uma análise baseada nos dados de arrecadação do período de 2015 a 2021¹

Estimating the Union's Active Debt Revenues with Machine Learning: An analysis based on tax revenues data from 2015 to 2021.

Estimación de los ingresos de la deuda activa de la Unión con Machine Learning: un análisis basado en datos de ingresos fiscales de 2015 a 2021.

Rubens Quaresma Santos

<https://doi.org/10.36428/revistadacgu.v14i26.529>

Resumo: Como entidade responsável pela cobrança da Dívida Ativa da União, a Procuradoria-Geral da Fazenda Nacional deve apresentar, ao final de cada exercício fiscal, os resultados alcançados e as previsões de arrecadação para o ano seguinte para a composição das Leis Orçamentárias vindouras. Atualmente essa estimativa é elaborada utilizando a técnica da suavização exponencial, que considera os ingressos passados para projetar a arrecadação futura. O presente trabalho procura avaliar a aplicação de algoritmos de aprendizado de máquina na elaboração dessas projeções, como meio de aprimoramento da gestão. Foram testados os algoritmos Regressão Linear, Árvore de Decisão, Floresta Randômica e Árvore de Decisão de Aumento de Gradiente. Os modelos foram alimentados com informações dos indicadores macroeconômicos IPCA, IGP-M, PIB, taxa de Câmbio e taxa SELIC, além de dados relativos a parcelamentos excepcionais e transações tributárias disponibilizados aos contribuintes pela Fazenda Nacional.

Palavras-chave: Dívida Ativa. Arrecadação Fiscal. Orçamento. Machine Learning. Ciência de Dados.

Abstract: As the entity responsible for collecting the Active Debt of the Union, the Attorney General's Office of the National Treasury must present, at the end of each fiscal year, the results achieved and the collection forecasts for the following year for the composition of forthcoming Budget Laws. Currently, this estimate is made using the exponential smoothing technique, which considers past income to project future revenue. This article seeks to evaluate the application of machine learning algorithms in the elaboration of these projections, as a means of improving management. Linear Regression, Decision Tree, Random Forest and Gradient Boosting Decision Tree algorithms were tested. The models were fed with information from the macroeconomic indicators IPCA, IGP-M, GDP, Exchange rate and SELIC rate, as well as data on exceptional installments and tax transactions made available to taxpayers by the National Treasury.

Keywords: Active debt. Tax Collection. Budget. Machine Learning. Data Science.

Resumen: La Procuraduría General del Tesoro Nacional, cómo ente encargado de recaudar la Deuda Activa de la Unión debe presentar, al cierre de cada ejercicio fiscal, los resultados alcanzados y las previsiones de recaudación del ejercicio siguiente para la composición de las próximas Leyes de Presupuesto. Actualmente, esta estimación se

1. Artigo submetido em 11/07/2022 e aceito em 01/12/2022.

realiza mediante la técnica de suavizamiento exponencial, que considera los ingresos pasados para proyectar los ingresos futuros. Este artículo busca evaluar la aplicación de algoritmos de aprendizaje automático en la elaboración de estas proyecciones, cómo medio para mejorar la gestión. Se probaron los algoritmos de regresión lineal, árbol de decisión, bosque aleatorio y árbol de decisión de aumento de gradiente. Los modelos fueron alimentados con información de los indicadores macroeconómicos IPCA, IGP-M, PBI, Tipo de Cambio y Tasa SELIC, así como datos de cuotas excepcionales y operaciones tributarias puestas a disposición de los contribuyentes por parte del Tesoro Nacional.

Palabras clave: Deuda activa. Recaudación de impuestos. Presupuesto. Aprendizaje automático. Ciencia de los datos.

1. INTRODUÇÃO

Na organização administrativa do Estado brasileiro, a Procuradoria-Geral da Fazenda Nacional (PGFN) é a instituição competente para representar a União na execução da sua dívida ativa, conforme expressa dicação do §3º, do art. 131, da Constituição Federal de 1988.

Sendo o Órgão incumbindo da apuração da liquidez, certeza e exigibilidade dos créditos encaminhados para inscrição no passível recebível da União, bem como sua respectiva cobrança, a PGFN é chamada, ao final de cada exercício financeiro, a estimar a arrecadação dos valores da Dívida Ativa da União - DAU para o ano seguinte.

Essa previsão é de suma importância para a composição da Lei Orçamentária Anual e, consequentemente, para a formulação de planos e projetos do Estado Brasileiro. Logo, a precisão e certeza dessa previsão é fundamental, uma vez que é através dela e dos prognósticos de outras fontes de receita que o Poder Executivo Federal elaborará o orçamento governamental e, em última análise, balizará as verbas a serem destinadas às diversas políticas públicas em curso.

Como é possível averiguar da Nota SEI nº 29/2021/COAGED/CDA/PGDAU/PGFN-ME², os cálculos atuais tomam por base os valores arrecadados no ano anterior e utilizam a metodologia de suavização exponencial para projetar as entradas para o ano seguinte. Este artigo procura avaliar se é possível, utilizando modelos de Machine Learning, aprimorar essas estimativas arrecadatórias.

As técnicas de aprendizado de máquina têm se mostrado eficientes em avaliar quantidades significativas de dados e encontrar padrões que podem ser utilizados para elaboração de projeções e previsões em

diversas áreas, não existindo, a priori, qualquer razão que justifique a Administração Pública não se valer dessas ferramentas para aprimorar sua gestão.

1.1. Suposições e Pressupostos

A premissa deste estudo é que a consideração de variáveis macroeconômicas pode servir de sinalizador dos montantes arrecadados pela DAU, porquanto elas registram características gerais da situação financeira dos agentes econômicos nacionais e a tributação significa, essencialmente, a captação de recursos por meio da incidência de exações sobre a atividade produtiva.

Por essa razão, na implementação dessa avaliação são considerados o Índice Nacional de Preços ao Consumidor - IPCA, elaborado pelo Instituto Brasileiro de Geografia e Estatística - IBGE; o Índice Geral de Preços do Mercado - IGP-M, apurado pela Fundação Getúlio Vargas - FGV; a flutuação cambial do Dólar; a taxa SELIC; e a variação do Produto Interno Bruto - PIB.

O IPCA registra a inflação da nossa moeda, é calculado com base na variação de preços de um conjunto de produtos e serviços comercializados no varejo e é um importante indicador do custo do capital em nossa sociedade (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2022a). O IGP-M, ao seu turno, acompanha a variação de preços no mercado e toma em consideração o valor de matérias primas agrícolas e industriais, constitui um indicativo da atividade econômica do país (FUNDAÇÃO GETÚLIO VARGAS, 2022).

A consideração da variação cambial é relevante por registrar um comparativo do valor de nossa moeda em face do Dólar, que a partir da conferência de Bretton Woods em 1944 passou a balizar as transações comerciais internacionais (KRUGMAN E OBSTFELD, 2005,

2. Documento obtido via da Lei de Acesso à Informação, pedido nº. 03005.291447/2022-40.

p. 408). Dado que parte significativa dos produtos e serviços produzidos e consumidos no Brasil são negociados com parceiros estrangeiros (BANCO CENTRAL DO BRASIL, 2022a), sua ponderação representa a consideração dos reflexos do comércio internacional nos custos produtivos e na inflação nacional.

A SELIC é a taxa básica de juros da economia definida pelo Banco Central do Brasil (BC). É um instrumento de política monetária, definidor do custo básico do crédito e utilizada como uma das ferramentas de controle da inflação (BANCO CENTRAL DO BRASIL, 2022b). É importante registrar que todas as dívidas inscritas em DAU são atualizadas monetariamente pela SELIC, por expressa determinação do artigo 30, da Lei nº 10.522, de 19 de julho de 2002.

O PIB é o medidor básico do crescimento econômico de um país, é o produto de toda a riqueza produzida em solo nacional (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2022b). Um PIB positivo e crescente aponta uma economia pujante.

Para além de serem alguns dos principais indicadores macroeconômicos, a escolha desses parâmetros se justifica, acima de tudo, porque suas projeções futuras são apuradas e divulgadas semanalmente pelo Banco Central do Brasil na publicação Relatório Focus (<https://www.bcb.gov.br/publicacoes/focus>). Logo, se o presente estudo busca indicar um caminho eficiente para a projeção da arrecadação, prudente se valer de indicadores cujas estimativas são calculadas pelos principais agentes da economia.

A arrecadação da DAU, entretanto, não decorre apenas de fatores externos, estranhos à atividade da Instituição responsável por sua cobrança. Há que se considerar, ainda, ações implementadas pelo Estado para ampliar a arrecadação.

De modo geral, a cobrança da dívida ativa é feita através de ações administrativas e do ajuizamento de execuções fiscais. A utilização de mecanismos de remissão e diferimento da dívida também são muitas vezes utilizados pela União para tentar ampliar a entrada de recursos. Tais medidas se dão, em sua maioria, pela instituição de parcelamentos extraordinários (PROCURADORIA-GERAL DA FAZENDA NACIONAL, 2021b). Mais recentemente, a Lei nº 13.988/2020 regulamentou e autorizou à PGFN a implantação de medidas de transação fiscal, que podem ser simplificadas como negociações individualizadas com o contribuinte devedor para adoção de medidas e planos personalizados de pagamento (PROCURADO-

RIA-GERAL DA FAZENDA NACIONAL, 2021c). Por essa razão, pertinente que essas ações também sejam consideradas na modelagem do estudo, dado seu potencial impacto na arrecadação.

2. METODOLOGIA

Segundo Sinan Ozdemir (2016, p. 202), o Aprendizado de Máquina ou Machine Learning se constitui numa ferramenta da Ciência de Dados preocupada com a capacidade de encontrar padrões de dados, mesmo que eles contenham erros e ruídos. Modelos de aprendizado de máquina são capazes de extrair conhecimento de amostras sem uma ajuda humana direta, o que lhes diferencia de algoritmos tradicionais. A maioria dos algoritmos de aprendizado de máquina focam em encontrar relacionamentos entre conjuntos de dados para identificar correlações que possam ser utilizadas para prever observações futuras.

Moreira, Carvalho e Horváth (2018) explicam que as tarefas de predição podem ser divididas em problemas de classificação e problemas de regressão. Na classificação, o objetivo primordial é que o algoritmo seja capaz de atribuir um rótulo (ex.: identificar, numa foto, se o que está representado é uma paisagem, um objeto, um animal etc.). Na regressão, busca-se a criação de uma função matemática que possa explicar uma saída a partir de um conjunto de atributos (ex.: estimar o tempo de viagem, dado um determinado caminho e as condições climáticas registradas).

Algoritmos de regressão têm sido usados em diversas áreas nos últimos anos, em alguns casos para estimar o valor futuro do mercado de ações (SANTOS, 2020), a evasão de clientes de um determinado produto (OLIVEIRA, 2021) e até a inflação da economia (FREITAS, 2019).

Portanto, para o objetivo pretendido, o mais adequado é o uso de algoritmos de regressão. Esse estudo é inovador em relação ao objeto, mas não quanto à técnica. Não se encontram trabalhos similares envolvendo a Dívida Ativa, mas os modelos utilizados são de amplo conhecimento, tendo sido testados em diversos conjuntos de dados, envolvendo áreas distintas.

Os arquivos das bases de dados utilizadas e os códigos de programação para treinamento e teste dos modelos estão disponíveis para consulta em repositório público no Github no endereço <https://github>.

[com/search?q=estimativa_dau](#). Nos tópicos seguintes são detalhados os algoritmos escolhidos e a forma de obtenção e respectivas fontes dos dados.

2.1. Modelos de Machine Learning Utilizados

Existem diversos métodos de aprendizado de máquina para análise de dados automatizada que podem ser utilizados com fins preditivos.

Kelleher e Tierney (2018, p. 70) explicam que para algumas áreas há razões teóricas, baseadas no tipo particular de relacionamento entre os atributos e o valor que se pretende prever, que sugerem a aplicação de algoritmos específicos, todavia, na ausência desse tipo de teoria para o domínio em discussão, constitui uma boa prática assumir que a forma de relacionamento mais simples – o relacionamento linear – é suficiente, só então sendo pertinente progredir para modelos mais complexos.

Seguindo essa premissa, optou-se por avaliar o uso dos algoritmos de predição: *Regressão Linear (Linear Regression)*, *Árvore de Decisão (Decision Tree)*, *Floresta Randômica* ou *Floresta Aleatória (Random Forest)* e *Árvore de Decisão de Aumento de Gradiente (Gradient Boosting Decision Tree)*.

A *Regressão Linear* procura estimar um valor, dado um conjunto de atributos fixos, a partir da identificação de uma função de regressão (KELLEHER E TIERNEY, 2018, p. 70), ou seja, uma equação linear do tipo:

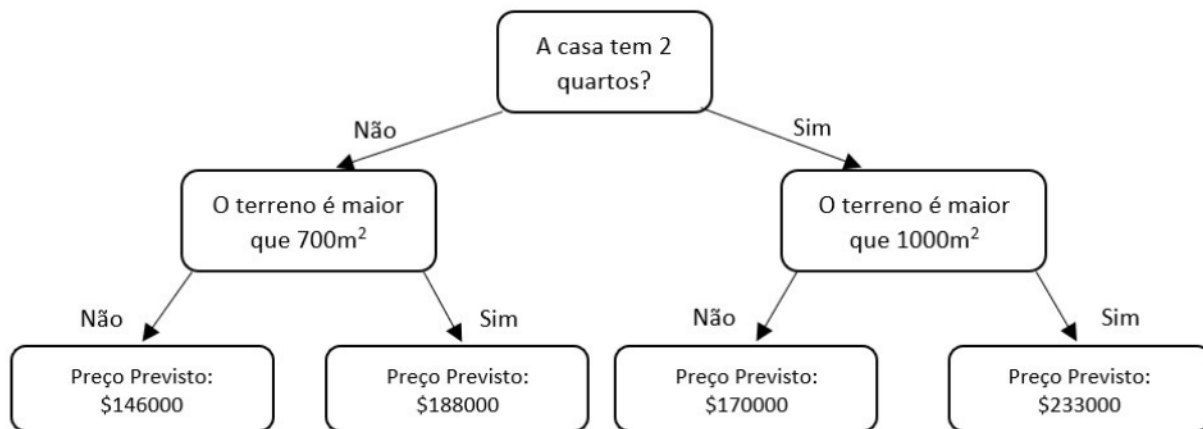
$$Y = w_0 + w_1X_1 + w_2X_2 + \dots + w_aX_a$$

Basicamente, trata-se de uma equação que considera uma ou mais variáveis (X_1, X_2, \dots, X_a) e seus respectivos coeficientes paramétricos ($w_0, w_1, w_2, \dots, w_a$) para encontrar o valor alvo (Y). O trabalho do algoritmo de aprendizado de máquina é, a partir do conjunto de dados, encontrar o coeficiente que influencia o peso de cada variável atributo da equação.

No modelo de *Árvore de Decisão*, para encontrar a fórmula preditiva são aplicadas uma série de classificações dos dados, camada a camada, de forma a melhor identificar a equação que se adequa aos atributos disponíveis (KELLEHER E TIERNEY, 2018, p. 82).

Utilizando o exemplo de Dan Becker (2017a), o modelo de árvore de decisão para previsão do valor de uma casa pode ser ilustrado da seguinte maneira:

FIGURA 1 - FUNCIONAMENTO DA ÁRVORE DE DECISÃO



Fonte: BECKER (2017a)

A última instância da árvore é denominada “folha” (BECKER, 2017a) e nela encontra-se o modelo preditivo que mais se adequa ao conjunto de dados, considerando as avaliações prévias feitas pelo algoritmo.

A *Floresta Randômica* faz uso de muitas árvores de decisão, cada uma treinada com um subconjunto dos dados disponibilizados para o algoritmo, a estimativa é feita calculando a média das previsões dessas várias árvores (KELLEHER E TIERNEY, 2018, p. 85).

Finalmente, a *Árvore de Decisão de Aumento de Gradiente* também faz uso de muitas árvores de decisão para conseguir um resultado satisfatório, mas de forma diferente do modelo *Floresta Randômica*. No GBDT (acrônimo para *Gradient Boosting Decision Tree*) as árvores de decisão são calculadas em sequência, cada uma levando em consideração o coeficiente de erro da árvore anterior, assim a precisão do modelo aumenta a cada nova árvore criada (ZHANG et al., 2021). Ao contrário da *Floresta Randômica*, em que as diversas árvores são calculadas de maneira independente e ao final a precisão é determinada pela ponderação da acurácia de cada uma das árvores geradas.

As breves explicações acima destinam-se a elucidar minimamente o funcionamento dos modelos escolhidos, porém não constitui escopo desse estudo analisar os teoremas matemáticos e a álgebra linear requerida para calcular e elaborar as previsões. Atualmente, com a popularização e simplificação dos modelos computacionais, é possível utilizar ferramentas gratuitas e de código aberto para executar essas análises.

Em razão do seu alto nível, simplicidade e ênfase na legibilidade (FURTADO, 2022), optou-se por gerar os modelos com a linguagem de programação *Python*, que tem na biblioteca de código aberto *Scikit-learn* todo o ferramental necessário para as projeções pretendidas (PEDREGOSA, 2011). O estudo se vale ainda da biblioteca *Pandas*, também gratuita e de código aberto, criada para a linguagem *Python*, focada na manipulação e análise estatística de dados (MCKINNEY, 2010).

Além disso, as simulações foram executadas no ambiente *Google Colaboratory*, serviço gratuito disponibilizando on-line para escrever e executar códigos de programação em ambiente de nuvem, que dispensa o uso de computadores com alta capacidade, já que todo o processamento é feito pelos servidores do Google (2022).

2.2. Levantamento, Tratamento e Exploração dos Dados

Esclarecidos os algoritmos, passemos a detalhar os dados e suas fontes. A análise proposta leva em conta os indicadores macroeconômicos projetados semanalmente no Relatório Focus divulgado pelo

Banco Central do Brasil: IPCA, IGPM, SELIC, taxa de Câmbio e PIB. Juntamente com esses dados foram considerados os parcelamentos excepcionais e as transações de créditos tributários efetivados pela PGFN.

Os indicadores macroeconômicos foram obtidos no site do BC, na ferramenta *Sistema Gerenciador de Séries Temporais – SGS* (<https://www3.bcb.gov.br/sgspub/>). Especificamente, foram capturadas as planilhas de códigos: “4380 - PIB mensal - Valores correntes (R\$ milhões) – Função: Variação Percentual”; “13522 - Índice nacional de preços ao consumidor - amplo (IPCA) - em 12 meses – Função: Linear”; “4189 - Taxa de juros - Selic acumulada no mês anualizada base 252 - % a.a. – Função: Linear”; “3698 - Taxa de câmbio - Livre - Dólar americano (venda) - Média de período - mensal - Função: Linear”; e “189 - Índice geral de preços do mercado (IGP-M) - Var. % mensal – Função: Linear”.

Ao contrário dos demais índices que são apresentados em percentual, o dólar foi considerado em valor monetário. Embora fosse possível apurar a variação percentual dos respectivos períodos, apropriada sua consideração em moeda corrente porque esta é a forma como o Banco Central estima seu valor, ou seja, as previsões semanais são em reais (R\$) e não percentuais.

Os valores arrecadados mensalmente pela dívida ativa, entre janeiro de 2015 e dezembro de 2021, e as quantidades de adesões aos parcelamentos e transações foram adquiridos diretamente da PGFN, através de solicitação via Lei de Acesso à Informação, pedido nº. 03005.291447/2022-40, pesquisável em <http://www.consultaesic.cgu.gov.br/busca/>.

Um fator peculiar em relação a esses diferimentos e benefícios concedidos aos contribuintes é que, em geral, eles são casuísticos, não existindo um padrão absoluto a ser seguido pelo legislador, que no momento de sua aprovação pode alterar a proposta apresentada pelo Poder Executivo, que por sua vez também não segue, de forma definitiva, padrões específicos. A listagem inicial obtida registra, no intervalo avaliado, um total de 43 tipos diferentes de modalidades de parcelamentos e transações, o que comprova essa diversidade:

TABELA 1 - PARCELAMENTOS ESPECIAIS E TRANSAÇÕES REGISTRADOS NO PERÍODO DE 2015 A 2021

1	PROIES	23	Transação excepcional - demais débitos
2	Parcelamento da recuperação judicial	24	Transação excepcional - débitos previdenciários
3	Parcelamento PACAL	25	Transação excepcional - crédito rural e fundiário
4	PROFUT	26	Transação excepcional - simples nacional
5	Parcelamento da lei nº. 12.810/2013	27	Transação na dívida ativa tributária de pequeno valor - demais débitos
6	Liquidação crédito rural e fundiário	28	Transação na dívida ativa tributária de pequeno valor - débitos previdenciários
7	Parcelamento de arrematação	29	Transação na dívida ativa tributária de pequeno valor - simples nacional
8	Parcelamento especial - simples nacional	30	Repactuação - transação excepcional - demais débitos
9	Programa de regularização tributária - PRT	31	Repactuação - transação excepcional - débitos previdenciários
10	Parcelamento especial débitos previdenciários dos entes federativos	32	Repactuação - transação excepcional - crédito rural e fundiário
11	Programa especial de regularização tributária - PERT - demais débitos	33	Repactuação - transação excepcional - simples nacional
12	Programa especial de regularização tributária - PERT - débitos previdenciários	34	Transação no contencioso tributário
13	Programa de recuperação tributária rural	35	Transação do setor de eventos - PERSE - demais débitos
14	Parcelamento especial de regularização tributária - lei 13.496/2017 - demais débitos	36	Transação do setor de eventos - PERSE - débitos previdenciários
15	Parcelamento da lei n. 13586, de 28 de dezembro de 2017	37	Transação do setor de eventos - PERSE - simples nacional
16	Programa especial de regularização tributária - PERT - débitos não quitados no parcelamento 12.996	38	Repactuação - transação do setor de eventos - PERSE - demais débitos
17	Programa especial de regularização tributária Simples Nacional - PERT SN	39	Repactuação - transação do setor de eventos - PERSE - débitos previdenciários
18	Parcelamentos diversos	40	Repactuação - transação do setor de eventos - PERSE - simples nacional
19	Parcelamento de honorários advocatícios	41	Transação para regularização fiscal do simples nacional
20	Transação - demais débitos	42	Transação - dívida ativa tributária de pequeno valor - simples nacional - edital 1/2022
21	Transação - débitos previdenciários	43	Parcelamento excepcional de débitos previdenciários para municípios - EC 113/2021
22	Transação individual		

Fonte: Elaborado pelo autor

Dado o objetivo proposto, de viabilizar, para o futuro, uma metodologia de estimativa para a arrecadação, considerar cada parcelamento e transação em si não é a melhor alternativa, uma vez que, pelo histórico observado, muitas dessas medidas não se repetirão *ipsis litteris*, ao passo que outras poderão surgir em seu lugar, com suas próprias peculiaridades.

Isso em mente, tornou-se essencial sistematizar um modelo de categorização, de forma a permitir uma fácil classificação dos diferimentos tributários. Avaliando a lista de medidas levadas a efeito pelo poder público, é possível constatar algumas características. Nota-se que parcelamentos e transações podem

atingir um setor específico da economia ou abranger um número genérico de contribuintes. Mas mesmo os genéricos podem ter sua aplicação direcionada, incidindo apenas sobre as empresas classificadas como integrantes do Simples Nacional ou abarcarem apenas débitos definidos como de pequeno valor.

Assim, propõem-se uma lista fechada de classificação que abarca todos os parcelamentos e transações que já foram efetivados, ao mesmo tempo, é capaz de enquadrar futuras medidas. Concebeu-se um sistema de representação em três níveis, considerando

o tipo de medida (parcelamento [P] ou transação [T]), sua abrangência (setorial [S] ou geral [G]) e sua setorialização.

No nível da setorialização, em sendo uma medida de abrangência geral, subclassifica-se em ampla [A] – sem qualquer restrição para adesão; simples nacional [SN] – destinada apenas às empresas enquadradas nessa modalidade de tributação; ou débitos de pequeno valor [PN] – abrangendo somente débitos considerados pequenos pelo Fisco.

Já na hipótese das medidas de natureza setorial, optou-se por utilizar o código de atividades econômicas do CNAE – Classificação Nacional de Atividades Econômicas, do Instituto Brasileiro de Geografia e Estatística, por se tratar de modelo bastante sedimentado

e utilizado no Cadastro Nacional de Pessoas Jurídicas da Receita Federal do Brasil (MINISTÉRIO DA ECONOMIA, 2014).

O próprio código CNAE é dividido em níveis, sendo a “seção” — designada por letras de A até U — destinada a representar o setor da atividade econômica desenvolvido pela empresa (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2022c). Utilizando especificamente essa informação, obtém-se então a abrangência necessária à classificação para permitir, no futuro, enquadrar novos tipos de diferimento tributário que possam ser criados.

Dessa forma, todos os parcelamentos e transações foram ordenados com base em um código de três símbolos:

TABELA 2 - ESQUEMA DE CATEGORIZAÇÃO DOS PARCELAMENTOS E TRANSAÇÕES

1º Símbolo	Tipo:	[P] parcelamento ou [T] transação
2º Símbolo	Abrangência:	[S] setorial ou [G] geral
3º Símbolo	Setorialização:	[SN] simples nacional, [PV] pequeno valor, [A] amplo ou [seção] código CNAE

Fonte: Elaborado pelo autor

O PROIES, por exemplo, sendo um parcelamento destinado a atender as empresas do setor de ensino, foi enquadrado no código P-S-P. A primeira letra referente ao tipo de medida (P – parcelamento), a segunda letra representando sua abrangência (S – setorial) e a terceira letra a seção CNAE a que pertence (P – educação). As transações viabilizadas para as empresas enquadradas no Simples Nacional, ao seu turno, receberam o código T-G-SN, ou seja, [T] de transação, [G] por ter abrangência geral e [SN] porque se destina apenas aos débitos do Simples Nacional.

Por essa sistemática é possível a criação de um total de 47 códigos, com os registros de todos os parcelamentos e transações detalhadas até o nível do setor da economia beneficiado. Contudo, grande parte dessas classes servem para o futuro, mas não possuem registros passados para alimentar os modelos. Assim, apenas as 12 classes que registraram dados, entre 2015 e 2021, foram usadas como atributo para os algoritmos de regressão:

TABELA 3 - CÓDIGO DE ENQUADRAMENTO DOS PARCELAMENTOS E TRANSAÇÕES

CÓD.	CATEGORIZAÇÃO
P-S-A	Parcelamento Setorial Agronegócio
P-S-B	Parcelamento Setorial Indústrias Extrativas
P-S-O	Parcelamento Setorial Administração Pública
P-S-P	Parcelamento Setorial Educação
P-S-R	Parcelamento Setorial Artes, Cultura, Esporte e Recreação
P-G-SN	Parcelamento Geral Simples Nacional
P-G-A	Parcelamento Geral Amplo
T-G-A	Transação Geral Ampla

CÓD.	CATEGORIZAÇÃO
T-G-SN	Transação Geral Simples Nacional
T-G-PV	Transação Geral Pequeno Valor
T-S-A	Transação Setorial Agronegócio
T-S-R	Transação Setorial Artes, Cultura, Esporte e Recreação

Fonte: Elaborado pelo autor

Foram agrupados em cada código os seguintes parcelamentos e transações:

TABELA 4 - AGRUPAMENTO DOS PARCELAMENTOS E TRANSAÇÕES POR CÓDIGO

CÓD.	PARCELAMENTOS OU TRANSAÇÕES AGRUPADOS
P-S-A	Parcelamento PACAL;
	Liquidação Crédito Rural e Fundiário;
	Programa de recuperação tributária rural;
P-S-B	Parcelamento da lei nº. 13.586/2017;
P-S-O	Parcelamento de Recuperação Judicial;
	Parcelamento da LEI nº. 12.810/2013;
	Parcelamento de Arrematação;
	Parcelamento Especial Débitos Previdenciários dos Entes Federados;
	Parcelamento de honorários advocatícios;
	Parcelamento excepcional de débitos previdenciários para municípios - EC 113/2021;
P-S-P	PROIES;
P-S-R	PROFUT;
P-G-SN	Parcelamento Especial – Simples Nacional;
	Programa especial de regularização tributária Simples Nacional - PERT SN;
P-G-A	Programa de Regularização Tributária – PRT;
	Programa especial de regularização tributária - PERT - demais débitos;
	Programa especial de regularização tributária - PERT - débitos previdenciários;
	Parcelamento especial de regularização tributária - lei 13.496/2017 - demais débitos;
	Programa especial de regularização tributária - PERT - déb. não quitados no parc. 12.996;
	Parcelamentos diversos;
T-G-A	Transação - demais débitos;
	Transação - débitos previdenciários;
	Transação individual;
	Transação excepcional - demais débitos;
	Transação excepcional - débitos previdenciários;
	Repactuação - transação excepcional - demais débitos;
	Repactuação - transação excepcional - débitos previdenciários;
	Transação no contencioso tributário;
T-G-SN	Transação excepcional - simples nacional;
	Repactuação - transação excepcional - simples nacional;
	Transação para regularização fiscal do simples nacional;

CÓD.	PARCELAMENTOS OU TRANSAÇÕES AGRUPADOS
T-G-PV	Transação na dívida ativa tributária de pequeno valor - demais débitos;
	Transação na dívida ativa tributária de pequeno valor - débitos previdenciários;
	Transação na dívida ativa tributária de pequeno valor - simples nacional;
	Transação - dívida ativa tributária de pequeno valor - simples nacional - edital 1/2022;
T-S-A	Transação excepcional - crédito rural e fundiário;
	Repactuação - transação excepcional - crédito rural e fundiário;
T-S-R	Transação do setor de eventos - PERSE - demais débitos;
	Transação do setor de eventos - PERSE - débitos previdenciários;
	Transação do setor de eventos - PERSE - simples nacional;
	Repactuação - transação do setor de eventos - PERSE - demais débitos;
	Repactuação - transação do setor de eventos - PERSE - débitos previdenciários;
	Repactuação - transação do setor de eventos - PERSE - simples nacional;

Fonte: Elaborado pelo autor

Esses atributos foram tratados como variáveis binárias, ou variáveis *dummy*, e nos meses em que havia registro de parcelamento ou transação disponível para adesão pelos contribuintes, de acordo com o código de classificação proposto, foi atribuído o valor 1, não estando disponível, recebeu o valor 0. Essa opção se deve ao fato de que a consideração da quantidade de contribuintes que aderiram a esses parcelamentos constitui uma explicação qualitativa e não quantitativa, pois não há uma relação perfeita entre a quantidade de aderentes e o valor total arrecadado pela DAU. O parcelamento de um único débito de uma grande empresa, como uma petrolífera ou uma instituição financeira, pode gerar o ingresso de um montante financeiro muito superior que o de milhares de contribuintes pessoas físicas. Levar em conta a quantidade de contribuintes aderentes poderia representar uma ponderação arbitrária, visto que estariam sendo tratadas igualmente variáveis em escalas quantitativas e qualitativas, o que é considerado um erro nas avaliações estatísticas (FÁVERO E BELFIORE, 2017, p. 541).

Ademais, sob a perspectiva da elaboração de projeções futuras, a PGFN não tem como precisar a quantidade de pessoas que irão aderir a um determinado parcelamento ou transação, pode apenas dizer se estará disponível num determinado mês.

Os modelos preditivos foram alimentados, então, com 17 atributos (IPCA, IGPM, SELIC, Taxa de Câmbio, PIB e 12 códigos de parcelamentos e transações) e tiveram como valor alvo o montante da arrecadação mensal da Dívida Ativa da União.

Considerando que os dados apresentavam métricas diferentes (percentual e valores absolutos), para que essa diferença não impactasse na previsão foi aplicada uma função de normalização, mais especificamente o método *min-max*, em que cada valor dos atributos é subtraído do valor mínimo do respectivo conjunto de dados e o resultado encontrado é dividido pela diferença entre o maior e o menor valor do conjunto (amplitude), isso traz todos os indicadores para valores entre 0 e 1 e nenhuma variável apresenta maior peso nos algoritmos somente em razão da discrepância de unidade de medida (MOREIRA *et al.*, 2018, p. 91).

Sumarizando os dados com métodos de estatística descritiva³, o período analisado é composto por 84 (oitenta e quatro) meses de registros, durante o qual as variáveis macroeconômicas consideradas apresentam as seguintes características:

3. Informações geradas com a função “describe” da biblioteca Pandas. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html>

TABELA 5 – ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS MACROECONÔMICAS - 2015 A 2021

	IPCA	IGP-M	SELIC	CÂMBIO	PIB
Média	5,74%	0,82%	8,17%	R\$ 4,02	0,59%
Desvio Padrão	2,85%	1,05%	4,29%	R\$ 0,88	3,76%
Menor Valor	1,88%	-1,10%	1,90%	R\$ 2,63	-10,80%
Maior Valor	10,74%	4,34%	14,15%	R\$ 5,65	10,50%
1º Quartil	3,26%	0,20%	4,98%	R\$ 3,26	-1,53%
2º Quartil	4,56%	0,65%	6,40%	R\$ 3,80	0,35%
3º Quartil	8,55%	1,19%	12,90%	R\$ 4,92	2,68%

Fonte: Elaborado pelo autor

Neste mesmo período, a arrecadação média mensal da dívida ativa foi de R\$ 1.945.842.649,48, com um desvio padrão de R\$ 801.734.014,24. A menor arrecadação em um único mês foi de R\$ 1.079.880.695,58, a maior R\$ 5.705.095.119,62 e o primeiro, segundo e terceiro quartis correspondentes a R\$ 1.314.979.048,29, R\$ 1.774.879.257,38 e R\$ 2.255.027.522,03, respectivamente.

Fávero e Belfiori (2017, p. 53 a 55) alertam que:

Um conjunto de dados pode conter algumas observações que apresentam um grande afastamento das restantes ou são inconsistentes. Estas observações são designadas por *outliers*, ou ainda por valores atípicos, discrepantes, anormais ou extremos.

Antes de decidir o que será feito com as observações *outliers*, devemos ter o conhecimento das causas que levaram a tal ocorrência. Em muitos casos, essas causas podem determinar o tratamento adequado dos respectivos *outliers*. As principais causas estão relacionadas a erros de medição, de execução e variabilidade inerente aos elementos da população.

(...)

Caso seja identificado apenas um *outlier* em

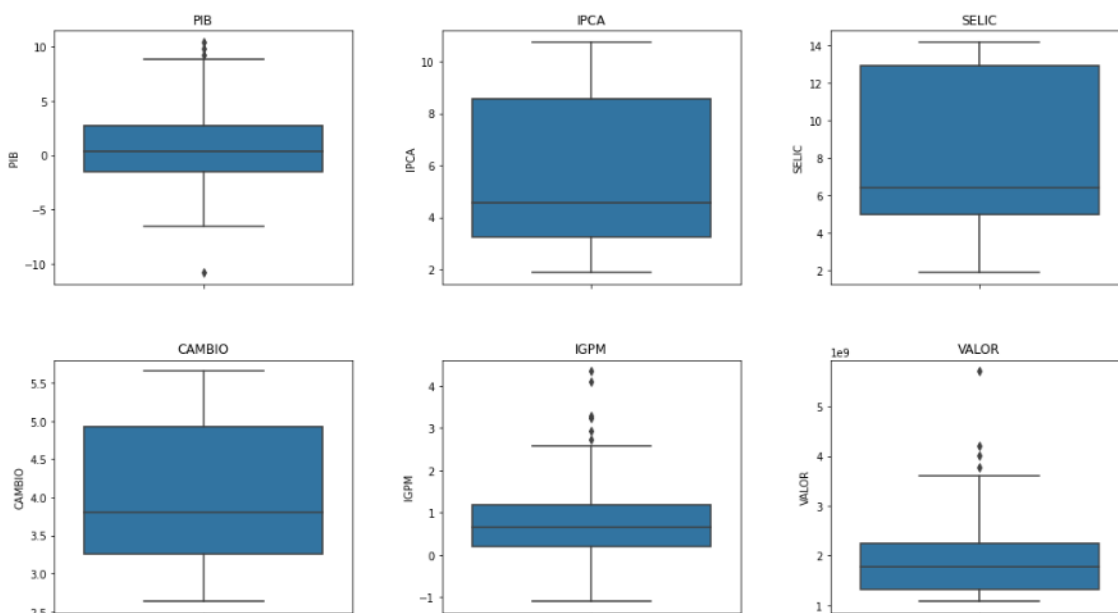
determinada variável, o pesquisador poderá tratá-lo por meio de alguns procedimentos existentes, por exemplo, a eliminação completa desta observação. Por outro lado, se houver mais de um *outlier* para uma ou mais variáveis individualmente, a exclusão de todas as observações pode gerar uma redução significativa do tamanho da amostra. (...).

(...), se a intenção for justamente analisar o comportamento dessas observações atípicas ou de criar subgrupos por meio de critérios de discrepância, talvez a eliminação dessas observações ou a substituição dos seus valores não seja a melhor solução.

Pela identificação da amplitude interquartil, que corresponde à diferença entre o terceiro e o primeiro quartil do conjunto de dados, foram considerados *outliers* os valores posicionados acima do terceiro quartil ou abaixo do primeiro quartil por uma margem superior a uma vez e meia a amplitude (FÁVERO E BELFIORE, 2017, p. 53).

Como mostram os gráficos do tipo *boxplot* abaixo, o conjunto de dados possui *outliers* na arrecadação (quatro registros superiores a R\$ 3,665 bi); no IGPM (seis registros superiores a 2,68%); e no PIB (três registros superiores a 9% e um registro inferior a -7,85%):

FIGURA 2 - BOXPLOT ATRIBUTOS QUANTITATIVOS



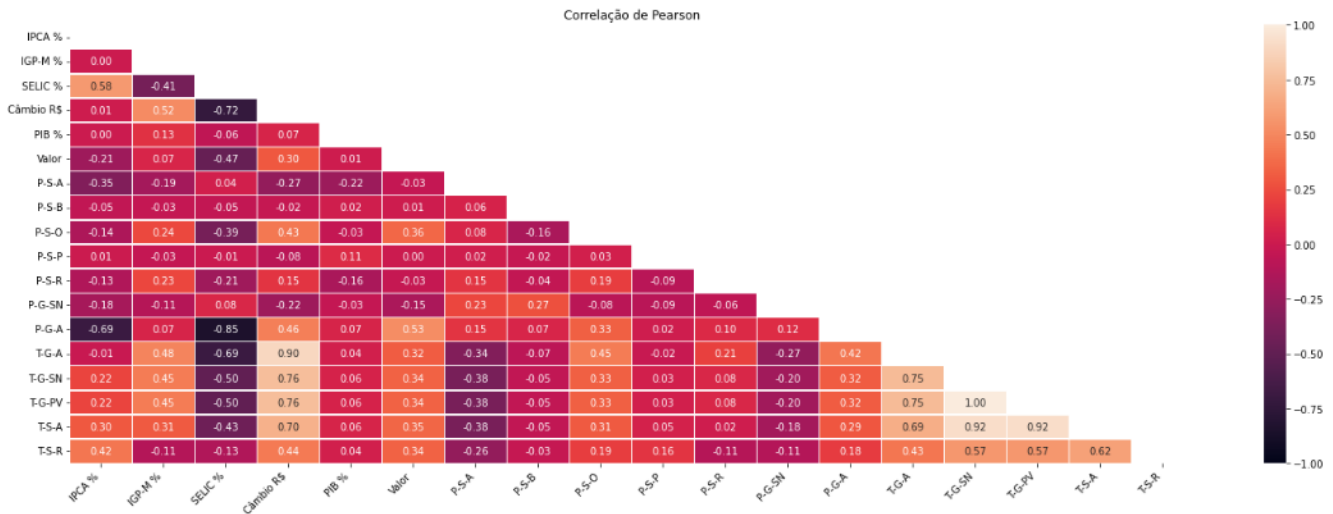
Fonte: Elaborado pelo autor

Esses valores, porém, não receberam qualquer tipo de tratamento, primeiro por se mostrarem registros válidos, constatações do mercado real, que apesar de incomuns não podem ser considerados equivocados; segundo, em razão da base de dados possuir uma quantidade limitada de informações, somente 84 meses, para a desconsideração desses valores seria necessária a exclusão de 14 meses da amostra, uma redução significativa e com provável impactando na precisão das estimativas.

Por derradeiro, calculou-se o *coeficiente de correlação de Pearson*⁴, que verifica o tipo de relação linear das variáveis. Sua medida varia entre -1 e 1, a aproximação dos extremos está vinculada a um aumento da intensidade da correlação, o sinal indica se a relação é direta ou inversa e quanto mais próximo de zero, menor a correlação (FÁVERO E BELFIORE, 2017, p. 118).

4. Coeficientes apurados com a função “.corr” do Pandas. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html>

FIGURA 3 - COEFICIENTE DE PEARSON



Fonte: Elaborado pelo autor

Seguindo o critério de avaliação qualitativa do grau de correlação entre as variáveis apresentado por Callegari-Jacques (2007, p. 90), em que um coeficiente 0 representa uma correlação nula; entre 0 e 0,3, fraca; acima de 0,3 até 0,6, regular; acima de 0,6 até 0,9, forte; acima de 0,9 até 1, muito forte; e 1 significa uma correlação plena ou perfeita; é possível afirmar que das variáveis macroeconômicas estudadas, a SELIC é a que apresenta uma correlação mais acentuada com a arrecadação (igual a -0,47), de grau regular e natureza inversa, isto é, o aumento da arrecadação está associado à redução da taxa básica de juros da economia. Câmbio, IGP-M e IPCA têm correlações fracas com a arrecadação (0,30, 0,07 e -0,21, respectivamente). Já com o PIB parece não existir qualquer correlação (0,01).

A visualização da correlação dos parcelamentos e transações com a arrecadação, ao seu turno, leva a constatações importantes. Enquanto todas as modalidades de transação apresentaram correlações regulares diretas com os movimentos da arrecadação, os parcelamentos excepcionais possuem comportamentos conflitantes. Há uma relação de grau regular entre os valores dos ingressos e os parcelamentos gerais amplos (0,53), os parcelamentos setoriais, contudo, não aparentam possuir relações significativas com a arrecadação, com exceção do parcelamento destinado ao setor público (P-S-O), que demonstrou um grau regular de correlação. O parcelamento geral

destinado aos débitos do simples nacional (P-G-SN) chegou a apresentar uma correlação negativa (-0,15), embora de grau fraco.

É fundamental lembrar, porém, que correlação não significa causalidade. O coeficiente de Pearson aponta o sentido e a intensidade com que as variáveis caminham na comparação mútua, mas não é possível afirmar, sem uma pesquisa mais aprofundada, o grau de determinação entre elas. Como alerta Paranhos (2014), acerca do perigo das correlações espúrias, dada a complexidade das interações sociais é difícil avaliar um fenômeno exclusivamente por meio de sua relação com uma única outra variável.

De todo modo, a partir dos dados analisados já é possível questionar o senso comum que atribui a todo e qualquer tipo de parcelamento excepcional uma ampliação da arrecadação federal, o que é comumente verbalizado quando das discussões desses tipos de diferimento no Congresso Nacional (AGÊNCIA SENADO, 2021). Neste sentido, as transações tributárias instituídas pela PGFN se mostraram mais consistentes, o que sugere que uma avaliação dos pontos convergentes e divergentes entre os dois institutos pode trazer insights para a formulação de políticas públicas mais profícuas nesta área.

3. RESULTADOS DOS MODELOS

Preliminarmente, salutar esclarecer que os modelos foram avaliados utilizando-se a técnica do *coeficiente de ajuste R^2* , que considera o quadrado da soma das distâncias entre os dados observados e a média desses dados (SQ_t), subtraído pelo quadrado da soma das distâncias entre os dados observados e os dados gerados pelo modelo de aprendizado de máquina (SQ_r). O número encontrado é dividido pelo quadrado da soma das distâncias entre os dados observados e a média desses dados (SQ_t), o que resulta num valor entre 0 e 1 (FÁVERO E BELFIORE, 2017, p. 522).⁵

O valor encontrado pode ser transformado em percentual bastando multiplicá-lo por 100. Esse percentual indica o quanto o modelo explica a variação dos dados, ou seja, um percentual de 75%, por exemplo, significa que o modelo preditivo explica a

variação dos dados em 75% das ocorrências, ou que é 75% mais preciso que simplesmente calcular a média dos valores observados.

Além disso, apurou-se o Erro Absoluto Médio (*Mean Absolute Error - MAE*)⁶, que é obtido pelo cálculo da média das diferenças absolutas entre os valores previstos pelo modelo e o valor correto (BECKER, 2017b). Portanto, no nosso caso, um MAE de 200.000.000 por exemplo significa que as previsões feitas pelo modelo estão, em média, até R\$ 200 mi distante do valor correto.

O treinamento dos modelos foi feito utilizando 75% da amostra (carga de treino) e a validação dos resultados considerou os 25% restantes (carga de teste). A separação dos dados entre as duas cargas se deu por meio de função automática da biblioteca *Scikit-learn*⁷, de modo randômico. A tabela a seguir condensa os melhores resultados obtidos:

TABELA 6 – RESULTADOS DOS TESTES DOS MODELOS

MODELO	R ²		ERRO MÉDIO ABSOLUTO
Regressão Linear	0,5228	52,28%	342.584.890,21
Árvore de Decisão	0,9323	93,23%	144.108.108,24
Floresta Randômica	0,8306	83,06%	119.490.899,62
Árvore de Decisão de Aumento de Gradiente	0,9476	94,76%	117.320.076,32

Fonte: Elaborado pelo autor (2022)

O modelo de Árvore de Decisão de Aumento de Gradiente (*Gradient Boosting Decision Tree*) se mostrou o mais bem sucedido em prever os valores arrecadados, sendo capaz de explicar a variação da arrecadação mensal em 94,76% das vezes e apresentou uma divergência média, para mais ou para menos, entre o valor de fato arrecadado e o previsto de aproximadamente R\$ 117,3 mi.

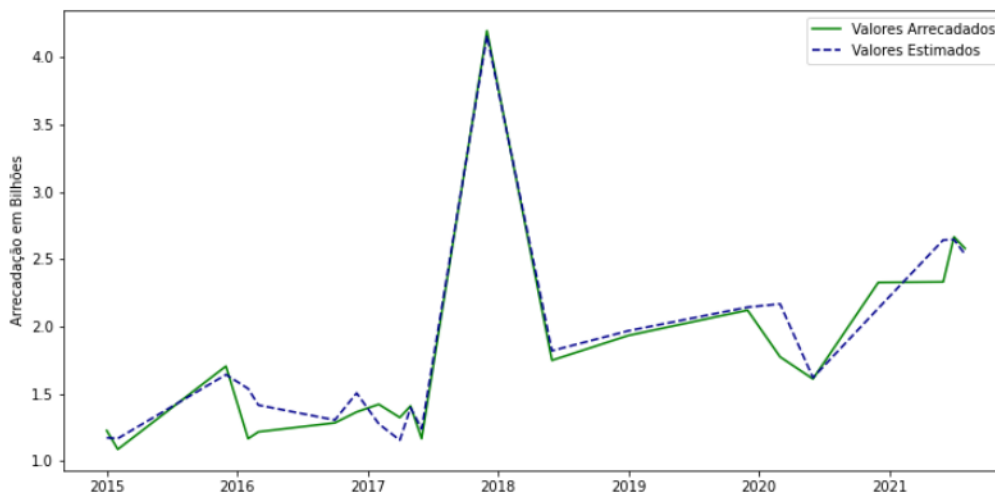
Considerando os 21 meses de dados utilizados na carga de teste, o total arrecadado pela DAU foi de R\$ 37.604.327.956,75, o total previsto pelo modelo, para os mesmos registros, foi de R\$ 38.634.416.006,69, uma diferença percentual de aproximadamente 2,57%. A comparação gráfica entre os valores previstos e o real se apresenta da seguinte maneira:

5. O cálculo dos coeficientes de ajuste R2 foi efetuado de maneira automatizada, com a função "r2_score" da biblioteca Scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html

6. Os erros médios foram obtidos de maneira automatizada, utilizando a função "mean_absolute_error" da biblioteca Scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html

7. Função "train_test_split". https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

FIGURA 4 - GRÁFICO RESULTADO ÁRVORE DE DECISÃO DE AUMENTO DE GRADIENTE



Fonte: Elaborado pelo autor (2022)

A título de comparação, valendo-se da suavização exponencial a PGFN, em suas últimas quatro estimativas, obteve uma diferença média entre o previsto e o real de 9,46%. Em sua melhor previsão, a divergência foi de 4,2% (PROCURADORIA-GERAL DA FAZENDA NACIONAL, 2021a).

4. CONCLUSÃO

Tendo como parâmetro de comparação os índices de acerto das estimativas de arrecadação realizadas pela PGFN utilizando o método de suavização exponencial, é possível afirmar que o uso dos algoritmos de *Machine Learning* Árvore de Decisão (*Decision Tree*) e Árvore de Decisão de Aumento de Gradiente (*Gradient Boosting Decision Tree*) mostraram-se bem-sucedidos, obtendo desempenho superior em prever os movimentos de ingressos da Dívida Ativa, quando alimentados com dados de indicadores macroeconômicos (PIB, IPCA, IGP-M, taxa de Câmbio e taxa SELIC) e das estratégias de arrecadação adotadas pela União (parcelamentos excepcionais e transações tributárias).

O modelo GBDT, além de apresentar o melhor coeficiente de previsão (coeficiente de ajuste $R^2 = 94,76\%$), também teve a menor média de erro entre o valor previsto e o valor de fato apurado (erro médio absoluto = 117.320.076,32). Para os 21 meses de dados que compuseram a carga de teste dos modelos, a diferença entre a soma dos valores previstos pela Árvore de Decisão de Aumento de Gradiente e a soma dos valores de fato arrecadados foi de 2,57%.

Todavia mais importante que os índices de acerto dos modelos, sua utilização revela-se mais vantajosa por levar em consideração, para a estimativa de ingresso de valores, variáveis que atualmente são desprezadas. Isso significa que os modelos propostos estão ancorados em parâmetros mais amplos e robustos, enquanto os cálculos atuais refletem somente os registros passados da própria arrecadação.

Ao elaborar as projeções de receita para a formulação da Lei Orçamentária Anual, a PGFN, com a metodologia proposta, poderá levar em conta os dados que os agentes econômicos preveem para a economia nacional, o que é mais consistente com as expectativas financeiras nacionais.

Outro ganho que sobressai está na possibilidade da Administração utilizar os algoritmos de *Machine Learning* para avaliar, a priori, os impactos arrecadatórios da instituição de parcelamentos e transações tributárias, o que viabiliza uma decisão pública mais bem amparada e pautada em fatores empiricamente simulados e testados.

Há que se observar, ainda, que uma análise descritiva dos dados demonstra eficácia maior das transações tributárias em detrimento dos parcelamentos excepcionais, ao menos do modo como veem sendo instituídos nos últimos seis anos, o que merece maiores estudos, notadamente visando identificar as razões desses resultados diversos com vista a aprimorar a construção de políticas públicas eficientes.

REFERÊNCIAS

- AGÊNCIA SENADO. (2021). *Projeto que reabre prazo para o Programa Especial de Regularização Tributária segue para a Câmara*. Brasília, BR. <https://www12.senado.leg.br/noticias/materias/2021/08/05/projeto-que-reabre-prazo-para-o-programa-especial-de-regularizacao-tributaria-segue-para-a-camara>
- BANCO CENTRAL DO BRASIL. (2022a). *Estatísticas do setor externo*. Nota para a imprensa - 29/04/2022. <https://www.bcb.gov.br/estatisticas/estatisticassetorexterno>
- _____. (2022b). *Taxa Selic*. <https://www.bcb.gov.br/controleinflacao/taxaselic>
- BECKER, Dan. (2017a). *How Models Work: The first step if you're new to machine learning*. Kaggle. <https://www.kaggle.com/code/dansbecker/how-models-work>
- _____. (2017b). *Model Validation: Measure the performance of your model, so you can test and compare alternatives*. Kaggle. <https://www.kaggle.com/code/dansbecker/model-validation>
- CALLEGARI-JACQUES, Sidia M. (2017). *Bioestatística: princípios e aplicações*. Porto Alegre, BR: Artmed.
- FÁVERO, Luiz Paulo; BELFIORE, Patrícia. (2017). *Manual de análise de dados*. 1. ed. Rio de Janeiro, BR: Elsevier.
- FUNDAÇÃO GETÚLIO VARGAS. (2022). *IGP-M: Resultados 2022*. FGV. <https://portal.fgv.br/noticias/igpm-resultados-2022>
- FURTADO, Paulo Augusto. (2022). *Por que escolher Python?*. In: *Jornada Python: uma jornada imersiva na aplicabilidade de uma das mais poderosas linguagens de programação do mundo*. Rio de Janeiro, BR: Brasport.
- FREITAS, Gabriel Belmino. (2019). *O uso de machine learning na modelagem da previsão de inflação: revisão bibliográfica*. 42 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências Econômicas) — Universidade de Brasília, Brasília, 2019. <https://bdm.unb.br/handle/10483/25328>
- GOOGLE. (2022). *Conheça o Colab*. https://colab.research.google.com/?utm_source=scs-index#scrollTo=Owu-xHmxllTwN
- GRUS, Joel. (2016). *Data Science do Zero*. Traduzido por Welington Nascimento. Rio de Janeiro, BR: Alta Books.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. (2022a). *IPCA - Índice Nacional de Preços ao Consumidor Amplo*. Portal do Governo Brasileiro. <https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=o-que-e>
- _____. (2022b). *Produto Interno Bruto - PIB*. Portal do Governo Brasileiro. <https://www.ibge.gov.br/explica/pcb.php>
- _____. (2022c). *CONCLA - Comissão Nacional de Classificação*. https://cnae.ibge.gov.br/?view=estrutura&tipo=cnae&versao_classe=7.0.0&versao_subclasse=9.1.0
- KELLEHER, John D.; TIERNEY, Brendan. (2018). *Data Science. The MIT Press essential knowledge series*. Cambridge, US: The MIT Press.
- KRUGMAN, Paul R; OBSTFELD, Maurice. (2005). *Economia internacional: teoria e política*. Tradutor técnico Eliezer Martins Diniz. São Paulo, BR: Pearson Addison Wesley.
- MCKINNEY, Wes. (2010). *Data structures for statistical computing in python, Proceedings of the 9th Python in Science Conference*, Volume 445. <https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>
- MINISTÉRIO DA ECONOMIA. (novembro, 2014). *Cadastro Nacional de Atividades Econômicas - CNAE*. <https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/cnpj/classificacao-nacional-de-atividades-economicas-2013-cnae/apresentacao>
- MOREIRA, J. M.; CARVALHO, A.; HORVÁTH, T. (2018). *A general introduction to data analytics*. Hoboken, US: Wiley.

OLIVEIRA, B. (2021). *Algoritmos de aprendizado de máquina na predição e avaliação de evasão de clientes em ambiente de produção*. 87 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Goiás. <http://repositorio.bc.ufg.br/tede/handle/tede/11522>

OZDEMIR, Sinan. (2016). *Principles of data science*. Birmingham, UK: Packt Publishing Ltd.

PARANHOS, R. et al. (2014). Desvendando os Mistérios do Coeficiente de Correlação de Pearson: o Retorno. *Leviathan (São Paulo)*, (8), 66-95. DOI: <https://doi.org/10.11606/issn.2237-4485.lev.2014.132346>

PEDREGOSA et al. (2011). *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research 12, pp. 2825-2830. <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>

PROCURADORIA-GERAL DA FAZENDA NACIONAL. (2021a). Nota SEI nº 29/2021/PGDAU-CDA-COAGED/PGDAU-CDA/PGDAU/PGFN-ME. *Presta informações sobre a dívida ativa da União para compor o anexo de riscos fiscais*. Ministério da Economia. http://www.consultaesic.cgu.gov.br/busca/dados/Lists/Pedido/Attachments/1637685/RESPOSTA_RECORSO_1_161568_SEI_ME_14626293_Nota.pdf

_____. (2021b). Nota SEI nº 7/2022/PGDAU-CDA-COAGED/PGDAU-CDA/PGDAU/PGFN-ME. *Boletim de Acompanhamento Gerencial - Edição Anual – 2021*. Ministério da Economia. <https://www.gov.br/pgfn/pt-br/assuntos/divida-ativa-da-uniao/estudos-sobre-a-dau/boletim-de-acompanhamento-gerencial-da-divida-ativa-da-uniao-e-do-fgts-edicao-anual-2021.pdf>

_____. (2021c). Nota Conjunta SEI nº 2/2021/PGDAU-CGR. *Analisa os resultados alcançados pelas modalidades de transação da dívida ativa da União e da transação do contencioso de pequeno valor*. Ministério da Economia. https://www.gov.br/pgfn/pt-br/assuntos/divida-ativa-da-uniao/estudos-sobre-a-dau/sei_me-17016922-nota-conjunta.pdf

SANTOS, Gustavo Carvalho. (2020). *Algoritmos de Machine Learning para previsão da B3*. 90 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Uberlândia, Uberlândia. DOI <https://doi.org/10.14393/ufu.di.2020.640>

ZHANG, Z. et. al. (julho, 2021). GBDT-MO: *Gradient-Boosted Decision Trees for Multiple Outputs* in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 7, pp. 3156-3167. <http://proceedings.mlr.press/v70/si17a.html>



Rubens Quaresma Santos

E-mail: rubens.quaresma@pgfn.gov.br

Instituição de vinculação: Procuradoria-Geral da Fazenda Nacional - PGFN

ID ORCID: <https://orcid.org/0000-0002-3170-4348>

Pós-graduado (lato sensu) em Administração Pública pela Fundação Getúlio Vargas (2016). Graduado em Direito pela Universidade Estadual de Feira de Santana (2006). Graduando em Engenharia de Software pelo Instituto de Educação Superior de Brasília. Procurador na Procuradoria-Geral da Fazenda Nacional (PGFN), tendo exercido as funções de Subprocurador-Regional da Fazenda Nacional na 1ª Região (2017-2019) e de Procurador-Regional da Fazenda Nacional na 1ª Região (2019-2021), atualmente integra a Divisão de Consultoria e Assessoramento Jurídico da Procuradoria-Regional da Fazenda Nacional na 1ª Região.

Ciência de Dados como Ferramenta de Apoio à Tomada de Decisão: Classificação Hierárquica Descendente de Pedidos de Acesso à Informação na Prefeitura de São Paulo¹

La Ciencia de Datos como Herramienta de Apoyo a la Toma de Decisiones: Clasificación Jerárquica Descendente de Solicitudes de Acceso a la Información en la Alcaldía de São Paulo²

Data Science as a Tool to Support Decision-Making: Descending Hierarchical Classification of Access to Information Requests in the Municipality of São Paulo

Claudio Henrique Fontenelle Santos e Ana Lúcia da Silva Romão

<https://doi.org/10.36428/revistadacgu.v14i26.544>

Resumo: Buscou-se compreender de que forma a ciência de dados e as tecnologias de mineração e classificação de textos podem contribuir para a tomada de decisões a partir de uma melhor compreensão agregada dos pedidos de acesso à informação. A pesquisa utilizou dados dos pedidos de acesso à informação feitos à Prefeitura Municipal de São Paulo (PMSP), de 2012 a 2019, disponíveis no Portal de Dados Abertos da municipalidade, propondo a identificação e classificação das principais questões apresentadas. Os 39.369 textos dos pedidos de acesso submetidos à PMSP foram reunidos em um corpus e submetidos a análise por meio de Classificação Hierárquica Descendente (CHD). Ao propor uma classificação de textos como uma metodologia para análise de dados textuais, reforçou-se um paradigma de que dados textuais não pertencem apenas ao campo qualitativo. Além disso, a consideração de apenas substantivos, excluídos verbos e advérbios; e os adjetivos mais ocorrentes serem usados como parte de expressões, permitiu uma otimização do contexto dos pedidos, proporcionando classificar os dados textuais de maneira mais objetiva, mitigando o viés dos investigadores. O artigo apresenta também outros estudos de caso relevantes para a pesquisa, com referências encontradas na análise de pedidos de acesso à informação, contribuindo para a compreensão de pedidos dos cidadãos de modo aglutinado e permitindo aos tomadores de decisões um melhor entendimento das demandas da sociedade, podendo resultar em políticas públicas mais focadas. Conclui-se que a análise dos dados através da CHD permite obter informações relevantes para a tomada de decisão baseada em dados e evidências e que a abordagem favorece a concretização de decisões fundamentadas e mais próximas das necessidades dos cidadãos.

Palavras-chave: tomada de decisão, pedidos de acesso à informação, mineração de texto, classificação hierárquica descendente.

Abstract: This article sought to understand how data science, text mining and text classification technologies can contribute to improved decision-making based on an overall understanding of access to information requests. Researchers used data from access to information requests made to the Municipality of São Paulo (PMSP), from

1. Artigo submetido em 18/07/2022 e aceito em 01/12/2022.

2. Artigo recebido em 18/06/2022 e aprovado em 01/12/2022.

2012 to 2019, available on the municipality's Open Data Portal, proposing presented the main issues identification and classification. 39,369 requests submitted to PMSP were gathered into a corpus and submitted for analysis through a Descending Hierarchical Classification (CHD). In proposing a classification of texts as a methodology for analyzing textual data, a paradigm has been reinforced: textual data does not belong only to the qualitative domain. In addition, considering only nouns, excluding verbs and adverbs, and the most frequent adjectives being used as part of vocabulary expressions, allowed a request context optimization, allowing to classify the textual data in a more objective way, mitigating researchers bias. We present main references found in access to information requests analysis, cases in Mexico, national studies, and in China, in the city of Beijing, and contributes to the understanding citizens requests in an aggregated way, allowing decision makers to a better understanding of society's demands, which may result in more focused public policies. It is concluded that data analysis through the CHD allows obtaining relevant information for decision-making based on data and evidence and that the approach favors the implementation of reasoned decisions that are closer to the citizens' needs.

Keywords: decision making, access to information requests, text mining, descending hierarchical classification.

Resumen: Este artículo buscó comprender cómo las tecnologías de ciencia de datos, minería de texto y clasificación de texto pueden contribuir a una mejor toma de decisiones a partir de la comprensión agregada de las solicitudes de acceso a la información. La investigación utilizó datos de solicitudes de acceso a la información realizadas al Municipio de São Paulo (PMSP), de 2012 a 2019, disponibles en el Portal de Datos Abiertos del municipio, proponiendo la identificación y clasificación de los principales problemas presentados. 39.369 textos de solicitudes enviados a PMSP fueron reunidos en un corpus y sometidos a análisis a través de una Clasificación Jerárquica Descendente (CHD). Al proponer una clasificación de textos como metodología para el análisis de datos textuales, se reforzó un paradigma de que los datos textuales no pertenecen únicamente al campo cualitativo. Además, considerar solo los sustantivos, excluyendo verbos y adverbios, y los adjetivos más frecuentes utilizados como parte de las expresiones del vocabulario, permitió una optimización del contexto de la solicitud, lo que permitió clasificar los datos textuales de forma más objetiva, mitigando el sesgo de los investigadores. Presentamos los principales referentes encontrados en el análisis de solicitudes de acceso a la información, casos en México, estudios nacionales y en China, en la ciudad de Beijing, y contribuye a la comprensión de las solicitudes ciudadanas de manera agregada, permitiendo a los tomadores de decisiones una mejor comprensión de las necesidades de la sociedad. demandas, lo que puede redundar en políticas públicas más focalizadas. Se concluye que el análisis de datos a través del CHD permite obtener información relevante para la toma de decisiones basadas en datos y evidencias y que el enfoque favorece la implementación de decisiones razonadas y más cercanas a las necesidades de los ciudadanos.

Palabras clave: toma de decisiones, solicitudes de acceso a la información, minería de textos, clasificación jerárquica descendente.

INTRODUÇÃO

A Lei de Acesso à Informação (LAI) completou, em 2022, 10 anos de existência no Brasil, e com ela surgiram, em todas as esferas de governo do Brasil, federal, estadual e municipal, portais de recebimento de pedidos de acesso à informação. A Escala Brasil Transparente 360 graus - 2ª edição identificou que todos os estados brasileiros e o Distrito Federal já haviam regulamentado a LAI até 2019 (Controladoria-Geral da União, 2019). Ao nível municipal, a Prefeitura de São

Paulo, maior cidade do Brasil, publicou em agosto de 2012 o Decreto nº 53.623 regulamentando a LAI no âmbito do município.

Os governos, através do acesso à informação, colhem subsídios para aprimorar suas políticas públicas, pois cada pedido representa uma solicitação, pessoal ou coletiva, e a compreensão desses pedidos, de modo individual e agregado, é muito importante para a tomada de decisão pública. A Ciência de Dados facilita essa compreensão, através da extração de valor dos

dados com a utilização de técnicas de tratamento de dados, como mineração de textos e inteligência artificial (Kotu e Deshpande, 2018; Soares, 2020).

O objetivo deste artigo passa por compreender de que forma as tecnologias de mineração e classificação de textos contribuem para a tomada de decisão baseada em dados e evidências, através da análise agregada de pedidos de acesso à informação. A pesquisa utilizou os dados dos pedidos feitos à Prefeitura Municipal de São Paulo (PMSP), de 2012 até 2019, propondo a identificação e classificação dos principais temas apresentadas.

Existem diversas técnicas de análise, entre elas, a Classificação Hierárquica Descendente (CHD), que, em linguagem simplificada, analisa um grande volume de textos, e, usando um algoritmo previamente criado e validado, apresenta de forma agregada, os principais assuntos abordados nos textos investigados.

Este artigo traz contributo ao utilizar a técnica de CHD na análise e classificação de pedidos de acesso à informação. Inova-se também ao considerar na análise apenas substantivos, palavras compostas, como “conselho municipal”, “rede municipal” e siglas, como “SME” e “SPTRANS”, e ao desconsiderar verbos, adjetivos e advérbios. Martin e Johnson (2015) concluíram que usar apenas substantivos pode trazer melhores resultados.

Esse artigo apresenta, primeiramente, na seção dois, um referencial teórico sobre ciência de dados, textos como dados e acesso à informação, seguido por uma revisão da literatura dos artigos que já realizaram investigações semelhantes. Na seção quatro é apresentada a metodologia e na seção cinco são mostradas e discutidas as análises realizadas, incluindo breves descrições estatísticas dos dados, segmentação por palavras e refinamento por classificação de tópicos. A última seção apresenta a conclusão do estudo, limitações e sugestões de estudos futuros.

REFERENCIAL TEÓRICO

Ciência de Dados

Estamos vivendo uma era de evolução do “mundo de dados”, continuamente conectados à nossa vida diária, ao nosso trabalho e à economia. Governos, ins-

tuições privadas, e acadêmicos estão diariamente empreendendo esforços para chegar a maneiras de converter dados em instrumentos de tomada de decisões, promover a pesquisa e o desenvolvimento da ciência (Cao, 2017). Os cientistas agora possuem, com mais facilidade, acesso a esse grande universo de dados, incluindo dados governamentais. Como resultado, pesquisas empíricas podem ser conduzidas em uma escala que não seria possível a uma ou duas gerações anteriores (Lane et al., 2022).

Para Kotu e Deshpande (2018), a ciência de dados envolve inferência e iteração de muitas hipóteses diferentes relacionadas aos dados disponíveis, sendo um dos seus aspectos-chave o processo de generalização de padrões a partir desses dados, generalização essa que deve ser válida, não só para a base de dados usada para observar o padrão, mas para novos dados ainda não acessados.

Em resumo, a Ciência de Dados busca aprimorar as tomadas de decisão baseando-as em “*insights*” extraídos das informações obtidas nos conjuntos de dados (Kelleher e Tierney, 2018).

Texto como Dados

O advento das novas tecnologias tem permitido a disponibilização de vasta quantidade de textos em formato digital por órgãos governamentais e formuladores de políticas públicas (Hollibaugh, 2019). Para os cientistas sociais, a informação codificada em texto é um rico complemento a suas pesquisas (Gentzkow et al., 2019).

As investigações usando textos como dados, auxiliadas por computador, estão se tornando relevantes em estudos nos campos da administração pública, políticas públicas e ciências políticas (Hollibaugh, 2019). Contudo, até 2022 ainda são escassos os estudos usando abordagens de análise de textos e modelagem de tópicos para extrair conhecimento de pedidos de acesso à informação. Em pesquisa realizada em junho de 2022 nas bases Scopus e Web of Science foram encontrados apenas quatro artigos que utilizaram metodologias do campo de ciências de dados para investigar pedidos de acesso à informação, conforme mostra o Quadro 1, a seguir:

QUADRO 1 – ESTADO DA PESQUISA

AUTORIA	PAÍS	ESFERA DE INVESTIGAÇÃO	PERGUNTAS DE PESQUISA	OBJETIVO	CONCEPÇÃO TEÓRICA	BASE DE DADOS	METODOLOGIA DE ANÁLISE DE DADOS	VARIÁVEIS/CATEGORIAS ANALÍTICAS	PRINCIPAIS CONCLUSÕES E CONTRIBUIÇÕES
Berliner, D., Bagozzi, B. E., & Palmer-Rubin, B. (2018)	México	Nacional	Que informação os cidadãos procuram? Como essas informações se relacionam com questões de importância pública?	Identificar os temas relacionados aos pedidos de acesso à informação feitos e como eles se relacionam com as questões públicas.	Modelo da <i>accountability</i> pública (principal agente), modelo iceberg (interesse privado) e transparência.	Pedidos feitos ao governo federal mexicano entre 2003-2015.	Análise não supervisionada de texto (LDA) e modelação de tópicos.	(dependente) Potencial de <i>accountability</i> do pedido.	Vinte tópicos (ver Quadro 2).
Bagozzi, B. E., Berliner, D., & Almqvist, Z. W. (2019)	México	Nacional	Quais são os temas de pedidos de acesso à informação feitos pelos cidadãos mexicanos que mais se associam com negação de respostas?	Identificar os temas dos pedidos mais associados a negações de respostas.	Motivação dos burocratas para responder ou negar respostas a pedidos de acesso à informação e transparência.	Pedidos feitos ao governo federal mexicano entre 2003-2015.	Análise supervisionada de texto (sLDA) e modelação de tópicos.	(dependente) Não responsividade.	Os cinco tópicos com maior probabilidade de receber uma negação apresentaram natureza investigativa, relacionados com atividades policiais, financeiras, políticas, compras governamentais.
Berliner, D., Bagozzi, B. E., Palmer-Rubin, B., & Erlich, A. (2021)	México	Nacional	Como o governo responde quando os cidadãos fazem questionamentos via pedidos de acesso à informação?	Compreender as motivações para responder ou não responder a pedidos de acesso à informação.	<i>Accountability</i> vertical (principal agente) e teoria da responsividade do governo.	Pedidos feitos ao governo federal mexicano entre 2003-2015.	Análise não supervisionada de texto (LDA) e modelação de tópicos.	(dependente) Não responsividade e (independente) alinhamento político ao nível municipal.	Pedidos feitos de locais onde o governo é bem votado recebem mais respostas.
Wang, Z., & Zhong, Y. (2020)	China	Local (Pequim)	Quais são os principais temas que os residentes de Pequim questionam ou sugerem ao governo?	Compreender os principais temas relacionados aos pedidos de acesso à informação feitos.	Governança	Solicitações feitas à Prefeitura de Pequim no ano de 2015.	Mineração de textos, análise de classes e modelação probabilística de tópicos.	(segmentação) Tópico do assunto.	Sete tópicos (ver Quadro 2).

Fonte: elaboração própria, com base nos estudos da pesquisa.



Berliner et al. (2018) investigaram, no México, no âmbito federal, que informações os cidadãos procuram, e como essas informações se relacionam com questões de importância pública. Os autores fizeram uso de um modelo generativo probabilístico de tópicos (*Latent Dirichlet Allocation* – LDA – ver Blei et al., 2003) para descobrir tópicos “latentes” no conjunto agregado de pedidos de acesso à informação recebidos entre 2003 e 2015. Concluíram que alguns tópicos e os pedidos que os compõem, como os relacionados a temas como “Meio Ambiente” ou “Gastos Públicos”, contribuem mais para a “*accountability* pública” do que outros, como os que contêm “Demandas Individuais” ou “Informações Comerciais”.

Bagozzi et al. (2019), dessa vez usando a técnica de (*Latent Dirichlet Allocation* supervisionada – sLDA – ver Mcauliffe e Blei, 2007) buscaram identificar os tópicos, e seus pedidos relacionados, mais associados a negações de respostas. Os tópicos com maior probabilidade de receberem uma “negação de resposta” foram de natureza investigativa, ligados a atividades policiais, financeiras, políticas e compras governamentais.

Berliner et al. (2021) investigaram como o governo responde às solicitações dos cidadãos. Segundo os autores, usando a responsividade como variável dependente, isto é, a motivação para responder os pedidos, recebem mais respostas favoráveis aqueles oriundos de localidades onde o partido do governo foi mais votado.

Um estudo realizado em Pequim por Wang e Zhong (2020), fazendo uso das técnicas de “mineração de textos”, “análises de *clusters*” e “modelação de tópicos”, investigou quais eram as principais questões apresentadas pelos cidadãos. Os autores concluíam serem mais frequentes os relacionados a registro de imóveis, construções ilegais, educação e moradia (Cf. Quadro 1).

Acesso à Informação

O direito ao acesso à informação surgiu em 1766, com a lei de liberdade de imprensa da Suécia, ao adotar que dados governamentais são, em princípio, abertos ao público (Rydholm, 2013). Em 1789, a Declaração dos Direitos do Homem e do Cidadão, aprovada na França, preconizou o direito dos cidadãos a acessar o orçamento (art. 14.^o) e a pedir contas à Administração Pública (art. 15.^o). Direito similar foi outorgado na Holanda, no ano de 1795, ao ser declarado que todos

têm o direito de exigir, de cada funcionário público, prestação de contas e justificativas de suas condutas (Banisar, 2006).

Após a II Guerra Mundial, com a criação da Organização das Nações Unidas (ONU), padrões internacionais de direitos humanos, entre eles o direito ao acesso à informação, começaram a serem propagados, inicialmente pela Declaração Universal dos Direitos Humanos (1948), que, em seu artigo 19.^o, diz que todos têm o direito de procurar, solicitar e receber informações (Banisar, 2006).

Para Walby e Larsen (2012), o acesso à informação e a liberdade de informação são relevantes mecanismos democráticos pelos quais a sociedade, organizações civis e jornalistas podem solicitar informações não publicadas pelos governos, em busca de mais detalhes sobre a atuação governamental. As leis de acesso à informação regulam a maneira como os órgãos do governo devem disponibilizar as informações ao público e o modo como devem lidar com os pedidos de informação (Angélico e Teixeira, 2012).

Savage e Hyde (2014) indicaram que leis de acesso à informação são uma poderosa ferramenta para os cientistas sociais, embora, àquela época ainda não haviam sido aproveitado todo esse potencial. Walby e Luscombe (2017) ratificaram, afirmando que pedidos de acesso à informação eram pouco utilizados como meios de produção de dados.

Os pedidos feitos pelos cidadãos aos governos, no âmbito das suas respectivas leis de acesso à informação possuem grande potencial, tanto ao nível prático, quanto ao nível teórico. Na prática, eles permitem acesso a dados, que não se encaixam facilmente como primários ou secundários, nem qualitativos ou quantitativos, sem compromissos epistêmicos ou antológicos, porém, susceptíveis de diversas análises com o uso das mais variadas lentes teóricas, contribuindo assim para o avanço teórico (Savage e Hyde, 2014). Walby e Luscombe (2017) argumentaram que, quando sistematicamente desenhadas e conduzidas, pesquisas com pedidos de acesso à informação terão credibilidade, ética, coerência e contribuição significativas.

Serviços de acesso à informação pública apresentam um número de oportunidades de pesquisa no desenvolvimento de modelos computacionais que puderam ajudar a compreender os cidadãos e suas necessidades (Flores et al., 2022). No entanto, a maioria dos pesquisadores parece não ter familiaridade com esses dados ou não lhe dá a devida importância, escrevendo

sobre eles com uma abordagem “jornalística”, em vez de produzir de forma sistemática dados qualitativos longitudinais sobre as práticas governamentais (Walby e Larsen, 2012).

O direito ao acesso à informação no Brasil foi outorgado com a promulgação da Lei n. 12.527 (2011), conhecida também como LAI, sendo uma resposta do legislador aos comandos constantes no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição da República Federativa do Brasil (1988/2022). Considerada um marco no processo de abertura das informações da administração pública brasileira, a LAI se tornou um dos principais instrumentos de promoção da transparência no âmbito das três esferas públicas, bem como no Distrito Federal (Angélico e Teixeira, 2012).

METODOLOGIA

Desenho do Estudo

Esta investigação, de natureza exploratória e descritiva, utiliza uma abordagem mista, qualitativa e quantitativa, com recurso à técnica de CHD, para categorizar os pedidos de informação feitos à Prefeitura Municipal de São Paulo. Este método permite a caracterização, de modo individual e agregado, do conteúdo de textos, reunindo-os em conjuntos de classes, apresentando os vocábulos mais associados a cada uma das classes, bem como às variáveis demográficas disponíveis.

A CHD, segundo Mendes et al. (2019), é uma técnica que categoriza as palavras em classes lexicais considerando a frequência e a posição relativa com que elas aparecem nos textos. Vários testes de qui-quadrado são realizados, até que são esgotadas as possibilidades de diferenciação entre as classes. O número inicial de classes é uma variável subjetiva a ser inserida pelo investigador, que pode fazer vários testes, e, a partir dos resultados, encontra a mais adequada para apresentar na análise. A força de associação entre as palavras e sua respectiva classe é representada pelo índice “qui-quadrado de Pearson”, o qual, quanto maior, mais provável é a hipótese de pertencimento da palavra na classe. A classificação final é representada por uma figura denominada “dendrograma”, contendo as classes e palavras mais representativas.

A amostra constou de 39.369 pedidos de acesso à informação realizados à Prefeitura de São Paulo (PMSP), a partir da instituição do acesso à informação no município, em agosto de 2012, até o final do ano de 2019. Os pedidos recebidos nos anos de 2020 a 2022 não estavam disponíveis à época da coleta de dados. A cidade de São Paulo foi escolhida por dois motivos: primeiro, por ser a maior cidade do Brasil, e o segundo, por ter a base de dados dos pedidos disponibilizada em dados abertos.

Além do conteúdo dos pedidos, foram também utilizadas as seguintes variáveis, todas oriundas dos conjuntos de dados abertos da PMSP: órgão ao qual o pedido foi endereçado, ano do pedido, tipo do solicitante (pessoa física ou jurídica), sexo (masculino ou feminino), e distrito de residência (pessoa física) ou localização (pessoa jurídica).

Procedimentos de Análise de Dados

Os dados foram analisados em duas etapas. Primeiramente, usando o programa SPSS, versão 27, foram analisados os dados demográficos, para identificação dos órgãos mais demandados e tipos de solicitantes, sexo e distritos mais frequentes. Esses dados serão apresentados mais adiante.

Na segunda etapa, os 39.382 pedidos foram reunidos em um arquivo texto e transformado em um “corpus”, termo utilizado na academia para identificar um conjunto de textos destinados à análise textual. O programa Iramuteq foi o escolhido para a análise dessa investigação. O programa usa os ambientes estatísticos das linguagens de programação R e Python, e oferece um conjunto de análises estatísticas de dados qualitativos, entre as quais a CHD.

O corpus textual

O arquivo contendo os pedidos e as respectivas variáveis demográficas dos solicitantes compreendeu o corpus textual que foi utilizado na análise³. Cada pedido foi precedido, em uma linha única, com as respectivas variáveis demográficas indicadas com “*”, como no exemplo a seguir. O exemplo inicia-se com quatro asteriscos, exigência do programa Iramuteq para indicar o início de um segmento de texto, no caso, um pedido de acesso à informação, seguido das variáveis. Foram elas: “o” – órgão, “a” – ano, “p” – pessoal solicitante, “s” – sexo, e “d” – distrito.

3. O arquivo está disponível em: https://github.com/chfsantos/RevistaCGU_Ciencia_de_Dados_PMSP/blob/main/corpus_couleur.docx.zip

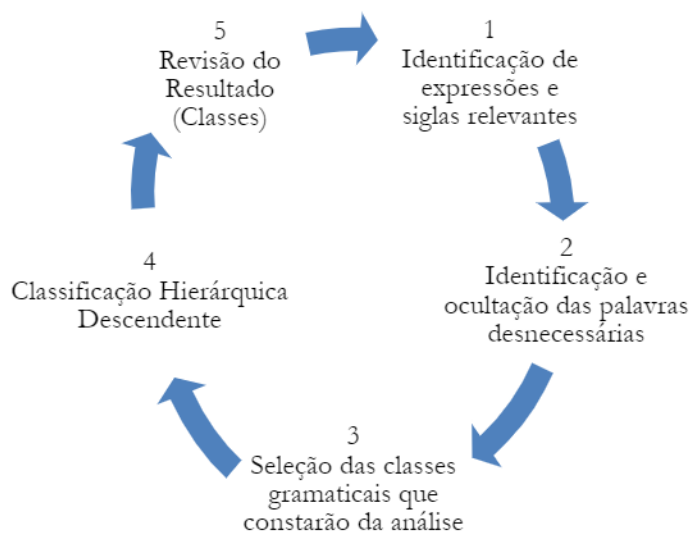
**** *o_19 *a_2013 *p_PF *s_M *d_cidadedutra

“gostaria de saber se existe alguma intervenção de canalização de córrego e ou remoções de moradias na rua samuel scott programada cep 04857060 grato”

Preparação do corpus textual

A preparação do *corpus* foi operacionalizada pelos pesquisadores em um ciclo de cinco etapas, conforme a Figura 1, a seguir:

FIGURA 1 – CICLO DE PREPARAÇÃO DO CORPUS



Fonte: elaboração própria.

Identificação de expressões e siglas relevantes

Na primeira etapa, o *corpus* foi analisado com apoio do programa RStudio e a função “textstat.collocations” do pacote R “Quanteda”. Nessa etapa foi possível identificar as “expressões compostas” mais frequentes, como “São Paulo” (12.048 ocorrências), “prefeitura regional” (2.433 ocorrências), e “servidores comissionados” (1.569 ocorrências). As expressões consideradas relevantes pelos pesquisadores foram inseridas no arquivo “expressões_pt.txt”, posteriormente usado na análise. A lista de expressões selecionadas está em arquivo complementar a esse artigo⁴. Ainda nessa etapa, foi executada uma análise CHD no corpus, para identificar as palavras ou expressões mais

ocorrentes. Dessa lista foram identificadas as siglas mais mencionadas, para serem inseridas no arquivo lexique_pt.txt, dicionário usado para auxílio na análise CHD.

Identificação e ocultação das palavras consideradas desnecessárias para a análise

Usando o resultado da análise CHD operacionalizada na primeira etapa, foram identificadas palavras e expressões muito usadas nos pedidos, porém consideradas desnecessárias para a análise. Por exemplo, palavras como. “informação” (5.765 ocorrências), “boa tarde” (3.782 ocorrências), e “não” (17.714 ocorrências), caso fossem consideradas na análise, não trariam significância suficiente para indicar o tema do pedido, afinal, todos eram pedidos de “informação”. Usando o

4. O arquivo está disponível em: https://github.com/chfsantos/RevistaCGU_Ciencia_de_Dados_PMSP/blob/main/expressoes_adicionadas.txt

editor de texto *Visual Studio Code*, com funções avançadas de busca e substituição, uma lista de palavras foi “ocultada” nos pedidos. Essa ocultação é feita sendo inserido o caractere “_” (*underline*) imediatamente à frente e após a palavra (exemplo: “_informação_”). A decisão de considerar uma palavra desnecessária é subjetiva aos pesquisadores, e depende do contexto da investigação. Na análise textual, alguns autores defendem a exclusão de palavras específicas em cada análise, como Schofield et al. (2017) e Sarica e Luo (2021). A lista de palavras e expressões ocultadas está disponível em arquivo suplementar a esse artigo⁵.

Seleção das classes gramaticais que constaram da análise

Com base em Martin e Johnson (2015) optou-se pela inclusão somente das palavras classificadas gramaticalmente como “substantivos”, excluídos os verbos, adjetivos, e advérbios, comumente considerados também das análises CHD. Segundo os autores, em comparação entre três análises: original, com lematização (ver Balakrishnan e Lloyd-Yemoh, 2014) e somente com substantivos, essa última se mostrou mais coerente pelos resultados apresentados.

A seleção das classes gramaticais foi feita no início da execução do programa Iramuteq, em um menu onde são selecionadas como “ativas”, “suplementares” ou “excluídas”. Conforme dito, somente substantivos, expressões e siglas foram marcados como ativas.

Realização da CHD

Passadas as etapas iniciais, operacionalizou-se várias CHD, com uso de lematização, também por opção dos autores. Na lematização, a partir de um arquivo “txt” contendo a lista de palavras, palavras com o mesmo radical gramatical, e significado associado, por exemplo, “valor” e “valores”, são consideradas uma só palavra, no caso, “valor”, com o objetivo de reduzir a variabilidade do vocabulário, permitindo uma maior homogeneidade a ser submetida ao tratamento lexicométrico (Sousa, 2021). Essas CHD operacionalizadas serviram para que fossem identificadas outras palavras desnecessárias, que também foram excluídas (ver nota de rodapé 6).

ANÁLISE E DISCUSSÃO DOS RESULTADOS

Estatística Descritiva

Os 39.364 pedidos de acesso à informação submetidos à Prefeitura Municipal de São Paulo entre os anos de 2012 e 2019 foram enviados, predominantemente, pela internet (96,9%) e tiveram o e-mail como canal prioritário para recebimento da resposta (97,4%).

Com relação à quantidade de pedidos, percebeu-se que o número vem crescendo à medida que a população toma conhecimento do direito que possui (Tabela 1). Proporcionalmente à população, o número ainda é inexpressivo, se considerarmos que a cidade possui 11.253.503 habitantes, com dados do último censo (Instituto Brasileiro de Geografia e Estatística, 2010).

TABELA 1 – PEDIDOS DE ACESSO À INFORMAÇÃO – SÃO PAULO/SP, 2012-2019

ANO	PEDIDOS	PROPORÇÃO EM RELAÇÃO AO ANO ANTERIOR
2012	285	-
2013	2.521	784,56% (*)
2014	2.404	-4,64%
2015	4.217	75,42%
2016	5.212	23,59%
2017	7.860	50,81%
2018	8.093	2,96%
2019	8.772	8,39%
Total	39.364	

Fonte: elaboração própria. (*) O sistema começou a funcionar em 24/08/2012.

5. O arquivo está disponível em: https://github.com/chfsantos/RevistaCGU_Ciencia_de_Dados_PMSP/blob/main/termos_excluidos_do_corpus.csv

Quanto aos solicitantes, há um predomínio de pessoas físicas (36.723) sobre pessoas jurídicas (2.641), o que é um resultado comum quando se trata de pedidos de acesso à informação (Bagozzi et al., 2019; Wang e Zhong, 2020). Quanto ao sexo, o percentual de pedidos de mulheres variou entre 32,7% (2012) e

45,1% (2015), enquanto de homens oscilou entre 54,9% (2015) e 67,3% (2012). Com relação aos distritos, de residência (pessoa física) ou localização (pessoa jurídica) de quem fez o pedido, são os seguintes os dez de onde mais se originaram pedidos de acesso à informação (Tabela 2)⁶:

TABELA 2 – DEZ DISTRITOS DE ONDE MAIS SE ORIGINARAM OS PEDIDOS DE ACESSO À INFORMAÇÃO – SÃO PAULO/SP, 2012-2019

DISTRITO	PEDIDOS	%	% ACUMULADO
Vila Mariana	616	4,6	4,6
Pinheiros	570	4,2	8,8
Bela Vista	546	4,0	12,8
Barra Funda	489	3,6	16,4
Jardim Paulista	426	3,2	19,6
Perdizes	419	3,1	22,7
Jabaquara	412	3,0	25,7
Itaim Bibi	347	2,6	28,3
Artur Alvim	346	2,6	30,9
Santa Cecília	328	2,4	33,3

Fonte: elaboração própria.

Com relação à profissão, destacou-se a quantidade de pedidos feitos por jornalistas (1.134, 2,9%), considerados por alguns como “intermediários de transparência” (Lane et al., 2022; Porumbescu et al., 2022). As pessoas que se declararam pesquisadores (349) e professores (267) representaram menos de 2% do total.

Quanto aos órgãos recebedores dos pedidos de acesso à informação, conforme os dados, dez deles receberam cerca de 50% das solicitações (Tabela 3).

TABELA 3 – PEDIDOS DE ACESSO À INFORMAÇÃO DOS DEZ ÓRGÃOS MAIS DEMANDADOS – SÃO PAULO/SP, 2012-2019

ÓRGÃO	PEDIDOS	% INDIVIDUAL	% ACUMULADO
SMS – Secretaria Municipal da Saúde	3.160	8,0	8,0
SME – Secretaria Municipal de Educação	3.059	7,8	15,8
CET – Companhia de Engenharia de Tráfego	2.873	7,3	23,1
SF – Secretaria Municipal da Fazenda	2.363	6,0	29,1
SPTTrans – São Paulo Transportes S/A	2.241	5,7	34,8
SMT – Secretaria Municipal de Mobilidade e Transportes	2.038	5,2	39,9
SMADS – Secretaria Municipal de Assistência e Desenv. Social	988	2,5	42,5
SMG – Secretaria Municipal de Gestão	911	2,3	44,8
SVMA – Secretaria Municipal do Verde e do Meio Ambiente	840	2,1	46,9
SMC – Secretaria Municipal de Cultura	825	2,1	49,0

Fonte: dados da pesquisa.

6. Foram consideradas apenas 13.499 respostas (34,3%) pois 25.883 (65,7%) solicitantes não informaram o distrito de residência, por ser um dado não obrigatório.

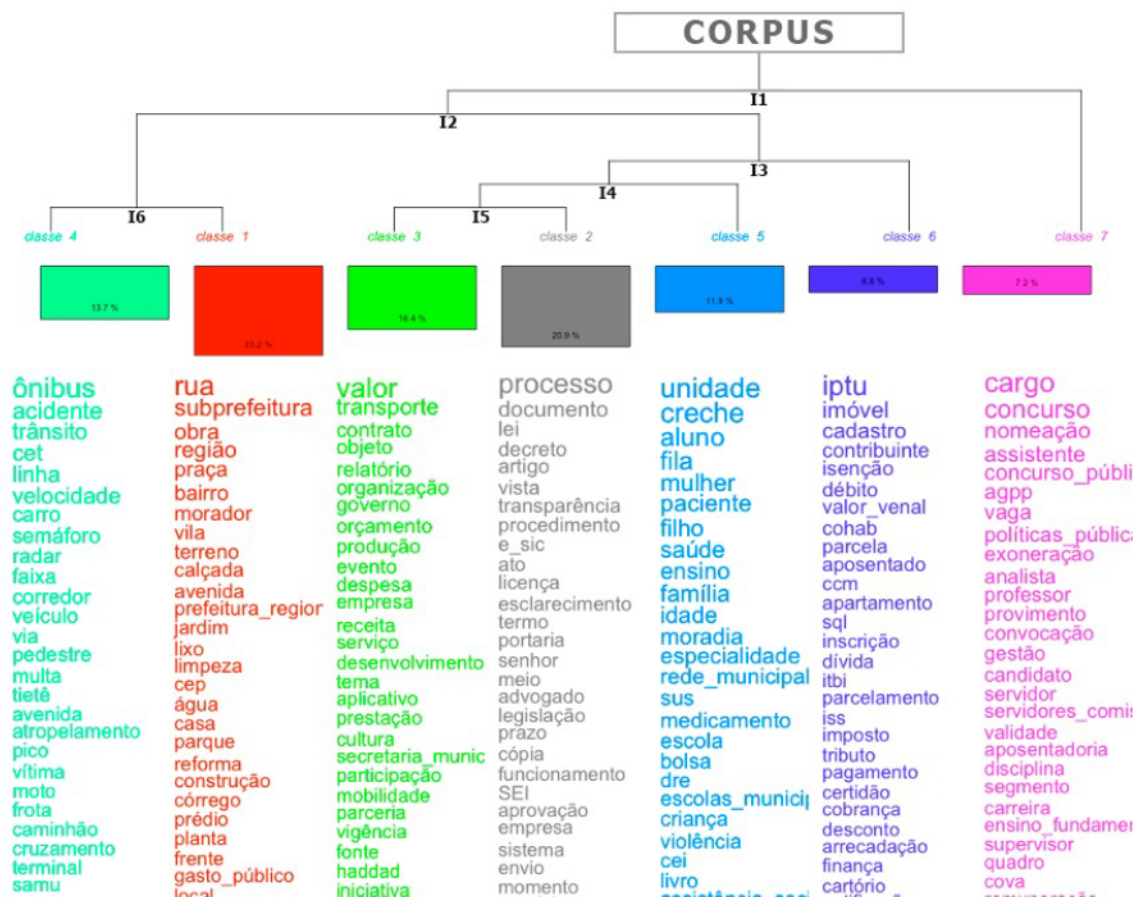
Cabe ser mencionado que é comum, quando há mudança de governo, também acontecerem mudanças nos nomes dos órgãos, o que dificulta o acompanhamento longitudinal dos dados. Por exemplo, a Secretaria Municipal de Habitação (SEHAB) possui pedidos a partir de 2018, sem dados anteriores, enquanto a Secretaria Municipal de Serviços (SES) possui pedidos até 2016, sem registros a partir de 2017.

Resultados da CHD

Após a execução na CHD resultaram 31.946 pedidos (81,1% da amostra) formados por 63.071 formas ativas (palavras, siglas ou expressões), com 3.102.034 ocorrências (vezes em que foram empregadas) e sete classes. 7.436 pedidos (18,9% da amostra) não apresentaram significância semântica para serem incluídos na CHD. O dendrograma apresentado na Figura 2 repre-

senta as iterações, as etapas da análise que resultaram em classe, e as principais palavras de cada uma delas, com base nos testes qui-quadrado (χ^2) realizados na CHD. Observa-se que na primeira iteração (I1), a classe 7 foi separada das demais, significando que essa classe possui formas ativas mais exclusivas para si, e menos utilizadas nas demais. Em seguida, aconteceu a iteração (I2) onde as classes 6, 5, 2 e 3 foram separadas das classes 1 e 4, significando que essas duas possuem formas ativas mais exclusivas, e, como na primeira iteração, menos frequentes das demais. Do mesmo modo aconteceram as duas iterações seguintes, (I3) e (I4) separando a classe 6 das classes 5, 2 e 3, e separando a classe 5 das classes 2 e 3. As últimas iterações ocorrem quando restam pares de classes, como ocorreu com as classes 2 e 3 (I5), 1 e 4 (I6).

FIGURA 2 – RESULTADO DA CHD – CLASSE E PALAVRAS



Fonte: elaboração própria, utilizando o programa Iramuteq.



As sete classes temáticas resultantes, cujos nomes foram dados pelos pesquisadores, assemelham-se às que emergiram das pesquisas de Wang e Zhong (2020) e de Berliner et al. (2018) (Quadro 2).

QUADRO 2 – VALIDAÇÃO EXTERNA E COMPARAÇÃO COM OUTROS ESTUDOS

ESTA PESQUISA	WANG E ZHONG (2020)	BERLINER ET AL. (2018)
São Paulo (Brasil)	Pequim (China)	México (País)
CLASSES TEMÁTICAS		
1) Bairros e Distritos;	1) Household Register;	1) Taxes and Finance;
2) Trâmite e documentos processuais;	2) Illegal Construction;	2) Environment and Land;
3) Contratações públicas;	3) Education;	3) Employees1: Salaries/Benefits;
4) Mobilidade Urbana;	4) Demolition;	4) Employees2: Functions/Qualif.;
5) Família: saúde, educação e assistência social;	5) City Management;	5) Employees3: Personnel;
6) Imóveis;	6) Housing;	6) Individual Needs;
7) Cargos e Concursos Públicos.	7) Traffic.	7) Commercial Information;
		8) Distributive Programs;
		9) Medical1: Contracts/Suppliers;
		10) Medical2: Purchases/Spending;
		11) Medical3: Inventories;
		12) Energy and Utilities;
		13) Health Statistics;
		14) Rules and Procedures;
		15) Education;
		16) Military, Police, and Crime;
		17) Budgets and Spending;
		18) Procurement1: Service Providers;
		19) Procurement2: Procedures/Docs;
		20) Procurement3: Anti-Corruption.

Fonte: elaboração própria.

Nos pontos a seguir as classes serão detalhadas, a partir dos significados agregados das palavras prevalentes em cada uma delas. As classes serão descritas, operacionalizadas e exemplificadas de forma sequencial, iniciadas pela classe 7, primeira à direita, para respeitar as iterações realizadas.

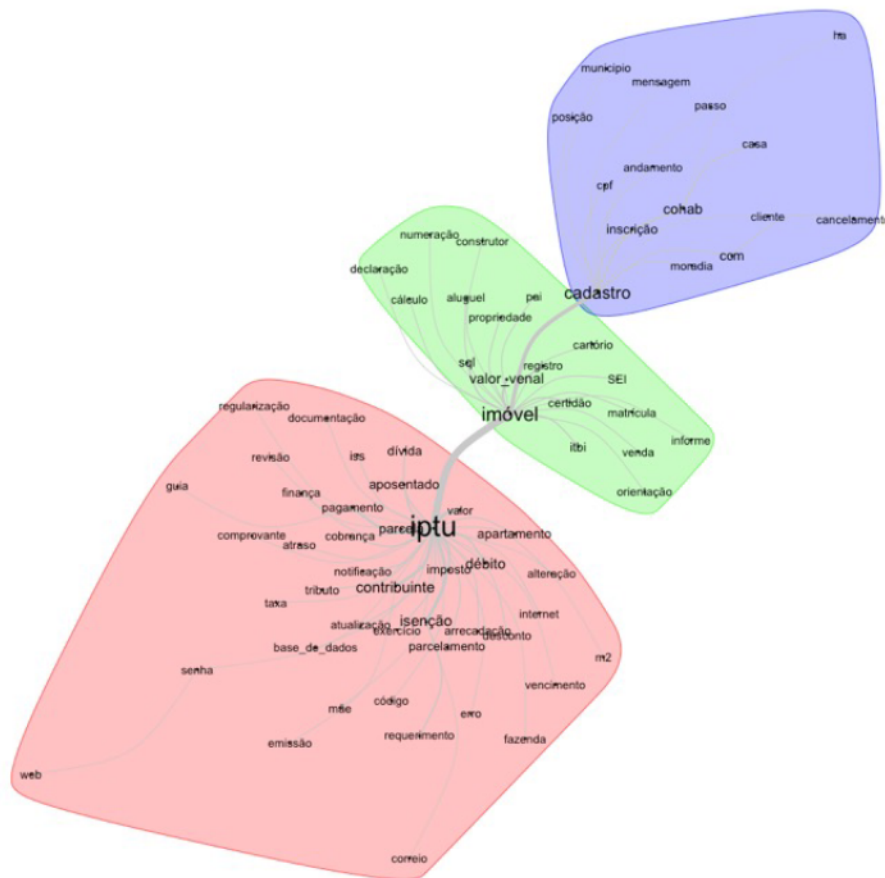
As classes foram nomeadas após a realização de uma análise amostral de 350 pedidos, correspondendo a 50 de cada uma (ver arquivo suplementar a esse artigo⁷). Observou-se que as palavras, expressões e siglas mais citadas geraram no seu entorno subclasses que podem ser consideradas para análise e tomada de decisões relacionadas aos temas.

7. O arquivo está disponível em: https://github.com/chfsantos/RevistaCGU_Ciencia_de_Dados_PMSP/blob/main/Amostra_das_classes.zip

Classe 6 – Imóveis

Composta por 2.171 pedidos (6,8%), formada por palavras, expressões e siglas como: IPTU ($X^2 = 9.781,26$), imóvel ($X^2 = 4.405,18$), cadastro ($X^2 = 3.130,69$), contribuinte ($X^2 = 1.944,35$) e valor venal ($X^2 = 1.540,36$).

FIGURA 4 – COCORRÊNCIA DAS FORMAS ATIVAS DA CLASSE 6

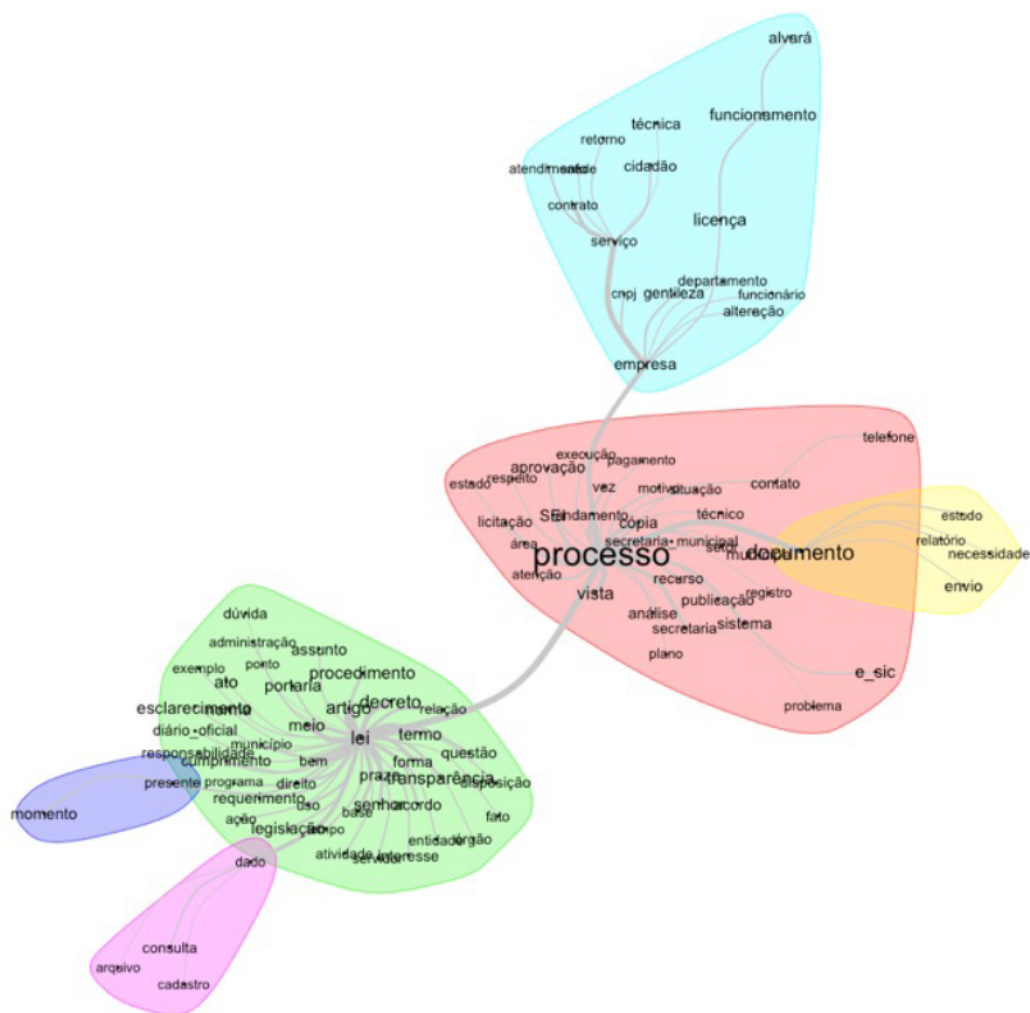


Fonte: elaboração própria

Classe 2 – Trâmite e documentos processuais

Composta por 6.665 pedidos (20,8%), constituída por palavras, expressões e siglas como: processo ($X^2 = 3.195,27$), documento ($X^2 = 1.303,76$), vista ($X^2 = 722,25$), processo administrativo ($X^2 = 321,35$) e despacho ($X^2 = 287,61$).

FIGURA 6 – COCORRÊNCIA DAS FORMAS ATIVAS DA CLASSE 2

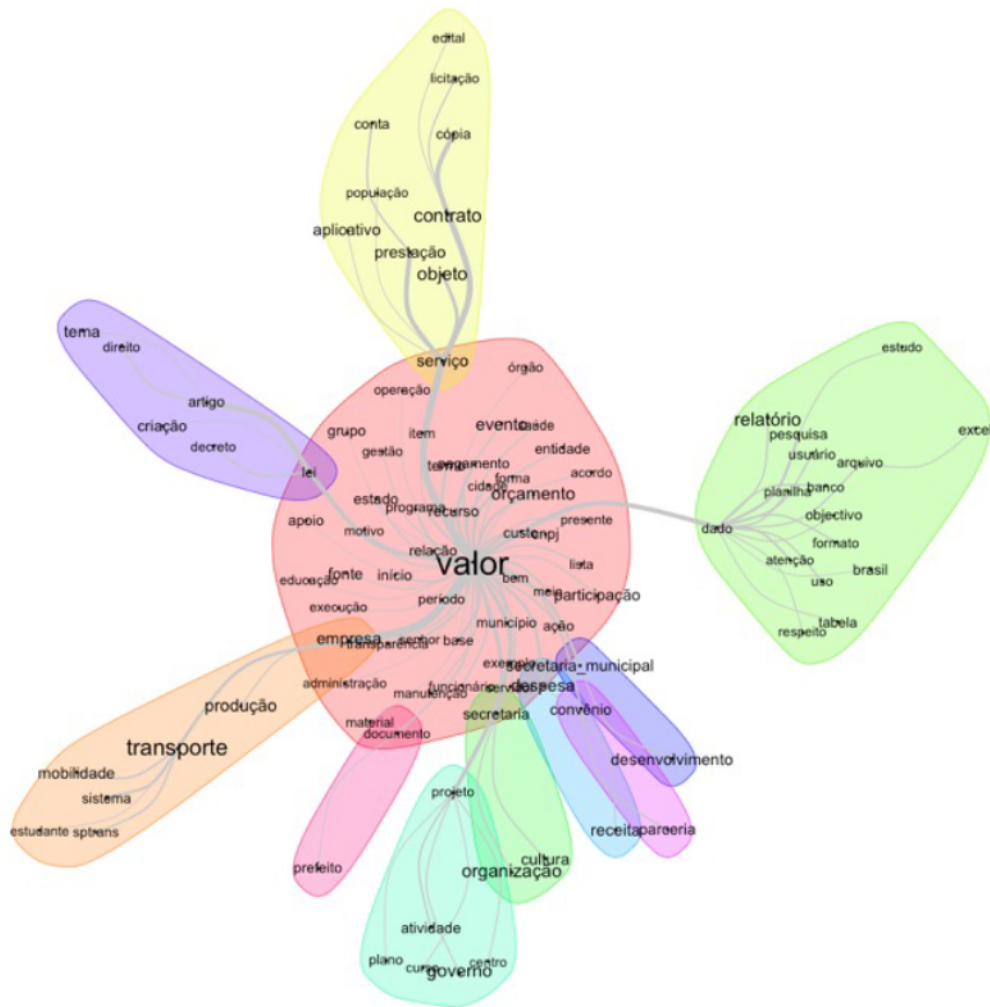


Fonte: elaboração própria

Classe 3 – Contratações públicas

Composta por 5.239 pedidos (16,4%), formada por palavras, expressões e siglas como: valor ($X^2 = 2.600,92$), transporte ($X^2 = 1.187,29$), contrato ($X^2 = 723,29$), objeto ($X^2 = 673,31$) e orçamento ($X^2 = 562,12$).

FIGURA 7 – COCORRÊNCIA DAS FORMAS ATIVAS DA CLASSE 3

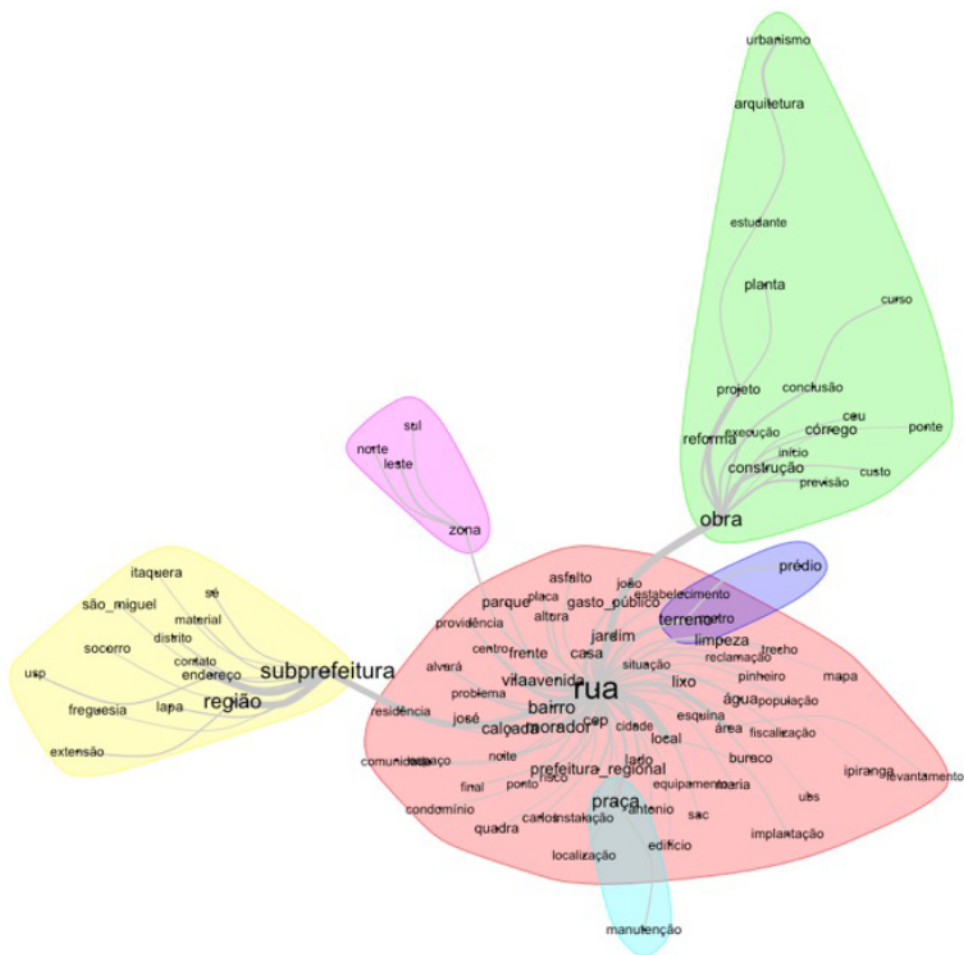


Fonte: elaboração própria

Classe 1 – Bairros e distritos

Composta por 7.404 pedidos (23,2%), constituída por palavras, expressões e siglas como: rua ($X^2 = 3.551,54$), obra ($X^2 = 1.531,77$), região ($X^2 = 1.523,25$), bairro ($X^2 = 1.079,21$) e prefeitura regional ($X^2 = 641,82$).

FIGURA 8 – COCORRÊNCIA DAS FORMAS ATIVAS DA CLASSE 1



Fonte: elaboração própria

Classe 4 – Mobilidade urbana

Composta por 4.372 pedidos (13,7%), formada por palavras, expressões e siglas como: ônibus ($X^2 = 4.926,60$), trânsito ($X^2 = 2.243,07$), CET ($X^2 = 2.198,87$), radar ($X^2 = 1.306,42$) e mobilidade urbana ($X^2 = 233,45$).

FIGURA 9 – COCORRÊNCIA DAS FORMAS ATIVAS DA CLASSE 4



Fonte: elaboração própria

Relação entre Classes e Variáveis

A partir do Corpus contendo os pedidos de acesso e suas variáveis foi possível se obter a relação entre as variáveis associadas aos pedidos de acesso à informação e as classes resultantes da CHD.

Classes x ano

Analisando a Tabela 4, que representa o qui quadrado (X^2) de associação dos anos com cada uma das classes, observa-se que a predominância das classes não foi a mesma ao longo dos anos.

TABELA 4 – ASSOCIAÇÃO ENTRE AS CLASSES DA CHD E ANOS

ANO	CLASSE 1	CLASSE 2	CLASSE 3	CLASSE 4	CLASSE 5	CLASSE 6	CLASSE 7
2012	-6,49	5,98	1,49	16,35	-1,04	-2,50	-15,90
2013	-2,69	4,25	18,56	6,04	-26,17	16,71	-57,34
2014	-5,98	4,17	-4,71	11,34	0,01	1,08	-2,87
2015	-0,47	-12,31	-0,91	52,77	5,13	3,95	-41,23
2016	4,55	-8,45	-15,17	21,13	-2,68	0,31	4,23
2017	3,67	-20,34	3,96	5,70	1,82	-59,16	13,96
2018	30,30	-3,80	-30,68	-18,39	13,26	0,65	5,63
2019	-30,38	69,71	35,17	-139,29	-3,98	3,57	13,84

Fonte: elaboração própria.

Por exemplo, a Classe 2 (Trâmite e Documentos Processuais) apresentou no ano de 2019 ($X^2 = 69,71$) uma maior associação, comparando com os demais anos, enquanto à Classe 4 (Mobilidade Urbana), observa-se um comportamento positivo desde 2012 ($X^2 = 16,35$), com um crescimento em 2015 ($X^2 = 52,77$), passando, podemos dizer, a uma desassociação média em 2018 ($X^2 = -18,39$) e alta em 2019 ($X^2 = -139,28$), denotando que, nesses dois anos, o tema foi menos mencionado. Do mesmo modo, a classe 6 (Imóveis), apresentou uma desassociação no ano de 2017 ($X^2 =$

-59,16). Observa-se que a Classe 7 (Cargos e Concursos Públicos) apresentou baixa associação com os anos 2013 ($X^2 = -57,34$) e 2015 ($X^2 = -41,22$). Essas variações indicam mudanças no interesse da população no assunto, que, associadas a outras informações, poderão colaborar na tomada de decisões relacionadas a políticas públicas afetas ao tema.

Classes x solicitantes

As associações entre as classes dos pedidos e os tipos de solicitantes estão representadas na Tabela 5.

TABELA 5 – ASSOCIAÇÃO ENTRE AS CLASSES DA CHD E SOLICITANTES

TIPO	CLASSE 1	CLASSE 2	CLASSE 3	CLASSE 4	CLASSE 5	CLASSE 6	CLASSE 7
Pessoa Física	-0,28	-138,81	-93,72	137,51	19,73	-0,01	148,56
Pessoa Jurídica	0,28	138,81	93,72	-137,51	-19,73	0,01	-148,56

Fonte: elaboração própria.

Observa-se que as classes 2 – Trâmite e documentos processuais ($X^2 = 138,81$) e 3 – Contratações públicas: projetos, compras, despesas e orçamentos ($X^2 = 93,72$) tiveram mais associação com pedidos de pessoas jurídicas, enquanto as classes 4 – Mobilidade Urbana ($X^2 = 137,50$) e 7 – Cargos e Concursos Públicos ($X^2 = 148,56$) mais associaram-se com pedidos de pessoas físicas.

Classes x Sexo dos Solicitantes

As associações entre as classes dos pedidos e os tipos de solicitantes estão representadas na Tabela 6.

TABELA 6 – ASSOCIAÇÃO ENTRE AS CLASSES DA CHD E O SEXO DO/A SOLICITANTE

SEXO	CLASSE 1	CLASSE 2	CLASSE 3	CLASSE 4	CLASSE 5	CLASSE 6	CLASSE 7
Feminino	1,86	-52,81	-86,94	-41,37	216,55	23,82	62,97
Masculino	-5,96	1,40	3,32	125,85	-140,14	-0,46	0,00

Fonte: elaboração própria.

Destacou-se uma associação entre a Classe 5 – Família: saúde, educação e assistência social ($X^2 = 216,55$) e solicitantes do sexo feminino, e uma associação entre a Classe 4 – Mobilidade Urbana ($X^2 = 125,80$) e solicitantes do sexo masculino. Associações como essas são relevantes, por exemplo, para os gestores públicos, por exemplo, decidirem como focar as decisões e mensagens relacionadas a tais políticas públicas.

Classes x órgãos destinatários dos pedidos de acesso à informação

A Tabela 7 mostra que os órgãos 60 ($X^2 = 7.132,52$) e 67 ($X^2 = 2.176,75$) foram destacadamente associados à Classe 4 (Mobilidade Urbana), enquanto os órgãos 10 ($X^2 = 1.961,22$) e 16 ($X^2 = 1.504,90$) tiveram maior associação à Classe 5 (Família: saúde, educação e assistência social). Continuando, temos que os órgãos 18 ($X^2 = 5.643,42$) e 62 ($X^2 = 1.801,91$) foram os com maior associação à Classe 6 (Imóveis), e os órgãos 10 ($X^2 = 1.285,89$) e 22 ($X^2 = 2.609,22$) representaram os mais associados à Classe 7 (Cargos e Concursos Públicos).

TABELA 7 - ASSOCIAÇÃO ENTRE OS ÓRGÃOS E AS CLASSES DA CHD

ÓRGÃO / NOME DO ÓRGÃO	CLASSE						
	1	2	3	4	5	6	7
67 - SPTrans - São Paulo Transportes S/A	-354,64	-80,30	91,52	2.176,75	-85,52	-122,21	-73,44
16 - SMS - Secretaria Municipal da Saúde	-65,64	-2,23	-8,31	-49,75	1.504,91	-175,35	-42,70
10 - SME - Secretaria Municipal de Educação	-285,86	-126,50	-33,96	-372,73	1.961,22	-151,88	1.285,89
22 - SG - Secretaria Municipal de Gestão	-228,64	0,05	18,73	-182,03	-40,66	-53,44	2.609,22
18 - SF - Secretaria Municipal da Fazenda	-432,02	25,06	0,27	-295,21	-182,20	5.643,42	-65,51
62 - COHAB - Companhia Metropolitana de Habitação	-18,68	-17,37	-76,70	-91,26	8,86	1.801,91	-38,34
60 - CET - Companhia de Engenharia de Tráfego	-148,63	-244,16	-157,73	7.132,52	-307,88	-197,05	-201,84

Informações como essas, somente obtida através de análises como a CHD, apresentam aos órgãos oportunidade de trabalharem em conjunto, tanto os pedidos, análises, decisões, respostas e transparência ativa, quanto as políticas públicas associadas a eles, em busca de soluções, também tomadas em parceria.

CONCLUSÃO

A pesquisa realizada permitiu evidenciar a relevância do texto como dado para pesquisas, principalmente quando se dispõe de métodos e técnicas como os que a ciência de dados tem nos trazido nas primeiras décadas do século XXI. Com a CHD, uma das técnicas de mineração e classificação de textos,

foi possível identificar, com mais clareza e de modo agregado, as temáticas mais abordadas nos pedidos de acesso à informação feitos à Prefeitura de São Paulo, entre os anos de 2012 e 2019, bem como dos órgãos destinatários das solicitações dos cidadãos. Destacaram-se os pedidos de acesso à informação relacionados a “Bairros e Distritos” (23,18%), “Trâmite e documentos processuais” (20,86%) e “Contratações Públicas” (16,40%). Sem a Ciência de Dados e suas ferramentas, não seria fácil interpretar e classificar quase 40.000 pedidos de acesso à informação, reunindo informações agregadas de apoio à tomada de decisão em contexto público.

Observando as Figuras 3 a 9 foi possível identificar, pelas diferentes cores, subclasses dentro de cada uma das classes, permitindo-se ir ainda mais fundo nos questionamentos apresentados, sendo cada cor uma dessas subclasses. Por fim, temos que a identificação de relações entre as classes temáticas e os órgãos a quem se destinaram os pedidos apresenta uma oportunidade para que o tratamento da informação e a tomada de decisão possa ser articulada em conjunto. Os problemas podem ser identificados em parceria, bem como as soluções podem ser buscadas em colaboração entre os órgãos.

A análise dos dados através da Classificação Hierárquica Descendente permite obter informação relevante para a tomada de decisão baseada em dados e evidências. Esta abordagem baseada em dados verificáveis e com qualidade favorece a existência de decisões fundadas e mais próximas das necessidades dos cidadãos.

REFERÊNCIAS

- Alves, M. S. D. (2012). Do sigilo ao acesso: Análise tópica da mudança de cultura. *Revista do Tribunal de Contas do Estado de Minas Gerais*, 85(esp.), 120-134. <https://revista1.tce.mg.gov.br/Content/Upload/Materia/1683.pdf>
- Angélico, F., & Teixeira, M. A. C. (2012). Acesso à informação e ação comunicativa: Novo trunfo para a gestão social. *Desenvolvimento em Questão*, 10(21), 7-27. <https://doi.org/10.21527/2237-6453.2012.21.7-27>
- Bagozzi, B. E., Berliner, D., & Almquist, Z. W. (2019). When does open government shut? Predicting government responses to citizen information requests. *Regulation & Governance*, 15(2), 280-297. <https://doi.org/10.1111/rego.12282>
- Balakrishnan, V., & Lloyd-Yemoh, E. (2014). Stemming and lemmatization: A comparison of retrieval performances. *Lecture Notes on Software Engineering*, 2(3), 262-267. <https://doi.org/10.7763/lmse.2014.v2.134>
- Banisar, D. (2006). *Freedom of information around the world 2006: A global survey of access to government information laws*. Privacy International. https://www.humanrightsinitiative.org/programs/ai/rti/international/laws_papers/intl/global_foi_survey_2006.pdf
- Berliner, D., Bagozzi, B. E., & Palmer-Rubin, B. (2018). What information do citizens want? Evidence from one million information requests in Mexico. *World Development*, 109, 222-235. <https://doi.org/10.1016/j.worlddev.2018.04.016>
- Berliner, D., Bagozzi, B. E., Palmer-Rubin, B., & Erlich, A. (2021). The political logic of government disclosure: Evidence from information requests in Mexico. *The Journal of Politics*, 83(1), 229-245. <https://doi.org/10.1086/709148>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Cao, L. (2017). Data science: A comprehensive overview. *ACM Computing Surveys*, 50(3), 1-42. <https://doi.org/10.1145/3076253>

Limitações do Trabalho e Sugestões de Investigações Futuras

Como limitações da investigação, tivemos a ausência de dados dos anos de 2020 a 2022, período em que o mundo enfrentou a pandemia da covid-19. Será relevante ver as prováveis alterações temáticas ocorridas. Percebeu-se também que alguns pedidos, tecnicamente, não são pedidos de acesso à informação nos termos da lei, mas pedidos de ajuda ou de solução de problemas. Outra limitação foi a existência de pedidos repetidos, para o mesmo órgão e para órgãos diferentes, que, apesar de terem sido identificados, não foram excluídos, podendo ter levado a ocorrência de algum viés (erro sistemático).

Para futuras explorações propomos a realização de estudos com dados de outras cidades e de outras esferas federativas, para verificar se há diferenças geopolíticas capazes de alterar a temáticas dos pedidos de acesso à informação, bem como para compreender a relação deste tipo de informação com os mecanismos de *accountability* disponíveis.

- Centre for Law and Democracy. (2017). *Global Right to Information Rating*. <http://www.rti-rating.org/country-data>
- Constituição da República Federativa do Brasil de 1988. (2022). Presidência da República. http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm
- Controladoria-Geral da União. (2019). *Escala Brasil Transparente 360º: Metodologia e critérios de avaliação*. Controladoria-Geral da União. <https://mbt.cgu.gov.br/static/Metodologia%20EBT.pdf>
- Convenção das Nações Unidas contra a Corrupção. (2003). Escritório das Nações Unidas contra Drogas e Crime. https://www.unodc.org/documents/lpo-brazil/Topics_corruption/Publicacoes/2007_UNCAC_Port.pdf
- Declaração dos Direitos do Homem e do Cidadão. (1789). Assembleia Nacional. <https://br.ambafrance.org/A-Declaracao-dos-Direitos-do-Homem-e-do-Cidadao>
- Declaração Universal dos Direitos Humanos. (1948). Organização das Nações Unidas. <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>
- Decreto nº 53.623, de 12 de dezembro de 2012. (2012, 13 de dezembro). Regulamenta a Lei Federal nº 12.527, de 18 de novembro de 2011, no âmbito do Poder Executivo, estabelecendo procedimentos e outras providências correlatas para garantir o direito de acesso à informação, conforme específica. Prefeitura de São Paulo. <http://legislacao.prefeitura.sp.gov.br/leis/decreto-53623-de-12-de-dezembro-de-2012>
- Duarte, J., & Theorga, A. B. (2012). O processo de implantação da Lei de Acesso à Informação em órgãos do Poder Executivo federal. *Comunicação & Informação*, 15(2), 66-79. <https://doi.org/10.5216/c&i.v15i2.24568>
- Flores, A. M., Pavan, M. C., & Paraboni, I. (2022). User profiling and satisfaction inference in public information access services. *Journal of Intelligent Information Systems*, 58(1), 67-89. <https://doi.org/10.1007/s10844-021-00661-w>
- Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535-574. <https://doi.org/10.1257/jel.20181020>
- Hollibaugh, G. E., Jr. (2019). The use of text as data methods in public administration: A review and an application to agency priorities. *Journal of Public Administration Research and Theory*, 29(3), 474-490. <https://doi.org/10.1093/jopart/muy045>
- Instituto Brasileiro de Geografia e Estatística. (2010). Censo 2010: São Paulo: panorama. *Portal do Governo Brasileiro*. <https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama>
- Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.
- Kotu, V., & Deshpande, B. (2018). *Data science: Concepts and practice* (2nd ed.). Morgan Kaufmann.
- Lane, J., Gimeno, E., Levitskaya, E., Zhang, Z., & Zigoni, A. (2022). Data inventories for the modern age? Using data science to open government data. *Harvard Data Science Review*, 4.2, 1-45. <https://doi.org/10.1162/99608f92.8a3f2336>
- Lei Modelo Interamericana sobre o Acesso à Informação Pública. (2010). Organização dos Estados Americanos. http://www.oas.org/dil/AG-RES_2607-2010_por.pdf
- Lei nº 12.527, de 18 de novembro de 2011. (2011, 18 de novembro). Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Presidência da República. http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/l12527.htm
- Martin, F., & Johnson, M. (2015, 8-9 December). More efficient topic modelling through a noun only approach. *Proceedings of the Australasian Language Technology Association Workshop 2015*, 13, 111-115. <https://aclanthology.org/U15-1013/>
- Mendes, A. M., Tonin, F. S., Buzzi, M. F., Pontarolo, R., & Fernandez-Llimos, F. (2019). Mapping pharmacy

- journals: A lexicographic analysis. *Research in Social & Administrative Pharmacy*, 15(12), 1464-1471. <https://doi.org/10.1016/j.sapharm.2019.01.011>
- Porumbescu, G.; Meijer, A.; Grimmelikhuijsen, S. *Government transparency: state of the art and new perspectives*. London: Cambridge University Press, 2022.
- Rydholm, L. (2013). China and the World's First Freedom of Information Act: The Swedish Freedom of the Press Act of 1766. *Javnost-The Public*, 20(4), 45-63. <https://doi.org/10.1080/13183222.2013.11009127>
- Sarica, S., & Luo, J. (2021). Stopwords in technical language processing. *Plos One*, 16(8), e0254937. <https://doi.org/10.1371/journal.pone.0254937>
- Savage, A., & Hyde, R. (2014). Using freedom of information requests to facilitate research. *International Journal of Social Research Methodology*, 17(3), 303-317. <https://doi.org/10.1080/13645579.2012.742280>
- Schofield, A., Magnusson, M., & Mimno, D. (2017, 3-7 April). Pulling out the stops: Rethinking stopword removal for topic models. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 2, 432-436. <https://aclanthology.org/E17-2069/>
- Soares, G. F. (2020). Ciência de dados aplicada à auditoria interna. *Revista da CGU*, 12(22), 196-208. <https://doi.org/10.36428/revistadacgu.v12i22.195>
- Sousa, Y. S. O. (2021). O uso do software Iramuteq: Fundamentos de lexicometria para pesquisas qualitativas. *Estudos e Pesquisas em Psicologia*, 21(4), 1541-1560. <https://doi.org/10.12957/epp.2021.64034>
- Walby, K., & Larsen, M. (2012). Access to information and freedom of information requests: Neglected means of data production in the social sciences. *Qualitative Inquiry*, 18(1), 31-42. <https://doi.org/10.1177/1077800411427844>
- Walby, K., & Luscombe, A. (2017). Criteria for quality in qualitative research and use of freedom of information requests in the social sciences. *Qualitative Research*, 17(5), 537-553. <https://doi.org/10.1177/1468794116679726>
- Wang, Z., & Zhong, Y. (2020). What were residents' petitions in Beijing-based on text mining. *Journal of Urban Management*, 9(2), 228-237. <https://doi.org/10.1016/j.jum.2019.11.006>



Claudio Henrique Fontenelle Santos

chfs@iscsp.ulisboa.pt

ORCID: <https://orcid.org/0000-0001-5237-2461>

Doutorando em Administração Pública pelo Instituto Superior de Ciências Sociais e Políticas da Universidade de Lisboa, Mestre em Administração Pública pela Universidade Federal da Bahia, Especialista em Comunicação, Publicidade e Propaganda pela Universidade de Fortaleza, Bacharel em Arquitetura e Urbanismo pela Universidade de Fortaleza e Bacharel em Administração de Empresas pela Universidade Estadual do Ceará. Pesquisa sobre transparência, Acesso à informação, governo aberto e ouvidorias públicas.



Ana Lúcia Romão

anaromao@iscsp.ulisboa.pt

ORCID: <https://orcid.org/0000-0003-2730-4007>

Doutorada em Economia, Investigadora no Centro de Administração e Políticas Públicas e Professora Auxiliar no Instituto Superior de Ciências Sociais e Políticas da Universidade de Lisboa. É Vice-Presidente do Centro de Administração e Políticas Públicas, Coordenadora Adjunta da Unidade de Coordenação de Administração Pública e Coordenadora Científica da Pós-Graduação em Contabilidade e Gestão Pública. Tem desenvolvido investigação na área da administração pública, nomeadamente no domínio da gestão pública, do controlo da gestão pública e da contabilidade pública.

Relato Técnico • A regulação baseada em dados: a era digital na agência reguladora de São Paulo¹

Data-based regulation: the digital era in the public services regulatory agency in the state of São Paulo

Regulación basada en datos: la era digital en la agencia reguladora de los servicios públicos del estado de São Paulo

Itamar Aparecido de Oliveira

<https://doi.org/10.36428/revistadacgu.v14i26.533>

Resumo: Este relato técnico apresenta a experiência da Diretoria Técnica de Saneamento Básico da Agência de Serviços Públicos do Estado de São Paulo (Arsesp) em seu trajeto de modernização das atividades regulatórias. O trajeto se estrutura em três frentes: criação de *data warehouse* que unifique as várias bases de dados; normatização do fluxo de dados entre regulados e regulador visando a padronização dos formatos e meios de envio; criação de grupo dedicado à análise de dados de atividades de suporte à elaboração de atos normativos e à capacitação de servidores em técnicas de análise de dados e uso de softwares específicos. Após quatro anos do início dessas ações, os resultados iniciais mostram a solidez do percurso: adoção de metodologias regulatórias alternativas, como a regulação por menus e a regulação Sunshine, baseadas em consumo de grandes bases de dados; elaboração de atos normativos fortemente embasados em análise de dados, resultando em redução de conflitos com o regulado; aplicação de técnicas de auditoria em bases de dados, induzindo o regulado a corrigir e atualizar suas fontes de informação; e a redução da assimetria de informação entre regulador e regulado, ampliando a percepção regulatória dos problemas enfrentados e das possíveis soluções.

Palavras-chave: Saneamento básico. Análise de dados. Digitalização. Modernização regulatória. Conhecimento.

Abstract: This technical report will present the experience of the Technical Director of Basic Sanitation of the Public Service Agency of the State of São Paulo – Arsesp in its paths to modernize regulatory activities. The path is structured on three fronts: the creation of a *data warehouse* that unifies its various databases; the standardization of the flow of data between the regulated and the regulator, promoting the standardization of formats and means of sending; the creation of a group dedicated to data analysis for activities to support the elaboration of normative acts and the training of civil servants in data analysis techniques and the use of specific software. Four years after these actions began, the initial results show the solidity of the path: adoption of alternative regulatory methodologies, such as menu regulation and Sunshine regulation, based on consumption of large databases; elaboration of normative acts strongly based on data analysis, resulting in a reduction of conflicts with the regulated; application of auditing techniques in databases, inducing the regulated to correct and update their sources of information and, finally, to reduce the asymmetry of information between the regulator and the regulated, expanding the regulatory perception of the problems faced and possible solutions.

Keywords: Basic sanitation. Data analysis. Digitalization. Regulatory modernization. Knowledge.

1. Artigo submetido em 15/07/2022 e aceito em 01/12/2022.

Resumen: Este reporte técnico presentará la experiencia del Director Técnico de Saneamiento Básico de la Agencia de Servicios Públicos del Estado de São Paulo – Arsesp en su camino de modernización de las actividades regulatorias. Éste se estructura en tres frentes: la creación de un almacén de datos que unifique sus distintas bases de datos; la estandarización del flujo de datos entre el regulado y el regulador, fomentando la estandarización de formatos y medios de envío; la creación de un grupo dedicado al análisis de datos para actividades de apoyo a la elaboración de actos normativos y la formación de funcionarios en técnicas de análisis de datos y uso de software específico. A cuatro años del inicio de estas acciones, los resultados iniciales muestran la solidez del camino: adopción de metodologías regulatorias alternativas, como la regulación del menú y la regulación Sunshine, basadas en el consumo de grandes bases de datos; elaboración de actos normativos fuertemente basados en el análisis de datos, resultando en una reducción de conflictos con lo regulado; aplicación de técnicas de auditoría en bases de datos, induciendo a los regulados a corregir y actualizar sus fuentes de información y, finalmente, a reducir la asimetría de información entre el regulador y los regulados, ampliando la percepción regulatoria de los problemas enfrentados y posibles soluciones.

Palabras-clave: Saneamiento básico. Análisis de datos. Digitalización. Modernización regulatoria. Conocimiento.

1. INTRODUÇÃO

Na década de 1990, durante o Programa Nacional de Desestatização, o país adotou a política do “Estado mínimo” por meio de privatizações, abandonando o histórico caráter intervencionista em detrimento da função regulatória nestes mesmos mercados. No lastro da evolução econômica, criaram-se as modernas agências reguladoras nas esferas federais, estaduais e municipais.

Agências reguladoras se constituíram em grandes consumidoras e/ou geradoras de dados relacionado ao mercado regulado. Todavia, a riqueza de dados comumente vem acompanhada de escassez de informação caso o regulador não disponha de infraestruturas adequadas a coleta, tratamento, armazenamento e análise que permitam transformar dados em informações que auxiliem no desenvolvimento de suas atividades.

Este relato técnico apresenta o caso da Agência Reguladora de Serviços Públicos do Estado de São Paulo (ARSESP), em especial da Diretoria Técnica de Saneamento Básico (DTSB). A Arsesp atua há 15 anos no setor, período em que se tornou um dos maiores detentores de dados relacionados a saneamento básico no estado. Porém, essa condição não refletiu imediata melhoria da qualidade regulatória, pois apresentava algumas fragilidades: (i) bases não eram padronizadas nem relacionais; (ii) inexistência de *data warehouse*;

(iii) não havia profissionais capacitados em análise de dados; (iv) ferramentas computacionais inadequadas para tratar os grandes bancos de dados existentes.

Em resposta, a DTSB realizou várias ações visando mitigar o problema, ainda que de forma não coordenada:

- (i) Estruturar *data warehouse* e permitir integração das bases de dados existentes na agência, adequando-as a modelos relacionais;
- (ii) Regulamentar meios de recebimento de novas bases de dados e detalhar formatos de variáveis e chaves primárias;
- (iii) Criar grupo dedicado à análise de dados como suporte aos estudos técnicos;
- (iv) Capacitar especialistas em técnicas de mineração e análise de dados, inteligência artificial e uso de softwares específicos a este fim.

Após a implementação das ações acima, a DTSB colheu os primeiros frutos: uso de novas metodologias regulatórias baseadas em consumo de grandes bases de dados, como a Regulação por Menus, reconhecida com o *Prêmio FGV Direito Rio – Melhores Práticas em Regulação 2021*²; desenvolvimento de deliberações fortemente apoiadas em análises de dados; aplicação de técnicas de auditoria nas bases de dados para garantir confiabilidade e exatidão, induzindo os prestadores a corrigirem e atualizarem seus sistemas.

2. O “Prêmio FGV Direito Rio – Melhores Práticas em Regulação”, tem como objetivo reconhecer e dar destaque às melhores práticas de regulação adotadas por agências reguladoras no Brasil, premiando iniciativas inovadoras em matérias como transparência, governança, participação, gestão e proteção aos direitos dos consumidores ou usuários. Cf.: <https://direitorio.fgv.br/premio-regulacao>.

2. DADO, INFORMAÇÃO, CONHECIMENTO: UM DESAFIO ESTRUTURANTE

A ARSESP, criada pela Lei Complementar n.º 1.025/2007, constitui-se numa autarquia de regime especial. Entre seus objetivos, destacam-se a regulação, controle e fiscalização dos serviços de saneamento básico de titularidade estadual ou de titularidade municipal delegados ao Estado de São Paulo pelos órgãos competentes, atividades que desenvolve atualmente em cerca de 350 municípios paulistas, incluindo a capital.

A DTSB se estruturou para realizar suas atividades considerando três estratégias principais:

- a) Realização de fiscalizações de campo para garantir que as normas editadas eram cumpridas pelo regulado, ao mesmo tempo em que imagens das estruturas fiscalizadas eram coletadas;
- b) Dados descrevendo os equipamentos de cada instalação ou a evolução de vários ser-

viços eram encaminhados rotineiramente pelos prestadores de serviços;

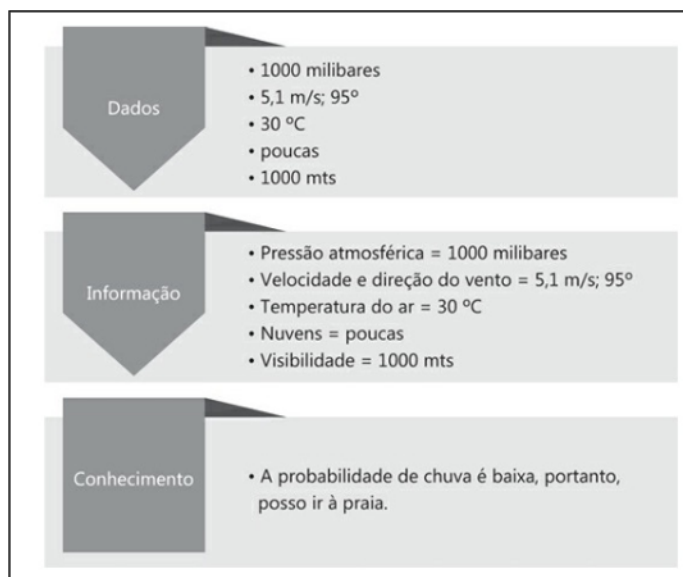
c) A partir dos dois eventos anteriores, a área normativa da Agência elaboraria regramentos para resolver os problemas identificados.

Tal estratégia, apesar de comum e observada na maior parte das Agências Reguladoras do Brasil, acarretou um problema no médio prazo: o acúmulo de dados não refletia aumento de informação por parte da Agência nem a construção de conhecimento a partir dos dados acumulados. Importa aqui estabelecer a diferença básica entre tais termos:

Os dados são símbolos ou signos não estruturados, sem significado, como valores em uma tabela, e a informação está contida nas descrições, agregando significado e utilidade aos dados [...]. Por fim, o conhecimento é algo que permite uma tomada de decisão para a agregação de valor. (CASTRO E FERRARI, 2016, p. 44)

O exemplo utilizado pelos autores auxilia a cristalizar a diferença entre as terminologias:

FIGURA 1 - DIFERENÇA ENTRE DADOS, INFORMAÇÃO E CONHECIMENTO



Fonte: Castro e Ferrari (2016, p. 45)

Ou seja, não obstante a Agência dispor de dados e informações acerca de cada equipamento dos prestadores de serviços regulados,³ a riqueza de dados não

se transformava em conhecimento estruturado. Mas o que impedia a transformação? Especialmente três problemas se apresentavam.

3. Por exemplo: dados sobre pressões, tipos de equipamento, potência, quantidade, volumes produzidos e consumidos, resultados de análises de qualidade, diâmetro, extensão e material de redes de abastecimento e de esgotamento, quantidade de vazamentos, etc.

O primeiro problema residia nos formatos de estrutura e armazenamento dos dados. Para cada fiscalização anual, uma base de dados atualizada era encaminhada em mídia eletrônica e anexada ao processo físico. Logo, tornou-se complexo unificar os dados num banco que permitisse comparar desempenhos entre anos para um mesmo município ou entre os vários municípios regulados. Além disso, comumente os dados estavam em formatos não estruturados (arquivos de imagens ou PDF, principalmente) ou semiestruturados, dificultando o trabalho com as informações.

Outro problema era a dispersão das bases de dados, mesmo entre aquelas que já dispunham de estrutura relacional para a análise. Havia bases em processos físicos e em desktops dos servidores responsáveis pelas fiscalizações ou atividades correlatas, o que impedia a visão geral dessas bases e do modo como seria possível relacioná-las.

Por fim, o terceiro obstáculo se dava pela ausência de profissionais capacitados para trabalhar com grandes bases de dados, fragilidade que se desdobrava em duas outras situações: de um lado, os profissionais não dominavam técnicas de análise de dados que permitissem a exploração controlada dos dados e informações disponíveis; de outro, não se disponibilizava aos profissionais softwares capazes de lidar com grandes bases. Obstáculo potencializado quando se considera a capacidade de geração de dados dos regulados.

Considerando apenas uma de suas 14 unidades de negócio, a Companhia de Saneamento Básico do Estado de São Paulo - Sabesp pretende “integrar cerca de 40 fontes de dados em seus diversos formatos, prevendo chegar a um volume aproximado de um 1 terabyte para a carga histórica dos dados. A previsão é alcançar 5 terabytes até 2025” (RENNER, 2021). Além disso, conforme noticiado por Ana Paula Lobo (2017), a empresa apresentou estratégia para a construção de um modelo de gestão e operação centrado nos conceitos da Saneamento 4.0 (baseados na Indústria 4.0 e IoT - Internet das Coisas). Mudanças estruturais no próprio modelo de negócio que exigirão do regulador rápida adequação para continuar atuando com excelência, visto que as mudanças e a necessidade de adequação se colocam como desafios, especialmente ao setor público.

Diante dos desafios, a DTSB concluiu pela necessidade de estruturar o consumo e produção de dados e informações e estabeleceu: i) organização das bases de

dados e criação do Sistema de Informações da Arsesp (SIA); ii) regulamentação dos meios de recebimento de novas bases de dados; iii) capacitação de servidores para atuar em análise de dados.

2.1. Bases de dados

Fase do trabalho desenvolvida em três frentes. Na primeira, a Gerência de Métodos, Controles e Dados reuniu as bases, organizando-as em diretório único. A centralização das bases e criação de fonte única se mostrou importante para reduzir a ocorrência de análises baseadas em bancos desatualizados e redução de duplicidades, inconsistências e conflito de variáveis. Ao todo, recolheu-se aproximadamente 33.131 bases de dados ou cerca de 4,28 GB de dados.

A segunda frente teve caráter institucional com a criação do Sistema de Informações da Arsesp (SIA), banco de dados com as informações críticas para rápida decisão regulatória dos gestores em situações de crise. O SIA congregará dados de todas as Diretorias da Arsesp, estruturando-os em modelo relacional, conforme descrito por Elsmari.

Os bancos de dados relacionais foram propostos originalmente para separar o armazenamento físico dos dados de sua representação conceitual e para fornecer uma base matemática para a representação e a consulta dos dados. O modelo de dados relacional também introduziu linguagens de consulta de alto nível, que forneciam uma alternativa às interfaces de linguagem de programação, tornando muito mais rápido a escrita de novas consultas. (ELMASRI, 2018, p. 21)

Não obstante os esforços empreendidos em ambas as fases, ainda restou sem tratamento cerca de 50 GB de dados históricos desestruturados ou semiestruturados. Para este conjunto, tem-se aplicado técnicas computacionais de conversão para bases relacionais objetivando recuperá-los.

Vencidos os desafios, ainda será necessário garantir a normalização dos dados para 3FN (3ª forma normal), garantindo a exclusividade das informações e evitando possíveis sobrecargas no gerenciador de banco de dados.

A normalização de dados pode ser considerada um processo de analisar os esquemas de relação dados com base em suas [dependências funcionais] e chaves primárias para

conseguir as propriedades desejadas de (1) minimização da redundância e (2) minimização das anomalias de inserção, exclusão e atualização [...], para fazer com que o projeto tenha uma qualidade cada vez melhor. (EL-MASRI, 2018, p. 430)

Também será necessário criar *data warehouse* que unifique frentes desenvolvidas até o momento, garantindo à DTSB o consumo fácil de todas as informações disponíveis. Ainda há muito trabalho à frente.

2.2. Padronização e Regulamentação

Paralelamente à organização das bases de dados, a Gerência de Regulação Técnica normatizou critérios de recebimento de novas bases de dados por meio da

deliberação ARSESP n.º 1.143/2021. A norma abordou níveis de criticidade, prazos de envio, meios de solicitação, sanções por descumprimento e estabeleceu o calendário anual de informações periódicas.

Este último documento se mostrou importante na estratégia de padronização das bases futuras, porque promoveu o alinhamento do regulado à nova estrutura de dados a fim de “evitar a duplicidade de informações, permitir o cruzamento de dados entre as instalações e municípios, com o objetivo de controlar e monitorar de maneira mais efetiva a prestação de serviço” (ARSESP, 2021, p. 8). O calendário estabelece ao regulado o que deverá encaminhar, bem como prazo, formatação, tipos de variáveis etc.

FIGURA 2 - EXEMPLO DE PADRONIZAÇÃO DE PEDIDO DE INFORMAÇÃO

Coluna	Tipo	Tamanho	Observação	Exemplo
codFormaAbastecimento	Texto	13 caract.	Código da instalação informado ao SISAGUA	S350010000001
nomeInstalacao	Texto	64 caract.	Nome Próprio da Instalação - deve caracterizar a ETA/UTA	ETA ADAMANTINA
tipoInstalacao	Texto	3 caract.	Caracterizar tipo de Sistema de Tratamento ETA/UTA	ETA
codSafi	Texto	15 caract.	Número do Cadastro da ETA no SAFI	RB0000000456
codMunicipio	Númerico	6 caract.	Código do município produtor no IBGE	350010
nomeMunicipio	Texto	64 caract.	Nome do município produtor no IBGE	Adamantina
codMunicipioAbast	Númerico	6 caract.	Código do município abastecido no IBGE.	355030
nomeMunicipioAbast	Texto	64 caract.	Nome do município abastecido no IBGE	São Paulo
unidadeNegocio	Texto	2 caract.	Unidade de negócio responsável pelo do sistema produtor	RB
nomeRegional	Texto	64 caract.	Nome da Regional responsável pela instalação	ADAMANTINA SABESP
dtAtualSISAGUA	dd/mm/aaaa	10 caract.	Data da atualização do cadastro no SISAGUA	01/01/2020
mes	mm/aaaa	7 caract.	Mês e ano do Cadastro	jan/20
numeroDeFiltros	Númerico	2 caract.	Número de filtros da instalação	4
tipoTratamento	Catagórico	-----	Ver Tabela 2	FILTRAÇÃO RÁPIDA
tipoManancial	Catagórico	-----	Ver Tabela 3	SUBTERRÂNEO
tempoMedioFuncionamento	hh.mm	5 caract.	Número médio de horas de funcionamento por dia	18:00
Volume Tratado	Númerico	13 caract.	Volume tratado disponibilizado à distribuição no mês em m ³	150000
latitude	Númerico	9 caract.	Latitude da instalação em grau decimal, 5 casas decimais	-21,69663
longitude	Númerico	9 caract.	Longitude da instalação em grau decimal, 5 casas decimais	-51,08095

Fonte: Calendário Anual de Informações Periódicas da Arsesp (2022).

Padronização e regulamentação mitigam a possibilidade de recebimento de dados em duplicidade e garantem uma base íntegra, condições para a manutenção do banco de dados, sua consistência e a persistência do modelo relacional.

2.3. Capacitação de servidores para análise de dados

Também se optou por criar um grupo específico que se responsabilizasse ou auxiliasse na análise de dados nos diversos estudos. Para estes a Agência ofe-

receu capacitação em técnicas básicas de análise de dados e uso do software *PowerBI*. Além do curso básico, alguns servidores da Agência receberam estímulo para pós-graduação nas áreas de análise de dados, inteligência artificial, *data mining* e geoprocessamento e no uso de softwares livres destinados à análise estatística (R e Python) ou de geoprocessamento (QGis).

A regulamentação baseada em dados simplifica o tempo, os custos e os processos. Métodos de análise inteligente aproveitam infraestruturas de computação baratas e seguras; dados com qualidade assegurada e pronta-

mente disponíveis; e uma força de trabalho crescente e bem treinada de cientistas de dados. (SEILER, 2017, tradução nossa)

Capacitar os servidores da Administração Pública em técnicas de análise de dados é essencial no enfrentamento dos problemas atuais e um caminho possível para que o poder público aloque os recursos de forma mais eficiente, indique soluções robustas e com isso aumente a satisfação do usuário em relação aos serviços ofertados.

3. RESULTADOS OBTIDOS

Mensurar o ganho obtido pela nova relação regulação/análise de dados é complexo, pois parte importante não é quantificável e envolve a estruturação de um ambiente propício à criação e à recepção de novas soluções analíticas. Elenca-se, contudo, três produtos entre os recentes trabalhos da Agência: adoção de metodologias baseadas em dados (Regulação por Menu e Regulação *Sunshine*); elaboração de dispositivos normativos baseados em análises de dados; elaboração de procedimentos de auditoria de bases volumosas.

3.1. Metodologias baseadas em dados

O aprimoramento no armazenamento, consumo e análise de dados permitiu que a Agência adotasse metodologias regulatórias alternativas ao modelo comando-controle, com melhores performance e alocação de recursos públicos, convergindo com o potencial de inovação possível a partir do uso de dados do setor público:

O uso de dados do setor público pode gerar produtos e serviços e, assim, contribuir de várias maneiras para melhorar a eficiência e a produtividade no setor público e em toda a economia. Os dados do setor público podem, assim, contribuir para criar sociedades e economias baseadas no conhecimento, onde os dados são um potencial impulsionador de crescimento, emprego, bem como de melhor prestação de serviços públicos e governança mais eficiente, transparente e participativa. (OECD, 2015, p. 406, tradução nossa)

O primeiro caso relatado se refere ao uso de Regulação por Menu, modelo desenvolvido pelos economistas Jean Tirolle e Jean-Jacques Laffont (1993) como solução ao desafio de se estabelecer metas regulatórias em ambientes com elevada assimetria de informação.

A metodologia de Regulação por Menu pauta-se pelo uso de indicadores e no consumo de bases de dados e pressupõe a capacidade de o regulador garantir fluxo de dados, armazenamento, auditoria e análise. A Regulação por Menu gerou ganhos consideráveis de desempenho nos indicadores e a iniciativa foi reconhecida com o **Prêmio FGV Direito Rio – Melhores Práticas em Regulação 2021**.⁴

A segunda metodologia a ser incorporada pela Arsesp foi a regulação *Sunshine* (Deliberação Arsesp nº 1.138/2021). Assim como a Regulação por Menu, a *Sunshine* se torna possível somente se o regulador tiver estruturado ambiente propício à análise de dados. Conforme Cunha (2005), a regulação *Sunshine* opera com a exposição dos regulados por meio de uma lista de melhores e piores em determinados serviços, pretendendo estimular os piores a melhorar seus desempenhos por meio do constrangimento. Vê-se a importância da análise de dados para garantir a assertividade do cálculo dos indicadores utilizados na comparação entre regulados.

Em ambos os casos, por meio da criação de base de dados decorrente das metodologias, o regulador pode associar técnicas descritivas a fim de melhor compreender os eventos que deseja controlar, reduzindo as assimetrias e refinando suas normas e metas.

3.2. Comandos baseados em dados

O uso de metodologias baseadas em dados é interessante não apenas como alternativa ao modelo comando-controle, mas também porque ao estruturar bases de dados confiáveis, permite que o regulador refine sua percepção do assunto regulado por meio das análises realizadas, gerando normas mais eficientes. Essa condição cria um ciclo virtuoso a todos os atores envolvidos na regulação: regulador, regulado, usuário, poder concedente e mercado.

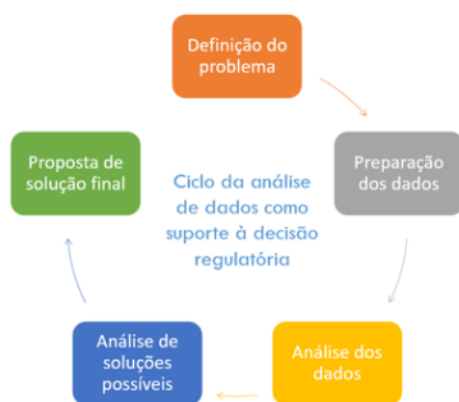
Um exemplo desse ciclo pode ser descrito na deliberação sobre prazos para reposição de pavimento, serviço assessorio que se encontra entre os principais objetos de reclamação por parte dos usuários. Na impossibilidade de se definir uma meta segura em de-

4. Para maiores informações acerca da Regulação por Menu, cf.: Oliveira, I. A. de, & Cruz, C. E. M. (2022). Índice Geral de Qualidade e a regulação por menus. *Revista De Direito Administrativo*, 281(1), 347–373. Recuperado em 10 julho de 2022, de: <https://bibliotecadigital.fgv.br/ojs/index.php/rda/article/view/85665>

corrência da elevada assimetria de informação, a Arsesp inseriu um indicador específico sobre o evento e o tratou na Regulação por Menus.

Conforme os bancos de dados se tornavam robustos e confiáveis, realizaram-se os processos referentes ao ciclo de análise exploratória para compreender o evento e definir parâmetros objetivos que deveriam ser cumpridos pelos regulados:

FIGURA 3 - FLUXO DA ANÁLISE EXPLORATÓRIA



Fonte: autoria própria

O Parecer Técnico nº 0041-2021⁵ que apresentou a proposta de solução final, embasando a Deliberação n. 1324/2022, foi objeto de elogios pelos atores regulados por entenderem que o novo procedimento atrelado em análise de dados é um significativo avanço na política de elaboração de atos normativos pela Agência Reguladora.

3.3. Procedimentos de auditoria

A intensificação do uso de dados para elaboração de diretrizes regulatórias, quer como aparato metodológico, quer como suporte aos atos normativos,

exige do regulador a capacidade de auditar as bases de dados utilizadas e garantir a confiança das fontes e a exatidão dos valores obtidos.

Para esse fim, a Arsesp desenvolveu metodologia para auditoria das bases volumosas, “procedimento escalonável e flexível estruturado em três fases principais (consistência, confiança e exatidão), podendo ser realizado com razoáveis prazo e equipe, garantindo a robustez nos resultados” (OLIVEIRA E MONTEIRO, 2021, p. 1008).⁶

5. Disponível em: http://www.arsesp.sp.gov.br/ConsultasPublicasBiblioteca/PARECER_TEC_S-0041-2021.pdf

6. O modelo de auditoria foi baseado no projeto Acertar, iniciativa do Ministério do Desenvolvimento Regional para auditar o Sistema Nacional de Informações de Saneamento – SNIS. Para maiores informações sobre o projeto Acertar, ver: <<http://www.snis.gov.br/projeto-acertar>>. Acesso em: 10. Jul. 2022.

FIGURA 4 - REPRESENTAÇÃO RESUMIDA DO PROCESSO MODULAR DE AUDITORIA ARSESP



Fonte: autoria própria

No primeiro ano de aplicação da metodologia, o ajuste nos resultados dos indicadores relacionados à Regulação por Menus representou o montante de 13 milhões de reais em favor dos usuários, conforme constante na Deliberação Arsesp nº 1150/2021. Além de garantir a equidade da ação regulatória, o desenvolvimento de auditorias nas bases de dados permite que o regulador se aproprie de todas as fases da geração do dado, desde a produção em campo, passando pelo armazenamento e chegando à extração nos sistemas do regulado, aproximação que reduz a assimetria de informação entre regulador-regulado e amplia a capacidade do regulador de elaborar normas assertivas.

4. RECOMENDAÇÕES

Há décadas vivenciamos um mundo cercado de informações, gerando um novo cenário para aqueles que dependem de dados no processo de tomada de decisão. No setor regulatório, a experiência não se deu de maneira distinta: regulados mantêm informações on-line, poderes públicos disponibilizam parte de seus bancos, agentes do setor têm milhares de bancos com informações geoespaciais, socioeconômicos, taxas, índices, séries históricas dos mais diversos setores.

Santos e Pozetti abordam a importância da regulação no contexto da economia digital e destacam que a “responsividade dos agentes a essas transformações

tem impacto profundo, porque, dado o caráter essencial das atividades, a morosidade ou incapacidade de adaptação pode afetar, inclusive, a sustentabilidade socioeconômica do país” (SANTOS E POZETTI, 2021, p. 27). Afirmativa que não deve se estreitar apenas ao contexto da economia digital. Ao contrário, deve ser entendida como uma premissa à modernização do poder público.

O uso prudente de recursos públicos pressupõe a capacidade de o órgão regulador analisar diferentes cenários, projetar soluções inovadoras, antecipar problemas, direcionar eficientemente sua força de trabalho, pressupostos que apenas têm validade em contexto de consumo, geração e análise dos dados disponíveis. Em todas as esferas da União, as modernas agências reguladoras já se apresentam como grandes consumidores e geradores de dados.

Resta, portanto, o esforço de estruturação de um novo ambiente que transforme tais dados e informações em conhecimentos, criando uma cultura de Regulação Baseada em Dados. O consumo deste imenso celeiro de dados permitirá melhores análises regulatórias e decisões mais refinadas, fortalecendo a segurança no ambiente regulado. Este é o caminho anunciado pelas ações da Diretoria de Saneamento da Arsesp.

5. REFERÊNCIAS

- Castro, L. N. & Ferrari, D. G. (2016). *Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações*. São Paulo: Saraiva.
- Elmasri, R. (2018). *Sistemas de banco de dados*. São Paulo: Pearson Education do Brasil.
- Laffont, J. J., Tirole, J. (1993). *A theory of incentives in procurement and regulation*. Cambridge, MA: MIT Press, 1993.
- Lobo, A. P. (2021). Sabesp faz projeto-piloto que une cloud, IoT e Inteligência Artificial. *Convergência Digital*. Recuperado em 11 de julho de 2022, de encurtador.com.br/mBLO5
- Marques, R. C. (2005). *Regulação de Serviços Públicos*. 1ª ed. Lisboa: Silabo.
- OECD (2015). *Data-Driven Innovation: Big Data for Growth and Well-Being*. Paris: OECD Publishing. Recuperado em 12 de julho de 2022, de https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page1
- Oliveira, I. A. & Monteiro, R. (2022). Procedimento de auditoria para bases volumosas de dados. *Anais do XII Congresso Brasileiro de Regulação*, p. 1007-1020. Recuperado em 10 de julho de 2022, de <https://abar.org.br/abar-divulga-anais-do-xii-congresso/>
- Oliveira, I. A. de, & Cruz, C. E. M. (2022). Índice Geral de Qualidade e a regulação por menus. *Revista De Direito Administrativo*, v. 281(1), p. 347-373. Recuperado em 10 de julho de 2022, de <https://bibliotecadigital.fgv.br/ojs/index.php/rda/article/view/85665>
- Renner, M. (2021). Sabesp: dados unificados com Engineering. *Baguete*. Recuperado em 11 de julho de 2022, de encurtador.com.br/qNP59
- Santos, A. C. K. A. dos & Pozetti, D. (2021). A importância da regulação no contexto da economia digital e da inovação. In Oliveira, C. R. & Vilarinho, C. M. R. *A regulação de infraestruturas no Brasil: Regulatory Beyond*. Santana do Parnaíba/SP: ABAR/KPMG.
- São Paulo (2021). Agência Reguladora de Serviços Públicos. *Deliberação nº 1150, de 08 de abril de 2021*. Recuperado em 26 de maio de 2022, de <http://www.arsesp.sp.gov.br/LegislacaoArquivos/ldl11502021.pdf>.
- São Paulo (2022). Agência Reguladora de Serviços Públicos. *Parecer Técnico n. 0041-2021*. Recuperado em 10 de julho de 2022, de <http://www.arsesp.sp.gov.br/SitePages/consultas-publicas.aspx>
- Seiler, E. (2017). *Smarter regulation through data science*. Recuperado em 10 de julho de 2022, de <https://fcw.com/digital-government/2017/08/smarter-regulation-through-data-science/257507/>



Itamar Aparecido de Oliveira

itamoliveira@sp.gov.br

Agência Reguladora de Serviços Públicos do Estado de São Paulo (ARSESP)

Lattes: <http://lattes.cnpq.br/1205260537760949>

Orcid: <https://orcid.org/0000000273548822>

Pós-graduado em Análise de Dados, Data Mining e Inteligência Artificial pela Fundação Instituto de Administração (FIA-SP). Especialista em Regulação e Fiscalização de Serviços Públicos III da Agência Reguladora de Serviços Públicos do Estado de São Paulo (ARSESP).

Relato Técnico • Proposta de *framework* de internacionalização no contexto da evolução da plataforma Lattes

José Francisco Salm Junior, Paulo Henrique de Assis Sant’ana e Geraldo Sorte

Evolução do Lattes, Fontes de Dados Curriculares, Ontologias de Currículo, Internacionalização, *Framework* de Cooperação

INTRODUÇÃO

Em um esforço para evoluir iniciativas ligadas à gestão da informação em ciência e melhorar os procedimentos de extração de dados na Plataforma Lattes e sua interoperabilidade com fontes de dados curriculares, o Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e a Fundação para a Ciência e a Tecnologia – FCT de Portugal estabeleceram, em novembro de 2021, um acordo para um programa de cooperação científica, tecnológica e de inovação focado no desenvolvimento conjunto de atividades de pesquisa e outras iniciativas em temas de comum interesse. Dentre as intenções do Acordo de Cooperação Internacional (ACI), previa-se o intercâmbio de pesquisadores(as) nos temas de gestão de dados científicos e repositórios de dados em pesquisas correntes.

Neste contexto, apresenta-se nesse relato técnico a proposta de intervenção, denominada *framework* CV-PTBR, que norteará o trabalho de assimilação de tecnologias e seus elementos para o uso nos projetos de aprimoramento das rotinas de importação de dados na Plataforma Lattes e a evolução das integrações de dados na Plataforma CIÊNCIAVITAE de Portugal. Tal *Framework* objetiva estabelecer um protocolo de trabalho conjunto, baseado em diretrizes, entre as equipes do CNPq e FCT. No caso do Currículo Lattes

(CV Lattes), a possibilidade de aprimoramento com a adoção do *Framework* dar-se-á pela apresentação de protótipo de interfaces e a proposição de cenários, em especial relativos à melhoria das rotinas de dados e ao uso de padronização internacional para sua manutenção e atualização.

No início de abril de 2022, iniciou-se pesquisa no âmbito do ACI, cujo escopo de atividades tratou da avaliação das ontologias de currículos e de cenários de evolução do CV Lattes, concentrando esforços na identificação e no estudo de modelos internacionais (i.e. VIVO *Ontology*¹, CERIF² e ORCID³) como potenciais referências para o projeto de evolução do modelo ontológico do referido CV Lattes. A pesquisa também contemplou o estudo de propostas de modelos narrativos de *Curricula Vitae* como um elemento de evolução do CV Lattes para aprimoramento de seu uso em procedimentos de avaliação ad hoc e do uso de recursos de extração de dados para o apoio à criação das narrativas, tanto para quem preenche seu CV quanto para as pessoas responsáveis por sua avaliação. Complementar ao plano de trabalho da pesquisa, elaborou-se um estudo com base no reaproveitamento de ativos colocados à disposição dos envolvidos no ACI. Esse estudo produziu recomendações e a proposição do *framework* CV-PTBR para trabalho conjunto, vi-

1. Conforme o site da VIVO, a Ontologia VIVO fornece um conjunto de tipos (classes) e relações (propriedades) para representar os dados sobre pesquisadores e o contexto em que trabalham. A Ontologia VIVO está vinculada a outras Ontologias, que fornecem um contexto complementar e significado em suas relações e tipos. Informações adicionais estão em: <https://wiki.lyrasis.org/display/VIVODOC111x/Ontology+Reference>.

2. Conforme o site do CERIF, o Formato Europeu Comum de Informação sobre a Pesquisa (CERIF) é o modelo de informação abrangente para o domínio da investigação científica. Destina-se a apoiar o intercâmbio de informações de pesquisa entre e com os Sistemas de Informação Sobre Pesquisas Correntes (CRIS). Mais informações estão disponíveis em: <https://eurocris.org/services/main-features-cerif>.

3. Conforme o site do ORCID, esse serviço serve como um centro de informações, permitindo que autores e revisores se conectem de forma confiável às suas contribuições e compartilhem informações de seus currículos registrados enquanto eles interagem com seu sistema. Mais informações estão disponíveis em: <https://info.orcid.org/what-is-orcid/>.

sando a sustentabilidade e a coprodução de ativos de software que serão adotados pelas organizações envolvidas no ACI.

Os elementos utilizados no diagnóstico incluíram o modelo de fontes e bases importadas ou sincronizadas pela Plataforma CIÊNCIAVITAE, a forma de estruturação dessas operações de dados e o modelo de resolução de conflitos ou inconsistências após a execução dessas rotinas. A proposta do integrador e sincronizador da Plataforma CIÊNCIAVITAE é semelhante a um pool de serviços de integração. Conceitualmente, se estabelece um barramento de dados, cujo objetivo é permitir interoperabilidade técnica, não tendo necessariamente base em um *data lake*.

Os objetivos esperados ao se adotar o *framework* CV-PTBR proposto nesse relatório são, além dos benefícios do modelo de internacionalização, o aumento e a simplificação do reaproveitamento dos ativos de interoperabilidade técnica. Também se espera, com a adoção do *framework* CV-PTBR, que se descentralizem as iniciativas de evolução das tecnologias ligadas ao currículo e a melhoria do compartilhamento dos dados associados aos currículos nos respectivos repositórios, tanto no contexto nacional, quanto no contexto do Acordo Binacional Brasil-Portugal.

Caso seja uma decisão dos representantes do país adotante do *Framework*, esse poderá apoiar o envolvimento de técnicos nas universidades para que contribuam com a evolução da tecnologia, provendo também apoio nas atividades de documentação técnica e no material de gestão de dados para os recursos comuns derivados do *framework* CV-PTBR.

Entende-se que os beneficiários deste *framework* terão diferentes expectativas em relação ao valor produzido por esse modelo. Inicialmente, foram considerados beneficiários todos aqueles representantes ligados à gestão da informação em ciência, tecnologia e inovação. De forma mais ampla, consideraram-se também os representantes da sociedade que desejem maior transparência das informações relativas aos resultados obtidos em decorrência do investimento público em formação e na realização de pesquisa. Dependendo da política de inovação estabelecida pelo país adotante do *Framework*, podem ser considerados beneficiários os representantes do segundo setor que desejarem prover serviços com os dados disponibilizados pela adoção do futuro *framework* CV-PTBR.

Diferentes representantes do governo em suas variadas esferas (i.e. nacional, regional ou estadual) terão a oportunidade de desenvolver iniciativas próprias de uso dos ativos de dados derivados do *framework* CV-PTBR. No caso do Brasil, esses ativos poderão ser disponibilizados aos agentes estaduais de CT&I, como as Fundações de Apoio à Pesquisa, ou no contexto Federal, bem como a outras agências de fomento que desejarem os dados para uso em seus processos internos. Aos agentes públicos será possível coordenar aplicações entre órgãos no uso desses ativos ou serviços, em um ciclo continuado de melhoria desses ativos decorrentes da aplicação do *Framework* proposto.

A seguir são apresentadas diretrizes de adoção do *Framework*, que deverão apoiar a tomada de decisão em relação ao seu uso nas ações de evolução dos ativos de software em plataformas de informação curricular.

Diretrizes da Adoção do *framework* CV-PTBR

As diretrizes para a adoção do *framework* representam elementos norteadores para projetos de manutenção dos ativos produzidos durante os ciclos compartilhados de engenharia das plataformas de informação curricular. As seis diretrizes reforçam as práticas de independência na execução dessas ações e a autonomia das equipes em cada país, estabelecendo acordos que apoiarão a condução dessas iniciativas conjuntas.

Assimilação e Adoção dos ativos produzidos a qualquer tempo

A criação de artefatos feita por uma organização, com base no *Framework*, deve permitir seu uso a qualquer tempo por parte das entidades participantes da cooperação, garantindo que esse artefato seja suficientemente modularizado para encaixe e agregação em ativos existentes na plataforma de informação curricular adotante.

Planejamento e Gestão de Projeto Desvinculados

A assimilação e adoção dos ativos não está associada a questões temporais ou à dependência de cronogramas entre as equipes participantes da cooperação. A adoção do *Framework* e seus elementos não obriga o alinhamento de cronogramas entre as organizações participantes.

Priorização de reaproveitamento dos ativos

No contexto da cooperação, os adotantes do *framework* poderão manifestar seu interesse em priorizar o reaproveitamento de novos ativos. Essa manifestação de interesse não gera obrigações aos demais participantes da cooperação em relação ao cronograma, como apontado pela diretriz de Planejamento e Gestão de Projeto Desvinculados.

Investimento e financiamento independentes

A organização participante da cooperação terá independência na execução de projetos para implementação de ativos compartilhados e cabe somente a essa participante a definição do volume de investimento em projetos, respeitando sua condição e autonomia de execução orçamentária.

Priorização da adoção de padrões internacionais

A organização participante da cooperação, sempre que possível, priorizará o uso de padrões internacionais adotados no contexto da informação curricular. Esses padrões podem representar ontologias, protocolos de comunicação, harvesting e padrões de identificadores persistentes entre outros.

Respeito à realidade cultural e nacional

Considerando que cada país estabelece suas normas e legislações, possuindo elementos culturais próprios, a adoção do *Framework* deve reforçar a configurabilidade e adaptabilidade dos ativos compartilhados, de modo a permitir sua configuração dentro de rotinas de operação da organização participante. Dentre esses aspectos, pode-se destacar a legislação sobre Acessibilidade Web, Lei de Proteção de Dados, Design systems locais e outros regramentos estabelecidos por iniciativas de governo eletrônico.

Estas diretrizes devem promover a redução de riscos relacionados à dependência de cronogramas, ao mesmo tempo que aumentam a autonomia de decisão na implementação das demandas regionais, sem obrigar o comprometimento prévio de orçamento por parte das organizações adotantes do *Framework*.

Na próxima sessão serão apresentados os elementos componentes do *Framework* de localização, as características do programa de construção do *framework* e suas fases.

Elementos do Framework de CV de língua Portuguesa – CV-PTBR

A internacionalização é entendida como um processo integral, onde a perspectiva universal permeia as informações curriculares, em especial aquelas ligadas às atividades acadêmicas por conta de sua natureza, abrangendo desde a pesquisa, os programas acadêmicos, a inovação e extensão, a formação continuada de professores e a evolução das práticas profissionais.

Características do Programa de Construção do framework CV-PTBR

A adoção do *framework* CV-PTBR será reavaliada a cada ciclo de tempo, permitindo assim a evolução desse modelo e a proposição de melhorias por parte das organizações que o adotarem. A possibilidade de execução em fases permitirá validar o modelo com casos práticos de criação de ativos baseados no *Framework*.

Recursos candidatos e melhorias desejáveis na adoção do framework CV-PTBR

A seguir são apresentados alguns recursos e melhorias desejáveis na adoção do novo *framework* CV-PTBR. Esses recursos são sugeridos nesse relatório, mas deverão passar pela revisão e avaliação das lideranças envolvidas no Acordo de Cooperação Internacional, podendo futuramente representar requisitos implementados em ativos de gestão de dados compartilhados.

- Fornecer uma atualização dos dados do perfil do CV, em duas vias, para outras APIs como, por exemplo, o ORCID. Ou seja, se existem dois sistemas diferentes, cada um deles com um perfil representando a mesma pessoa, que haja um mecanismo para sincronizar esses dois perfis.
- Fornecer uma atualização dos dados dos perfis nos currículos, em duas vias, para a dupla CIÊNCIA-VITAE/Lattes (dependendo da origem da requisição de atualização).
- Fornecer uma atualização de perfil unidirecional com o Oasisbr (Portal Brasileiro de Acesso Aberto à Informação Científica), LattesData (Repositório de dados abertos de pesquisa da Plataforma Lattes), PICC (Plataforma Integrada Carlos Chagas de gestão do fomento no CNPq) e Plataforma Sucupira (Plataforma

de avaliação de cursos na CAPES).

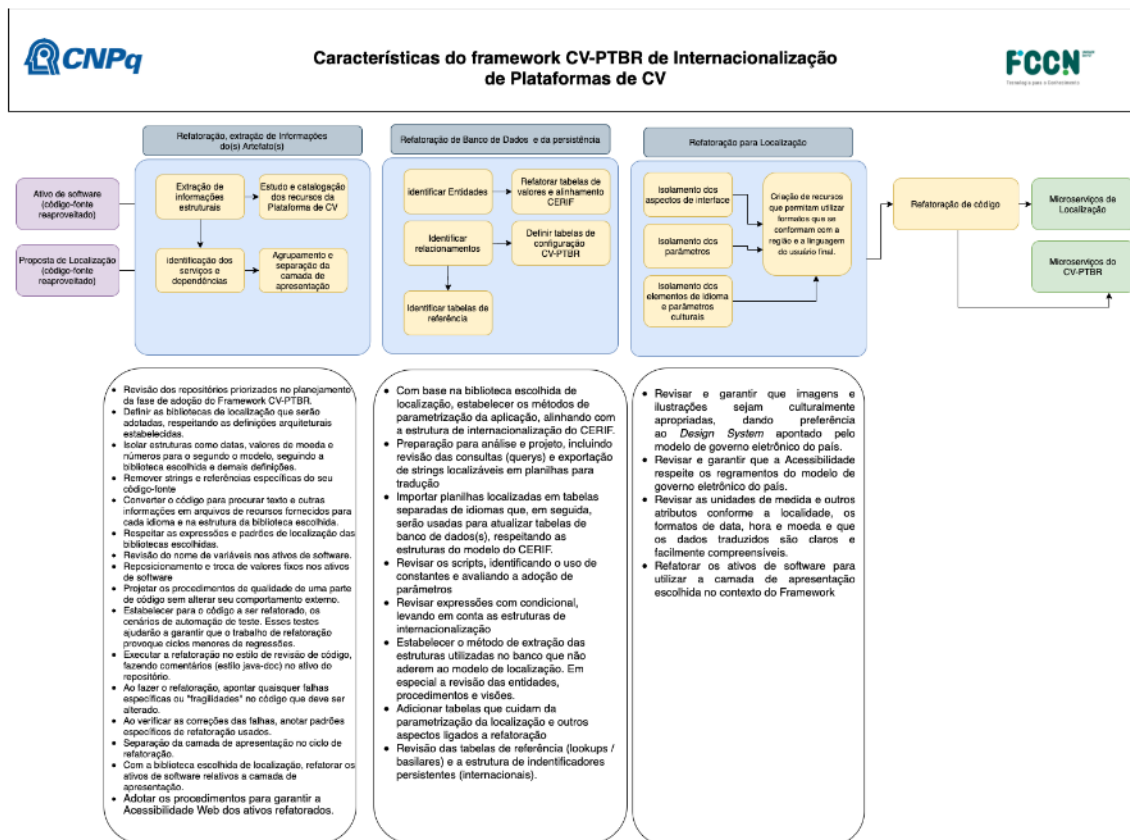
- Fornecer uma rotina de *upload* e *download* dos trabalhos de uma pessoa para/de o formato BibTex.
- Fornecer uma rotina de *upload* e *download* de registro de produções (i.e. técnicas, científicas e artísticas) e projetos de uma pessoa para/de em formato de preferência aberto.
- Fornecer uma carga de metadados do Google Scholar, Mendeley, Zotero ou aplicações semelhantes.
- Prover agrupamento de suporte de qualquer parte dos dados do CV — definindo uma ação de grupo. Esse recurso viabilizaria análise em redes de pesquisa e formaria uma estrutura de apoio para iniciativas como o Diretório de Grupos de Pesquisa do CNPq.
- Fornecer rotinas de indexação e a criação de um portal de pesquisa unificado Lattes (*Lattes Research Portal*) com uma interface de busca, permitindo que o resultado da busca utilize classes (conceitos), como por exemplo, busca por produção científica, tecnológica e artística, busca por organizações de pesquisa, fomento de projetos e permitir a criação de subdomí-

nios do portal para uso pelas universidades.

- Viabilizar suporte para análise de impacto de pesquisa. Prover extensões que utilizem a ontologia do CV e suas saídas e associá-los aos dados de impacto de pesquisa.
- Permitir a visualização dos resultados da busca com representação desses pontos de dados em um mapa. Permitir associar as informações curriculares que possuem datas em uma linha do tempo.
- Fornecer capacidade de indexação dos dados para que seja possível encontrar especialistas, incluindo buscas que selecionem especialistas com perfil semelhante ao do que caracteriza determinado projeto de pesquisa.

Como proposição da intervenção, a seguir apresenta-se as características do *framework* CV-PTBR divididas em: aspectos de refatoração na extração de informações dos artefatos existentes; aspectos de refatoração considerados para o banco de dados e persistência dos ativos existentes e aspectos de refatoração geral que promovam a localização e as mudanças necessárias na camada de apresentação dos dados (ver Figura 1).

FIGURA 1 - CARACTERÍSTICAS DO FRAMEWORK CV-PTBR



Aspectos de refatoração considerados na adoção do Framework e na extração de informações do(s) artefato(s) existentes:

- Revisão dos repositórios priorizados no planejamento da fase de adoção do *framework* CV-PTBR.
- Definir as bibliotecas de localização que serão adotadas, respeitando as definições arquiteturais estabelecidas.
- Nas rotinas que fazem acesso ao banco de dados, isolar estruturas tais como datas, valores e tipos de moeda e números para a nova versão do artefato de software (i.e. resultado da refatoração), seguindo a biblioteca escolhida e demais definições. Cuidar da completude dos dados, realizando rotinas de verificação antes de iniciar a revisão dos objetos no banco de dados.
- Projetar os procedimentos de qualidade de dados sem alterar o comportamento da rotina.
- Estabelecer, para o código a ser utilizado, os cenários de automação e testes. Esses testes ajudarão a garantir que o trabalho de reaproveitamento de ativos

provoque ciclos menores de ajustes.

- Executar a refatoração no estilo de revisão de código-fonte, fazendo comentários que expliquem o que foi modificado (estilo java-doc) no ativo de software do repositório.
- Apontar quaisquer falhas específicas ou “fragilidades” no código que devam ser alteradas durante os procedimentos de readequação e reaproveitamento do ativo de software.
- Anotar padrões específicos de refatoração usados.
- Adotar os procedimentos para garantir a Acessibilidade Web dos ativos de software utilizados.

Aspectos de refatoração considerados na adoção do framework para o Banco de Dados e da persistência para os ativos existentes:

- Com base na biblioteca de localização escolhida, estabelecer os métodos de parametrização da aplicação, alinhando-os com a estrutura de internacionalização do CERIF.
- Preparação para análise e projeto, incluindo revisão das consultas (*queries*) e exportação de textos (*strings*)

localizáveis em planilhas para tradução.

- Importar planilhas localizadas em tabelas separadas de idiomas que, em seguida, serão usadas para atualizar tabelas de banco de dados, respeitando as estruturas do modelo do CERIF / VIVO e CIÊNCIAVITAE.
- Revisar os scripts, identificando o uso de constantes e avaliando a adoção de parâmetros.
- Estabelecer o método de extração das estruturas utilizadas no banco que não aderem ao modelo de localização. Em especial, a revisão das entidades, procedimentos e visões materializadas no sistema gerenciador de banco de dados (SGBD).
- Adicionar tabelas que cuidem da parametrização da localização e outros aspectos ligados ao ciclo de revisão dos ativos de software.
- Revisão das tabelas de referência (*lookups* / basilares) e a estrutura de identificadores persistentes (internacionais).

Aspectos de revisão considerados na adoção do *framework* para os elementos de localização na camada de apresentação:

- Revisar e garantir que imagens e ilustrações sejam do *Design System* apontado pelo modelo de governo eletrônico do país.
- Revisar e garantir que a Acessibilidade Web e a legislação de proteção de dados (e.g. LGPD e GDPR) atendam aos regramentos do modelo de governo eletrônico do país.
- Revisar as unidades de medida e outros atributos conforme a localidade, os formatos de data, hora e moeda e que os dados traduzidos sejam claros e facilmente compreensíveis.
- Refatorar os ativos de software para utilizar a camada de apresentação escolhida no contexto do *Framework* CV-PTBR.

Os aspectos apresentados acima não devem ser considerados como uma lista exaustiva de elementos ou mesmo uma lista finalizada para adoção do *fra-*

mework CV-PTBR. Como já informado, o *Framework* deve ser validado e adaptado após a realização do primeiro ciclo de adoção.

Cenário de importação e sincronização dos dados na base Lattes

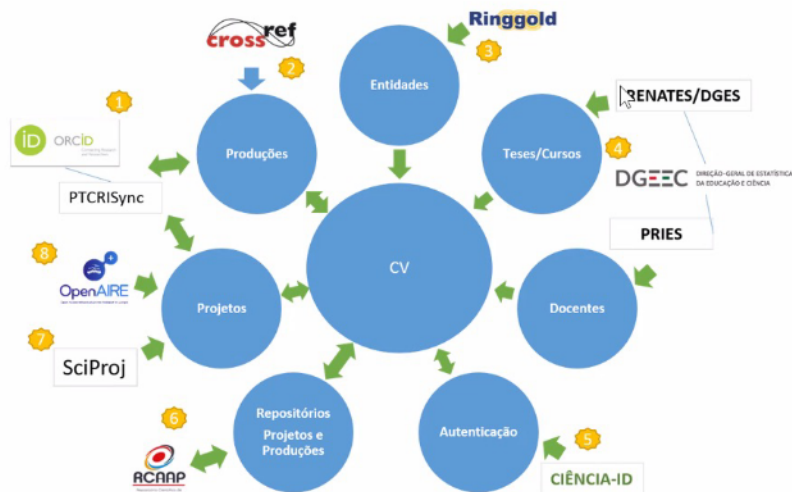
Nesse cenário, avaliaram-se três aspectos relacionados às rotinas de importação e sincronização de dados implementados no CIÊNCIAVITAE e seu modelo de serviços. Essa avaliação foi realizada por meio revisão documental, revisão simples de código-fonte e entrevistas com o engenheiro de software responsável na Fundação para a Computação Científica Nacional - FCCN de Portugal. Nessa parte do estudo, verificou-se a possibilidade de reaproveitamento dos artefatos de software mantidos pela FCCN e a possível aplicação do *framework* CV-PTBR na criação de um módulo isolado na Plataforma Lattes para importação e sincronização de dados.

Elementos utilizados na avaliação do Cenário

Os elementos utilizados no presente cenário incluíram o modelo de fontes e bases importadas ou sincronizadas pelo CIÊNCIAVITAE, a forma de estruturação dessas duas operações e o modelo de resolução de conflitos ou inconsistências de dados após a execução dessas rotinas, considerando o *framework* CV-PTBR.

Como já apresentado, a proposta do integrador e sincronizador de dados do CIÊNCIAVITAE e PTCRIS (*Current Research Information System* de Portugal) é semelhante a um pool de serviços de integração e as fontes de informação são apresentadas na Figura 2.

FIGURA 2 - MODELO DE IMPORTAÇÃO/SINCRONIZAÇÃO DO CIÊNCIAVITAE



Fonte: FCCN

O modelo de importação/sincronização do CIÊNCIAVITAE estabelece uma visão de “central de serviços” que cuidam da importação ou sincronização de conteúdos para o currículo. No caso específico do ORCID, após o primeiro ciclo de importação, é possível manter a sincronização de dados com a base CIÊNCIAVITAE. Nessa sincronização, é facultado ao operador realizar a gestão de ocorrências e/ou inconsistências dos dados entre essas fontes sincronizadas.

No modelo do CIÊNCIAVITAE, cada serviço de importação depende de um conjunto de regras/rotinas de preparação de dados. Um exemplo disso foi a preparação feita pela equipe da FCCN, em conjunto com as Instituições de Ensino Superior Portuguesas, estabelecendo uma rotina de dados cadastrais dessas Instituições. Atualmente, os dados são atualizados semanalmente por uma rotina de extração, transformação e carga (ETL) entre a base do *Ringgold* e as bases do CIÊNCIAVITAE e de outras aplicações da FCCN. O

Ringgold é uma fonte de informações sobre a estrutura organizacional de Instituições e Ensino Superior e Instituições de Pesquisa.

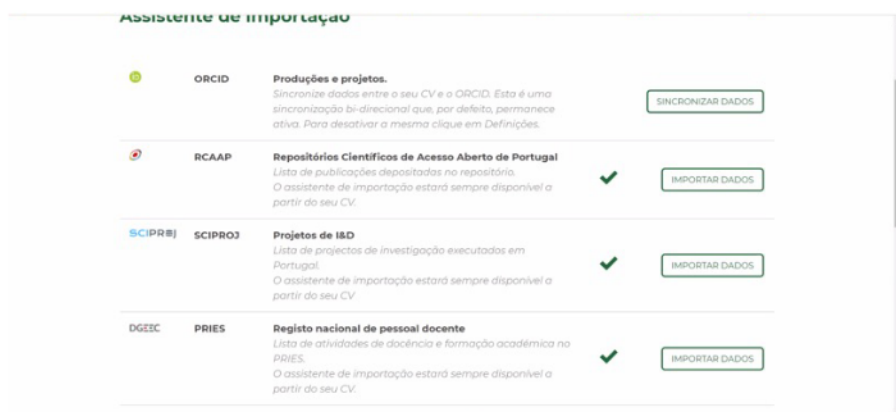
Em outra situação de importação de dados por parte do CIÊNCIAVITAE, para a sincronização de dados curriculares do sistema ORCID, a implementação do CIÊNCIAVITAE utiliza um cliente REST para as APIs que foi derivado de um pacote publicado em licença de código aberto. A rotina realiza a comparação entre a produção (output) informada no CIÊNCIAVITAE e o que está informado no ORCID e notifica quando as fontes não estão sincronizadas. Cabe ao operador do CIÊNCIAVITAE aceitar ou não a sincronização dos dados. A Figura 3 demonstra a interface do assistente de importação do CIÊNCIAVITAE, com a opção de sincronização de dados do ORCID e importação de dados do RCAAP⁴, SCIPROJ⁵ e DGEEC⁶ entre outras.

4. Conforme o site do portal RCAAP (Repositório Científico de Acesso Aberto de Portugal), esse portal tem como objetivo a coleta, agregação e indexação dos conteúdos científicos em acesso aberto (ou acesso livre) existentes nos repositórios institucionais das entidades nacionais de ensino superior, e outras organizações de I&D. Mais informações em: <https://www.rcaap.pt/about.jsp>

5. Base de dados com o registro nacional de financiamento de ciência e tecnologia de Portugal.

6. A DGEEC disponibiliza as bases de dados da área da Educação e da Ciência e Tecnologia para fins de investigação, de acordo com o Protocolo estabelecido com o Instituto Nacional de Estatística e com a Fundação para a Ciência e a Tecnologia de Portugal (FCT).

FIGURA 3 - INTERFACE DO ASSISTENTE DE IMPORTAÇÃO DO CIÊNCIAVITAE PARA ESCOLHA DAS FONTES



Fonte: FCCN

Como já mencionado, em um esforço para a adoção de identificador persistente, a equipe da FCCN fez a preparação e a atualização dos dados das IES na fonte do *Ringgold*. O *Ringgold Identifier* é um identificador persistente para organizações que inclui cerca de quinhentos mil IDs representando organizações e consórcios que utilizam conteúdos de publicações e outras fontes acadêmicas. Para manter as informações atualizadas no CIÊNCIAVITAE e outros projetos que utilizam esses dados, a equipe da FCCN realiza periodicamente o *download* e a atualização dos dados do *Ringgold*. O processo utiliza-se de uma rotina em conjunto com a ferramenta de ETL da Talend e inclui o *download* do arquivo com os dados via FTP. O processo está estruturado na ferramenta de ETL onde o script cuida da descompactação do arquivo e das rotinas de atualização dos dados na base operacional. A rotina é executada a cada ciclo de tempo para garantir que a atualização feita pelas organizações no *Ringgold* possa estar refletida nas bases da FCCN.

A rotina do assistente de importação também se utiliza da interface do *OpenAIRE*⁷, sendo nesse caso uma iniciativa de projeto em andamento. A identificação de publicações é feita com 28 repositórios portugueses de dados de pesquisa (repositórios da Universidade do Minho, entre outros, hospedados pela rede

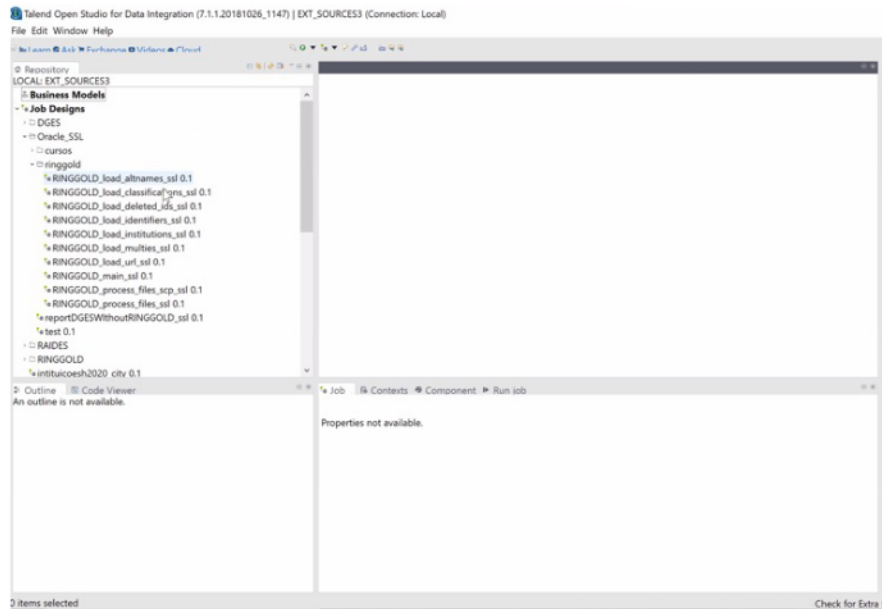
RCAAP) a partir de um conjunto de 10 mil projetos financiados pela FCT. Essa abordagem sugere um modelo descentralizado de dados para alimentar a informação curricular nacional.

Outra integração é com a base do SCIPROJ, um repositório de dados sobre projetos financiados pela FCT e desenvolvido seguindo os padrões estabelecidos pela OpenAIRE CERIF, tendo sido implementado uma interface cliente para realizar sua importação de dados. Outras fontes podem ser consideradas candidatas ao projeto conjunto com o currículo Lattes, em especial a integração com o RCAAP. Para aproveitamento desses ativos criados pela FCT, a intervenção proposta neste relatório novamente sugere a adoção do *framework* CV-PTBR.

A seguir são apresentadas imagens das rotinas e ferramentas utilizadas pelos scripts de atualização dos dados da base do *Ringgold*. Para tal, a Figura 4 demonstra o ambiente do *software* Talend e a edição dos scripts do *Ringgold*; a Figura 5 apresenta a rotina de sincronização do *Ringgold* com notificação via e-mail e a Figura 6 demonstra uma parte do procedimento (*Script Oracle*) para atualização da visão materializada do *Oracle* para o *Ringgold*. Essas figuras representam diferentes aspectos e elementos da rotina de atualização da base.

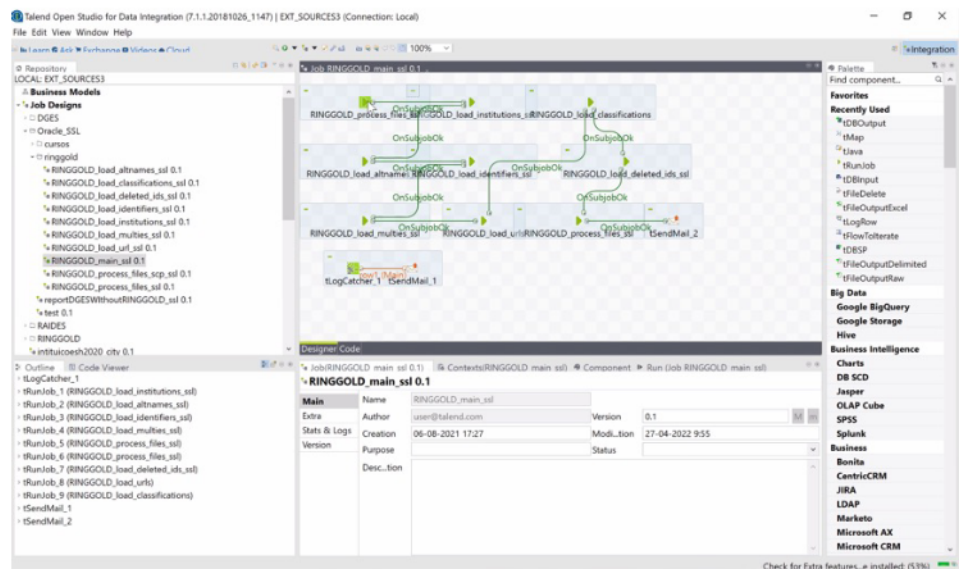
7. Conforme informado pelo site, o OpenAIRE (Open Access Infrastructure for Research in Europe) tem como missão mudar a comunicação científica para a abertura e a transparência e facilitar formas inovadoras de comunicar e monitorizar a investigação. Informações adicionais estão disponíveis em: <https://www.openaire.eu/mission-and-vision>.

FIGURA 4 - IMAGEM DO TALEND NA EDIÇÃO DOS SCRIPTS DO RINGGOLD



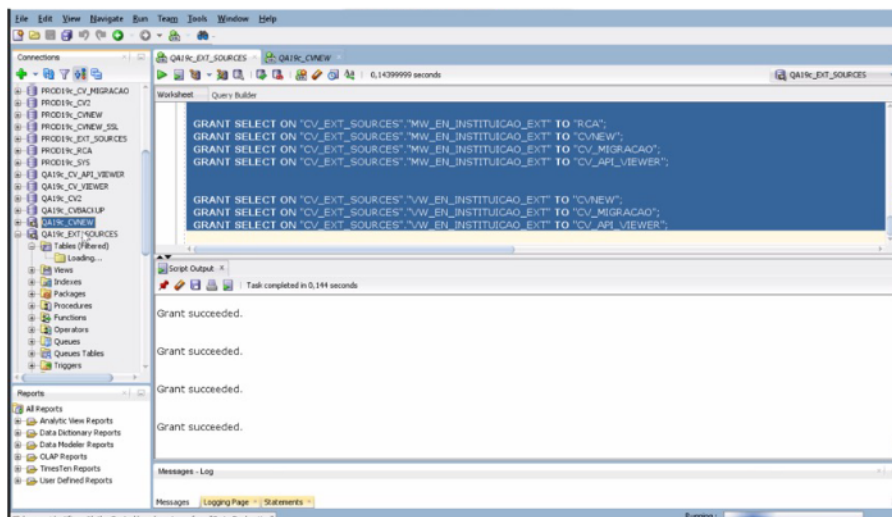
Fonte: FCCN

FIGURA 5 - IMAGEM DA ROTINA DE SINCRONIZAÇÃO DO RINGGOLD COM NOTIFICAÇÃO VIA E-MAIL



Fonte: FCCN

FIGURA 6 - SCRIPT ORACLE PARA ATUALIZAÇÃO DA VISÃO MATERIALIZADA RINGGOLD

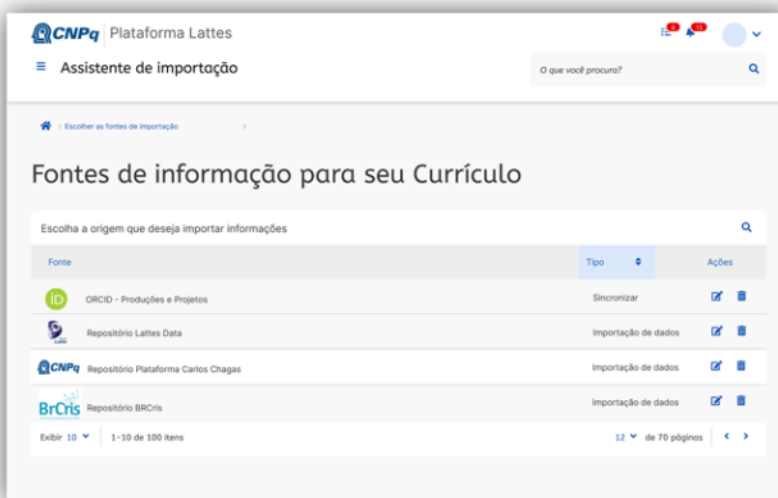


Fonte: FCCN

Os elementos utilizados no presente cenário tratam da escolha das fontes de informação participantes das rotinas de importação e sincronização. Também está demonstrado a visualização dos objetos resultantes da importação e sincronização e o seu respectivo status.

As figuras 7, 8, 9, 10 e 11 ainda demonstram a prototipação da complementação de conteúdo em registros de produção científica, a complementação de conteúdo em projeto lançado nas Plataformas PICC e na Sucupira⁸ e a complementação de conteúdo em projeto lançado no Lattes Data⁹.

FIGURA 7 - PROTÓTIPO DO ASSISTENTE DE IMPORTAÇÃO (MÓDULO INDEPENDENTE NA PLATAFORMA LATTES)



Fonte: Proposta do serviço de importação feito por José Francisco Salm Jr. (2022)

8. A Plataforma Sucupira é um sistema de coleta de informações, análises e avaliações a serem utilizadas como base padronizadora do Sistema Nacional de Pós-Graduação brasileira. A Plataforma disponibiliza informações, processos e procedimentos que a CAPES realiza no SNPG para toda a comunidade acadêmica.

9. O Lattes Data é o repositório de dados de pesquisa, que foi criado em parceria entre Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

FIGURA 8 - RESULTADOS E PENDÊNCIAS NO MÓDULO DE IMPORTAÇÃO

The screenshot displays the 'Assistente de importação' (Import Assistant) interface on the Lattes platform. At the top, there is a search bar with the placeholder text 'O que você procura?'. Below this, the main heading is 'Resultados e pendências' (Results and pending items). A sub-heading reads 'Acompanhe o resultado das produções importadas.' (Follow the result of imported productions.) To the right of this text is a button labeled 'Pendências de Todos' (Pending items of all).

A legend defines the status icons:

- ✔ *importado* - Cargo foi realizada com sucesso.
- ⚠ *incompleto* - É necessário complementar conteúdo
- ✘ *falha* - Não foi possível importar.

The main section, titled 'Minhas pendências' (My pending items), contains a table with the following columns: Status, Pendência, Permissões, Tipo de produção, No. do PID, Data, and Ação. The table lists five items, each with a status icon, a title, permission icons, production type, PID number, date, and a logo (BrCris, ID, or PTCRIS).

Status	Pendência	Permissões	Tipo de produção	No. do PID	Data	Ação
✔	Sentiment analysis using support...	🔍 🗑️ 🔄	Produção bibliográfica	12.1.000000077-4	16/02/2021	BrCris
✔	Use of support vector machines in ...	🔍 🗑️	Produção bibliográfica	-	15/02/2021	ID
⚠	An experiment in integrating sentim...	🔍 🗑️	Produção bibliográfica	12.1.000001117-2	10/02/2020	Lattes
✘	De-confated semantic representat...	🔍	Produção bibliográfica	12.1.000008888-9	12/12/2020	ID
✘	Uncovering text mining: A survey of...	🔍 🗑️ 🔄	Produção bibliográfica	12.1.000000000-3	12/12/2020	PTCRIS

At the bottom of the table, there are pagination controls: 'Exibir 10 de 1-10 de 100 Itens' and '12 de 70 páginas'.

Fonte: Módulo de serviço de importação proposto na pesquisa pós-doutoral de José Francisco Salm Jr. (2022)

FIGURA 9 - PROTÓTIPO DE TELA PARA COMPLEMENTAÇÃO DE CONTEÚDO EM PRODUÇÃO CIENTÍFICA

Plataforma Lattes

Assistente de importação

Complementar informações sincronizadas

Categoria (obrigatório): Publicações

Tipo (obrigatório): Livro

Identificador DOI (Identificador de Objetos Digitais) da produção:

Importar

Referência

Sentiment analysis using support vector machines with diverse information sources

Sentiment analysis using support vector machines with diverse information sources

Detalhes:

Título: Sentiment analysis using support vector machines with diverse information sources

Volume: Edição: Número da página:

Complemento:

Entre as 5 produções mais relevantes (obrigatório):

Sim Não

Cancelar Importar

Fonte: Módulo de serviço de importação proposto na pesquisa pós-doutoral de José Francisco Salm Jr. (2022)

FIGURA 10 - PROTÓTIPO DE TELA PARA COMPLEMENTAÇÃO DE CONTEÚDO EM PROJETO LANÇADO NAS PLATAFORMAS PICC E NA SUCUPIRA

Completar informações importadas de Projetos (PICC ou CAPES)

Tipo de projeto (obrigatório):
Projeto de pesquisa

Referência (Nome do projeto)

Atualização da Plataforma Lattes com foco em Ontologias internacionais de CV
 Comunidades de padronização semântica de dados de bases científicas

Nome do Projeto:
Atualização da Plataforma Lattes com foco em Ontologias internacionais de CV

Descrição:
Atualmente, a Plataforma Lattes possui mecanismos heterogêneos de interoperabilidade e serviços que conectam diferentes fontes de informação de CBT. Com os avanços dos serviços e o surgimento de novas interfaces de aplicação (Science data APIs), faz-se necessário avaliar novas abordagens de barramentos de serviços, permitindo que os pesquisadores completem seu currículo de forma estruturada e apresentem suas contribuições mais importantes para a ciência em narrativas breves, e não apenas como listas de publicações. Ao avaliar a iniciativa do CENCI/ITAD, pode-se entender como a Plataforma Lattes poderá evoluir nos aspectos do uso de barramentos.

Situação: Em andamento Ano início: 2022 Ano fim:

É um projeto de cooperação instituição de pesquisa e empresa? Sim Não
O projeto possui potencial de inovação de produtos, processos ou serviços? Sim Não

Instituição de execução:

Órgão/Unidade:

Equipe (obrigatório):

Cancelar Importar

Fonte: Módulo de serviço de importação proposto na pesquisa pós-doutoral de José Francisco Salm Jr. (2022)

FIGURA 11 - PROTÓTIPO DE TELA PARA COMPLEMENTAÇÃO DE CONTEÚDO EM PROJETO LANÇADO NO LATTES DATA

Completar informações importadas de Projetos (PICC ou CAPES)

Tipo de projeto (obrigatório):
Projeto de pesquisa

Referência (Nome do projeto)

As Dimensões da Biodiversidade em Lepidoptera Neotropicais
 As Dimensões da Biodiversidade em Lepidoptera Neotropicais

Nome do Projeto:
As Dimensões da Biodiversidade em Lepidoptera Neotropicais

Descrição:
O projeto propõe três linhas de pesquisa com Lepidoptera Neotropical: 1) Taxonomia e filogenia; 2) Ecologia de comunidades e conservação; e 3) Estrutura filogenética e funcional de comunidades. O objetivo do projeto é que estas três linhas contemplem seus resultados para que padrões mais amplos de diversificação de borboletas neotropicais sejam demonstradas. Os resultados obtidos não são de interesse acadêmico, pois irão contribuir para uma melhor compreensão de padrões e processos de diversificação de borboletas neotropicais, mas também devem ser úteis para subsidiar ações de conservação de espécies e habitats ameaçados nos Neotrópicos, que incluem os sistemas.

Situação: Em andamento Ano início: 2022 Ano fim:

É um projeto de cooperação instituição de pesquisa e empresa? Sim Não
O projeto possui potencial de inovação de produtos, processos ou serviços? Sim Não

Instituição de execução:
UNICAMP

Órgão/Unidade:

Equipe (obrigatório):

Cancelar Importar

Dataverse
Data: 12022-06-24
Assunto: Medicina, Saúde e Ciências da Vida
Palavra-chave: Ecologia de comunidades, Borboletas, Evolução, Conservação, Estrutura filogenética, Taxonomia
Agência financiadora: CNPq
Número do processo: 303834/2015-3
Depositante: André Victor Lucchi Freitas
Data do depósito: 2022-06-24

Fonte: Módulo de serviço de importação proposto na pesquisa pós-doutoral de José Francisco Salm Jr. (2022)

RECOMENDAÇÕES

Cabe destacar que o modelo adotado pela FCT/FCCN para o importador e sincronizador atende a todas as características contemporâneas de interoperabilidade técnica e está aderente às boas práticas de reuso das fontes de informação no contexto de sistemas RIMS¹⁰ (*Research Information Management Systems*). Dessa forma, recomenda-se aplicar o *framework* CV-PTBR no modelo de assistente de importação e sincronização para a criação de um módulo complementar à Plataforma Lattes que seja capaz de incorporar a interoperabilidade técnica e os serviços de dados em fontes internacionais de C&T,I.

RECONHECIMENTO E FINANCIAMENTO

Esse trabalho é fruto do projeto de pós-doutorado, cuja bolsa de estudo foi concedida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, sendo que o termo de outorga está associado ao processo de N^o: 400286/2022-0 (Pós-Doutorado no Exterior – PDE). Os autores agradecem ao CNPq, à FCT/FCCN e à Universidade do Estado de Santa Catarina pelo apoio no desenvolvimento desse projeto.

REFERÊNCIA BIBLIOGRÁFICA

- ANDON, P. I.; REZNICHENKO, V. A.; CHISTYAKOVA, I. S.. Mapping of Description Logic to the Relational Data Model. *Cybernetics And Systems Analysis*, [S.L.], v. 53, n. 6, p. 963-977, nov. 2017. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10559-017-9998-3>.
- CHUJAI, Pasapitch; KERDPRASOP, Nittaya; KERDPRASOP, Kittisak. On Transforming the ER Model to *Ontology* Using Protégé OWL Tool. *International Journal Of Computer Theory And Engineering*, [S.L.], v. 6, n. 6, p. 484-489, dez. 2014. IACSIT Press. <http://dx.doi.org/10.7763/ijcte.2014.v6.914>.
- GONÇALVES, Adriano de Oliveira; JACYNTHO, Mark Douglas de Azevedo. Um método para publicação semântica Linked Data de bases de dados convencionais e um estudo de caso real de artigos acadêmicos. *Transinformação*, [S.L.], v. 32, n. 0, p. 0-0, 2020. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1678-9865202032e180051>.
- KRAFFT, Dean; CAPPADONA, Nicholas; CARUSO, Brian; CORSON-RIKERT, Jon; DEVARE, Medha; LOWE, Brian Vivo. (2010). *VIVO: Enabling National Networking of Scientists*. *Proc of Web Sci Conf*.
JSON-SCHEMA <https://json-schema.org/understanding-json-schema/UnderstandingJSONSchema.pdf>
- KHUZAYEM, Lama Al; MCBRIEN, Peter. OWLRel: Learning Rich Ontologies from Relational Databases. *Baltic Journal Of Modern Computing (Bjmc): ISSN 2255-8950 (Online), ISSN 2255-8942 (Print)*. Acta Universitatis Latviensis, p. 466-482. 30 abr. 2016.
- MAHRIA, Bilal Ben; CHAKER, Ilham; ZAHI, Azeddine. A novel approach for learning *Ontology* from relational database: from the construction to the evaluation. *Journal Of Big Data*, [S.L.], v. 8, n. 1, p. 2-22, 28 jan. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/s40537-021-00412-2>.
- MOGOTLANE, Kgotatso Desmond; FONOU-DOMBEU, Jean Vincent. AUTOMATIC CONVERSION OF RELATIONAL DATABASES INTO ONTOLOGIES: A COMPARATIVE ANALYSIS OF PROTÉGÉ PLUG-INS PERFORMANCES. *International Journal Of Web & Semantic Technology*. Pretoria, p. 21-40. out. 2016.
- SEQUEDA, Juan F.; TIRMIZI, Syed Hamid; CORCHO, Oscar; MIRANKER, Daniel P.. Survey of directly mapping SQL databases to the Semantic Web. *The Knowledge Engineering Review*, [S.L.], v. 26, n. 4, p. 445-486, dez. 2011. Cambridge University Press (CUP). <http://dx.doi.org/10.1017/s0269888911000208>

10. Os sistemas de informação de pesquisa (RIMSs) podem ser definidos como os tipos de sistemas de informação que permitem gerenciar e fornecer acesso às informações de conteúdo e identidade de autoria dos pesquisadores. Editores, bibliotecas, universidades, mecanismos de busca e agregadores de conteúdo criaram diferentes RIMSs, todos com diferentes modelos de dados, cobertura e qualidade.

SEGUNDO, José Eduardo Santarém; CONEGLIAN, Caio Saraiva; LUCAS, Elaine Rosangela de Oliveira. Conceitos e tecnologias da Web semântica no contexto da colaboração acadêmico-científica: um estudo da plataforma vivo. *Transinformação*, [S.L.], v. 29, n. 3, p. 297-309, dez. 2017. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/2318-08892017000300007>.

SMITH, Barry; CEUSTERS, Werner. Ontological realism: a methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, [S.L.], v. 5, n. 3-4, p. 139-188, 2010. IOS Press. <http://dx.doi.org/10.3233/ao-2010-0079>.

SURE, York; BLOEHDORN, Stephan; HAASE, Peter; HARTMANN, Jens; OBERLE, Daniel. The SWRC *Ontology* – Semantic Web for Research Communities. *Progress In Artificial Intelligence*, [S.L.], p. 218-231, 2005. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/11595014_22.



José Francisco Salm Junior

Universidade do Estado de Santa Catarina – UDESC

jose.salmjunior@udesc.br

<https://orcid.org/0000-0002-8492-1645>

Possui Doutorado em Engenharia do Conhecimento pela Universidade Federal de Santa Catarina. É professor permanente do Programa de Pós-Graduação em Gestão da Informação da Universidade do Estado de Santa Catarina e pertence ao Departamento de Administração Pública do Centro de Ciências da Administração e Socioeconômicas - ESAG. Criou a Comunidade que auxiliou o CNPq na padronização da Plataforma Lattes. Atuou como pesquisador e consultor no Programa Star Metrics do Escritório de Políticas Públicas em Ciência e Tecnologia da Casa Branca (OSTP - Executive Office of the President of the United States). É membro do Grupo Consultivo Técnico de Evidências e Inteligência para Ação em Saúde das Américas na OPAS em Washington, DC.



Paulo Henrique de Assis Santana

Servidor aposentado do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

santanapha@gmail.com

<https://orcid.org/0000-0002-9674-3401>

Engenheiro mecânico pela Escola de Engenharia de São Carlos – EESC/USP e Mestre em Ciências pelo Instituto Nacional de Pesquisas Espaciais - INPE. Foi gestor de TI no Ministério da Agricultura, no Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, no Ministério do Meio Ambiente – MMA e no Ministério do Desenvolvimento Social e Combate à Fome – MDS. Mais recentemente, foi Coordenador Geral de Gestão e Inovação no Ministério de Ciência, Tecnologia e Inovação – MCTI e Coordenador de Informação e Estudos Internacionais no CNPq. Hoje, aposentado, é consultor independente. Coordenou o desenvolvimento do Currículo Lattes e sua disseminação pelas universidades brasileiras e agências de fomento, o projeto da Rede ScienTI, que disseminou o padrão do currículo Lattes internacionalmente, o Acordo de Cooperação Internacional entre o CNPq e a Fundação para a Ciência e a Tecnologia de Portugal para evolução conjunta de suas plataformas curriculares e o Acordo de Cooperação Técnica entre CNPq e IBICT para integração entre o CV Lattes e o Portal OasisBR. Coordenou, até maio de 2022, as atividades em Ciência Aberta do CNPq e o desenvolvimento do Repositório Aberto de Dados de Pesquisa do CNPq – LattesData.



Geraldo Sorte

Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

gsorte@cnpq.br

<https://orcid.org/0000-0002-9281-8859>

Graduado em Processamento de Dados pela Universidade de Brasília (1986), possui 34 anos de experiência na área de TI, sendo 17 anos coordenando áreas de TI na Administração Pública Federal. Participou da idealização e implementação da Plataforma Lattes e Plataforma Integrada Carlos Chagas do CNPq. Coordenou projetos de implantação de modelos de gestão por serviços na área de TI no CNPq e Fiocruz. Coordenou o projeto de implantação da infraestrutura tecnológica e de serviços de TI do Centro Hospitalar da COVID-19 da Fiocruz, bem como do novo sistema de gestão acadêmica da Fundação. Atualmente é Coordenador Geral de Tecnologia da Informação e Membro do Comitê de Governança Digital do CNPq. Atua nas áreas de Ciência da Computação, Modelos de Gestão de TIC, Ciência de Dados, Data Warehouse e Análise de Redes.

Relato Técnico • Ciência de Dados no Controle Social em Educação¹

Data Science in Social Control in Education

Ciencia de Datos en Control Social en Educación

Gleddson Fryttys Menezes Leite e Hugo Leonardo Tosto Cuoco

<https://doi.org/10.36428/revistadacgu.v14i26.551>

Resumo: O presente relatório técnico descreve o processo de desenvolvimento utilizado para a implementação da ferramenta de *Business Intelligence* sobre os dados das despesas com educação básica, utilizando como base os dados do Sistema de Informações sobre Orçamentos Públicos em Educação – SIOPE. Apresentam-se, ainda, as oportunidades que surgiram diante da utilização de uma metodologia ágil e como isso pode trazer vantagens nos casos em que não há um grande domínio sobre o negócio, facilitando, inclusive, a comunicação entre as partes. Por fim, descreve a ferramenta no estágio atual e apresenta os subprodutos que foram construídos com os resultados do pré-processamento.

Palavras-chaves: SIOPE. Educação Básica. *Business Intelligence*. Análise Comparativa. Educação Básica. Auditoria Governamental.

Abstract: This technical report describes the development process used to implement the *Business Intelligence* tool on basic education expenditure data, using data from the Information System on Public Budgets in Education. It also presents the opportunities that arose from the use of an agile methodology and how it might bring advantages in cases where there is not a great deal of knowledge about the business, even facilitating communication between the parts of the deal. Finally, it describes the tool at its current stage and presents the by-products that were built with the pre-processing results.

Keywords: SIOPE. Basic education. *Business Intelligence*. Comparative Analysis. Basic education. Government Audit.

Resumen: El presente informe técnico describe el proceso de desarrollo para la implementación de la herramienta *Business Intelligence* sobre datos de gasto en educación básica, utilizando datos del Sistema de Información de Presupuestos Públicos en Educación. También presenta las oportunidades que surgieron del uso de una metodología ágil y cómo puede traer ventajas en casos donde no hay mucho control sobre el negocio, incluso facilitando la comunicación entre las partes. Finalmente, describe la herramienta en su etapa actual y presenta los subproductos que se construyeron con los resultados del preprocesamiento.

Palavras clave: SIOPE. Educación básica. *Business Intelligence*. Análisis comparativo. Educación básica. Auditoría Gubernamental.

1. Artigo submetido em 05/07/2022 e aceito em 02/12/2022.

1. INTRODUÇÃO

A Lei nº 11.494/2007, que Regulamenta o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação – Fundeb, previu a criação de Conselhos de Acompanhamento e Controle Social - CACS no âmbito dos entes subnacionais, institucionalizando a participação da sociedade civil em um sistema de controle composto por unidades gestoras dos 3 níveis de governo e por órgãos de controle.

Ao longo dos anos, fiscalizações da CGU nos municípios registram impropriedades e irregularidades recorrentes nas aplicações de recursos do Fundeb, evidenciando dificuldades dos atores envolvidos no sistema de controle em mitigar perdas, desvios ou fraudes. Ao mesmo tempo, as fiscalizações também evidenciaram fragilidades relevantes no funcionamento dos CACS, os quais ocupam, em razão de sua proximidade com a comunidade escolar, papel central no sistema de controle, pressionando a gestão local e alertando eventuais irregularidades aos órgãos de controle.

O Relatório de Auditoria CGU nº 201900353 (BRASIL, 2019), sobre o sistema de controle do Fundeb, evidenciou que os mecanismos federais de monitoramento e avaliação das aplicações dos respectivos recursos pelos entes subnacionais não geram as informações necessárias para o funcionamento dos conselhos, sendo essa a principal causa das fragilidades detectadas nos CACS pelas fiscalizações nos municípios. Não se verifica a sinergia esperada entre os atores do sistema, comprometendo o resultado do controle.

A União, para cumprimento de suas atribuições constitucionais para a educação, dispõe de bases de dados capazes de atender também as necessidades

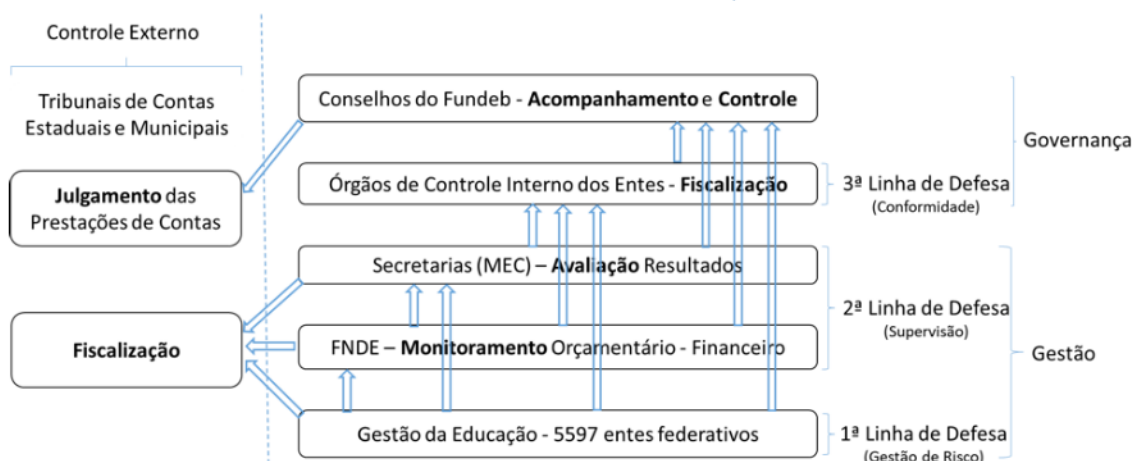
de informação dos CACS. A disponibilização integral desses dados, como de fato acontece nos órgãos federais, não significa qualquer apoio ao controle social, uma vez que não é razoável esperar dos conselheiros que cruzem bases e produzam suas próprias informações. Tão somente disponibilizar as informações produzidas para a gestão também não atende às necessidades dos conselhos. Trata-se, portanto, de a União assumir os CACS como clientes e produzir informações com características específicas, a fim de fazer funcionar o sistema de controle do Fundeb.

O relatório supramencionado concluiu que os controles instituídos pela Lei nº 11.494/2007 não produzem os resultados esperados, uma vez que os CACS não dispõem das informações gerenciais necessárias para sua atuação.

A auditoria parte de um conjunto de 262 fiscalizações da CGU sobre aplicações de recursos do Fundeb no período de 2013 a 2018, que detectaram, de forma recorrente, irregularidades em licitações, em contratos e com despesas inelegíveis, por exemplo. Ao mesmo tempo, 48% dessas fiscalizações relataram deficiências na atuação dos CACS, incluindo ausência de capacitação para os conselheiros, sonegação de informações por parte da gestão local, inexistência de suporte por parte da prefeitura, entre outros.

A atuação incipiente dos CACS impacta diretamente na capacidade das gestões locais e dos órgãos de controle em mitigar perdas, desvios e fraudes nas aplicações dos recursos do Fundeb. Conforme explicitado na Figura 1, o sistema de controle desenhado para o Fundeb depende de atividades de controle específicas (monitoramento, avaliação, fiscalização, acompanhamento e prestação de contas) de diversos atores, com centralidade para a atuação dos conselhos:

FIGURA 1: FLUXO DE INFORMAÇÕES DAS APLICAÇÕES DE RECURSOS DO FUNDEB CONFORME RELATÓRIO DE AUDITORIA Nº 201900353 (BRASIL, 2019)



Os CACS devem se valer da proximidade com a comunidade escolar para acompanhar o atendimento ao interesse público na educação. A ideia é que as percepções dos conselhos acerca dos serviços prestados pelas redes de ensino alertem a própria gestão e os órgãos de controle, potencializando os resultados da atuação estatal.

A atuação dos conselhos, por outro lado, depende da disponibilidade, em transparência ativa, de informações gerenciais em linguagem cidadã. Quanto melhores forem as informações disponíveis aos conselhos, mais capazes serão os conselheiros de identificar e acompanhar as questões mais relevantes e de maior risco, agregando qualidade aos alertas emitidos.

De posse de informações gerenciais produzidas pela administração, os conselhos devem usar suas prerrogativas (representar junto ao legislativo e aos órgãos de controle, requisitar documentos, convocar o secretário de educação para esclarecimentos, realizar visitas e inspeções *in loco*) para confirmar eventuais inconsistências e encaminhá-las para providências. Não é razoável esperar dos conselheiros que fiscalizem processos na busca de inconsistências, como fazem os órgãos de controle.

Considerando que a composição do CACS não pressupõe conhecimentos técnico-administrativos dos conselheiros, as informações necessárias à sua atuação devem ser disponibilizadas. Isso significa que a gestão pública deveria trabalhar os dados disponíveis com o objetivo específico de informar a sociedade civil, que atua institucionalmente por meio dos conselhos.

Por decorrência da competência da União de ordenar a política nacional de educação, os órgãos federais de educação dispõem de bases federais que consolidam dados de todos os entes subnacionais. O conjunto de dados coletados, sobre aprendizagem, atendimentos, infraestrutura das redes, dispêndios, entre outros, devem ser tratados e transformados em informações úteis à atuação federal junto aos entes, de forma a assegurar a universalização, a qualidade e a equidade do ensino obrigatório.

As mesmas bases de dados federais podem ser tratadas para gerar informações acessíveis e de interesse da sociedade civil, de forma a reduzir a assimetria de informação existente entre as gestões e os conselhos. Além do dever legal de promover a transparência ativa das informações públicas, comum aos 3 níveis de governo, a União tem a oportunidade de viabilizar ganho de efetividade na atuação dos CACS em todo o país.

Ainda que as informações relativas aos gastos com a educação básica tenham transparência ativa, entende-se que os dados disponibilizados não possuem significados práticos que possam ser utilizados para reforçar o processo de acompanhamento. Tais informações, de maneira geral, resumem-se a relatórios demonstrativos da execução da despesa com filtros bastante detalhados e específicos, tornando dispendiosa e até mesmo impeditiva a comparabilidade das informações relativas às despesas.

A comparabilidade pode aumentar a utilidade da informação prestada (SARQUIS, 2021), permitindo que conselheiros identifiquem e compreendam simi-

laridades e diferenças entre entes ou daquele mesmo ente ao longo do tempo, dando, assim, um maior significado aos dados quantitativos.

A construção dessas referências é factível para as diversas despesas relacionadas à educação, por meio da construção de relação entre os dados do Sistema de Informações sobre Orçamentos Públicos em Educação – SIOPE (FNDE, 2022), do Censo Escolar (INEP, 2022), do Índice de Desenvolvimento da Educação Básica – IDEB, do Indicador de Nível Socioeconômico das Escolas de Educação Básica – INSE, da Taxa de Rendimento Escolar (Aprovação, Reprovação e Abandono), do Índice de Desenvolvimento Humano – IDH e Dados Demográficos (IBGE).

2. O SISTEMA DE INFORMAÇÕES SOBRE ORÇAMENTOS PÚBLICOS EM EDUCAÇÃO – SIOPE

O Siope, instituído pela Portaria Ministerial (MEC) nº. 06, de 20 de junho de 2006, é um sistema informatizado para coleta, processamento, disseminação e acesso público às informações referentes aos orçamentos públicos de educação da União, dos Estados, do Distrito Federal e dos Municípios, no intuito de acompanhar os investimentos vinculados ao ensino. A grande vantagem de se utilizar o Siope é que o sistema é alimentado com todas as receitas e despesas dos orçamentos em educação dos entes independente da fonte de recurso e da natureza de receita(despesa?) orçamentária.

Na busca de informações comparáveis e de forma a produzir uma linguagem cidadã, utilizou-se 3 estratégias importantes no processo de ETL:

- 1 - Redução das despesas totais pagas em termos de gasto por aluno, utilizando-se, para tanto, a quantidade de matrículas do Censo Escolar. As similaridades e as diferenças nos gastos das redes de ensino com materiais didáticos ou gêneros alimentícios, por exemplo, tornam-se perceptíveis.
- 2 – Desmembramento do campo “grupo de despesa”, composto por uma aglutinação entre a classificação funcional-programática e por fonte/destinação de recursos, sendo possível identificar a despesa, por exemplo, por nível, modalidade e etapa de ensino, ou pelo programa ao qual está vinculado o gasto;
- 3 – Substituição das contas da natureza da despesa pelo campo “Tipo de Gasto”, que re-

presenta um agrupamento das contas de natureza semelhante em 17 tipos: Alimentação; Auxílios (Bolsas); Benefícios Previdenciários; Escolas Conveniadas; Material de Consumo; Material Didático; Mobiliário; Obras e Instalações; Obrigações Patronais; Remuneração; Sentenças Judiciais; Serviços; Terceirização; Transferências Entidades Privadas; Transferências Voluntárias; Transporte; Vagas Privadas 1ª Infância.

3. A FERRAMENTA DE BUSINESS INTELLIGENCE SOBRE OS DADOS DAS DESPESAS COM EDUCAÇÃO BÁSICA

Neste sentido, foi idealizado no âmbito da CGU uma ferramenta de *Business Intelligence*, tendo como base as informações do Siope e do Censo Escolar, com o objetivo de simplificar o processo de análise das despesas em educação. Em linhas gerais, a ferramenta objetivou proporcionar o seguinte:

- 1 - Visualização sistematizada dos gastos declarados no Siope pelos entes da federação (estados e municípios);
- 2 - Comparabilidade por meio dos gastos por aluno entre 2 entes diferentes ou entre 2 exercícios para um mesmo ente (valores sem atualização monetária), de forma que os gastos destoantes sejam acompanhados;
- 3 – Comparabilidade entre diversos entes da federação, conforme a escolha de categorias ou de intervalos de valores provenientes dos variados filtros disponíveis, incluindo medidas de estatística descritiva e visualização gráfica da distribuição e valores discrepantes (outliers) dos dados;
- 4 – Aplicação de técnica de agrupamento nos dados de forma a identificar características ou propriedades relevantes nos gastos em educação que não são mapeados apenas com os índices ou indicadores existentes (FACELI, 2021).

Como estratégia, uma vez que a documentação referente à base de dados era escassa e foi necessário constante envolvimento com a área finalística para compreensão dos dados, foi aplicado uma metodologia de desenvolvimento ágil (Carter, 2019). Desta

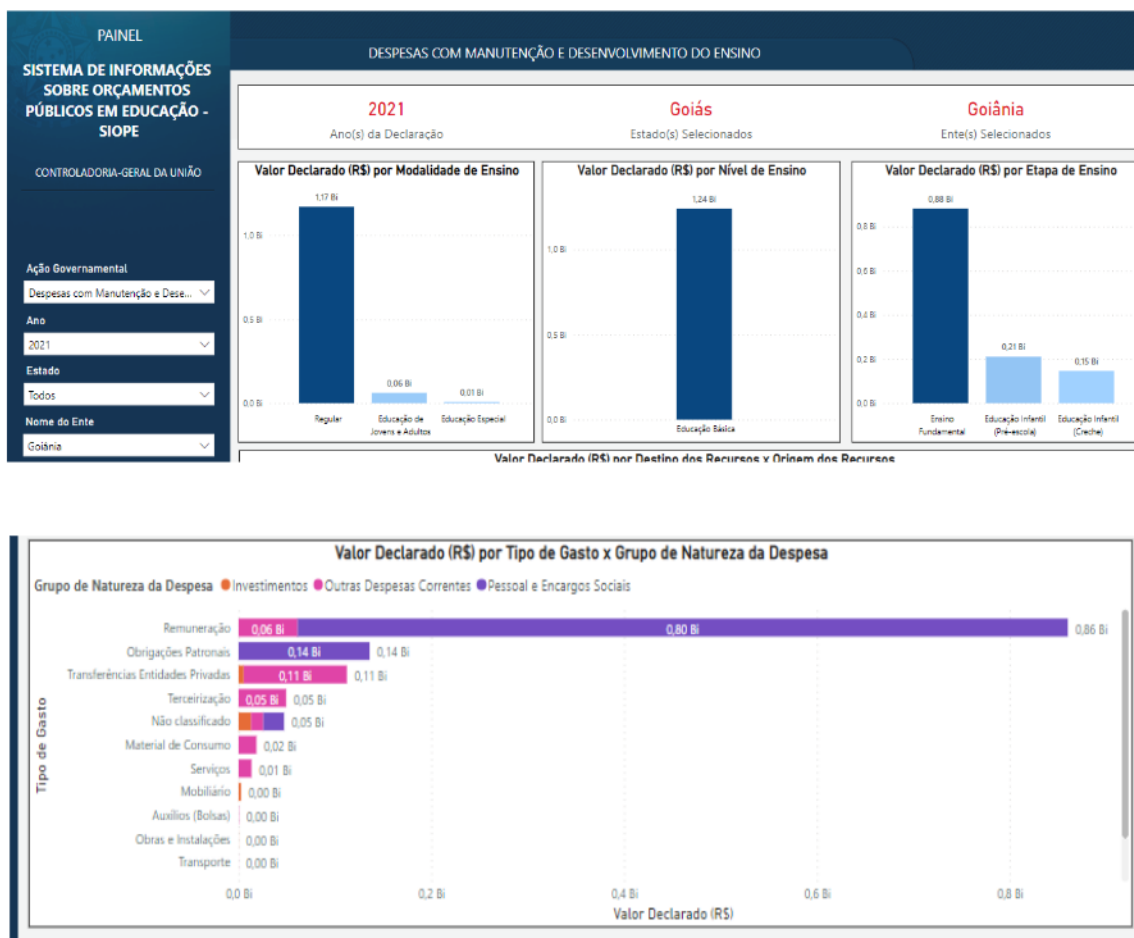
forma, optou-se por, desde o início do projeto, proporcionar uma visualização do *dataset* em elaboração. A cada iteração os dados eram refinados e validados.

Na ferramenta, nomeada como “Painel Siope”, foi incluído o *dataset* com os dados do Siope relativo aos anos de 2017 a 2021, bem como os índices, indicadores, dados demográficos e do Censo Escolar.

Na seção “Relatório” do Painel Siope, resultado das primeiras iterações, já foi possível observar de forma consolidada ou analítica, por ente da federação e por exercício, os valores totais das despesas pagas relacionadas à educação. Neste sentido, ressalta-se que a utilização do Siope proporcionou uma visão

não apenas centrada no Fundeb, objetivo de acompanhamento dos CACS, mas também de diversas outras fontes de recurso. Além disso, pode-se visualizar gráficos adicionais, que agrupam os valores de despesas pelos níveis, etapas e modalidades de ensino e outros atributos para reconhecimento do gasto. Um importante marco para o “Relatório” era a compreensão dos dados e o estabelecimento de definições que permeariam os demais produtos do objetivo. A exemplo, a definição dos 17 tipos de gasto só foi possível após o envolvimento de diversos especialistas em finanças públicas que utilizaram, como norteador, os dados já tratados e visualizados no painel.

FIGURA 2: PAINEL SIOPE – SEÇÃO “RELATÓRIO” –RECURSOS APLICADOS COM MANUTENÇÃO E DESENVOLVIMENTO DO ENSINO PARA O MUNICÍPIO DE GOIÂNIA/GO NO EXERCÍCIO DE 2021



Na Figura 2, observa-se as despesas com manutenção e desenvolvimento do ensino para o município de Goiânia/GO no ano de 2021. Já é possível verificar o montante aplicado por modalidade de ensino, nível de ensino e etapa de ensino, além dos valores aplicados nos tipos de gasto específico. Ainda que não seja possível obter insights com a seção “Relatório”, a visualização já apresenta avanços em relação à consulta disponibilizada pelo FNDE mediante transparência ativa, a qual não permite manipular os dados disponíveis, apresentando-os de maneira fragmentada, com recortes específicos, a exemplo dos indicadores como “Percentual de aplicação das receitas de impostos e transferências vinculadas à educação em MDE (mí-

nimo de 25% para estados, DF e municípios)” ou “Percentual das despesas com pessoal e encargos sociais da área educacional em relação à despesa total com MDE”.

Já a seção “Comparativo”, apresenta uma visualização sumarizada dos gastos anuais médios (por aluno) de cada Tipo de Gasto, de forma simples, fácil e intuitiva, entre dois entes federativos distintos (sejam estados ou municípios), ou de exercícios distintos de um mesmo ente federativo, em valores nominais. Nesta seção, já é possível realizar a comparação em termos de valores totais, médios e os principais indicadores e índices da educação básica (Figura 3 e 4).

FIGURA 3: PAINEL SIOPE – SEÇÃO “COMPARATIVO” – COMPARABILIDADE ENTRE O MUNICÍPIO DE GOIÂNIA/GO E CUMARU – PE, NO EXERCÍCIO DE 2019

Ensino Fundamental									
Goiânia - GO 2019					Cumarú - PE 2019				
Indicadores Socioeconômicos					Indicadores Socioeconômicos				
População (IBGE 2020)	IDH	IDH Educação	IDH Longevidade	IDH Renda	População (IBGE 2020)	IDH	IDH Educação	IDH Longevidade	IDH Renda
1.536.097	0,80	0,74	0,82	0,84	10.192	0,57	0,47	0,55	0,72
Indicadores Educacionais					Indicadores Educacionais				
Escolas Públicas (Qtd. Matrículas)	Escolas Conveniadas (Qtd. Matrículas)	IDEB (EF Anos Iniciais)	IDEB (EF Anos Finais)	IDEB (Ensino Médio)	Escolas Públicas (Qtd. Matrículas)	Escolas Conveniadas (Qtd. Matrículas)	IDEB (EF Anos Iniciais)	IDEB (EF Anos Finais)	IDEB (Ensino Médio)
72.350	564	5,90	5,00		2.087	0	4,50	3,60	
Taxas de Rendimento (%) - Escolas Municipais					Taxas de Rendimento (%) - Escolas Municipais				
Aprovação			Reprovação			Abandono			
AI	AF	EM	AI	AF	EM	AI	AF	EM	
99,1	98,0		0,5	0,7		0,4	1,3		
Indicador de Nível Socioeconômico - Rede Municipal					Indicador de Nível Socioeconômico - Rede Municipal				
Ano Referência	Qtd. Alunos	Média INSE	Nível_1_3 (%)	Nível_4_8 (%)	Ano Referência	Qtd. Alunos	Média INSE	Nível_1_3 (%)	Nível_4_8 (%)
2019	10.276	5,07	22,88	77,11	2019	284	4,28	64,44	35,56
Taxas de Rendimento (%) - Escolas Estaduais					Taxas de Rendimento (%) - Escolas Estaduais				
Aprovação			Reprovação			Abandono			
AI	AF	EM	AI	AF	EM	AI	AF	EM	
99,4	97,1	93,5	0,5	2,3	5,3	0,1	0,6	1,2	
Indicador de Nível Socioeconômico - Rede Estadual					Indicador de Nível Socioeconômico - Rede Estadual				
Ano Referência	Qtd. Alunos	Média INSE	Nível_1_3 (%)	Nível_4_8 (%)	Ano Referência	Qtd. Alunos	Média INSE	Nível_1_3 (%)	Nível_4_8 (%)
2019	15.496	5,11	22,30	77,70	2019	120	3,93	75,99	24,01

FIGURA 4: PAINEL SIOPE – SEÇÃO “COMPARATIVO” – GASTO TOTAL E MÉDIO PARA TIPOS DE GASTO ENTRE O MUNICÍPIO DE GOIÂNIA/GO E CUMARU – PE, NO EXERCÍCIO DE 2019

Tipos de Gastos			Tipos de Gastos		
Tipos de Gastos	Gasto Médio por Aluno (R\$)	Total (R\$)	Tipos de Gastos	Gasto Médio por Aluno (R\$)	Total (R\$)
Alimentação	136,43	9.079.359,0	Alimentação	6.754,81	12.259.985,0
Auxílios (Bolsas)	0,00	0,0	Auxílios (Bolsas)	0,00	0,0
Benefícios Previdenciários	0,00	0,0	Benefícios Previdenciários	6.007,76	10.904.085,0
Escolas Conveniadas	0,00	0,0	Escolas Conveniadas	0,00	0,0
Material de Consumo	92,83	6.178.255,4	Material de Consumo	11.092,57	20.133.009,0
Material Didático	0,00	0,0	Material Didático	0,00	0,0
Mobiliário	0,00	0,0	Mobiliário	0,00	0,0
Obras e Instalações	2,55	169.921,7	Obras e Instalações	29.243,27	53.076.533,0
Obrigações Patronais	1.052,84	70.068.306,1	Obrigações Patronais	55.405,19	100.560.411,0
Remuneração	8.526,15	567.432.089,2	Remuneração	369.677,39	670.964.458,0
Sentenças Judiciais	0,00	0,0	Sentenças Judiciais	0,00	0,0
Serviços	245,75	16.355.046,9	Serviços	61.899,14	112.346.937,0
Terceirização	461,27	30.698.669,7	Terceirização	0,00	0,0
Transferências Entidades Privadas	55.574,54	31.344.037,8	Transferências Entidades Privadas	0,00	0,0
Transferências Voluntárias	0,00	0,0	Transferências Voluntárias	0,00	0,0
Transporte	79,53	5.293.059,3	Transporte	23.997,27	43.555.052,0
Vagas Privadas 1ª Infância	0,00	0,0	Vagas Privadas 1ª Infância	0,00	0,0

Nos exemplos das Figuras 3 e 4, pode-se observar que, apesar da diferença populacional e estudantil, é possível estabelecer um comparativo entre os dois municípios. Portanto, nesta ferramenta já é possível obter *insights* sobre os dados que auxiliam tanto conselheiros quanto auditores no acompanhamento e monitoramento dos gastos. Por exemplo, pode-se verificar se as distorções entre os gastos médios têm alguma implicação quanto às diferenças.

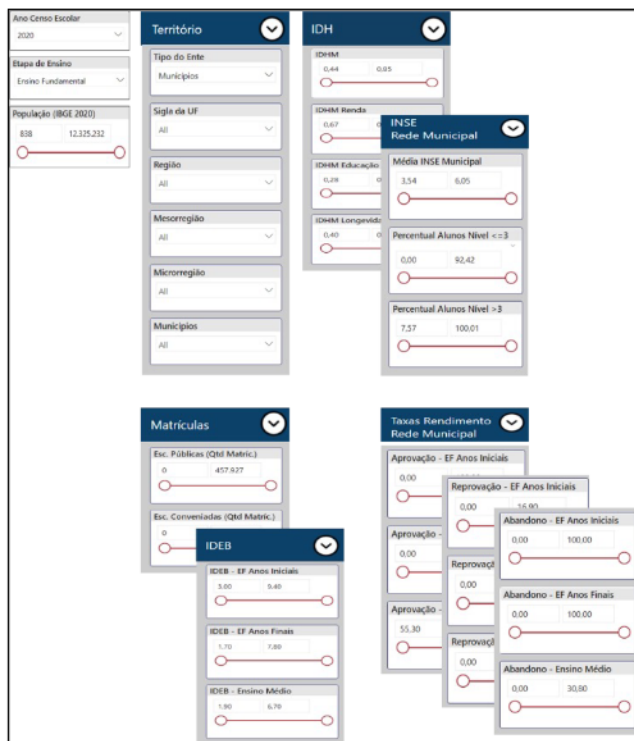
Em trabalhos já desenvolvidos, pode-se verificar, por exemplo, que um mesmo município teve seu gasto médio com transporte escolar elevado entre o exercício de 2019 e 2020, ano em que a pandemia ocasionada pelo Covid-19 obrigou a suspensão das aulas presenciais.

Em termos de linguagem cidadã, vislumbrou-se que os requisitos são atendidos, pois as informações comparáveis a respeito da execução orçamentária e financeira dos entes estão resumidos de forma aces-

sível. Não se exigindo, portanto, conhecimentos contábeis ou orçamentários de seus usuários. Todavia, a limitação em se trabalhar com apenas dois entes ou dois exercícios impede que sejam realizados trabalhos comparativos mais refinados, para os quais haja a necessidade de múltiplas delimitações de escopo. Desta forma, nas iterações de desenvolvimento do projeto, surgiu, junto aos usuários da ferramenta, a necessidade de implementação de painel com informações de filtro mais amplas, para tanto foi idealizado a seção nomeada de “Filtros”.

A seção “Filtros” considera um conjunto específico de entes federativos que atendam a diversos critérios de seleção (Figura 5) para, em seguida, apresentar uma série de visualizações consolidadas com informações sobre estes entes e seus gastos, bem como visualizações gráficas que explicitam quais são os intervalos de valores dos gastos médios por aluno, em cada tipo de gasto.

FIGURA 5: POSSIBILIDADES DE SELEÇÃO DA SEÇÃO “FILTROS” DO PAINEL SIOPE



Outro diferencial da seção “Filtros” é a apresentação de medidas de estatística descritiva sobre os dados filtrado e de ferramentas gráficas que permitem visualizar a distribuição e valores discrepantes dos dados (*boxplot*) (GALARNYK, 2018), favorecendo ainda mais a comparabilidade entre os dados. Conforme apresentado na Figura 6, é possível de imediato verificar aqueles entes que se destoam em termos de média de gasto por aluno no conjunto de uma seleção.

FIGURA 6: BOXPLOT DO TIPO DE GASTO “ALIMENTAÇÃO ESCOLAR” PARA UMA SELEÇÃO ESPECÍFICA



Outro fato que favorece a comparabilidade são as visualizações dos conjuntos de informações de entes contidos na seleção, que permite uma avaliação em larga escala, conforme demonstrado nas Figuras 7 e 8.

FIGURA 7: COMPARAÇÃO DOS TIPOS DE GASTO ENTRE OS ENTES – SEÇÃO “FILTROS” DO PAINEL SIOPE

Ano Censo Escolar	Etapa de Ensino	Sigla da UF	Nome do Ente	Alimentação Escolar	Amortização	Atividades Pedagógicas	Sentenças Judiciais	Transferências	Transporte Escolar
2019	Ensino Fundamental	GO	Abadia de Goiás	6,00	0,00	15,90	0,00	0,00	653,95
			Adelândia	778,11	0,00	0,00	0,00	0,00	887,80
			Americano do Brasil	59,69	0,00	0,82	0,00	0,00	22,73
			Amorinópolis	325,34	0,00	0,00	0,00	0,00	1.079,21
			Anápolis	184,65	0,00	5,76	0,00	0,00	25,14
			Anicuns	236,90	0,00	48,12	0,00	0,00	717,94
			Aparecida de Goiânia	282,93	0,00	0,00	0,00	0,00	15,71
			Araçu	186,34	0,00	11,16	0,00	0,00	187,01
			Aragoiânia	93,24	0,00	0,00	0,00	0,00	160,88
			Aurilândia	121,18	0,00	66,28	0,00	0,00	1.600,22
			Avelinópolis	99,30	0,00	0,00	0,00	0,00	236,62
			Barro Alto	235,27	0,00	189,70	0,00	0,00	10,08
			Bela Vista de Goiás	536,18	0,00	9,76	0,00	0,00	372,48
			Bonfinópolis	298,96	0,00	31,16	0,00	0,00	266,06
			Brazabrantes	153,63	0,00	2,01	0,00	0,00	836,63
			Buriti de Goiás	688,13	0,00	52,21	0,00	0,00	802,04
			Cachoeira de Goiás	52,08	0,00	81,08	0,00	0,00	588,20
			Caldazinha	177,51	0,00	25,15	0,00	0,00	391,07
			Campo Limpo de Goiás	273,27	0,00	10,59	0,00	0,00	285,14
			Carmo do Rio Verde	30,45	0,00	0,00	0,00	0,00	377,08

FIGURA 8: COMPARAÇÃO DOS DADOS SOCIODEMOGRÁFICOS, INDICADORES DE PERFORMANCE EDUCACIONAIS E QUANTIDADE DE MATRÍCULAS DE CADA ENTE – SEÇÃO “FILTROS” DO PAINEL SIOPE

Região	Sigla UF	Nome do Ente	População (IBGE 2020)	Ano Censo Escolar	IDHM	Ano Referência IDEB - EF Anos Iniciais	Nota IDEB - EF Anos Iniciais	Aprov. EF Anos Iniciais (Municipal)	Aprov. EF Anos Finais (Municipal)	Média INSE (Estadual)	Percentual Alunos - Nível 1 a 3 (Estadual)	Percentual Alunos - Nível 4 a 8 (Estadual)
CO	GO	Abadia de Goiás	8.958	2019	12,04	2019	5,90	97,50	90,50	4,68	43,19	56,81
CO	GO	Adelândia	2.516	2019	11,93	2019	5,20	92,00		5,08	11,29	88,72
CO	GO	Americano do Brasil	6.164	2019	11,90	2019	6,20	98,90		4,94	23,71	76,30
CO	GO	Amorinópolis	3.069	2019	11,58	2019	6,60	98,90		5,11	13,89	86,12
CO	GO	Anápolis	391.772	2019	12,53	2019	6,30	96,50	92,60	5,26	16,90	83,10
CO	GO	Anicuns	21.981	2019	12,14	2019	7,50	100,00	100,00	5,14	18,98	81,02
CO	GO	Aparecida de Goiânia	590.146	2019	12,21	2019	5,60	97,00	86,30	4,96	26,23	73,76
CO	GO	Araçu	3.486	2019	11,78	2019	6,80	100,00		4,76	38,19	61,81
CO	GO	Aragoiânia	10.496	2019	11,63	2019	5,70	97,30		4,85	27,83	72,17
CO	GO	Aurilândia	3.058	2019	11,90	2019	6,00	100,00		4,78	40,63	59,39
CO	GO	Avelinópolis	2.409	2019	11,22	2019	6,90	100,00	100,00	4,85	32,08	67,92
CO	GO	Barro Alto	11.408	2019	12,61	2019	7,00	99,30	96,20	4,82	31,37	68,63
CO	GO	Bela Vista de Goiás	30.492	2019	12,17	2019	6,00	97,10		4,85	28,74	71,26
CO	GO	Bonfinópolis	9.919	2019	11,61	2019	5,70	99,30		4,88	29,16	70,83
CO	GO	Brazabrantes	3.746	2019	11,92	2019	6,30	94,80		4,80	31,64	68,36
CO	GO	Buriti de Goiás	2.476	2019	11,68	2019	6,70	99,40	92,30	4,83	26,94	73,07

Na concepção da ferramenta, o objetivo principal era produzir um dataset com os dados dos gastos em educação relacionados com índices, indicadores e performance dos entes federais, de forma que a aplicação do agrupamento (cluster) forneça insights que indiquem características ou propriedades relevantes nos gastos em educação, que não são facilmente observáveis numa visão não sistematizada. Todavia, conforme já relatado, no decorrer do estudo do negócio e do domínio da aplicação e na fase de pré-processamento, surgiram vários subprodutos que não haviam sido previstos pelos gestores da ferramenta, resultando na ferramenta que já está em produção, o que demonstra a vantagem da metodologia ágil. Se por um lado houve alterações nos objetivos inicialmente pactuados, postergando os trabalhos de aplicação de machine learning, que num dado momento poderão oferecer outros insights sobre o gasto em educação, por outro houve um ganho no domínio negócio por parte dos idealizadores do projeto.

Após etapas de apresentação ao FNDE, órgão gestor do Siope, houve apoio da instituição quanto à liberação externa da ferramenta que auxiliará os CACS no processo de acompanhamento e controle dos gastos com a educação básica. Foram também incluídos recursos de exportação dos dados tratados e filtrados para que seja possível, também, a disseminação com vistas à elaboração de estudos no âmbito da educação pública. Por fim, houve diretrizes às superintendências regionais da CGU, emanada no 2º Encontro de Dirigentes 2022 da SFC, realizada no mês de julho de 2022, para que as equipes de auditoria

incluam análises do Painel Siope para delimitação de escopo dos trabalhos de auditoria na utilização das verbas para a educação.

4. CONCLUSÃO

No processo de concepção da ferramenta de *Business Intelligence*, percebeu-se que o modelo ágil pode trazer diversos benefícios em projeto de dados. De uma concepção inicial, os objetivos foram reformulados ao longo da execução do projeto assim que entregáveis eram apresentados aos usuários, o que, além de ajudar na compreensão do negócio por todos os envolvidos, ajudou a identificar oportunidades na concepção da ferramenta. Com isso, o *dataset* produzido na fase de pré-processamento dos dados se tornou um importante insumo na concepção do Painel Siope.

Quando do desenho da ferramenta, não se conseguiu verificar uma solução semelhante que auxiliasse os diversos atores envolvidos no controle da aplicação das verbas públicas em educação e que contribuísse na consecução de suas atividades, o que dá um tom de inovação à solução.

Por fim, apesar de a ferramenta estar se mostrando viável nas estratégias de acompanhamento da aplicação dos recursos públicos na educação básica, ainda não foram vislumbrados ajustes nas políticas públicas ocasionadas pela utilização da ferramenta. Contudo, a utilização ainda está incipiente por conta da não implantação para os entes externos à CGU, o que deve ocorrer brevemente.

5. REFERÊNCIAS

- BRASIL. Controladoria Geral da União. Relatório de Avaliação nº 201900353. 2019. Disponível em: <<https://auditoria.cgu.gov.br/>>. Acesso em: 18 jul. 2022.
- GALARNYK, M. Understanding Boxplots. 2018. Disponível em: <<https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>>. Acesso em 18 jul. 2022.
- FACELI, K. et al. Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina, 2ª Ed, Rio de Janeiro : LTC, 2021.
- SARQUIS, R. W., Afinal, o que se busca é comparabilidade ou uniformidade? Revista Mineira de Contabilidade, Belo Horizonte, v. 22, n. 2, 2021.
- CARTER, Eric et al., Agile Machine Learning, New York: Apress, 2019.



Gleddson Fryttys Menezes Leite

<https://orcid.org/0000-0002-2624-3874>

Controladoria-Geral da União - CGU

gleddson.leite@cgu.gov.br

Auditor Federal de Finanças e Controle da Controladoria-Geral da União, ocupando o cargo de chefe da divisão de ciência de dados da Diretoria de Auditoria de Políticas Sociais e de Segurança Pública da Secretaria Federal de Controle Interno. Possui graduação em Ciência da Computação pela Pontifícia Universidade Católica de Goiás (2003), com especialização em *Business Intelligence* pela Universidade Federal de Goiás - UFG e em Ciência de Dados pela Universidade de São Paulo - USP.



Hugo Leonardo Tosto Cuoco

<https://orcid.org/0000-0001-6962-4477>

Controladoria-Geral da União - CGU

hugo.cuoco@cgu.gov.br

Auditor Federal de Finanças e Controle da Controladoria-Geral da União, ocupando o cargo de chefe da divisão de Educação Básica da Coordenação-Geral de Auditoria das Áreas de Educação Básica, Cidadania e Turismo. Possui graduação em Filosofia pela Universidade de Brasília (2002).

Relato Técnico • Alice: Desafios, resultados e perspectivas da ferramenta de auditoria contínua de compras públicas governamentais com uso de inteligência artificial¹

Alice: Challenges, results and perspectives of the continuous auditing tool for government procurement using artificial intelligence

Alice: Desafíos, resultados y perspectivas de la herramienta de auditoría continua para compras gubernamentales utilizando inteligencia artificial

Andre Luiz Monteiro da Rocha, Matheus Scatolino de Rezende e Tiago Chaves Oliveira

<https://doi.org/10.36428/revistadacgu.v14i26.530>

Resumo: As compras governamentais são realizadas para implementar políticas e prestar serviços públicos de forma satisfatória à sociedade. As compras realizadas pelo Governo Federal Brasileiro correspondem a 1,5% do Produto Interno Bruto nacional em média (R\$ 119 bilhões em 2021). Além do grande volume de recursos, o processo de compras públicas é percebido como muito vulnerável a fraudes (OECD, 2018). Considerando a relevância, os valores e os riscos envolvidos a Controladoria-Geral da União criou a Analisadora de Licitações, Contratos e Editais (Alice) com o objetivo agregar valor à gestão pública com uma atuação preventiva e tempestiva em relação às compras públicas. O sistema coleta diariamente, de forma automática, informações sobre os processos em curso nas principais plataformas de compras públicas do Governo Federal, avalia um conjunto de riscos e emite alertas para direcionar a atenção dos auditores e dos gestores envolvidos a situações que fogem ao padrão referencial. É, portanto, uma ferramenta de Auditoria Contínua, com funcionamento de automação robótica de processos, que emite alertas sobre possíveis problemas em compras públicas. A Alice acumula resultados positivos cotidianamente. Já são mais de R\$ 9,7 bilhões em compras suspensas a partir de suas indicações, apenas pela CGU. No caminho até o alcance desses resultados a CGU vivenciou grandes desafios descritos nesse relato. O ponto crucial para o sucesso da ferramenta foi a integração de seus resultados ao processo de trabalho ordinário da Casa, com forte patrocínio da alta administração. A ferramenta agora evolui para adição de mais funcionalidades que usam técnicas de inteligência artificial e para a agregação de novas bases de dados.

Palavras-chave: Automação robótica de processos, Auditoria Contínua, Alice, Compras públicas e alertas.

Abstract: Government procurements are carried out to implement policies and provide public services in a satisfactory manner to society. Purchases made by the Brazilian Federal Government correspond to 1.5% of the national Gross Domestic Product on average (R\$119 billion in 2021). In addition to the large volume of resources, the public procurement process is perceived as very vulnerable to fraud (OECD, 2018). Considering the relevance, the total amount of expenses, and the risks involved, the Office of the Comptroller General created the Bidding, Contracts, and Tenders Analyzer (Analisador de Licitações, Contratos e Editais – Alice) with the objective of adding value to public management acting in a preventive and timely manner in relation to public procurement.

1. Artigo submetido em 17/07/2022 e aceito em 01/12/2022.

The system automatically collects daily information about the processes underway in the Federal Government's main public procurement information systems, evaluates a set of risks, and raises red flags to direct the attention of the auditors and managers about situations that deviate from the reference standard. It is, therefore, a Continuous Auditing tool, with robotic automation of processes, which raises red flags about possible problems in public procurements. Alice accumulates positive results daily. More than R\$ 9.7 billion in purchases have already been suspended based on its indications, only by the CGU. On the way to achieving these results, the CGU has experienced great challenges, as described in this report. The crucial point for the tool's success was the integration of its results into the ordinary work process of the CGU, with strong top management sponsorship. The tool is now evolving to add more functionalities that use artificial intelligence techniques, and to add new databases.

Keywords: robotic process automation, Continuous Auditing, Alice, Public procurement and red flags.

Resumen: Las compras gubernamentales se realizan para implementar políticas y prestar servicios públicos de forma satisfactoria para la sociedad. Las compras realizadas por el Gobierno Federal de Brasil corresponden al 1,5% del Producto Interno Bruto nacional en promedio (119 mil millones de reales en 2021). Además del gran volumen de recursos, el proceso de contratación pública se percibe como muy vulnerable al fraude (OCDE, 2018). Teniendo en cuenta la relevancia, los valores y los riesgos involucrados, la Contraloría General creó el Analizador de Licitaciones, Contratos y Concursos (Analisador de Licitações, Contratos e Editais - Alice) con el objetivo de agregar valor a la gestión pública actuando de manera preventiva y oportuna en relación con las compras públicas. El sistema recoge automáticamente información diaria sobre los procesos en curso en los principales sistemas de información de la contratación pública del Gobierno Federal, evalúa un conjunto de riesgos y emite alertas sobre problemas para dirigir la atención de los auditores y gestores sobre las situaciones que se desvían de la norma de referencia. Es, por tanto, una herramienta de Auditoría Continua, con automatización robótica de los procesos, que emite alertas sobre posibles problemas en las contrataciones públicas. Alice acumula diariamente resultados positivos. Más de R\$ 9,7 mil millones en compras ya fueron suspendidos con base en sus indicaciones, sólo por la CGU. En el camino para lograr estos resultados, la CGU ha experimentado grandes retos, como se describe en este informe. El punto crucial para el éxito de la herramienta fue la integración de sus resultados en el proceso de trabajo ordinario de la CGU, con un fuerte patrocinio de la alta dirección. La herramienta está evolucionando para añadir más funcionalidades que utilizan técnicas de inteligencia artificial, y para añadir nuevas bases de datos.

Palabras clave: Automatización robótica de procesos, Auditoría continua, Alice, contratación pública, Alertas

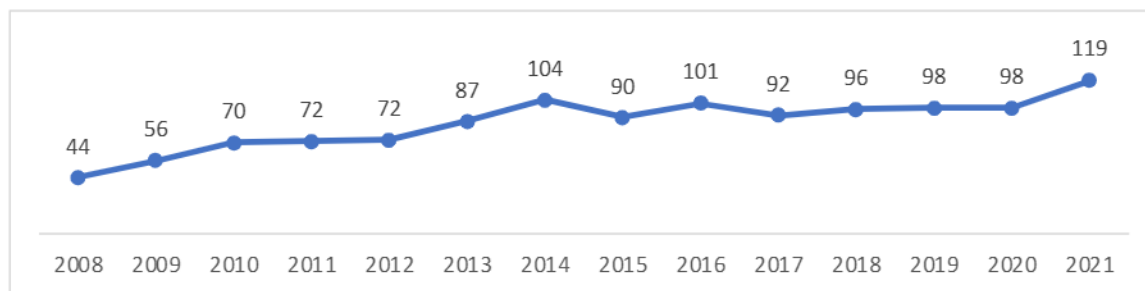
1. INTRODUÇÃO

A Controladoria-Geral da União (CGU) realiza o mandamento constitucional de avaliação da execução dos programas de governo e dos resultados da gestão dos recursos públicos no âmbito da Administração Pública Federal. A Secretaria Federal de Controle Interno (SFC) é a Unidade da CGU responsável pela atividade de Auditoria Interna Governamental (AIG) no Poder Executivo Federal (CGU, 2022). A atividade de auditoria governamental tem o objetivo de adicionar valor e melhorar as operações das organizações públicas. Entre outras, a AIG deve promover a prevenção, a detecção e a investigação de fraudes na utilização de recursos públicos federais (BRASIL, 2017). Entre essas competências, destaca-se a análise sobre as compras governamentais.

As compras governamentais são aquisições de bens e contratação de serviços realizadas pelos governos para implementar políticas e prestar serviços públicos de forma satisfatória. A crise do COVID-19 explicitou que as estratégias, práticas e sistemas de contratação pública afetam diretamente a qualidade de vida e o bem-estar dos cidadãos (OECD, 2021).

Nos países da Organização para Cooperação e Desenvolvimento Econômico (OCDE), observou-se que as despesas com contratos públicos em relação ao Produto Interno Bruto (PIB), aumentaram na última década, de 11,8% em 2008 para 12,6% em 2019. Com a pandemia de COVID-19, o valor alcançou 14,9% em 2020 (OECD, 2021). Estudo do Banco Mundial indica que o Brasil pode ter alcançado 20% de seu PIB em compras públicas em 2018 (Bosio & Djankov, 2020).

FIGURA 1 - COMPRAS DO GOVERNO FEDERAL (R\$ EM BILHÕES)



Fonte: Elaborado pelos autores com base em Extração do Tesouro Gerencial

No Governo Federal, o volume de compras é correspondente a 1,5% do PIB nacional² em média.

Apesar da elevação gradual dos recursos destinados a compras públicas, a CGU observou diminuição sem sua capacidade operacional no período de 2012 a 2022, período em que não houve novos concursos para ingresso de servidores. Nesse período observou-se 864 vacâncias, redução de 28% da força de trabalho³ (SFC/CGU, 2021).

Pela relevância do assunto, a OCDE indica a necessidade de controles internos e de auditorias regulares, realizadas frequentemente, para garantir conformidade legal e regulatória das compras e para o devido monitoramento do desempenho do sistema de compras públicas (OECD, 2018).

Nesse contexto, considerando a relevância das compras públicas, os riscos envolvidos e o crescente volume monetário aplicado, surgem no mundo as ferramentas de emissão de alertas em compras públicas. Os alertas indicam possíveis fraudes e desconformidades ao longo do processo de compras (Tátrai & Németh, 2018).

Paralelamente, amplia-se o uso de ferramentas como a automação robótica de processos (*Robotic Process Automation* – RPA). Trata-se de robôs treinados para executar tarefas manuais e repetitivas, reduzindo os custos, a carga de trabalho e os erros dos humanos. Assim, as pessoas podem ser alocadas em questões mais importantes, que exigem percepção e decisão humana (Fantina, Storozhuk, & Goyal, 2022).

Na mesma linha do RPA, surge a Auditoria Contínua que possibilita o alcance de resultados simultaneamente ou em um curto período após a ocorrência de eventos relevantes, de forma mais abrangente, robusta e célere (Chan, Chiu, & Vasarhelyi, 2018).

Nesse contexto, o Analisador de Licitações, Contratos e Editais (Alice) foi elaborado pela CGU em 2014 com o objetivo de possibilitar aos auditores uma atuação preventiva e tempestiva em relação às compras públicas. O sistema coleta diariamente, de forma automática, informações sobre os processos em curso nas principais plataformas de compras públicas do Governo Federal, avalia um conjunto de riscos e emite alertas para direcionar a atenção dos auditores e dos gestores envolvidos. É, portanto, uma ferramenta de Auditoria Contínua, com funcionamento de RPA que emite alertas sobre possíveis problemas em compras públicas.

A Alice acumula resultados positivos cotidianamente. Já são mais de R\$ 9,7 bilhões em compras suspensas a partir de suas indicações, apenas pela CGU. No entanto, para alcançar este resultado a CGU percorreu um longo caminho. Nesse sentido, esse relato objetiva descrever a ferramenta, seus resultados e o processo percorrido desde a sua criação até junho de 2022.

Na próxima seção é detalhado o funcionamento da Alice e descritos os seus resultados. Na sequência apresenta-se o percurso histórico percorrido até o alcance desses resultados. A última seção apresenta conclusões e encaminhamentos propostos.

2. Considerando extração do Tesouro Gerencial de 22/04/2022 com dados de 2008 a 2021, considerando apenas aplicações diretas do Governo (modalidades de aplicação 90 e 91) e os elementos de despesa previstos por Ribeiro & Inácio Júnior (2019) acrescidos dos seguintes: 30, 33, 35, 36, 37, 39, 40, 61 e 62. Dados do PIB Nacional extraídos de <http://www.ipeadata.gov.br/exibeserie.aspx?serid=38415>, acesso em 25/04/2022.

3. Força de trabalho disponível na CGU em: Jan/2012 – 2884 servidores, Dez/2020 - 2060 servidores.

2. A ALICE E SEUS RESULTADOS

A ferramenta possibilitou o alcance de resultados expressivos, no sentido do combate à corrupção, mesmo com a diminuição na quantidade de servidores na CGU, o que é um paradoxo em relação ao entendimento de que se deve aumentar a quantidade de auditores para garantir o combate à corrupção (Fortini & Motta, 2016).

A Alice é uma ferramenta de auditoria contínua, segundo definições da AICPA (2015) e de Chan, Chiu e Vasarhelyi (2018), por:

- produzir resultados de forma simultânea em relação ao processo de negócio (compras públicas) e de forma tempestiva após a realização dos atos relacionados com as compras públicas;
- Com o uso do sistema, a CGU é capaz de economizar recursos financeiros já que evita a necessidade de contratação de servidores para tarefas realizadas pela ferramenta;
- O sistema avalia todas as transações realizadas, não apenas uma amostra delas, o que reduz a possibilidade de que licitações irregulares não sejam detectadas; e
- As tarefas que exigem o julgamento humano são alertadas como exceções a um padrão e direcionadas a um auditor para análises. Dessa forma, implementa a auditoria por exceção, outra característica da auditoria contínua.

Seguindo os entendimentos do IIA (2009), a Alice possibilita maior efetividade da análise de dados para a detecção de fraudes porque compara dados e transações de diversas origens e avalia falhas de controle que existem entre diferentes sistemas, aplicando diferentes avaliações sobre a ocorrência de fraude de forma regular e tempestiva.

A Alice já foi objeto de outros estudos acadêmicos no passado. Costa & Bastos (2020) descreveram como se deu a implantação, a utilização e a evolução da Alice depois que foi doado pela CGU ao Tribunal de Contas da União (TCU) em 2016. Os autores ressaltaram que o TCU contabilizou R\$3,8 bilhões em benefícios em 2019, auferidos a partir de indicações da Alice.

Panis, et al. (2021), analisaram o processo de inovação seguido pela CGU para a construção da Alice sob a ótica do Modelo Integrado de Gestão Pública Inovadora.

Riedel & Ferreira (2021) vincularam a utilização da Alice com o fortalecimento do compliance das instituições às regras de contratações públicas, garantindo precisão e otimização de tempo na prevenção de irregularidades.

Oliveira, et al. (2021) analisaram a eficácia da Alice na detecção de fraudes na aplicação de recursos direcionados ao combate à pandemia de Covid-19. Os autores destacaram que a ferramenta foi útil como meio de detecção de vícios nos editais que pudessem ser utilizados como oportunidades para o cometimento de fraudes.

2.1. Funcionamento da Alice

O sistema Alice está implementado para funcionar em 4 grandes etapas:

1. **Coleta de dados sobre transações de compras:** Atualmente, os dados são coletados do sistema Compras.gov.br; do sistema Licitações-e; do Diário Oficial da União (DOU), de onde se buscam dados sobre inexigibilidades e dispensas de licitações; e de bases de dados de despesas do Estado e dos Municípios de São Paulo. Além do Poder Executivo Federal, essas fontes contêm informações de compras realizadas por diversas instituições federais de outros Poderes da União e de outros entes federados. Todos esses dados são processados pela ferramenta diariamente;
2. **Análises computadorizadas sobre os conteúdos obtidos:** As análises incluem avaliações textuais sobre os editais e outros documentos das licitações e cruzamentos de dados para identificação de riscos e de indícios de irregularidades;
3. **Comunicação:** Etapa em que o sistema, diariamente, comunica por e-mail para a CGU e para o próprio órgão comprador a relação de todas as aquisições que foram analisadas, com os possíveis alertas correspondentes.
4. **Auditoria:** Em relação às compras consideradas mais críticas e arriscadas de acordo com critérios de materialidade, criticidade e relevância pré-definidos, a Alice abre, automaticamente, auditorias chamadas de Avaliações Preventivas de Contratações. Essas tarefas são realizadas para que auditores confirmem ou descartem os alertas informados.

O trabalho garante tempestividade à atuação da CGU porque ocorre em apenas oito dias corridos a partir da data de publicação do edital ou do ato relacionado com a compra. Com essa agilidade de análise, possíveis recomendações podem ser implementadas pelos gestores antes da realização da sessão pública do pregão, por exemplo. O objetivo principal do trabalho é agregar valor à gestão pela in-

dicação de riscos sobre a futura contratação. Como resultado a CGU garante a conformidade dos certames e contabiliza benefícios financeiros para a União, em geral, pela adequação das estimativas dos custos ou da volumetria das compras e pelo cancelamento de pregões com indícios de fraudes ou desnecessários (que possuem os requisitos de viabilidade técnica e econômica necessários).

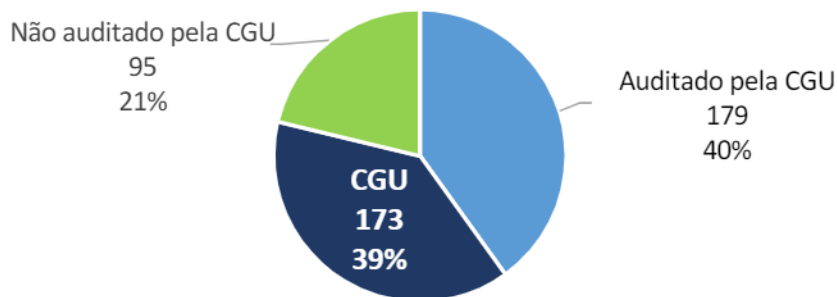
FIGURA 2 - PROCESSO DA ALICE NA CGU



Fonte: Elaborado pelos autores

O sistema envia informes diários com alertas para 447 usuários sendo 39% de auditores da CGU, 40% de unidades de auditoria interna de órgãos que estão dentro do escopo de avaliação da CGU (Poder Executivo Federal) e 21% de unidades de auditoria interna de órgãos fora da abrangência das competências da CGU (outros Poderes ou em unidades do Poder Executivo que são acompanhadas por uma Secretaria de Controle Interno – Ciset, a exemplo a Presidência da República, o Ministério da Defesa e o Ministério das Relações Exteriores.

FIGURA 3 - QUANTIDADE DE USUÁRIOS POR RELAÇÃO COM A CGU

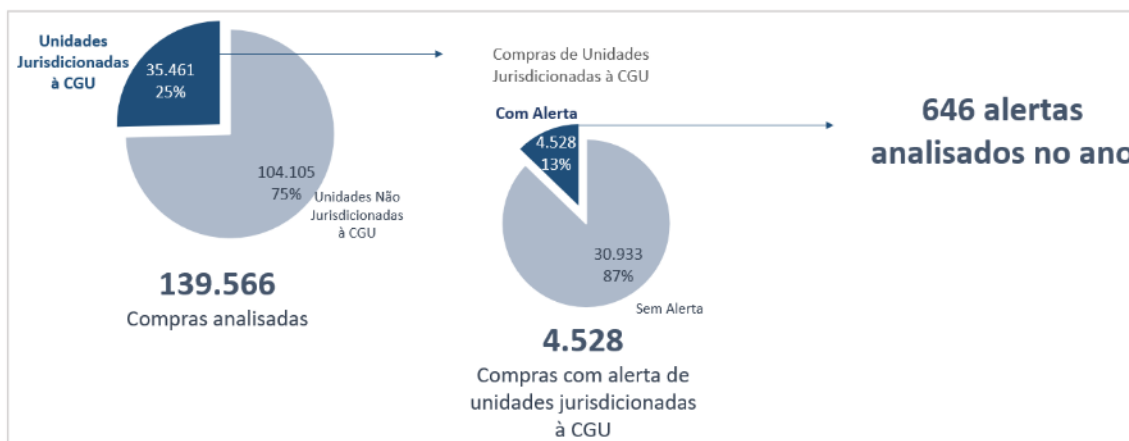


Fonte: Elaborado pelos autores

Atualmente, o sistema está preparado para realizar 40 diferentes avaliações sobre os riscos que envolvem os processos de compras públicas.

Considerando o ano de 2021 e os dados obtidos do Compras.gov.br, do Licitações-e e do Diário Oficial da União (dispensas e inexigibilidades de licitação), a Alice avaliou 139.566 diferentes processos de compras. Desse total, 35.461 (25%) foram de unidades jurisdicionadas da CGU. Dentre essas, 4.528 (13%) alertas foram gerados, dos quais, há registros no sistema e-Aud de análises de 646.

FIGURA 4 – ANÁLISES REALIZADAS PELA ALICE EM 2021

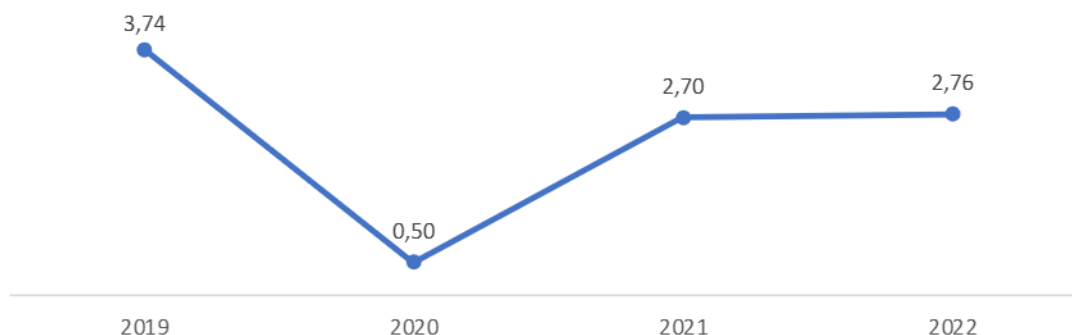


Fonte: Elaborado pelos autores

Além dos registros feitos no sistema e-Aud, há a possibilidade de que problemas pontuais indicados pela Alice tenham sido resolvidos pelos gestores a partir de contatos dos auditores ou das unidades de controle interno da própria instituição.

A partir dessa atuação, de 2019 a junho de 2022, a CGU registrou R\$ 9,7 Bilhões em licitações suspensas ou canceladas com indícios de fraude ou erro.

FIGURA 5 – LICITAÇÕES SUSPENSAS OU CANCELADAS A PARTIR DE INDICAÇÕES DA ALICE (VALORES EM BILHÕES R\$)



Fonte: Elaborado pelos autores com base em dados do sistema e-Aud (anos de 2020 a 2022) e de registros de registros históricos (ano de 2019).

Em termos de benefícios financeiros aprovados conforme a sistemática de contabilização de benefícios seguida pela CGU⁴, observa-se expressivo aumento a partir de 2020, com a maior institucionalização do tratamento dos alertas na CGU.

3. HISTÓRICO DE IMPLEMENTAÇÃO DA ALICE

3.1. Motivação

Antes da sua criação, a análise preventiva de editais somente ocorria se direcionada por denúncias ou por decisão ad hoc dos auditores. (Panis, et al., 2021). Não havia acompanhamento sistemático, em tempo real, de todas as licitações na CGU.

Usualmente, as auditorias de licitações eram executadas durante a execução ou após o término contratual. Essa forma de atuação pode acarretar dificuldade

de recuperação administrativa dos recursos públicos pagos indevidamente, necessidade de abertura de processos judiciais, de Processo Administrativo Disciplinar (PAD) e de processos Tomadas de Contas Especiais (TCE). Todos esses processos são demorados e caros, por consumirem a capacidade operacional dos auditores que poderiam estar concentrados na agregação de valor em outras situações.

Pode, também, gerar sensação de falta de controle em casos de licitações com falhas grosseiras, prejudicando a imagem da CGU perante a sociedade. Além disso, a atuação intempestiva tem baixa capacidade de mudar eventuais regras equivocadas definidas nos editais das contratações já realizadas, com consequências negativas e duradouras nas políticas públicas relacionadas.

FIGURA 6 – OPORTUNIDADE CONSTRUTIVA VERSUS INTERVENÇÃO DESTRUTIVA



Fonte: Adaptado de Vargas (2009, 55)

Nesse sentido, aplicando os conceitos de Vargas (2009), ilustrados na figura 5, no contexto deste relato, quanto maior o lapso temporal entre o processo licitatório e a realização da auditoria, maior será o custo para a correção de eventuais problemas e menor o potencial para a auditoria agregar valor à gestão.

3.2. A origem

Em 2012, a SFC criou um núcleo de auditoria de TI. Essa estrutura objetivava realizar auditorias com TI, ou seja, aplicando TI para melhorar a efetividade

dos trabalhos, e realizar auditorias de TI, avaliando, por exemplo, processos licitatórios de TI, segurança da informação, governança de dados e serviços públicos digitais. Esse núcleo, nos anos posteriores, originou a Coordenação-Geral de Auditoria de TI (CGATI) e Coordenação-Geral de Prospecção e Inovação (CGPRI), responsáveis pelas auditorias de TI e com TI respectivamente.

Em 2014, a Controladoria Regional da União no Estado do Mato Grosso (CGUMT) disponibilizou uma lista com 21 formas de restrição da competitividade de certames usualmente detectados em editais.

4. Conforme Portaria nº 1.976, de 20 de agosto de 2021 (Institui a sistemática de quantificação e registro dos benefícios decorrentes das ações executadas pela Controladoria-Geral da União).

Com a lista, o auditor, ao ler um edital, saberia identificar essas inconformidades. Assim, a Alice nasceu para substituir ou reduzir a necessidade de um auditor ler e analisar todo o edital em busca dessas inconformidades.

3.3. Desafios iniciais

De fevereiro e agosto de 2014, três grandes desafios foram enfrentados para viabilizar a ferramenta: dificuldade de obtenção dos dados, inexistência de um ambiente de desenvolvimento e a implementação do algoritmo em si.

Era necessário obter dados gerais (metadados) sobre as licitações e do texto dos editais no momento de sua publicação no Comprasnet. Entretanto, estava apenas disponível um banco de dados de licitações, atualizado mensalmente, e sem o texto dos editais, o *datawarehouse* do sistema SIASG (DWSIASG).

Nesse momento, verificou-se que os metadados e o texto dos editais estavam disponíveis publicamente e em tempo real no portal Comprasnet. Para construir uma massa de testes, vislumbrou-se a possibilidade de desenvolver um *software* “*scraper*” (raspador) para efetuar o *download* automático de alguns arquivos de editais. Ocorre que há no sistema um mecanismo de segurança para evitar os *downloads* automatizados, o captcha, que apresenta um problema de fácil resolução para humanos, mas de difícil resolução para máquinas.

Com esse impedimento, solicitou-se ao Ministério do Planejamento, Orçamento e Gestão (MPOG) a disponibilização de todos os arquivos de editais ou a liberação do acesso sem o captcha. O MPOG não se opôs, mas o pleito exigiria uma demanda ao Serviço Federal de Processamento de Dados (SERPRO), empresa responsável pelo desenvolvimento do Comprasnet. A empresa cobraria pela demanda e a colocaria em uma fila de implementação. Considerando o ineditismo e as incertezas relacionadas, decidiu-se que o investimento não seria oportuno.

A solução encontrada foi contratar serviços de resolução de captcha por humanos. Dessa forma, o “*scraper*” encaminhava as imagens do captcha ao serviço, que utilizava um humano para solucionar e devolver a resposta. Cabe destacar que o custo desse serviço era irrisório (cerca de US\$ 0,50 por 1000 captchas). Assim, considerando o baixo custo, o alto nível de incertezas e a complexidade do processo de contratação, o valor foi custeado por um auditor envolvido

no desenvolvimento da Alice. Pouco tempo depois foi desenvolvida um algoritmo para resolver o captcha automaticamente, sem a necessidade do serviço.

Em relação ao ambiente de desenvolvimento da solução, naquele momento, o uso de ferramentas de desenvolvimento de *software* era restrito a algumas áreas da CGU. Dessa forma, os auditores desenvolvedores da ferramenta usaram seus computadores pessoais até conseguirem as autorizações necessárias para usarem o ambiente da CGU.

Em relação à implementação do algoritmo, naquele momento, o processamento de linguagem natural (*Natural Language Processing* – NLP) era pouco difundido. As principais bibliotecas de NLP estavam disponíveis apenas na linguagem Python, que era considerada exótica na época. Após o estudo de técnicas de NLP, elaborou-se um algoritmo que encontrava palavras-chave a partir de expressões regulares e analisava o contexto da região do texto em que essa palavra-chave estava inserida, diminuindo os casos de falsos positivos.

3.4. Resultados preliminares

Com o uso do *scraper*, formou-se uma massa de testes com 140.888 licitações do período de setembro/2012 até agosto/2014, com valores totais estimados em cerca de R\$ 100 bilhões. A quantidade de arquivos analisados totalizou 442.557, considerando que cada certame poderia ter mais de um documento.

Considerando que os algoritmos implementados podem resultar em falsos positivos, isto é, apontar inconformidades em editais adequados, realizou-se prova de conceito com a aplicação de algoritmos para detectar 13 dos 21 tipos de achados de auditoria mapeados. Dessa forma, mediu-se o sucesso inicial da Alice pelo cálculo da taxa de acertos para cada tipo de achado implementado a partir de uma pesquisa com intervalo de confiança de 95% e margem de erro de 2%. Após a pesquisa, os algoritmos que apresentaram pior desempenho foram deixados de lado para serem melhorados no futuro.

3.5. A evolução

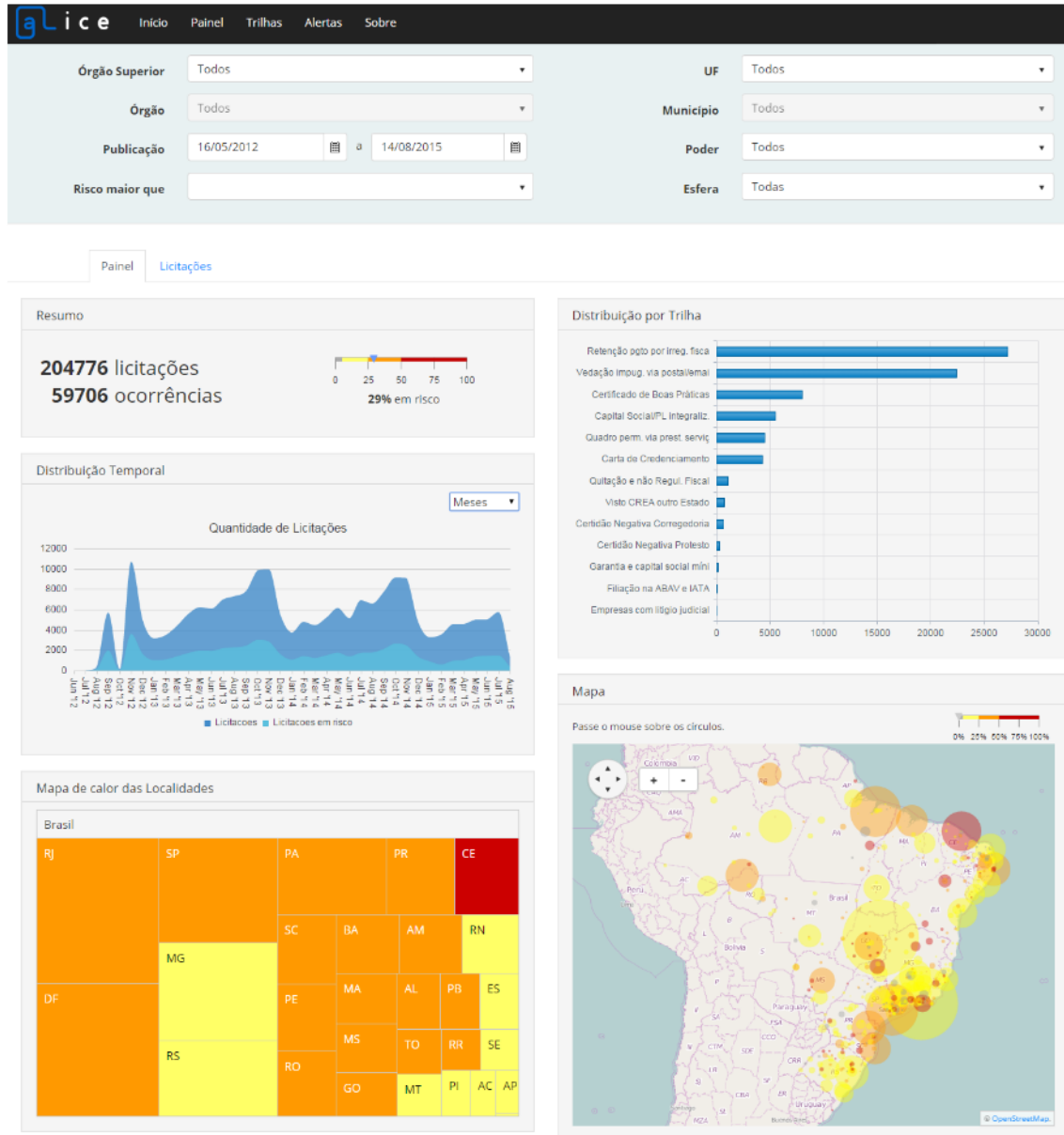
A primeira versão da ferramenta Alice foi disponibilizada em setembro de 2014 para testes pelos auditores, com os seguintes módulos:

- *Scraper*: realiza o *download* automático dos editais e obtém os metadados das licitações

- publicadas no Comprasnet diariamente. Esse módulo foi desenvolvido em C# (.NET) e o banco de dados utilizado foi o MySQL;
- Analytics: executa a análise textual dos editais com base nos tipos de achados pré-definidos. Este módulo foi desenvolvido em Python;

- Web: apresenta graficamente as licitações analisadas, com painéis gerenciais, permitindo detectar localidades e órgãos com maior risco. O módulo foi inicialmente desenvolvido em C#.

FIGURA 7 - TELA INICIAL DA PRIMEIRA VERSÃO DA ALICE



Fonte: Sistema Alice

Durante os testes, surgiram críticas sobre o possível direcionamento incorreto da Alice para determinadas licitações em detrimento de outras que poderiam ter problemas, prejudicando o julgamento profissional do auditor. Outra preocupação foi a possível impossibilidade de realização das análises necessárias dado o comprometimento da capacidade operacional com auditorias obrigatórias, como a Auditoria Anual de Contas e o Programa de Fiscalização por Sorteios de Municípios.

Nesse contexto, em 2015, a utilização da Alice foi liberada para todos os auditores. Entretanto, a utilização por cada área era facultativa, cabendo a decisão ao Coordenador-Geral de auditoria da SFC.

Em 2016, conforme descrito por Costa & Bastos (2020), a CGU cedeu ao TCU da Alice. O Tribunal, então, adotou, evoluiu e usa a ferramenta até os dias atuais, também com ótimos resultados.

Em 2018, verificou-se que a Alice estava em desuso na CGU. O diagnóstico era de que a parte tecnológica era adequada, mas não havia um processo formalizado e padronizado a ser seguido obrigatoriamente por todas as áreas. Nesse sentido, a CGATI retomou o projeto com o redesenho das atividades do processo, com a inclusão de novos tipos de achados de auditoria, adição do uso de técnicas de inteligência artificial para classificação das compras e com a integração com o sistema e-Aud.

Como visto, entre 2019 e 2021, a ferramenta voltou a gerar resultados expressivos com a ampliação gradual do uso na CGU e em diversas outras instituições. Destaca-se que em 2021, realizou-se um acordo de cooperação técnica o Ministério da Economia para integrar a Alice ao Compras.gov.br, permitindo que o gestor seja alertado de eventuais fragilidades ainda na etapa de planejamento da aquisição.

Por fim, em 2022, a CGPRI assumiu o projeto com a missão de institucionalizar o processo em toda a CGU, estabelecendo critérios objetivos para determinar a realização de auditorias em licitações com alto risco de irregularidade. Foi, também, criado um grupo de auditores especialistas em compras e contratações, a Comissão Alice, para monitorar o uso e direcionar as evoluções da ferramenta de forma inclusiva e abrangente. Em complemento, novas formas de uso de inteligência artificial estão sendo exploradas para aperfeiçoar a eficiência do uso da Alice. Nesse ano a ferramenta foi apresentada em reunião da Organização das Nações Unidas como exemplo do uso de tecnologia para combater a corrupção, despertando o interesse e a curiosidade de diversos países.

4. CONCLUSÕES E ENCAMINHAMENTOS PROPOSTOS

O desenvolvimento da Alice demonstrou a grande capacidade dos servidores da CGU de enfrentarem desafios para cumprirem com maestria o seu objetivo de agregar valor à gestão pública.

A incorporação da Alice nas rotinas cotidianas da CGU abriu uma série de oportunidades de aperfeiçoamento de processos, aumento de eficiência e mitigação de riscos à administração pública, traduzidas nos belos resultados aqui apresentados.

Para o futuro observa-se a oportunidade ampliação do uso de inteligência artificial e elaboração de novas tipologias em artefatos anteriores ao edital, permitindo que a auditoria possa contribuir nos estágios iniciais da licitação. A incorporação de análises informações de compras de Estados e Municípios também é oportuna para ampliar nacionalmente os benefícios a serem gerados.

REFERÊNCIAS BIBLIOGRÁFICAS

AICPA. (2015). *Audit Analytics and Continuous Audit: Looking Toward the Future*. New York: American Institute of Certified Public Accountants, Inc.

Bosio, E., Djankov, S., Glaeser, E., & Shleifer, A. (2020). Public Procurement in Law and Practice. *American Economic Review*, 112(4), 1091-1117. doi:DOI 10.3386/w27188

BRASIL. (2017, 11 22). Decreto nº9.203 de 22/11/2017. *Dispõe sobre a política de governança da administração pública federal direta, autárquica e fundacional*. Brasília. Retrieved 04 14, 2022, from http://www.planalto.gov.br/ccivil_03/ato2015-2018/2017/decreto/d9203.htm

- CGU. (2017). *Manual de Orientações Técnicas da Atividade de Auditoria Interna Governamental do Poder Executivo Federal*. Brasília: Controladoria-Geral da União. Retrieved Abr 12, 2022, from https://www.gov.br/defesa/pt-br/arquivos/lai/auditoria/ciset/manuais-normativos-e-pop/manual_orient_tecnica_atividade_auditoria_interna_governamental.pdf
- CGU. (2022). *Relatório de Gestão CGU 2021*. Brasília: Controladoria-Geral da União. Retrieved Abr 12, 2022, from https://repositorio.cgu.gov.br/bitstream/1/67793/8/Livro_Relatorio_Gestao_2021_v2.pdf
- Chan, D., Chiu, V., & Vasarhelyi, M. (2018). *Continuous Auditing: Theory and Application*. Bingley: Emerald Publishing Limited.
- Costa, M., & Bastos, P. (2020). Alice, Monica, Adele, Sofia, Carina e Ágata: o uso da inteligência artificial pelo Tribunal de Contas da União. *Revista do Tribunal de Contas do Estado de Goiás*, 11-34.
- Fantina, R., Storozhuk, A., & Goyal, K. (2022). *Introducing Robotic Process Automation to Your Organization: A Guide for Business Leaders*. New York: Apress. doi:<https://doi.org/10.1007/978-1-4842-7416-3>
- Fortini, C., & Motta, F. (2016). Corrupção nas licitações e contratações públicas: sinais de alerta segundo a Transparência Internacional. *Revista de Direito Administrativo & Constitucional*, Ano 16 n. 64 (Abril/Junho - 2016), 93-113.
- IIA. (2009). *Global Technology Audit Guide (GTAG) 13 - Fraud Prevention and Detection in an Automated World*. Altamore Springs: The Institute of Internal Auditors Inc.
- OECD. (2018). *The MAPS Analytical Framework*. Retrieved Abril 27, 2022, from Methodology for Assessing Procurement Systems (MAPS): <https://www.mapsinitiative.org/methodology/>
- OECD. (2021). *Government at a Glance 2021*. Paris: OECD Publishing. doi:<https://doi.org/10.1787/1c258f55-en>
- Oliveira, E., Pedrosa, C., Silva, S., Costa, W., & Silva, J. (2021). O Uso de Tecnologias Para Detecção de Fraudes na Pandemia da Covid-19. *Revista Contabilidade e Controladoria*, 13(1), 156-177.
- Panis, A., Isidro, A., Carneiro, D., Montezano, L., Resende Junior, P., & Sano, H. (2021). Inovação em compras públicas: Atividades e resultados no caso do robô Alice da Controladoria-Geral da União. *Cadernos Gestão Pública e Cidadania (CGPC)*, 27(86), 1-19. doi:<https://doi.org/10.12660/cgpc.v27n86.83111>
- Ribeiro, C. G., & Inácio Júnior, E. (2019). *O Mercado de Compras Governamentais Brasileiro (2006-2017): mensuração e análise*. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada – Ipea. Retrieved from https://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=34850
- Riedel, L., & Ferreira, L. (2021). Aplicabilidade do compliance com o suporte da inteligência artificial no combate a corrupção nos processos licitatórios. In A. Vasconcelos, *Ciências Jurídicas: Fundamentação, participação e efetividade* (pp. 78-90). Ponta Grossa: Atena Editora. doi:10.22533/at.ed.906211506
- SFC/CGU. (2021). *Relatório Gerencial: Capacidade operacional da atividade de auditoria interna governamental da CGU em 2021*. Brasília: Secretaria Federal de Controle Interno.
- Tátrai, T., & Németh, A. (2018, Mai 14). Improving red flag instruments for public procurement. *ERA Fórum*, 267-285. doi:<https://doi.org/10.1007/s12027-018-0513-8>
- VARGAS, R. (2009). *Gerenciamento de projetos - Estabelecendo diferenciais competitivos* (7a ed.). Rio de Janeiro: Brasport Livros e Multimídia Ltda.

**Andre Luiz Monteiro da Rocha**

Controladoria-Geral da União

andre.rocha@cgu.gov.brORCID: <https://orcid.org/0000-0003-3177-8889>

Engenheiro de Computação, pela Universidade Católica Dom Bosco. Pós-graduando em Prevenção e Combate a Desvios de Recursos Públicos, pela Universidade Federal de Lavras. Auditor Federal de Finanças e Controle da Controladoria-Geral da União desde 2006. Atuou como Coordenador-Geral de Auditoria de Tecnologia da Informação da CGU, entre 2018 e 2020, e como Coordenador-Geral de Inteligência de Dados em 2021, em Brasília-DF. Atualmente ocupa o cargo de chefe do Núcleo de Ações Especiais da CGU-Regional Mato Grosso do Sul.

**Matheus Scatolino de Rezende**

Controladoria-Geral da União

matheus.rezende@cgu.gov.br<http://lattes.cnpq.br/8893510894308192>ORCID: <https://orcid.org/0000-0002-6437-5819>

Auditor Federal de Finanças e Controle desde 2012. Especialista em Ciência de Dados aplicada a Políticas Públicas pela Escola Nacional de Administração Pública (ENAP). Bacharel em Ciência da Computação pela PUC Minas e Administração pela UFMG. Certificado em Auditoria de Sistemas da Informação (CISA-ISACA) e em Gestão de Riscos (C31000). Atuou como Coordenador-Geral de Auditoria de Tecnologia da Informação (CGATI) no período entre 2020 e 2022. Atualmente, ocupa o cargo de Chefe de Divisão da CGATI.

**Tiago Chaves Oliveira**

Controladoria-Geral da União

tiagocoliveira@gmail.com<https://www.linkedin.com/in/tiago-chaves-oliveira/>ORCID: <https://orcid.org/0000-0002-8273-1194>

Auditor Federal de Finanças e Controle, Coordenador-Geral de Prospecção e Inovação da Secretaria Federal de Controle Interno da Controladoria-Geral da União (CGU). Mestre em Governança e Desenvolvimento pela Escola Nacional de Administração Pública (ENAP). Bacharel em Ciência da Computação com MBA em Gestão de Projetos de Engenharia de *Software*. Certificado em Auditoria Governamental (CGAP-IIA) e em Gerenciamento de Projetos (PMP-PMI). Atua nas temáticas Gerenciamento de Projetos e Auditoria Interna como praticante, professor e facilitador.

Artigos

Contribuições das denúncias e representações autuadas em 2020 para o controle externo realizado pelo Tribunal de Contas do estado de Santa Catarina¹

Contributions of complaints and representations filed in 2020 to the external control performed by the Court of Auditors of the state of Santa Catarina

Contribuciones de quejas y denuncias presentadas en 2020 al control externo realizado por el Tribunal de Cuentas del estado de Santa Catarina

Fernanda Camila De Carli e Fabiano Maury Raupp

<https://doi.org/10.36428/revistadacgu.v14i26.579>

Resumo: O objetivo do artigo é identificar as contribuições do controle social por meio de denúncias e representações na atividade de controle externo realizada pelo Tribunal de Contas do Estado de Santa Catarina. A partir da concepção teórica do Novo Serviço Público e das noções de *accountability* diagonal, a pesquisa explora as denúncias e representações autuadas no TCE/SC no ano de 2020. O estudo aplicou uma abordagem quantitativa descritiva. Os principais resultados da análise demonstram que: i) a maioria dos processos propostos pelo controle social possuem como objeto procedimentos licitatórios; ii) são apresentadas majoritariamente por empresas privadas; iii) se referem à esfera municipal em detrimento da esfera estadual; e, iv) possuem baixa efetividade nos termos em que a maioria dos processos originados das denúncias e representações acabam sendo improcedentes, extintos ou não conhecidos. Apesar de positiva a possibilidade da interação entre o controle social e o controle externo, observa-se que ela ainda é limitada, necessitando de maior desenvolvimento.

Palavras-chave: Controle social. Denúncias. Representações. Controle Externo.

Abstract: The objective of the article is to identify the contributions of social control through complaints and representations in the external control activity carried out by the Court of Auditors of the State of Santa Catarina. Based on the theoretical conception of the New Public Service and the notions of diagonal *accountability*, the research explores the complaints and representations filed in the TCE/SC in the year 2020. The study applied a descriptive quantitative approach. The main results of the analysis demonstrate that: i) most of the processes proposed by social control have bidding procedures as their object; ii) are presented mostly by private companies; iii) refer to the municipal sphere to the detriment of the state sphere; and, iv) they have low effectiveness in the terms in which most of the processes originated from the complaints and representations end up being groundless, extinct or not known. Despite the positive possibility of interaction between social control and external control, it is observed that it is still limited, requiring further development.

Keywords: Social control. Complaints. Representations. External control.

1. Artigo submetido em 27/07/2022 e aceito em 01/12/2022.

Resumen: El objetivo del artículo es identificar las contribuciones del control social a través de denuncias y gestiones en la actividad de control externo realizada por el Tribunal de Cuentas del Estado de Santa Catarina. Con base en la concepción teórica del Nuevo Servicio Público y las nociones de rendición de cuentas diagonal, la investigación explora las denuncias y gestiones presentadas en el TCE/SC en el año 2020. El estudio aplicó un enfoque cuantitativo descriptivo. Los principales resultados del análisis demuestran que: i) la mayoría de los procesos propuestos por el control social tienen como objeto los procedimientos de licitación; ii) son presentados en su mayoría por empresas privadas; iii) referirse al ámbito municipal en detrimento del estatal; y, iv) tienen baja efectividad en los términos en que la mayoría de los procesos originados a partir de las quejas y representaciones terminan siendo infundados, extinguidos o desconocidos. A pesar de la posibilidad positiva de interacción entre el control social y el control externo, se observa que aún es limitada y requiere mayor desarrollo.

Palabras clave: Control social. quejas Representaciones. Control externo.

1. INTRODUÇÃO

A administração pública pode ser analisada a partir da perspectiva de modelos. Denhardt e Denhardt (2015) apresentam a corrente do Novo Serviço Público (NSP) que se trata de um novo modelo de gestão com ênfase na participação democrática e cidadania, expandindo o interesse público para além do conjunto dos interesses individuais, situando os modelos de comunidade e sociedade civil. O modelo do Novo Serviço Público no paradigma da Administração Pública tem seus pilares na participação social dos atores da sociedade para reafirmação de valores como o interesse público, o processo de governança e a expansão da cidadania democrática. Nesse sentido, constata-se que o surgimento do Estado democrático de direito instituído a partir da Carta Magna de 1988 prevê também o maior controle das atividades administrativas, seja por um meio institucional, seja por meio da presença mais próxima da participação popular, através de um controle social.

Dentre os meios institucionais, o sistema de controle da administração pública é composto pelo controle judicial, o controle interno, o controle externo e o controle social, nos moldes designados pela Constituição Federal. Assim, o controle da Administração Pública pode ocorrer mediante ação no Judiciário para avaliar a legalidade da ação administrativa; mediante o controle interno, tangenciando a autotutela da própria administração; mediante controle externo, a cargo do Poder Legislativo e Tribunais de Contas; e mediante controle social, realizado pelo cidadão, grupo de cidadãos ou sociedade. Os diferentes tipos de controle se interligam através de contribuições mútuas a fim de favorecer a transparência e fortalecer o sistema de controle das ações da Administração Pública, com o po-

der-dever de comunicar as irregularidades verificadas em seu âmbito para as outras instâncias de controle (SERRA; CARNEIRO, 2011).

O controle da Administração Pública realizado especificamente pelo Tribunal de Contas recai sobre a fiscalização contábil, financeira, orçamentária, operacional e patrimonial quanto à legalidade, legitimidade, economicidade, aplicação das subvenções e renúncia de receitas do entes que estão sob sua jurisdição administrativa. No caso do Estado de Santa Catarina, tal função cabe ao Tribunal de Contas do Estado de Santa Catarina (TCE/SC) que foi criado em 4 de novembro de 1955, pela Lei Estadual n. 1.366, sancionada pelo governador Irineu Bornhausen, durante a Constituição Federal de 1946 (BRASIL, 1947).

A Corte de Contas também se abre para apropriar a participação social em suas atividades, o que pode ser denominado de controle social. Neste trabalho a diferenciação dos conceitos de participação social e controle social será feita considerando o controle social como uma espécie do gênero da participação social. Assim, entende-se que a participação social inclui o aspecto fiscalizatório que se traduz no controle social (SANTOS, 2002). A adoção desse posicionamento se revela na percepção empírica acerca da farta doutrina que engloba o controle social para tratar do escopo da fiscalização realizada pela sociedade sobre o Estado, na medida em que o termo “participação social” é utilizada em diferentes assuntos.

Acerca dos instrumentos ao dispor da sociedade para sua manifestação, as principais ferramentas são as ouvidorias, canais de recepção para denúncias, representações e consultas junto aos controles interno e externo da Administração (SERRA; CARNEIRO, 2011). Neste sentido, a existência de mecanismos

é fundamental para que o cidadão possa exercer ou acionar as instituições de controle, promovendo o controle social.

Dentre as ferramentas citadas, ressalta-se que as ouvidorias recebem as solicitações e informações sobre atos de gestão com indícios de irregularidades praticados por agentes públicos ou por demais responsáveis pela aplicação de recursos públicos, como também recebem críticas, sugestões e concedem demais informações que lhe são solicitadas. Já os canais para recepção de denúncias, representações e consultas são mais específicas e necessitam do preenchimento de requisitos de admissibilidade sobre os fatos apresentados e qualificação daquele que denuncia.

Dessa forma, o Tribunal de Contas é incumbido de propiciar a provocação da ação fiscalizatória autônoma, por meio da sociedade, tornando-se um instrumento da cidadania ativa, sendo indispensável ao funcionamento do regime democrático. Conforme o Relatório de Atividades constante no site do TCE/SC, o quantitativo de denúncias e representações recebidas no ano de 2020 evidencia que houve um crescimento na apresentação desses instrumentos perante o processo de fiscalização e controle externo da Administração Pública, conforme se extrai dos dados disponibilizados no portal da Corte de Contas (TCE/SC, 2020). Apresenta-se, por meio da Tabela 1, o número de denúncias e representações no período 2016-2020.

TABELA 1 - NÚMERO DE DENÚNCIAS E REPRESENTAÇÕES NO PERÍODO 2016-2020

ANO	TOTAL
2016	685
2017	748
2018	576
2019	509
2020	830

Fonte: TCE/SC (2020).

Apesar do aumento na apresentação das denúncias e representações, é prudente verificar as especificidades destes processos, de modo a ser possível uma avaliação, ainda que inicial, da relevância dessa participação como efetiva manifestação da sociedade e comunicação legítima entre o Estado e a sociedade, O foco, então, deve se pautar não apenas na quantidade, mas sim na qualidade e efetividade das denúncias e representações autuadas perante a Corte de Contas.

Nesse rumo, o objetivo geral do presente trabalho é identificar as contribuições do controle social por meio de denúncias e representações na atividade de controle externo realizada pelo Tribunal de Contas do Estado de Santa Catarina. Em termos específicos busca-se: a) descrever a tramitação dos processos de denúncias e representações autuados no TCE/SC; b) caracterizar os processos de denúncias e representações autuados em 2020 no TCE/SC; c) analisar as decisões decorrentes dos processos de denúncias e representações e as respectivas medidas tomadas pelo TCE/SC.

A justificativa para o desenvolvimento do estudo envolve a necessidade de elucidar de que forma as representações e denúncias contribuem nas atividades

do controle externo. O recorte de pesquisa considera o ano de 2020. Neste ano também se iniciou a pandemia do coronavírus e igualmente houve um aumento relevante na autuação de processos por provocação de agentes externos ao Tribunal de Contas (TCE/SC, 2020).

Assim, esse estudo contribuiu para o campo da Ciência da Administração Pública na medida em que procurou esclarecer a forma com que o controle social repercutiu nas atividades de controle externo. Paralelamente, buscou esclarecer quais medidas foram tomadas pelo órgão de controle externo nessa seara. Dentre as contribuições práticas vislumbradas, avaliou-se o monitoramento constante da população na promoção da democracia e cidadania perante as ações do controle externo. Igualmente, realizou-se um esclarecimento sobre a forma de atuação do TCE/SC com relação às provocações externas que recebe, o que afeta a legitimidade das instituições democráticas.

2. FUNDAMENTAÇÃO TEÓRICA

A partir das premissas do modelo de Administração Pública do Novo Serviço Público, desenvolvido pelos autores Denhardt e Denhardt (2015), baseadas na cidadania democrática e no reconhecimento da complexidade da *accountability*, o estudo se aprofunda sobre conceitos de *accountability* e, em especial, aborda a *accountability* diagonal que revela as interações entre instituições de controle da administração pública, como a atividade de controle externo desenvolvida pelo Tribunal de Contas, e o controle social, exercido pela sociedade. Assim, pretende-se conectar a existência das denúncias e representações como manifestação do controle social exercido por meio do controle externo e que contribui para a concretização da *accountability* diagonal, sob a perspectiva do modelo do Novo Serviço Público.

2.1. Novo serviço público

Esta abordagem se constrói sobre a ideia do interesse público, da valorização do cidadão e do diálogo aberto, com estabelecimento de um espaço democrático para servir a sociedade (SANTOS; SELIG, 2014). No Novo Serviço Público, a cidadania passa a ser encarada de forma mais ampla e considerada a partir da natureza da condição do indivíduo enquanto membro da comunidade política, incluindo questões como direitos e responsabilidades dos cidadãos, o que implica em um engajamento ativo do cidadão na vida política. Dentro dessa perspectiva, os administradores devem ver os cidadãos não meramente como eleitores, mas como agentes capazes de compartilhar a autoridade das decisões, confiando na eficácia da colaboração para que haja maior responsividade e correspondente aumento na confiança do cidadão na gestão (DENHARDT; DENHARDT, 2015).

A abordagem do Novo Serviço Público tem seu foco principal voltado para o espaço democrático da sociedade e requer cidadãos mais ativos, cidadãos na prática e não apenas no papel, com efetiva disposição para exercer a sua cidadania. Nesse cenário, os cidadãos aceitam que cada direito assegurado requer uma responsabilidade intrínseca e, por isso, estão presentes para participar ativamente das questões cívicas (SANTOS; SELIG, 2014). Dessa forma, passa-se de um processo deliberativo fechado de gestão a um aberto, em que a decisão passa a ser composta ao invés de imposta. Além disso, o governo não é mais a única so-

lução institucional para revelar os rumos da sociedade, cedendo espaço para um método alternativo de solução negociada, próprio de sociedades abertas. Com isto, os diálogos político e administrativo se expandem e interligam como forma de conectar o governo com o seu entorno social no processo decisivo, conferindo voz a novos agentes, sobretudo os cidadãos. Assim sendo, ainda há o intervencionismo estatal, porém sob uma nova dinâmica branda e interativa (COUTINHO, 2020).

Um nível de participação política baixo e pouca confiança sobre meios e motivações dos políticos em suas atuações é o reflexo de um abismo entre os líderes e os cidadãos. Também deve ser considerada a complexidade pura e simples da sociedade que torna difícil o envolvimento cívico pois há muitos encargos na vida moderna, e o envolvimento na esfera pública toma tempo. Entretanto, se há problemas numa sociedade que não é democraticamente ativa, a solução é continuar a educar e a informar os cidadãos sobre o sistema político existente e seu papel basilar (DENHARDT; DENHARDT, 2015). Sob a visão do Novo Serviço Público, em que pese as dificuldades enfrentadas, a maior participação dos cidadãos traz benefícios como: identificar quais as expectativas dos cidadãos; aumentar a qualidade da política pública; ajudar na implementação dessa política, posto que os agentes têm mais participação nos resultados; responder aos pedidos de maior transparência e responsabilização do governo; ajudar a enfrentar os desafios da emergente sociedade da informação; e resultar em melhor informação pública (BALBE, 2013). O reforço da abertura do governo para a interação com a sociedade de forma abrangente possui raízes na teoria da cidadania democrática, cuja base se pauta em uma cidadania revigorada, mais ativa e engajada (SANTOS; SELIG, 2014).

A função dos servidores públicos e administradores públicos vai além de solucionadores de problemas ou empregados do governo, pois são considerados cidadãos que estendem responsabilidades da cidadania ao trabalho. Os administradores públicos devem assumir responsabilidades especiais e conhecimentos específicos sobre responsividade e *accountability*, que são inerentes à ideia da moralidade democrática. Nesse sentido, os servidores públicos têm obrigação ética de facilitar a participação pública no processo político por todas as maneiras possíveis, ainda que possa redundar em atrasos e confusão na execução do seu trabalho. O tempo envolvido no en-

gajamento dos cidadãos deve ser visto como um dever do gestor público de envolver cidadãos no trabalho da democracia (DENHARDT; DENHARDT, 2015).

Nesse rumo, os administradores públicos devem contribuir para a construção de uma noção coletiva e compartilhada do interesse público. Há uma necessidade de contrapartida da ação governamental na medida em que os cidadãos assumem seu papel ativamente. Por isso, os gestores devem estar dispostos a ouvir e considerar as necessidades dos cidadãos na tomada de decisão, o que implica na demonstração de comprometimento com os valores democráticos (IOCKEN, 2018). Dessa forma, a participação pública encara cada indivíduo como não mero cliente mas como agente ativo na sociedade. Este novo entendimento acerca da participação cidadã resgata uma perspectiva proativa de cidadania presente em uma democracia substantiva, pela qual a ação estatal se legitima continuamente por intermédio da presença dos indivíduos nas várias fases de lapidação das políticas públicas (COUTINHO, 2020).

O modelo do Novo Serviço Público sugere, ainda, que o governo não deve responder aos interesses egoístas de curto prazo dos cidadãos de forma prioritária e exclusiva, cabendo a demonstração de preocupação com a comunidade de forma mais ampla, com disposição para assumir responsabilidade pessoal pelo que ocorre em suas vizinhanças, bairros e na comunidade. Afinal, essa é a cidadania efetiva e responsável, e o governo, para acompanhar, deve ser responsivo às necessidades e aos interesses dos cidadãos, estimulando cada vez mais pessoas a cumprirem suas responsabilidades enquanto cidadãos (DENHARDT; DENHARDT, 2015).

Desta maneira, as reformas administrativas devem se conformar aos anseios dos cidadãos, com abertura e participação na gerência da coisa pública. Entretanto, para que haja uma transformação efetiva, as reestruturações conduzidas pelos líderes dependem de um consenso alcançável apenas mediante a ampla participação dos cidadãos nos processos deliberativos. Igualmente, a estabilidade da transformação do Estado depende de uma articulação política que transcenda os setores burocráticos, sob o risco de haverem alterações apenas paliativas e que atendam a interesses corporativistas, o que é possível a partir da instrumentalização de meios participativos para obstar o tradicional isolamento das instituições burocráticas (COUTINHO, 2020)

Busca-se resgatar a relação de confiança e de colaboração entre os cidadãos e o governo através do exercício contínuo dos valores de responsabilidade e de moralidade. É necessária a construção de indicadores específicos sobre o Novo Serviço Público a partir dos estudos nesse campo em que existe um debate acadêmico que estimula construção e o fomento das diretrizes deste modelo pautado na cidadania democrática (IOCKEN, 2018). O Novo Serviço Público reconhece a importância e complexidade da *accountability* e sustenta que as medidas de eficiência ou padrões baseados em mercados não conseguem medir ou estimular um comportamento responsável do gestor. As correntes anteriores simplificaram demais a questão (DENHARDT; DENHARDT, 2015).

Novos rumos apresentam-se para que haja o aumento da permeabilidade social, da transparência e de novos instrumentos controladores. Isso demanda governos mais transparentes e proativos na prestação de informações aos cidadãos e agências de controle com acompanhamento concomitante e minucioso, exercido direta ou indiretamente por parte da sociedade, considerando que a controle e responsabilidade são faces de uma mesma moeda (COUTINHO, 2020). Assim, para uma aplicação com sucesso dos pressupostos do serviço público, os fatores-chave são o diálogo, as redes de colaboração e a capacidade institucional. Paralelamente, cabe estabelecer uma cultura de participação cívica, pois os cidadãos devem perceber que a dinâmica da administração pública não diz respeito apenas ao governo, mas sim a todos os interessados na esfera política, econômica e social (PEREIRA; CORREIA, 2020).

2.2. *Accountability*

O Novo Serviço Público tem como premissa o reconhecimento que a *accountability* não é simples, considerando a gama de responsabilidades na atuação dos servidores públicos que devem resguardar as normativas do direito constitucional e estatutário, dos valores da comunidade, das normas políticas, dos padrões profissionais e dos interesses dos cidadãos. A *accountability* vai além do foco em apenas num conjunto restrito de medidas de desempenho, de forças de mercado ou de uma expertise neutra. Trata-se de um agir a partir dos pressupostos da democracia, do papel da cidadania e de um serviço público dedicado a servir a cidadãos no interesse público (DENHARDT; DENHARDT, 2015). O condicionamento das ações

dos representantes do Estado aos anseios dos representados é uma das prerrogativas do cidadão, uma vez que os agentes públicos atuam em nome e em razão da sociedade. Este condicionamento pode ser exercido direta ou indiretamente, tendo em vista que a prestação de contas dos assuntos de interesse público está sujeita a múltiplas formas de responsabilização e controle (COUTINHO, 2020).

Accountability, em um sentido amplo, pode ser conceituada como a obrigação de justificar uma conduta a terceiro, indo além de uma mera responsabilidade formal. A *accountability* pode ser observada sob o aspecto passivo, em que se refere à imputação de um comportamento a alguém e, sob a ótica ativa, que alude à assunção da imputação. Em termos práticos, envolve a identificação e a responsabilização de quem tenha poder, na seguinte ordem: i) prestação de contas; ii) controle e; iii) responsabilização, quando for o caso (ALENCAR, 2018).

A *accountability* pode ser classificada em *accountability* vertical, em que os atores da sociedade controlam a agentes estatais especialmente por meio do processo eleitoral; e também em *accountability* horizontal, realizada pela existência de agências estatais que têm o direito e o poder legal e que estão de fato dispostas e capacitadas para realizar ações que visam monitorar e sancionar condutas delituosas. Inclusive, há uma dependência entre a *accountability* vertical e horizontal, dependendo uma da outra para que haja um reforço de sua aplicação. (O'DONNELL, 1998). Para o presente trabalho, considera-se que a articulação entre a sociedade civil e as instituições estatais de controle e fiscalização pode ser denominada de *accountability* diagonal (PELIZZO; STAPENHURST, 2013). Há alguns princípios fundamentais que norteiam o exercício da *accountability* diagonal, assim elencados por Willeman (2019, p.76):

De acordo com a doutrina especializada, os princípios fundamentais que devem nortear o exercício da *accountability* diagonal são:

(i) participação, segundo o qual o eixo horizontal de *accountability* deve colocar-se aberto à inserção e à participação da cidadania, em lugar de se criarem novas e separadas instâncias de controle;

(ii) fluxo informacional, que busca assegurar o acesso da cidadania a informações de interesse público que, em regra, somente esta-

riam disponíveis às instâncias horizontais de fiscalização, bem como garantir o acesso às decisões e deliberações adotadas pelas próprias agências de *accountability* horizontal;

(iii) e, finalmente, agregação de autoridade, segundo o qual o exercício da *accountability* diagonal favorece e, ao mesmo tempo, aproveita-se da autoridade legalmente reconhecida às entidades do eixo horizontal tanto para exigirem respostas, quanto para aplicar em sanções formais a agentes estatais submetidos à sua fiscalização.

O sentido da *accountability* a ser tratado neste trabalho aborda apenas o controle social organizado que interage com o controle externo do tribunal de contas que é institucionalizado, caracterizando uma forma de manifestação da *accountability* diagonal.

2.3. Controle externo

Dentre as formas de controle da administração pública, o controle externo se refere à fiscalização de uma instituição sobre outra, a ser realizada num determinado órgão público, por funcionários de fora desse órgão, pertencentes a outro setor, situados dentro do sistema estatal. No atual sistema brasileiro, o controle externo é atribuição do Poder Legislativo, que o exerce com o auxílio do Tribunal de Contas, o qual realiza a fiscalização financeira, patrimonial e contábil em todos os órgãos dos três Poderes estatais (SIMÕES, 2014).

As funções do Tribunal de Contas podem ser sistematizadas em sete categorias: fiscalizadora, opinativa, julgadora, sancionadora, corretiva, consultiva, informativa, ouvidora e normativa. A função fiscalizadora compreende as ações relativas ao exame e à realização de diligências, auditorias e outras atividades de fiscalização. A função opinativa se constitui dos pareceres prévios sobre as contas dos Chefes dos Poderes. A função julgadora se refere ao julgamento das contas e infrações de competência da Corte (LIMA, 2019). As decisões na Corte de Contas são formadas por intermédio dos votos tomados por um colegiado formado por Conselheiros que são responsáveis pela compatibilização entre meios e fins e também pelos trabalhos executados (SIMÕES, 2014). Em um aspecto mais amplo, o controle se insere no sistema de *accountability*.

As denúncias e representações são mecanismos que permitem o exercício do controle social. Conectando-o ao controle externo, dentro de um sistema de *accountability* diagonal diante da complexidade reconhecida pela corrente do Novo Serviço Público. No Tribunal de Contas do Estado de Santa Catarina, a legitimidade para apresentação das denúncias e representações é de qualquer cidadão, partido político, associação ou sindicato. Esses atores possuem tanto o direito de denunciar atos ilegais, ilegítimos e anti-econômicos praticados por gestores públicos, quanto o dever de apontar essas irregularidades (TCE/SC, 2020).

2.4. Controle social

No Estado Democrático de Direito, qualquer atuação administrativa está exposta ao crivo social desde a sua formulação até os seus efeitos. Logo, o controle realizado em relação ao ato administrativo é indispensável, tal qual também são os respectivos mecanismos viabilizadores do controle (FRANÇA, 2016). Assim sendo, verifica-se que o controle social é dotado de representatividade e legitimidade. Entretanto, necessita que haja elementos na gestão pública que adotem mecanismos que ajudem na apuração e responsabilização dos agentes irresponsáveis (BRAGA, 2020).

Para aferir a extensão do exercício de controle democrático do Estado, é necessário avaliar o nível de difusão de informação disponibilizada ao cidadão de forma qualitativa e quantitativamente, seja por meio da publicidade dos atos da administração pública ou por meio de canais de comunicação, como ouvidorias ou centros de atendimentos aos cidadãos, abertos para denúncias que devem ser devidamente apuradas e implicar em sanções efetivas (FRANÇA, 2016). Desta maneira, percebe-se que a transparência e participação são conceitos complementares, sendo um meio para fomentar e qualificar o controle social dentro de um governo democrático e uma sociedade complexa (BRAGA, 2020).

Alguns estudos procuram identificar quais são os aspectos do controle social diante do controle externo, exercido pelos Tribunais de Contas, analisando aspectos como: eficácia de apuração de denúncias, interação entre controle social e o controle externo, formas de instrumentalização do controle social diante do controle externo e análise quantitativa de denúncias. Dentre eles, cabe destacar Azevedo (2015), por meio de artigo, realizou uma pesquisa quantitativa sobre as

denúncias formuladas e apreciadas pelo Tribunal de Contas do Estado de Minas Gerais no período de 2009 a 2014. Apurou-se que o sistema necessita de melhoras na divulgação e deliberação das denúncias. Todavia, houve o crescimento das autuações e das apreciações das denúncias no período analisado, o que revelou que as denúncias são um meio valioso de colaboração em prol da democracia e do atendimento ao interesse público primário.

Castro (2016) objetivou apresentar ferramentas para o exercício do controle social e as maneiras como a sociedade pode induzir e fortalecer a atuação do controle externo através de uma abordagem sócio jurídica. Depreendeu-se do estudo que deve haver parceria entre tais controles para o aperfeiçoamento da gestão pública, pontuando-se a necessidade de expansão da atuação.

Rocha, Zuccolotto e Teixeira (2020), analisaram a permeabilidade dos Tribunais de Contas brasileiros à sociedade, por meio de 3 perspectivas: a) transparência dos portais; b) transparência da atividade de fiscalização; e c) participação cidadã. A metodologia utilizou-se de um roteiro de observações, a partir de recomendações de organismos internacionais e após foram elaborados mapas perceptuais para cada uma das perspectivas analisadas. Os resultados evidenciam que os tribunais de contas não são transparentes, o que limita a *accountability* social, e são muito resistentes à participação, limitando-se a disponibilizar os mecanismos de ouvidoria, portal da transparência e serviço de atendimento ao cidadão, tornando-os ainda, muito insulados, burocratizados e pouco abertos à participação popular. Por outro lado, restou identificada uma forte influência da Lei de Responsabilidade Fiscal (LRF) e da Lei de Acesso à Informação (LAI), apontando futuras mudanças, ainda que incrementais, na direção da atuação dos tribunais de contas como aliados para o aprimoramento da *accountability* vertical.

Diante dos estudos sobre o tema, verifica-se alguns aspectos relevantes. Em síntese, nota-se que os trabalhos apresentados apontam em comum a necessidade de aprimorar a interação entre controle social e o controle externo, uma vez que se trata de instrumento essencial para aprimorar a democracia na esfera brasileira.

Nesse sentido, o presente trabalho busca verificar como está sendo realizada esta interação no âmbito do Tribunal de Contas catarinense, para que se

identifique as medidas que derivam do controle social e contribuem para o aperfeiçoamento da gestão pública. Dessa forma, a pesquisa avança em relação aos estudos anteriores ao entrelaçar o paradigma do Novo Serviço Público com as noções de *accountability* diagonal, observando paralelamente a construção jurídica do controle externo e controle social, para analisar o escopo quantitativo e qualitativo das denúncias e representações apresentadas pela sociedade em 2020 no âmbito da Corte de Contas catarinense.

3. PROCEDIMENTOS METODOLÓGICOS

A pesquisa possui como objeto as representações e denúncias autuadas no TCE/SC no ano de 2020. Para sua compreensão, foram identificadas as decisões expedidas pela Corte de Contas em relação a tais processos. O estudo aplicou uma abordagem quantitativa descritiva. O modelo de análise que orienta o presente estudo contempla, para cada objetivo específico, as dimensões, categorias e a respectiva base legal, conforme apresentado por meio do Quadro 1.

QUADRO 1 - MODELO DE ANÁLISE

OBJETIVOS ESPECÍFICOS	DIMENSÕES	CATEGORIAS	BASE LEGAL
Descrever a tramitação dos processos de denúncias e representações autuados no TCE/SC	Espécie	Denúncia	Lei Complementar n.202/2000; Regimento Interno- Resolução n. TC-06/2001; Instrução Normativa n. TC-0021/2015
		Representação	
Caracterizar os processos de denúncias e representações autuados em 2020 no TCE/SC	Origem	Cidadão	Lei Complementar n.202/2000; Regimento Interno- Resolução n. TC-06/2001; Instrução Normativa n. TC-0021/2015.
		Conselheiro	
		Entidades Coletivas	
		Empresas Privadas	
	Assunto	Atos de Pessoal	
		Licitações e Contratações	
		Outros assuntos	
	Unidade Gestora	Estado	
Municípios			
Analisar as decisões decorrentes dos processos de denúncias e representações e as respectivas medidas tomadas pelo TCE/SC	Decisão	Não conhecimento	Lei Complementar n.202/2000; Regimento Interno- Resolução n. TC-06/2001; Instrução Normativa n. TC-0021/2015
		Extinção do processo sem julgamento de mérito	
		Procedência	
		Improcedência	
		Em andamento	
	Medidas tomadas	Suspensão do certame/contrato	
		Aplicação de multa	
		Outras medidas- Determinação/ Recomendação/Alertas	
		Sem medidas	
		Multa e outras medidas	
Conversão do processo em tomada de contas especial			

Fonte: Elaborado pela autora (2022).

4. ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS

Foram analisados 430 processos de denúncias e representações consideradas derivadas do controle social. Analisando os processos quanto à dimensão de espécie, percebe-se que a maioria dos processos são de representações apresentadas com base na lei de licitações, representando um total de 81,86% dos processos. Em seguida, os processos de denúncias representam 9,53% da amostra, e os demais processos, originários de representações de Conselheiro, equivalem a 8,60% da amostra. Ou seja, a maioria dos processos propostos pelo controle social possuem a temática envolvendo procedimentos licitatórios.

TABELA 2 - FREQUÊNCIA DA DIMENSÃO ESPÉCIE

CATEGORIAS	TOTAL	FREQUÊNCIA RELATIVA
Representação - art. 113, § 1º, da Lei nº 8.666/93	352	81,86%
Denúncia	41	9,53%
Representação de Conselheiro	37	8,60%
Total	430	100%

Fonte: Elaborado pela autora (2022).

Verifica-se que as demandas impulsionadas a partir do controle social implicam em sua maioria no questionamento das contratações públicas o que demonstra que há uma participação mais ativa nesse tema.

Analisando os processos quanto à dimensão de origem, verifica-se que as empresas privadas correspondem ao maior número de agentes provocadores do controle externo, representando 54,19% da

amostra. Logo em seguida, os cidadãos possuem sua participação representando 32,56% do controle social. Em terceiro lugar, restam a origem “Conselheiro” que corresponde a 8,60% e abrange aquelas denúncias e representações recebidas via ouvidoria. Por fim, as entidades coletivas, assim compreendidas qualquer associações, sindicatos ou organizações sociais, representaram 4,65% do controle social.

TABELA 3 - FREQUÊNCIA DA DIMENSÃO ORIGEM

CATEGORIAS	TOTAL	FREQUÊNCIA RELATIVA
Empresas Privadas	233	54,19%
Cidadão	140	32,56%
Conselheiro	37	8,60%
Entidades Coletivas	20	4,65%
Total	430	100%

Fonte: Elaborado pela autora (2022).

Ressalta-se que os processos com origem na ouvidoria muitas vezes são feitos de forma anônima e, portanto, são assumidas pelo Conselheiro ouvidor para iniciar efetivamente um processo de fiscalização dentro do controle externo. Por outro lado, essa sistemática não permite classificar a origem própria do agente que apresentou a denúncia. Ademais, não causa estranheza que as empresas privadas figurem

com maior frequência de processos pois estão diretamente envolvidas nos processos licitatórios, vindo a apresentar impugnações no TCE/SC que não restaram acolhidas pela unidade gestora no certame.

Em seguida, foram agrupados os dados relativos à variável assunto, que pode se tratar de atos de pessoal, licitações e contratações ou demais assuntos que envolvam a Administração Pública. Verifica-se

que o assunto de maior frequência é sobre licitações e contratações, representando 88,14% da amostra. Essa identificação era esperada em congruência com a maior origem ser de empresas privadas e também a

maior espécie ser relativa a representações com base na lei de licitações. Em seguida, outros assuntos assumem a relevância de 6,74%. Por último, o assunto atos de pessoal foi frequente em 5,12% da amostra.

TABELA 4 - FREQUÊNCIA DA DIMENSÃO ASSUNTO

CATEGORIAS	TOTAL	FREQUÊNCIA RELATIVA
Licitações e Contratações	379	88,14%
Outros assuntos	29	6,74%
Atos de Pessoal	22	5,12%
Total	430	100%

Fonte: Elaborado pela autora (2022).

Posteriormente, na dimensão unidade gestora, verifica-se que a maior demanda das representações e denúncias se refere aos municípios, contando com a frequência relativa de 82,33% dos processos de controle social. Esse dado é esperado tendo em vista que há um número muito maior de atos ocorrendo nos municípios do que no Estado, que possui a frequência relativa de 17,67% nos processos analisados.

TABELA 5 - FREQUÊNCIA DA DIMENSÃO UNIDADE GESTORA

CATEGORIAS	TOTAL	FREQUÊNCIA RELATIVA
Municípios	354	82,33%
Estado	76	17,67%
Total	430	100%

Fonte: Elaborado pela autora (2022).

Diante dos resultados apresentados frequência da dimensão origem, assunto e unidade gestora, é possível estabelecer um perfil do controle social. Isto é, as empresas privadas são as que mais apresentaram denúncias e representações, e estas abordam majoritariamente o assunto de licitações e contratações que ocorrem nos diversos municípios catarinenses. Isso demonstra que a própria concorrência nas licitações e contratações faz com que as empresas privadas se movimentem para fiscalizar os certames e apresentar suas impugnações junto ao controle externo. Por outro lado, há pouca contribuição do controle social por parte das entidades coletivas e por meio do canal da ouvidoria, quando há representação por parte do Conselheiro, nos assuntos que fogem ao tema de licitações e contratações.

Esse contexto verificado vai ao encontro da literatura que afirma sobre a necessidade de aprimorar a interação entre o controle social e controle externo

de forma a fomentar os meios de atuação da sociedade. Eis que a participação do controle social se revela focada em apenas um tema sem que seja possível afirmar que exista efetividade dos preceitos que estão abarcados na democracia representativa, da *accountability*, e que se integrem com as premissas da cidadania democrática abordada no paradigma do Novo Serviço Público.

Na sequência os dados relativos à dimensão decisão demonstram que a maior parcela dos processos se encontra em andamento, mais especificamente 26,98% desses processos. Próximo a este patamar, estão os processos com extinção sem julgamento do mérito, que representam 26,74% da amostra. Na sequência, os processos improcedentes se revelam com frequência de 23,49%. Posteriormente, encontram-se os processos com procedência, que representam 15,81% dos processos. Por fim, os processos em que não houve conhecimento são de 6,98%.

TABELA 6 - FREQUÊNCIA DA DIMENSÃO DECISÃO

CATEGORIAS	TOTAL	FREQUÊNCIA RELATIVA
Em andamento	116	26,98%
Extinção do processo sem julgamento de mérito	115	26,74%
Improcedência	101	23,49%
Procedência	68	15,81%
Não conhecimento	30	6,98%
Total	430	100%

Fonte: Elaborado pela autora (2022).

Somando os grupos de ‘não conhecimento’, ‘improcedência’ e ‘extinção do processo sem julgamento de mérito’ obtém-se a porcentagem de 57,21% dos processos sem que a irregularidade tenha sido analisada ou confirmada. Assim, há um custo para movimentação da estrutura do Tribunal de Contas, sem que haja efetiva confirmação de irregularidades em mais da metade dos processos. Esta constatação não representa necessariamente um problema uma vez que há um custo da democracia. Isto é, ao cidadão deve ser permitido a movimentação das instituições por meio de denúncias e representações ainda que a suposta irregularidade noticiada não tenha se efetivada. O simples fato de haver a provocação do TCE/SC por meio do controle social já revela que há uma contribuição

por parte da sociedade na fiscalização da gestão pública. Por outro lado, pode-se também questionar em que medida o controle social está servindo como um instrumento de competição empresarial. Como evidenciado, não parte do cidadão a maioria das denúncias e representações e a maior parte dos temas envolve licitação.

Em relação à frequência das medidas tomadas, a maior parte dos processos não possui medidas tomadas em 70,93% dos casos. Nesta categoria estão os processos em andamento e também aqueles de improcedência, não conhecimento e extinção do processo sem julgamento de mérito, mas que não tiveram nenhuma medida tomada ainda pelo TCE/SC.

TABELA 7 - FREQUÊNCIA DA DIMENSÃO MEDIDAS TOMADAS

CATEGORIAS	TOTAL	FREQUÊNCIA RELATIVA
Sem medidas	305	70,93%
Outras medidas - Determinação/Recomendação/Alertas	98	22,79%
Suspensão do certame/contrato	11	2,56%
Aplicação de multa	9	2,09%
Multa e outras medidas	7	1,63%
Conversão do processo em tomada de contas especial	0	0,00%
Total	430	100%

Fonte: Elaborado pela autora (2022).

Em seguida, verifica-se que as outras medidas como determinações, recomendações e alertas, representam a frequência de 22,79% dos processos, englobando qualquer categoria de decisão. Na sequência, a medida mais frequente, no patamar de 2,56% é a suspensão de certame e/ou contrato quando o processo está em andamento e é deferida uma medida cautelar. Já a aplicação de multa foi verificada como

uma medida tomada em 2,09% dos processos, quando se verificou alguma irregularidade e a sanção foi imposta. Além disso, houve aplicação de multa em conjunto com outras medidas na frequência de 1,63% dos processos. Já em relação à conversão do processo em tomada de contas especial para que haja imputação de débito, não se verificou nenhuma ocorrência.

Conforme verificado nos dados da pesquisa, não houve nenhuma conversão em tomada de contas especial nos processos analisados. Isso implica dizer que não se verificou a ocorrência de desfalque, desvio de bens ou outra irregularidade de que resulte dano ao erário nos processos trazidos pelo controle social, e, portanto, não resultou em nenhuma imputação de débito. Então, pode-se observar a situação encontrada a partir do prisma do controle externo, que frente às demandas exerce um julgamento com poucas sanções efetivamente aplicadas; quanto a partir do prisma do controle social, que acaba trazendo denúncias e representações não qualificadas diante das medidas que poderiam ser aplicadas pelo TCE/SC.

5. CONCLUSÃO

Este trabalho possibilitou identificar as contribuições do controle social por meio de denúncias e representações no controle externo realizado pelo Tribunal de Contas do Estado de Santa Catarina. Com isso, pôde-se perceber o nível de interação destas espécies de controle incidentes sobre a gestão pública que se inserem dentro do paradigma do Novo Serviço Público e das noções de *accountability* diagonal. Os principais resultados da análise se pautam na observação de que a maioria dos processos propostos pelo controle social possuem a temática envolvendo procedimentos licitatórios; são apresentadas majoritariamente por empresas privadas, em primeiro lugar, e em segundo lugar pelos cidadãos cidadãos; e se referem à esfera municipal em detrimento da esfera estadual.

Paralelamente, identificou-se que a maior parcela dos processos se encontra em andamento, isto é, ainda sem uma decisão definitiva, considerando o lapso de análise de aproximadamente um ano após a autuação. Em seguida, restam de maior representatividade os processos com extinção sem julgamento do mérito e processos improcedentes. Isto significa que há um custo para movimentação da estrutura do Tribunal de Contas, sem que haja efetiva confirmação de irregularidades em mais da metade dos processos.

Em consonância com esta realidade, percebe-se que a maior parte dos processos não possui medidas tomadas. Dentre aqueles processos com medidas tomadas, majoritariamente aplicam-se determinações, recomendações e alertas. Em menor frequência a me-

didada de suspensão de certame e/ou contrato é utilizada e também aplicação de multa, isoladamente ou em conjunto com outras medidas.

Desta forma, conclui-se que a interação entre o controle social e o controle externo, a partir da análise das denúncias e representações autuadas em 2020 no TCE/SC, são demandas que tratam de processos licitatórios que, em sua maioria, acabam sendo improcedentes, extintos ou não conhecidos. Tais constatações podem reforçar a tese de que os procedimentos de denúncia e representação podem estar sendo usados como mecanismos de competição empresarial.

Verifica-se que há uma grande espaço para o avanço de políticas públicas para o fomento da atividade de controle social e também de simplificação e maior abertura do controle externo diante da participação popular em suas atividades. Ainda, percebe-se que apesar da licitação ser um grande alvo do controle social, as medidas de fiscalização do contrato, em relação à sua execução em si, no que tange aos serviços propriamente prestados, não aparece como alvo das denúncias e representações do controle social. Diante desse cenário, vislumbra-se uma possibilidade de avanço na interação do controle social junto ao controle externo para que haja uma dinâmica mais incisiva no fornecimento de ferramentas na efetiva fiscalização sobre as rotinas diárias da Administração Pública. Apesar de positiva a possibilidade da interação entre o controle social e o controle externo, observa-se que ela ainda é limitada, necessitando de maior desenvolvimento de ações – educativas, de apoio etc - para que os cidadãos possam exercê-la da forma apropriada.

Paralelamente, sugere-se para estudos futuros a construção de um conjunto de indicadores de *accountability* diagonal, para que hajam métricas que possam ser monitoradas para fortalecer esta realidade. Ao mesmo tempo, sugere-se para investigações futuras a formulação de uma metodologia para identificar a expansão do controle social perante a atividade de controle externo, cabendo aplicar o mesmo modelo de análise utilizado no presente estudo como uma ferramenta replicável em outras esferas e outros tribunais a partir dos dados fornecidos pela instituição que devem ser públicos e até mesmo fazer parte da política de dados abertos. Também é possível pesquisar sobre as expectativas dos atores que se comunicam com as atividades de controle externo e qual o retorno esperado para o controle social, apontando sua dimensão educativa e orientativa.

REFERÊNCIAS

- ALENCAR, Leandro Zannoni Apolinário de. *O novo Direito Administrativo e governança pública*: responsabilidade, metas e diálogo aplicados à Administração Pública do Brasil. Belo Horizonte: Fórum, 2018.
- AZEVEDO, Pedro Henrique Magalhães. *Participação popular e controle externo*: análise quantitativa das denúncias formuladas e apreciadas pelo Tribunal de Contas do Estado de Minas Gerais no período de 2009 a 2014. *Revista do Tribunal de Contas do Estado de Minas Gerais*. [s.l.: s.n.]. Disponível em: <http://www.renatojanine.pro.br/FiloPol/versus>. Acesso em: 8 fev. 2021.
- BALBE, Ronald da Silva. *Controle Interno e o Foco nos Resultados*. 1. ed. Belo Horizonte: Fórum, 2013.
- BRAGA, Marcus Vinícius de Azevedo. *Tudo Sobre Controle*. 1a ed. Belo Horizonte: Fórum, 2020.
- BRASIL. *Constituição Federal de 1947*. 1947. Disponível em: <https://www2.camara.leg.br/legin/fed/consti/1940-1949/constituicao-1946-18-julho-1946-365199-publicacaooriginal-1-pl.html>. Acesso em: 20 dez. 2020.
- CASTRO, Davi Assunção Salvador Nery De. *A atuação articulada do controle social e do controle externo e o fortalecimento do controle da administração pública*. 2016. IDP/ EDB, [S. l.], 2016. Disponível em: <https://repositorio.idp.edu.br/handle/123456789/2266>. Acesso em: 8 fev. 2021.
- COUTINHO, Doris de Miranda. *Prestação de contas do governo*. Belo Horizonte: Fórum, 2020.
- DENHARDT, Janet V.; DENHARDT, Robert B. *The new public service: Serving, Not Steering*. 4a ed. NOVA YORK & LONDRES: Routledge, 2015.
- FRANÇA, Phillip Gil. *Controle da administração pública*. 4a edição ed. São Paulo: Saraiva, 2016.
- IOCKEN, Sabrina Nunes. *Controle Compartilhado das Políticas Públicas*. Belo Horizonte: Editora Fórum, 2018.
- LIMA, Henrique. *Controle Externo - Teoria e Jurisprudência para os Tribunais de Contas*. [s.l.] : Grupo GEN, 2019. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9788530987077/cfi/6/38!/4/2/4@0:90.5>. Acesso em: 26 jan. 2021.
- MARQUES PEREIRA, Sandra Patrícia; ALVES RIBEIRO CORREIA, Pedro Miguel. Movimentos Pós-Nova Gestão Pública: o Novo Serviço Público. *Lex Humana*, v. 12, n. 1, p. 69–85, 2020.
- O'DONNELL, G. Horizontal *accountability* and new polyarchies. *Working Paper of the Helen Kellogg Institute for International Studies*, n. 253, p. 24, 1998.
- PELIZZO, Riccardo; STAPENHURST, Frederick. Government *accountability* and legislative oversight. *Government Accountability and Legislative Oversight*, n. November, p. 1–111, 2013. DOI: 10.4324/9781315850610.
- ROCHA, Diones Gomes da; ZUCCOLOTTO, Robson; TEIXEIRA, Marco Antonio Carvalho. Insulados e não democráticos: a (im)possibilidade do exercício da social *accountability* nos Tribunais de Contas brasileiros. *Revista de Administração Pública*, v. 54, n. 2, p. 201–219, 2020.
- SANTOS, Jair Lima. O TCU e os controles estatal e social da Administração Pública. *Revista do TCU*, n. 94, 2002.
- SANTOS, Paloma Maria; SELIG, Paulo Maurício. Indicadores para o novo serviço público: uma análise bibliométrica e sistêmica. *Perspectivas em Ciência da Informação*, v. 19, n. 3, p. 82–97, 2014.
- SERRA, Rita de Cássia Chió; CARNEIRO, Ricardo. As Interfaces Entre O Controle Social E Os Controles Externo E Interno Na Administração Pública Brasileira. In: XXXV ENCONTRO DA ANPAD 2011, Rio de Janeiro. *Anais* [...]. Rio de Janeiro: EnANPAD, 2011.
- SIMÕES, Edson. *Tribunais de contas : controle externo das contas públicas*. 1a ed. São Paulo: Saraiva, 2014.

TCE/SC. *Regimento interno do tribunal de contas de Santa Catarina* – Resolução n. Tc-06/2001. 2001. Disponível em: http://www.tce.sc.gov.br/sites/default/files/regimento_interno_consolidado_08112018.pdf Acesso em: 20 dez. 2020.

TCE/SC. *Lei Complementar n. 202/2000*. 2000. Disponível em: <https://www.tcsc.tc.br/sites/default/files/LEI-ORGANICA-CONSOLIDADA.pdf>. Acesso em: 20 dez. 2020.

TCE/SC. *Instrução Normativa n. TC-0021/2015*. 2015. Disponível em: https://www.tcsc.tc.br/sites/default/files/leis_normas/INSTRU%C3%87%C3%83O%20NORMATIVA%20N%2021-2015%20CONSOLIDADA.pdf. Acesso em: 20 dez. 2020.

TCE/SC. *Relatórios de Atividades :: 2020* | Tribunal de Contas SC. 2020. Disponível em: <http://www.tce.sc.gov.br/relatorio-atividades/2020>. Acesso em: 20 dez. 2020.

WILLEMANN, Marianna Montebello. *Accountability Democrática e O Desenho Institucional dos Tribunais de Contas No Brasil*. Belo Horizonte: Fórum, 2019.



Fernanda Camila De Carli

fe.camilla@gmail.com

Udesc

<https://orcid.org/0000-0003-1193-0257>

Mestrado em Administração pela Universidade do Estado de Santa Catarina (2022), graduação em Direito pela Universidade Federal de Santa Catarina (2014), e atualmente ocupa o cargo de Auditora Fiscal de Controle Externo no Tribunal de Contas do Estado de Santa Catarina (2018).



Fabiano Maury Raupp

fabianoraupp@hotmail.com

Udesc

<https://orcid.org/0000-0001-9533-2574>

Doutorado em Administração pela Universidade Federal da Bahia (2011), Mestrado em Administração pela Universidade Federal de Santa Catarina (2003) e Graduação em Ciências Contábeis pela Universidade Federal de Santa Catarina (2001). É professor associado na Universidade do Estado de Santa Catarina, atuando no Departamento de Administração Empresarial, no Mestrado Profissional em Administração, no Mestrado Acadêmico em Administração e no Doutorado em Administração do Centro de Ciências da Administração e Socioeconômicas. É autor de livros, capítulos de livros, artigos em periódicos e trabalhos socializados em anais de eventos científicos.

Dados abertos: uma análise da Empresa Brasileira de Serviços Hospitalares e de sua rede de hospitais universitários federais¹

Open data: an analysis of the Empresa Brasileira de Serviços Hospitalares and its network of federal university hospitals

Datos abiertos: un análisis de la Empresa Brasileira de Servicios Hospitalares y su red de hospitales universitarios federales

Gabriela de Oliveira Silva, Renata Mendes de Araujo e Fernando Filgueiras

<https://doi.org/10.36428/revistadacgu.v14i26.527>

Resumo: Dados abertos têm sua relevância reconhecida mundialmente em diferentes setores para maior transparência, controle social, eficiência de serviços, inovação e crescimento econômico, com grande potencial na área de saúde pública. No Brasil, a Política Nacional de Dados Abertos (PNDA) busca contribuir para que esses benefícios sejam alcançados, porém ainda há dificuldades em sua implantação, de modo que os dados publicados pelas instituições públicas não atendem a todos os princípios para serem considerados abertos. Com este estudo, buscou-se analisar o grau de alinhamento da Empresa Brasileira de Serviços Hospitalares (Ebserh) e de sua rede de hospitais universitários federais aos princípios da PNDA. O levantamento dos dados abertos pelas 39 unidades da rede Ebserh foi realizado no portal Gov.Br, registrando os tipos de dados e suas características. O grau de alinhamento da rede Ebserh aos princípios de dados abertos foi calculado de acordo com a pontuação obtida pelas unidades avaliadas nas questões definidas segundo a PNDA. Verificou-se que somente 21 das unidades da rede Ebserh (51%) publicam dados abertos, totalizando 854 conjuntos de dados. O grau de alinhamento identificado foi de 35%, considerando toda a rede Ebserh, e de 67%, considerando somente as unidades da rede que publicam dados abertos. Observa-se que as unidades da rede Ebserh encontram-se em níveis de maturidade distintos quanto à abertura de dados, sendo que nenhuma unidade atende a todos os princípios PNDA, o que pode dificultar a busca, reutilização e interoperabilidade dos dados. A qualidade dos dados abertos da rede precisa ser melhorada para que possam, de fato, serem úteis para diferentes propósitos de reuso. Aponta-se para a importância do reconhecimento dos dados como um ativo estratégico na rede Ebserh, do esclarecimento das questões quanto à Lei Geral de Proteção de Dados Pessoais e do estabelecimento de uma política de governança de dados.

Palavras-chave: Dados abertos. Governança de Dados. Hospitais Universitários. Proteção de Dados. Ciência de Dados.

Abstract: Open data has its relevance recognized worldwide in different sectors for greater transparency, social control, service efficiency, innovation, and economic growth, with great potential in public health. In Brazil, the National Open Data Policy (PNDA) seeks to contribute to these benefits being achieved, but there are still difficulties in its implementation, so that the data published by public institutions do not meet all the principles to be considered open. With this study, we sought to analyze the degree of alignment of the *Empresa Brasileira de Serviços Hospitalares* (Ebserh) and its network of federal university hospitals with the principles of the PNDA.

1. Artigo submetido em 26/07/2022 e aceito em 01/12/2022.

The survey of open data published by the 39 units of the Ebserh network was carried out on the Gov.Br portal, recording the types of data and their characteristics. The degree of alignment of the Ebserh network with the principles of open data was calculated according to the score obtained by the units evaluated in the questions defined according to the PNDA. It was found that only 21 of the Ebserh network units (51%) publish open data, totaling 854 data sets. The degree of alignment identified was 35%, considering the entire Ebserh network, and 67%, considering only the network units that publish open data. It appears that the units of the Ebserh network are at different maturity levels regarding data opening, and no unit meets all the PNDA principles, which can make it difficult to search, reuse, and interoperate data. The quality of open network data needs to be improved so that it can be useful for different reuse purposes. It points to the importance of recognizing data as a strategic asset in the Ebserh network, clarifying issues regarding the General Law for the Protection of Personal Data and establishing a data governance policy.

Keywords: Open Data. Data Governance. University Hospitals. Data Protection. Data Science.

Resumen: Los datos abiertos tienen su relevancia reconocida a nivel mundial en diferentes sectores para una mayor transparencia, control social, eficiencia del servicio, innovación y crecimiento económico, con gran potencial en el área de la salud pública. En Brasil, la Política Nacional de Datos Abiertos (PNDA) busca contribuir para que se alcancen estos beneficios, pero aún existen dificultades en su implementación, de modo que los datos publicados por las instituciones públicas no cumplen con todos los principios para ser considerados abiertos. Con este estudio, buscamos analizar el grado de alineación de la *Empresa Brasileira de Serviços Hospitalares* (Ebserh) y su red de hospitales universitarios federales con los principios del PNDA. El relevamiento de los datos abiertos por las 39 unidades de la red Ebserh fue realizado en el portal Gov.Br, registrando los tipos de datos y sus características. El grado de alineamiento de la red Ebserh con los principios de los datos abiertos se calculó según la puntuación obtenida por las unidades evaluadas en las preguntas definidas según el PNDA. Se encontró que solo 21 de las unidades de la red Ebserh (51%) publican datos abiertos, totalizando 854 conjuntos de datos. El grado de alineamiento identificado fue del 35%, considerando toda la red Ebserh, y del 67%, considerando solo las unidades de red que publican datos abiertos. Se observa que las unidades de la red Ebserh se encuentran en diferentes niveles de madurez en cuanto a apertura de datos, y ninguna unidad cumple con todos los principios del PNDA, lo que puede dificultar la búsqueda, reutilización e interoperabilidad de datos. Es necesario mejorar la calidad de los datos abiertos de la red para que puedan ser realmente útiles para diferentes propósitos de reutilización. Señala la importancia de reconocer los datos como un activo estratégico en la red Ebserh, aclarando aspectos de la Ley General de Protección de Datos Personales y estableciendo una política de gobernanza de datos.

Palabras clave: Datos abiertos. Gobernanza de Datos. Hospitales Universitarios. Protección de Datos. Ciencia de los Datos.

1. INTRODUÇÃO

Segundo a *Open Knowledge Foundation*, dados abertos são aqueles que qualquer pessoa pode usar, reutilizar e compartilhar livremente para qualquer propósito, bastando, no máximo, que credite a sua autoria e os compartilhe com a mesma licença aberta (OKF, 2015). Quando esses dados são gerados, coletados ou custodiados pela administração pública e disponibilizados em formato aberto, são denominados dados abertos governamentais (DAG) (TCU, 2015).

A importância dos DAG tem sido reconhecida mundialmente em diversos setores para maior transparência, *accountability* e eficiência de serviços, avaliação de políticas públicas, desenvolvimento de novas tecnologias, inovação, crescimento econômico e agregação de valor público (Borgesius et al., 2015; Cruz-Rubio, [s.d.]; D'Agostino et al., 2018; Pires, 2015). Na área de saúde, DAG possuem potencial para reduzir lacunas de informação, fortalecer pesquisas científicas e tradução do conhecimento, favorecer a formulação de políticas públicas baseadas em evidências, e melhor prever ou detectar precocemente possíveis surtos e epidemias (D'Agostino et al., 2017; Heijlen & Crom-

pvoets, 2021). Em emergências de saúde pública, a exemplo das associadas ao zika vírus e ao novo coronavírus, DAG auxiliaram no mapeamento da disseminação dos vírus e de casos das doenças, na identificação de potenciais riscos e na sua prevenção (OKBr, 2021a; Santos et al., 2017).

No Brasil, a Política Nacional de Dados Abertos (PNDA) visa promover a publicação de DAG, possibilitar o acesso pela sociedade a esses dados, aprimorar a cultura de transparência pública e o controle social, facilitar o intercâmbio de dados entre diferentes entidades, fomentar a pesquisa científica de base empírica sobre a gestão pública e promover a oferta de serviços públicos digitais de forma integrada (Decreto n. 8.777, 2016). Tal Política faz parte das estratégias para o governo aberto e congrega do seu princípio de tratar o acesso à informação como regra e o sigilo como exceção. Assim, também se associa aos princípios constitucionais de acesso à informação e da publicidade de atos administrativos, regulados pela Lei de Acesso à Informação (LAI) (Lei n. 12.527, 2011).

A gestão da PNDA é coordenada pela Controladoria Geral da União (CGU) por meio da Infraestrutura Nacional de Dados Abertos (INDA) com um “mecanismo de governança multiparticipativa, transparente, colaborativa e democrática” (Decreto n. 8.777, 2016). Sua implantação é executada por meio de Planos de Dados Abertos (PDA) elaborados por cada instituição pública do Poder Executivo Federal, que orientam as ações relacionadas à implementação e promoção da abertura de dados, a fim de facilitar a sua compreensão e reutilização.

No âmbito da Empresa Brasileira de Serviços Hospitalares (Ebserh) e da sua rede de hospitais universitários federais (HUF), a institucionalização da PNDA pode trazer benefícios para a própria rede, a administração pública de forma geral, outras entidades e a sociedade. Criada em 2011, essa empresa pública é vinculada ao Ministério da Educação (MEC) e possui contratos de gestão firmados com 32 instituições federais de ensino superior para a gestão de 40 HUF com atendimento integral ao Sistema Único de Saúde (SUS) (Lei n. 12.550, 2011). Trata-se de uma rede complexa que associa processos de gestão, assistência à saúde, ensino, extensão, pesquisa e inovação, utilizando diversos tipos de plataformas e tecnologias, com elevada coleta, gerenciamento e produção de dados.

A publicação de DAG pode favorecer a organização de dados nas unidades e o benchmarking entre os HUF. Pode também contribuir para a transparência e o controle social sobre os indicadores de desempenho dos planos de reestruturação dos HUF, para a avaliação da qualidade da prestação de serviços, da prática de ensino, do desenvolvimento de pesquisas, da geração de inovação e do impacto dos contratos de gestão estabelecidos com a Ebserh. De forma mais abrangente, pode contribuir para a ampliação do conhecimento sobre as políticas de ensino e o SUS e possibilitar a identificação de possíveis problemas e soluções por outras entidades e pesquisadores. Outras plataformas podem também ser desenvolvidas com o uso dos dados, com a demonstração de indicadores, quer seja apoiando a gestão ou a criação de novos produtos e serviços (Filgueiras, Palotti & Nascimento, 2022). A publicação de DAG também gera mais confiança das partes interessadas e melhora a confiabilidade dos dados, além de ser uma exigência legal pela LAI para todas as instituições públicas (Campagnucci, 2021).

Apesar de haver legislação já estabelecida e dos benefícios dos DAG, ainda é incipiente o entendimento quanto à abertura de dados na administração pública, tendo em vista que muitos dados não são publicados adequadamente (Avelino et al., 2021; Pires, 2019; Silva & Domínguez, 2017). Além da falta de clareza quanto ao processo integral de gestão dos dados, ao conceito e aos princípios de DAG, há dificuldades de compreensão ou interpretações equivocadas da Lei Geral de Proteção de Dados Pessoais (LGPD) que dificultam a abertura de dados governamentais no Brasil (Campagnucci & Belisário, 2021; Langenegger, 2021).

Nesse sentido, estudos têm sido conduzidos para definir métricas e indicadores para avaliar o alinhamento de repositórios desses dados aos princípios de DAG e monitorar a transparência e reuso de DAG (Nascimento-Silva & Kerr-Pinheiro, 2019; Silva, 2018). Cardoso et al. (2020), avaliando repositórios de dados de sistemas de saúde pública e prisionais, identificaram um grau de alinhamento de 39% do repositório nacional de dados do SUS (DATASUS) aos princípios de DAG, valor bem inferior ao alcançado por repositórios internacionais, como o da cidade de Nova York (100%). Pelo Índice de Transparência da Covid-19, a *Open Knowledge Brasil* (2021a) verificou que 90% dos entes governamentais avaliados em 04/2020 não publicavam dados suficientes para monitorar a propagação da Covid-19, sendo identificados problemas

quanto aos DAG nas diferentes esferas (omissão e inconsistência de dados, dificuldade de acesso a bases de dados, falta de padronização, não disponibilização de microdados etc.). As evidências serviram de base para melhorias nos padrões dos DAG e, assim, no monitoramento da pandemia. Análises de outras instituições quanto à abertura de dados podem também contribuir para detecção de falhas e oportunidades de melhorias.

Com este estudo, objetivou-se analisar o grau de alinhamento da rede Ebserh aos princípios dos DAG, utilizando como parâmetro questões definidas de acordo com os princípios da PNDA. Buscou-se, assim, verificar quais dados são abertos pelas unidades da rede e se estes podem ser considerados, de fato, DAG. Por meio dessa análise, inicia-se a discussão sobre a importância do reconhecimento dos DAG como um ativo estratégico, de estabelecimento de uma política de governança de dados e do clareamento das questões quanto à LGPD.

2. DELIMITANDO CONCEITOS

Os DAG se associam a dois grandes movimentos da atualidade que apresentam interface com a Ebserh, entidade da administração pública federal e gestora de uma rede de HUF envolvidos com o ensino, a pesquisa e a inovação em saúde: o da ciência aberta e do governo aberto (Santos et al., 2017). Embora liderados por frentes distintas, esses movimentos se integram e conjugam alguns de seus objetivos. Não se pretende apresentar todo o histórico desses movimentos, mas, sim, delimitar os seus conceitos, a fim de demonstrar que os DAG se inserem em um contexto mais amplo.

Ciência aberta envolve o movimento, vanguardado pela Iniciativa de Acesso Aberto de Budapeste (*Budapest Open Access Initiative/BOAI*), para tornar todo o processo da pesquisa científica, os dados e sua disseminação acessíveis para todos os níveis da sociedade, permitindo o compartilhamento do conhecimento entre comunidade científica, governo, empresas e cidadãos, e a integração de diferentes atores para a produção e o desenvolvimento de inovações (Santos et al., 2017). Dados abertos figuram-se como um dos elementos da ciência aberta, envolvendo diferentes tipos de dados, incluindo tanto dados de pesquisa quanto DAG (FOSTER, 2015; OKF, [s.d.]).

No Brasil, há iniciativas para os diferentes eixos da ciência aberta, mas ainda se carece de uma política integrada nacional específica. Uma dessas iniciativas

é a Parceria para Governo Aberto (*Open Government Partnership*), da qual o Brasil faz parte desde 2011, a fim de difundir e fomentar práticas governamentais para maior transparência, participação social, accountability e inovação, trazendo o cidadão como foco central para as políticas públicas. Uma das obrigações dos parceiros é publicar Planos de Ação Nacional para Governo Aberto, sendo que o Brasil já está no 5º Plano de Ação Nacional para Governo Aberto, que traz, dentre seus 12 compromissos para 2021 a 2022, a implantação de novos mecanismos de avaliação para o avanço da ciência aberta, assim como vários compromissos relacionados à melhoria da qualidade dos DAG em diferentes setores (CGU, 2021). A PNDA surgiu em decorrência das iniciativas dessa parceria, assim como a Política Nacional de Governo Aberto, que traz, como uma de suas diretrizes, o aumento da disponibilidade de dados e informações sobre as atividades governamentais (Decreto n. 10.160, 2019).

Apesar de os conceitos de DAG e governo aberto estarem relacionados, eles não são intercambiáveis (Cruz-Rubio, 2015). DAG podem ser identificados como um dos mecanismos utilizados para maior transparência e *accountability* em um governo aberto, mas este último vai muito além do uso desses dados para tal. Na perspectiva proposta por Meijer et al. (2012), em que o governo aberto envolve a conexão das dimensões de visão (transparência, acesso aberto às informações governamentais) e voz (participação, acesso aberto às arenas de tomadas de decisão), os DAG podem ser entendidos como elementos da primeira dimensão, que proporcionam, de forma ativa, acesso aberto às informações governamentais e que, de forma sinérgica, influenciam a segunda dimensão.

Segundo a PNDA, dados abertos são:

dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte (Decreto n. 8.777, 2016).

Suportam a fundamentação desse conceito, princípios e diretrizes estabelecidos na PNDA que estão relacionados aos oito princípios fundamentais dos DAG: (1) completos, todos os dados públicos (que não tenham restrições de privacidade, segurança ou de acesso) devem ser disponibilizados; (2) primários, coletados na fonte, com o maior nível possível

de granularidade; (3) atualizados, disponibilizados o mais prontamente possível para preservar o seu valor; (4) acessíveis, devendo estar disponíveis para a mais ampla gama de usuários e propósitos; (5) processáveis por máquina, devendo ser razoavelmente estruturados para permitir o processamento automatizado; (6) acesso não discriminatório, disponíveis para qualquer pessoa, sem necessidade de registro; (7) formato não proprietário, sobre o qual nenhuma entidade tem controle exclusivo; e (8) sem licença, não devendo estar sujeitos a regulamento de direitos autorais, patentes, marcas registradas ou segredos comerciais (restrições razoáveis de privacidade, segurança e privilégios podem ser permitidas) (O'Reilly et al., 2007).

Há ainda as três leis em que se pautam os DAG, propostas por David Eaves (2009): (1) se o dado não pode ser encontrado e indexado na Internet, ele não existe; (2) se não estiver disponível em formato aberto e legível por máquina, ele não pode ser reaproveitado; e (3) se algum dispositivo legal não permitir seu reuso, ele não é útil.

Os DAG podem ser conectados e permitir maior articulação, inovação e cooperação entre diferentes atores, gerando ganhos potenciais para ampliar o entendimento de problemas, contribuir para o encontro de soluções e para apoiar a tomada de decisão. Dados conectados (*linked data*) são definidos como um conjunto de boas práticas para publicação e conexão de dados estruturados na Internet a fim de se criar uma Rede de Dados (Bizer et al., 2006).

DAG conectados seguem os princípios propostos por Tim Berners-Lee (2006), conhecidos como sistema de 5 estrelas, que classifica o grau de abertura dos dados, sendo o maior número de estrelas representativo da maior abertura e maior facilidade de conexão dos dados:

- ★ disponíveis na web e com uma licença aberta;
- ★★ estruturados legíveis por máquina;
- ★★★ em formato não proprietário;
- ★★★★ em padrões abertos; e
- ★★★★★ conectados a outros gerando um contexto.

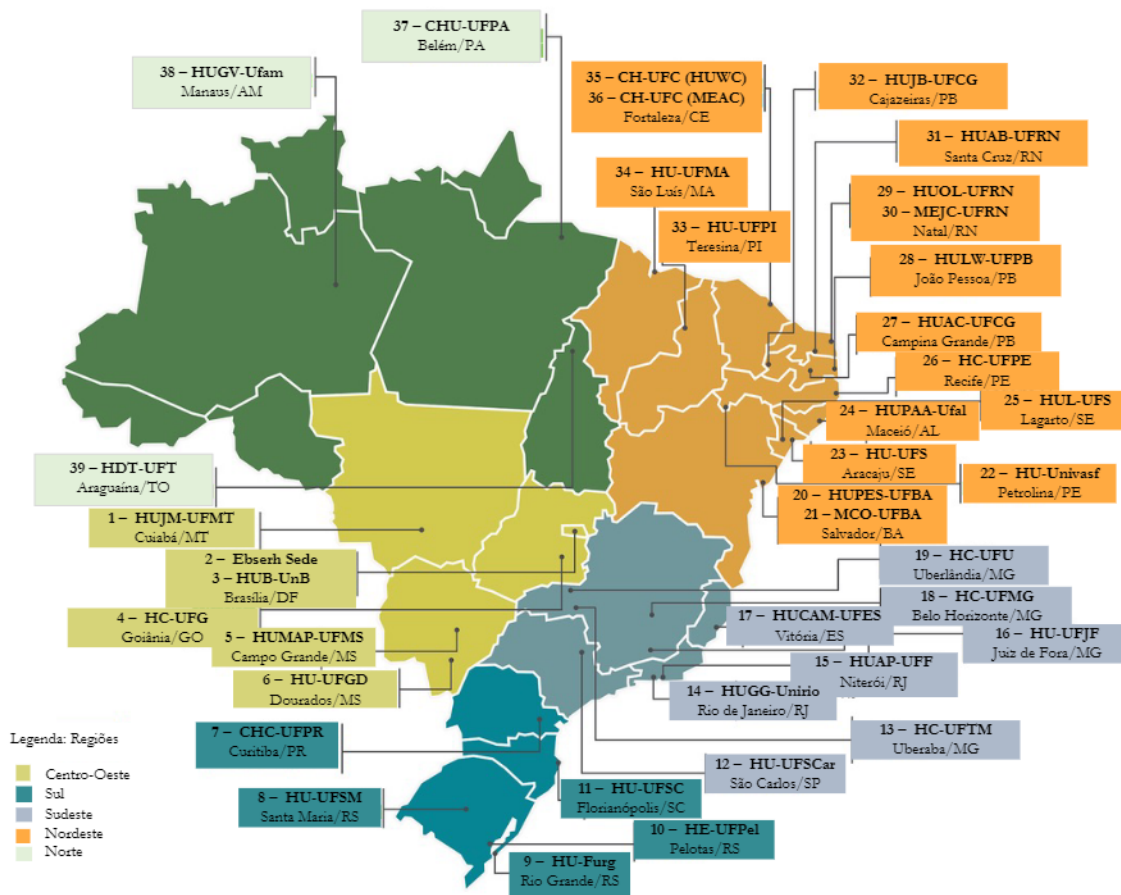
3. MÉTODO

Trata-se de uma pesquisa exploratória, documental, com abordagem quali e quantitativa que busca responder a seguinte questão de pesquisa: qual o grau de alinhamento da rede Ebserh aos princípios da PNDA?

3.1. Amostra

Como rede Ebserh, considerou-se a administração central da Ebserh (Sede), os 36 HUF e os dois complexos hospitalares, totalizando 39 unidades analisadas (Figura 1).

FIGURA 1 - UNIDADES DA REDE EBSERH POR REGIÃO DO BRASIL



Fonte: adaptado de Ebsersh (2022).

3.2. Fonte de dados

Utilizou-se, como fonte de dados, a página da Ebsersh: <https://www.gov.br/ebsersh/pt-br>. A escolha dessa fonte deve-se ao fato de o portal Gov.Br ser o portal único do governo federal, estabelecido em 2019 para disponibilizar as informações e serviços públicos de maneira centralizada (Decreto n. 9.756, 2019). Esse portal integra um dos elementos da plataforma Gov.br, que busca facilitar o acesso e promover integrações mais efetivas entre governo e sociedade, por meio da “plataformização” da estrutura governamental no âmbito do governo digital (Filgueiras et al., 2022).

O portal único permitiu que os diferentes sites dos HUF fossem integrados à página da Ebsersh Sede e que fosse estabelecida uma estrutura padronizada para a publicação dos conteúdos de todas as unidades da rede. As páginas de todas as unidades contam com caixas de conteúdo sobre Governança, Ensino e Pes-

quisa, Saúde, Comunicação e Acesso à Informação. Nessa última, há um tópico padronizado que dá acesso à página “Dados Abertos”. Assim, por meio da página da Ebsersh, foram acessadas as páginas de acesso à informação da Sede e de cada HUF e levantados os tipos de DAG publicados e suas características.

3.3. Coleta de dados

A coleta de dados foi realizada em 05/2022, manualmente por uma das autoras, com registro em planilha estruturada com as seguintes categorias: região, nome da unidade, sigla, descrição dos conjuntos de DAG (conforme informado no portal), período dos dados, formato do arquivo, fonte, responsável pelos dados, contato, período do PDA da instituição e endereços eletrônicos para acesso ao plano e aos DAG.

3.4. Categorização e análise dos dados

Cada conjunto de dados foi categorizado segundo a dimensão de processos da rede Ebserh a que correspondem os DAG: gestão hospitalar, assistência à saúde ou ensino, extensão, pesquisa e inovação.

Em seguida, os DAG identificados foram analisados quanto ao conteúdo do arquivo, correspondência do conteúdo ao informado na página, completude ou falta de dados, clareza da descrição, atualização, padronização e granularidade.

3.5. Análise do grau de alinhamento

A partir do levantamento realizado, procedeu-se à análise do grau de alinhamento da rede Ebserh aos princípios da PNDA. Para cada princípio ou diretriz da PNDA, foram definidas questões para proceder a tal análise, utilizando como base a pesquisa de Cardoso et al. (2020) com adaptações ao objeto deste estudo (**Quadro 1**).

QUADRO 1 – QUESTÕES DE ANÁLISE DO ALINHAMENTO DA REDE EBSERH AOS PRINCÍPIOS DA PNDA

PRINCÍPIOS DA PNDA	QUESTÃO DE ANÁLISE
1. observância da <u>publicidade das bases de dados</u> como preceito geral e do sigilo como exceção	1.1. O PDA está disponível? 1.2. Há DAG publicados? 1.3. Há disponibilização de DAG referentes às dimensões de processos de assistência, gestão, ensino, extensão, pesquisa e inovação?
2. garantia de <u>acesso irrestrito</u> às bases de dados, as quais devem ser legíveis por máquina e estar disponíveis em formato aberto	2.1. O acesso aos DAG é livre, sem necessidade de identificação ou registro? 2.2. Os arquivos são disponibilizados em formatos legíveis por máquina? 2.3. Os arquivos estão disponíveis em formato aberto (não proprietário)?
3. descrição das bases de dados, com <u>informação suficiente</u> para a compreensão de eventuais ressalvas quanto à sua qualidade e integridade	3.1. Há descrição clara do que compõem os conjuntos de dados (título, descritivo, formato do arquivo, período dos dados)? 3.2. Há indicação da periodicidade de publicação/atualização dos dados?
4. <u>permissão irrestrita de reuso</u> das bases de dados publicadas em formato aberto	4.1. Os dados estão livres de direitos autorais, patentes e marcas comerciais?
5. <u>completude e interoperabilidade</u> das bases de dados, as quais devem ser disponibilizadas em sua <u>forma primária</u> , com o <u>maior grau de granularidade</u> possível, ou <u>referenciar as bases primárias</u> , quando disponibilizadas de forma agregada	5.1. As fontes dos dados são informadas? 5.2. Os dados são disponibilizados em sua forma primária ou, quando disponibilizados de forma agregada, há indicação da fonte primária? 5.3. Os conjuntos de dados estão disponíveis de forma completa? 5.4. Há uma estrutura padrão de conjuntos de dados do mesmo tipo para permitir a interoperabilidade?
6. <u>atualização periódica</u> , de forma a garantir a perenidade dos dados, a padronização de estruturas de informação e o valor dos dados à sociedade e atender às necessidades de seus usuários	6.1. Os dados são atualizados periodicamente?
7. designação clara de responsável pela publicação, atualização, evolução e manutenção de cada base de dados aberta, incluída a <u>prestação de assistência</u> quanto ao uso de dados.	7.1. O responsável pelos dados é informado de forma clara? 7.2. O contato desse responsável é disponibilizado?

Fonte: elaboração própria (2022) com princípios e diretrizes extraídos do Decreto n. 8.777 (2016).

As questões 1.1, 1.2, 2.1 e 4.1 foram analisadas de forma geral para cada unidade da rede, conferindo valor 1 para respostas afirmativas e valor 0 para respostas negativas. Para a questão 1.3, foi concedido o valor 0,33 para cada dimensão atendida, totalizando 0,99. As demais questões foram avaliadas para cada conjunto de dados divulgado, calculando-se, em seguida, a proporção em relação ao total de conjuntos de dados da unidade.

Na sequência, calculou-se o somatório dos valores obtidos nas questões por cada unidade, dividindo-se pelo número total de questões e multiplicando por 100, para se obter o grau de alinhamento de cada unidade aos princípios da PNDA. A partir da média dos valores das unidades, identificou-se o grau de alinhamento geral da rede Ebserh. Para cada uma das 16 questões de análise definidas, verificou-se ainda o percentual de atendimento da rede (n=39) ao princípio ou diretriz. De modo a verificar o alinhamento, em separado, das unidades que já publicam DAG, calculou-se também o grau de alinhamento somente dessas unidades (n=20), isolando as unidades que ainda não possuem DAG.

3.6. Apresentação dos resultados

Os resultados estão apresentados em formato de tabelas e gráfico, seguidos de narrativa, no tópico a seguir. Os *insights* gerados a partir da análise do grau de alinhamento identificado constam no tópico “Recomendações” e visam o alinhamento à PNDA pela rede Ebserh.

4. RESULTADOS E DISCUSSÃO

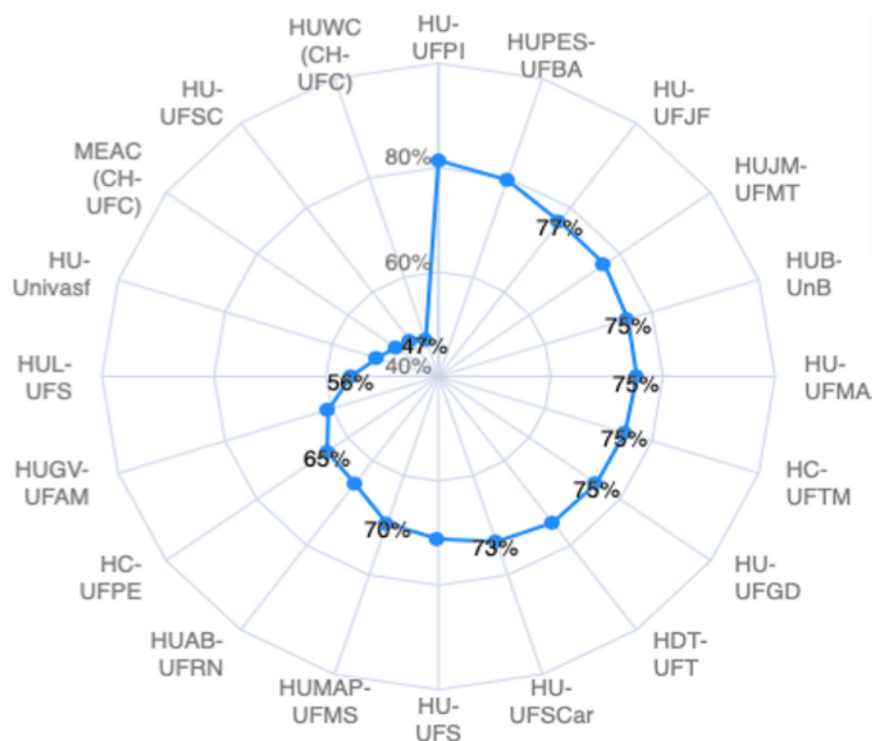
O grau identificado de alinhamento da rede Ebserh aos princípios da PNDA foi de 35%, considerando toda a rede, e de 67%, considerando somente as unidades que publicam DAG (Tabela 1). Verifica-se, pois, um grau de alinhamento baixo da rede aos princípios da PNDA, inclusive dentre aquelas unidades que já adotaram PDA. Observa-se uma variação entre o grau de alinhamento entre as unidades, sendo que algumas se afastam mais do atendimento total aos princípios da PNDA (Figura 2).

TABELA 1 – GRAU DE ALINHAMENTO DA REDE EBSERH AOS PRINCÍPIOS DA PNDA

PRINCÍPIOS	QUESTÃO	GRAU DE ALINHAMENTO			
		TODA A REDE		UNIDADES QUE PUBLICAM DAG	
Publicidade das bases de dados	1.1	46%	48%	90%	93%
	1.2	51%		100%	
	1.3	46%		90%	
Acesso irrestrito, legíveis por máquina, formato aberto	2.1	51%	47%	100%	92%
	2.2	46%		91%	
	2.3	44%		85%	
Descrição das bases	3.1	30%	18%	59%	34%
	3.2	5%		10%	
Permissão irrestrita de reuso	4.1	51%	51%	100%	100%
Completude e interoperabilidade, forma primária ou referência às bases primárias, maior grau de granularidade	5.1	12%	27%	24%	52%
	5.2	34%		66%	
	5.3	29%		57%	
	5.4	32%		61%	
Atualização periódica	6.1	8%	8%	15%	15%
Designação do responsável e prestação de assistência	7.1	36%	33%	70%	65%
	7.2	31%		60%	
TOTAL		35%		67%	

Fonte: elaboração própria (2022).

FIGURA 2 – GRAU DE ALINHAMENTO DAS UNIDADES* DA REDE EBSEH AOS PRINCÍPIOS DA PNDA



*Não constam no gráfico as unidades que não publicam DAG (49%): Sede, HC-UFG, HUAC-UFCG, HUIB-UFCG, HULW-UFPB, HUOL-UFRN, HUPAA-UFAL, MCO-UFBA, MEJC-UFRN, CHU-UFPA, HC-UFMG, HUCAM-UFES, HC-UFU, HUAP-UFF, HUGG-Unirio, CHC-UFPR, HE-UFPE, HU-Furg e HUSM-UFMS.
Fonte: elaboração própria (2022).

Somente 20 unidades da rede Ebserh (51%) publicam DAG, sendo que, destas, 18 possuem PDA publicados. Totalizam 854 conjuntos de dados publicados em toda a rede.

A Ebserh Sede está dentre as unidades que não publicam DAG. A mensagem publicada em sua página indica que a obrigatoriedade da publicação de DAG se restringe aos órgãos e entidades da Administração Pública Federal direta, autárquica e fundacional, e que, sendo a Ebserh uma empresa pública, pertencente à Administração Indireta, não estaria abrangida pelo disposto no Decreto n. 8.777/2016 (EBSEH, 2020). Apresenta ainda que tem atendido a todos os pedidos de acesso à informação, que tem priorizado a publicação de seus conteúdos em formato aberto e não-proprietário, e indica o link para acesso ao Fala.Br para solicitação de informações (EBSEH, 2020).

Cabe ressaltar que tanto a administração direta quanto a indireta estão subordinadas à LAI. Portanto, a Ebserh, sendo uma empresa pública, suas filiais devem seguir integralmente o disposto nesse regra-

mento, incluindo a divulgação, independentemente de requerimento, em local de fácil acesso, de informações produzidas ou custodiadas pela rede Ebserh que sejam de interesse público (Lei n. 12.527, 2011). Ainda, conforme esse dispositivo, há requisitos a serem cumpridos nesse sentido, tais como possibilitar a gravação de relatórios em formatos abertos e não proprietários e o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina. Apesar de o decreto que instituiu a PNDA se voltar ao Poder Executivo federal e de dispor como um de seus objetivos a promoção da publicação de dados contidos em bases de dados de “órgãos e entidades da administração pública federal direta, autárquica e fundacional sob a forma de dados abertos” (Decreto n. 8.777, 2016), a PNDA auxilia na efetivação e implementação da LAI por qualquer entidade pública, estabelecendo os objetivos, princípios, diretrizes e as estruturas de governança dos DAG.

Entretanto, observa-se que não há o reconhecimento do potencial dos DAG na administração central da Ebserh, bem como da importância da transparência colaborativa e de se utilizar efetivamente a estrutura centralizada de publicação de conteúdo nesse formato, já organizada para esse fim. Conforme consulta às normativas publicadas no portal Gov.br relacionadas a dados e informações, esse cenário também se associa ao escopo das políticas atualmente existentes na Ebserh, que não tratam, em qualquer ponto, sobre DAG e governança de dados:

1. Política de Divulgação de Informações: dá enfoque ao sigilo e traz o rol mínimo de informações de transparência ativa que, obrigatoriamente, devem ser divulgadas no portal. Cabe observar que, dentre essas informações, não estão as relacionadas à assistência, ensino, extensão, pesquisa e inovação, que são afetadas diretamente ao propósito da Ebserh;
2. Política de Proteção de Dados Pessoais da Ebserh: visa à conformidade com a LGPD e indica que outros documentos definirão procedimentos adicionais, os quais não foram identificados no portal;
3. Política de Classificação de Informação, Sigilo e Temporalidade: dá um enfoque maior aos pedidos de acesso à informação e respostas pela Ouvidoria; e
4. Política de Segurança da Informação e Comunicações: orienta sobre a proteção de dados, gestão de riscos e tratamento de incidentes.

Durante o levantamento dos DAG, identificou-se, na página de um dos HUF, que a Ebserh possuía um PDA, atualizado em 2018 (EBSERH, 2018). Portanto, reconheceu-se anteriormente a possibilidade e importância de atendimento à PNDA, porém posteriormente houve uma alteração quanto a esse entendimento, retrocedendo e deixando de publicar DAG. Tal direcionamento da administração central da Ebserh pode influenciar negativamente as demais unidades da rede e dificultar a abertura de dados. Tanto é que 15 das outras 18 unidades que não publicam DAG (83%) direcionam suas páginas de DAG para essa página da Sede ou apresentam o mesmo texto em suas páginas. As outras três unidades que não publicam DAG (HU-JB-UFMG, MEJC-UFRN e HUGG-Unirio) indicam, en-

tretanto, perspectiva de estabelecimento de um PDA, apesar de a previsão apresentada já estar com data retroativa.

Quanto aos tipos de DAG, não há uma padronização na rede Ebserh do que deva ser publicado. Considerando as três dimensões de processos a que se relaciona a rede, era de se esperar que houvesse DAG relativos à assistência à saúde, gestão hospitalar e ensino, extensão, pesquisa e inovação. Porém, cinco unidades (25%) publicam DAG somente de uma dessas dimensões ou de duas delas. Os conjuntos de dados são, em sua maioria (52%), relativos à assistência à saúde, seguidos pela gestão hospitalar (28%) e ensino, extensão, pesquisa e inovação (20%).

Todas as unidades que publicam DAG os disponibilizam diretamente no portal Gov.br, sem necessidade de identificação ou registro do usuário, sendo um ponto positivo e fundamental para a garantia de **acesso universal** aos dados.

A maioria dos conjuntos de dados (90%) estão em **formato aberto** (csv, ods e odt), os demais estão em formatos proprietários (docx, xlsx, pdf, jpeg e Power BI). Nem todos os conjuntos de dados são **legíveis por máquina**, sendo identificados arquivos com dados não estruturados, incluindo relatórios, contratos, boletins informativos, entre outros documentos. A forma de divulgação de dados está diretamente relacionada ao seu potencial de uso, reuso e compartilhamento, sendo fundamental que o formato aberto e estruturado seja adotado. Quando isso não ocorre, o esforço para coleta e processamento é aumentado, tornando menos efetiva a reutilização dos dados (Cardoso et al., 2020). Nesse quesito, já se verifica um alinhamento da rede, apesar de haver a necessidade de melhorias em algumas unidades.

Considerando o sistema de classificação de cinco estrelas de Tim Berners-Lee, nenhuma unidade alcança a 4ª e 5ª estrelas (Tabela 2). A maioria das unidades que publicam DAG alcançam as três estrelas, mas há aquelas que estão no padrão duas estrelas ou uma estrela. Apesar dos avanços, há passos a serem percorridos para a melhoria da qualidade dos DAG na rede Ebserh, para que possam ser transformados em DAG conectados.

TABELA 2 – UNIDADES* DA REDE EBSEERH SEGUNDO SISTEMA DE CLASSIFICAÇÃO DE CINCO ESTRELAS DE TIM BERNERS-LEE

UNIDADES DA REDE EBSEERH	CLASSIFICAÇÃO 5 ESTRELAS	FORMATOS DOS ARQUIVOS COM DADOS ABERTOS DISPONIBILIZADOS
HUB-UnB	★ ★ ★	70 arquivos em formato csv
HUJM-UFMT	★ ★ ★	65 arquivos em formato csv
HUMAP-UFMS	★ ★ ★	33 arquivos em formato csv
HUWC (CH-UFC)	★ ★ ★	1 arquivo em formato csv
MEAC (CH-UFC)	★ ★ ★	1 arquivo em formato csv
HUAB-UFRN	★ ★ ★	10 arquivos em formato csv
HUPES-UFBA	★ ★ ★	36 arquivos em formato csv
HU-UFPI	★ ★ ★	116 arquivos em formato csv
HU-UFS	★ ★ ★	20 arquivos em formato csv
HU-Univasf	★ ★ ★	21 arquivos em formato csv e 2 odt
HU-UFGD	★ ★	106 arquivos em formato csv e 1 xlsx
HC-UFTM	★ ★	144 arquivos em formato csv e 1 xlsx
HC-UFPE	★ ★	21 arquivos em formato xlsx
HUL-UFS	★ ★	16 arquivos em formato xlsx e 1 csv
HDT-UFT	★ ★	10 arquivos em formato xlsx, 1 ods e 41 csv
HU-UFSC	★	Dados disponibilizados na página web
HU-UFMA	★	34 arquivos em formato csv, 1 pdf e 2 odt
HUGV-UFAM	★	3 arquivos no formato pdf, 3 docx e 12 csv
HU-UFSCar	★	3 arquivos no formato pdf e 20 csv
HU-UFJF	★	36 arquivos no formato ods, 5 jpeg, 1 Power BI e 17 pdf

*Não constam na tabela as unidades que não publicam DAG: Sede, HC-UFG, HUAC-UFMG, HUJB-UF-CG, HULW-UFPPB, HUOL-UFRN, HUPAA-UFAL, MCO-UFBA, MEJC-UFRN, CHU-UFPA, HC-UFMG, HUCAM-UFES, HC-UFU, HUAP-UFF, HUGG-Unirio, CHC-UFPR, HE-UFPEl, HU-Furg e HUSM-UFMS.

Fonte: elaboração própria (2022).

A categoria referente à **descrição das bases de dados** alcançou um grau de alinhamento médio de 34% entre as unidades que publicam DAG. Somente em 10% dos conjuntos de dados havia indicação da periodicidade de sua atualização e em 59% havia descrição clara do que os compunham (título, descritivo, formato do arquivo, período dos dados). Caso fôssemos realizar uma avaliação mais aprofundada, considerando apresentação de dicionário de dados, nenhuma unidade atenderia. O dicionário de dados descreve a estrutura de um conjunto de dados, indicando do que se trata cada variável, ou seja, uma coleção de metadados com definições de cada elemento dos DAG, que auxilia na compreensão dos dados. Os metadados também não são apresentados de forma

estruturada na página de DAG. Para a maioria dos HUF, para cada conjunto de dados, há a apresentação do título e de um descritivo resumido com *hiperlink* para cada arquivo.

A **permissão irrestrita de reuso** dos dados foi a única categoria de princípios atendida por 100% das unidades que publicam DAG. Aponta-se, entretanto, que tal avaliação foi realizada a partir da constatação de que não havia restrição indicada na página de publicação, nem nos arquivos dos dados. Apesar disso, também não há indicação literal da permissão irrestrita de reuso.

Os princípios relativos à **completude e interoperabilidade** apresentaram média de 52% de alinhamento nas unidades que publicam DAG. Cabe observar que

não foram avaliadas, nessa categoria, as variáveis que deveriam conter cada conjunto de dados, voltando-se a análise à verificação do que foi apresentado quanto ao que foi proposto para ser publicado pela unidade. O menor grau de alinhamento dessa categoria referiu-se à fonte dos dados. Somente 24% dos conjuntos de dados apresentam indicação da fonte. Dentre esses, identificam-se sistemas e painéis padronizados para a rede Ebserh como fontes dos dados (Censo Hospitalar, AGHU, Relatório de Produção, Gestão à Vista, Vigihosp), bem como sistemas específicos utilizados por algumas unidades (Sistema Smart, MV Soul) e plataformas externas à rede, do governo federal (SIA-DATASUS, TABWIN-DATASUS, SIGP, Tesouro Gerencial, Sistema e-Ouv, Compras Governamentais).

Em 66% dos conjuntos de dados, os dados estão disponibilizados em sua forma primária. Os conjuntos restantes apresentam dados totalizados ou valores de indicadores, sem indicação dos números primários utilizados para seu cálculo, o que dificulta análises mais qualificadas. Tal fato pode estar associado às dúvidas que os gestores possuem quanto à publicação de microdados e receio de afetar a privacidade do cidadão a que os dados se referem (OKBr, 2021b). Entretanto, há soluções para mitigar os riscos associados, que envolvem os níveis de detalhamento e agregação, pseudonimização dos dados (ex.: substituição do CPF por outro código que não remeta ao indivíduo) e publicações em separado dos microdados (subdividir os dados) (Belisário, 2021). Essas soluções impedem a vinculação direta dos dados a indivíduos específicos (D'Agostino et al., 2018). Cabem ainda discussões organizacionais, estabelecimento de padrões e treinamentos na temática para enfrentamento desse problema.

Também se observa que há considerável falta de dados, seja de alguns períodos não disponibilizados ou não informados, dados faltantes de algumas variáveis, diferenças entre descrição dos dados e conteúdo do arquivo, entre outros. Evidenciou-se que 57% dos conjuntos de dados estavam publicados de forma completa.

Quanto à padronização intraunidades dos conjuntos de dados do mesmo tipo, há pontos de melhoria. Em 61% dos conjuntos de dados, há uma estrutura padrão que permite a comparabilidade. Nos demais conjuntos, há variáveis e formas de apresentação distintas, o que dificulta a comparabilidade e interoperabilidade. Observa-se ainda que, entre as unidades, não há padronização dos tipos de conjuntos de DAG e das variáveis utilizadas em cada conjunto de forma a permitir a interoperabilidade. A **Tabela 3** apresenta os conjuntos de dados que se repetem em duas ou mais unidades, apesar de haver diferenças tanto de descritivo, período e estruturação dos dados. Além desses conjuntos de dados, há ainda outros 36 que não se repetem entre as unidades. A maior correspondência de tipos de conjuntos de DAG entre as unidades se dá na dimensão de assistência, sendo estes relativos a indicadores de qualidade assistenciais.

TABELA 3 – CONJUNTOS DE DADOS ABERTOS POR DUAS OU MAIS UNIDADES DA REDE EBSERH POR DIMENSÃO (% DE HOSPITAIS UNIVERSITÁRIOS FEDERAIS*)

ASSISTÊNCIA À SAÚDE	ENSINO, EXTENSÃO, PESQUISA E INOVAÇÃO	GESTÃO
<ul style="list-style-type: none"> • Nº de consultas ambulatoriais por especialidade (90%) • Nº de exames de imagem ou análises clínicas realizados por tipo (90%) • Nº de internações por especialidade médica ou na UTI (85%) • Nº de cirurgias realizadas por especialidade (70%) • Taxa de ocupação de leitos hospitalares ou de UTI (65%) • Taxa de mortalidade: internação hospitalar ou UTI (65%) • Nº de óbitos da internação hospitalar ou na UTI (45%) • Relação de comissões, comitês e núcleos instituídos (35%) • Relação de especialidades médicas ou multiprofissionais (25%) • Nº de partos cesarianos e normais realizados (15%) • Nº de pacientes com infecção hospitalar (15%) • Taxa de infecção hospitalar (15%), infecção de sítio cirúrgico em cirurgias limpas (10%) • Taxa de permanência (15%) • Nº de altas hospitalares (10%) 	<ul style="list-style-type: none"> • Relação de preceptores das residências médica, multiprofissional e uniprofissional por especialidade (60%) • Relação de alunos da residência médica, multiprofissional e uniprofissional por especialidade (50%), do estágio curricular por curso (40%) e do internato rural (10%) • Relação de docentes do estágio curricular (45%), da residência médica, multiprofissional e uniprofissional em saúde (45%) • Relação de projetos de pesquisas (30%) • Nº de residentes por especialidade (25%) • Relação de editais de processo seletivo das residências médica e multiprofissional (20%) • Vagas por Programa de Residência Médica (10%) 	<ul style="list-style-type: none"> • Relação de empregados do regime CLT por cargo (80%), de servidores do regime RJU por cargo (75%), de ocupantes de função gratificada ou em cargo comissionado (60%) e de colaboradores terceirizados (45%) • Relação de contratos realizados (50%), de licitações e pregões (30%), de processos de dispensa de licitações (20%), dos pagamentos efetuados (20%) e de atas de registro de preços (10%) • Execução orçamentária e financeira (25%), movimentação financeira (10%) • Nº de manifestações registradas na Ouvidoria (25%) • Ações de manutenção e infraestrutura realizadas no período (10%) • Relações das matérias publicadas no sítio eletrônico do hospital (10%)

*Ntotal=20.

Fonte: elaboração própria (2022).

A **atualização periódica** dos dados envolve a categoria de princípios com menor alinhamento da rede Ebserh. Somente 15% dos conjuntos de dados estavam atualizados. Além disso, a maioria das páginas traz dados de períodos de 2020 adiante, não trazendo um maior histórico. Somente uma unidade traz alguns conjuntos de dados de 2017, duas trazem de 2018 e três de 2019. O acesso tempestivo aos dados é crucial para que muitos dos objetivos dos DAG sejam alcançados, sendo algo que precisa de atenção especial para ser melhorado na rede. Ademais, a linha histórica permite a análise da evolução e de melhorias de desempenho.

Para a categoria de **responsabilidade e prestação de assistência**, apurou-se um grau de alinhamento médio de 65% entre as unidades que publicam DAG. Para 70% dos dados publicados, havia a indicação do responsável pelos dados, e, para 60%, o contato desse responsável era informado (e-mail ou telefone). Cabe observar, entretanto, que não é disponibilizado mecanismo direto na página de DAG para que a sociedade possa interagir com o responsável pelos dados.

Na perspectiva dos novos modos de comunicação fomentados para plataformas digitais e da experiência do usuário, trata-se de uma oportunidade de melhoria quanto à prestação de assistência ao usuário. Isso também pode auxiliar no reconhecimento das necessidades dos usuários e de elementos para melhorias contínuas da publicação das bases de dados.

5. RECOMENDAÇÕES

A qualidade dos DAG da rede precisa ser melhorada para que possam, de fato, serem úteis para diferentes propósitos de reuso. Nisso, incluem-se a atualização tempestiva do portal Gov.br, a disponibilização de dados completos, de todos os períodos, com indicação da fonte, e com a maior granularidade possível. Atenção também deve ser dada aos tipos de DAG, incluindo a abertura de outras bases de dados disponíveis.

O portal Gov.Br já auxilia sobremaneira a integração das páginas dos diferentes HUF da rede e a estruturação da forma de apresentação de conteúdos, sendo uma importante ferramenta para um catálogo central de DAG da rede. Além da disponibilização de DAG, a criação de painéis dinâmicos com as bases de dados pode auxiliar na visualização dos dados e comunicação com a sociedade.

Identifica-se, sobretudo, a necessidade de estabelecimento de uma política de governança de dados para a rede Ebserh, com a liderança da Sede e MEC e apoio do Ministério da Saúde, que aborde os DAG de saúde pública, com a identificação clara dos benefícios potenciais para a sociedade, os riscos envolvidos e como gerenciá-los. Nesse sentido, é fundamental o mapeamento de riscos e estratégias para a adequada divulgação de DAG de acordo com cada tipo de dado, considerando aspectos de privacidade, segurança e qualidade (Borgesius et al., 2015; D'Agostino et al., 2017). Tal política deve definir as regras relacionadas ao controle e uso dos dados gerados nos HUF e na Sede da empresa, indicando o que, como e por quem deve ser aberto (Campagnucci, 2021). A partir disso, será possível discutir uma mudança de cultura na rede, que reconheça os dados como um ativo estratégico, bem como a importância e o valor de sua abertura, seguindo os princípios da PNDA, que trará benefícios tanto intra quanto extra rede.

Uma compreensão da LGPD é premente e necessária, partindo do pressuposto fundamental de que a lei não proíbe a divulgação de dados, mas, sim, estabelece regras para o seu compartilhamento. Bases de dados que envolvam dados pessoais necessitam de cuidados adicionais, mas isso não deve ser um obstáculo para a abertura de dados quando houver interesse público (Langenegger, 2021). Borgesius et al. (2015) apontam sobre a necessidade de se identificar o ponto de equilíbrio que respeite a privacidade e os interesses públicos relacionados, sem prejudicar indevidamente os benefícios dos DAG. Essa discussão também precisa ser levada a outras entidades, em âmbito nacional, para promoção da abertura de dados de saúde pública e integração a outras bases de dados. Os HUF possuem papel estratégico nesse contexto, ao integrarem ensino, extensão, pesquisa e inovação em saúde, e terem uma relação direta com a sociedade.

6. LIMITAÇÕES

Entre as limitações desta pesquisa, aponta-se que este estudo foi realizado com base nos dados publicados, não se investigando o tema do ponto de vista interno das organizações estudadas. Ademais, a análise foi conduzida na perspectiva dos autores, podendo outros pesquisadores terem outras interpretações por meio de métodos distintos.

Registre-se também que a busca de DAG foi apenas no portal Gov.br nas páginas das unidades da rede Ebserh, não tendo sido realizada a comparabilidade dos DAG dessas páginas com os publicados no Portal Brasileiro de Dados Abertos (dados.gov.br). Esperava-se encontrar uma associação entre essas páginas, porém isso não ocorreu. Há tanto a catalogação no Portal Brasileiro quanto a publicação na página institucional da unidade, com formatos de apresentação diferenciados e sem conexão entre páginas.

Entretanto, tais limitações não invalidam os resultados encontrados. Futuros estudos podem trazer essa comparação, utilizar outros métodos e buscar compreender outros fatores de contexto intervinientes à abertura de dados na rede Ebserh.

7. CONSIDERAÇÕES FINAIS

A rede Ebserh tem disponibilizado DAG, mesmo que parcialmente, mas há considerável proporção de unidades que não publicam DAG (49%). O ponto de equilíbrio entre abertura e proteção de dados ainda não foi alcançado. Verifica-se que as unidades da rede se encontram em níveis de maturidade distintos quanto à abertura de dados (grau de alinhamento aos princípios da PNDA de 35% da rede e 67% das unidades que publicam DAG), o que pode ser indicativo da ausência de governança de dados e políticas da Ebserh Sede ou do MEC que guiem essa estratégia. Muitas unidades nem iniciaram o desenvolvimento de seus PDA, outras já iniciaram o seu desenvolvimento, mas não o publicaram. Entre as que já publicam DAG, há também diferenciações, sendo que nenhuma unidade atende a todos os princípios da PNDA. Fato este que pode dificultar a busca, reutilização e interoperabilidade dos dados.

A produção e o gerenciamento de dados na rede Ebserh são realizados de maneira isolada e não interoperável por cada unidade, sem uma estratégia comum para padronização e organização do processo de aber-

tura de dados. A Ebserh Sede, como ente central da rede, falha ao não fornecer um direcionamento para todas as unidades para a abertura e integração de dados. Assim, nem todos os dados estão facilmente disponíveis para uma análise agregada ou comparativa da rede.

Diante desse cenário, recomendações foram apresentadas para maior alinhamento da rede Ebserh à PNDA, incluindo melhorias da qualidade dos DAG, abertura de outras bases, instituição de uma política de governança de dados e aprimoramento da compre-

ensão quanto à LGPD. Os resultados apresentados neste estudo podem contribuir para melhorias na governança e abertura de dados na rede Ebserh. O método utilizado, de análise do grau de alinhamento aos princípios da PNDA, pode também ser aplicado a outros órgãos e entidades da administração pública, de modo a colaborar com avaliações quanto aos DAG que têm sido publicados e o seu potencial para alcance dos benefícios esperados de transparência, controle social, inovações e crescimento econômico.

REFERÊNCIAS

- Avelino, D. P. de, Pompeu, J. C., & Fonseca, I. F. da. (2021). TD 2624 - Democracia digital: Mapeamento de experiências em dados abertos, governo digital e ouvidorias públicas. *Texto para Discussão*. Instituto de Pesquisa Econômica Aplicada, 1–52. <https://doi.org/10.38116/td2624>
- Belisário, A. (2021). *Pseudonimização de dados com editores de planilha*. Open Knowledge Brasil. <https://escolade-dados.org/tutoriais/pseudonimizacao-de-dados-com-editores-de-planilha/>
- Berners-Lee, T. (2006). *Linked Data—Design Issues*. <https://www.w3.org/DesignIssues/LinkedData.html>
- Bizer, C., Heath, T., & Berners-Lee, T. (2006). Linked Data—The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*.
- Borgesius, F. Z., Eechoud, M. van, & Gray, J. (2015). *Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework* (SSRN Scholarly Paper No 2695005). Social Science Research Network. <https://papers.ssrn.com/abstract=2695005>
- Campagnucci, F. (2021). Formulando políticas públicas baseadas em dados. Em *Publicadores de dados [recurso eletrônico]: Da gestão estratégica à abertura* (p. 10–18). Open Knowledge Brasil. <https://www.ok.org.br/wp-content/uploads/2021/05/PublicadoresDeDados2.pdf>
- Campagnucci, F., & Belisário, A. (2021). *Publicadores de dados [recurso eletrônico]: Da gestão estratégica à abertura*. Open Knowledge Brasil. <https://www.ok.org.br/wp-content/uploads/2021/05/PublicadoresDeDados2.pdf>
- Cardoso, R. A. L., Carneiro, G. F., & Menezes, J. E. X. (2020). Dados Abertos Governamentais no contexto de Políticas Públicas de Saúde e Sistemas Prisionais: Realidade ou Utopia? *Olhares das ciências sobre as questões sociais*, 19(2), 65–80. <https://doi.org/10.48550/ARXIV.2103.00541>
- Controladoria Geral da União (2021). *5º Plano de Ação Nacional em Governo Aberto*. *Open Government Partnership*. https://www.opengovpartnership.org/wp-content/uploads/2022/01/Brazil_Action-Plan_2021-2023_PT.pdf
- Cruz-Rubio, C. N. ([s.d.]). *Hacia el gobierno abierto: Una caja de herramientas*. Organización de los Estados Americanos. Recuperado em 29 maio, 2022, de https://www.gigapp.org/administrator/components/com_jresearch/files/publications/FINAL%20Caja%20de%20Herramientas.pdf
- Cruz-Rubio, C. N. (2015). O que é (e o que não é) governo aberto? Uma discussão conceitual. *Rev. Temas de Administração Pública*, 10(1), 129–148.
- D'Agostino, M., Marti, M., Mejía, F., Cosio, G. de, & Faba, G. (2017). Estrategia para la gobernanza de datos abiertos de salud: Un cambio de paradigma en los sistemas de información. *Rev Panam Salud Publica*, 41, 1–6.

- D'Agostino, M., Samuel, N. O., Sarol, M. J., Cosio, F. G., Marti, M., Luo, T., Brooks, I., & Espinal, M. (2018). Open data and public health. *Revista Panamericana de Salud Pública*, 42. <https://doi.org/10.26633/RPSP.2018.66>
- Decreto n. 8.777, de 11 de maio de 2016. (2016). Institui a Política de Dados Abertos do Poder Executivo federal. Recuperado de http://www.planalto.gov.br/ccivil_03/ato2015-2018/2016/decreto/d8777.htm
- Decreto n. 9.756, de 11 de abril de 2019. (2019). Institui o portal único “gov.br” e dispõe sobre as regras de unificação dos canais digitais do Governo federal. Recuperado de http://www.planalto.gov.br/ccivil_03/ato2019-2022/2019/decreto/d9756.htm
- Decreto n. 10.160, de 9 de dezembro de 2019. (2019). Institui a Política Nacional de Governo Aberto e o Comitê Interministerial de Governo Aberto. Recuperado de https://www.planalto.gov.br/ccivil_03/Ato2019-2022/2019/Decreto/D10160.htm#art13
- Eaves, D. (2009). The Three Laws of Open Government Data. *Eaves.Ca*. <https://eaves.ca/2009/09/30/three-law-of-open-government-data/>
- EBSERH. (2018). *Plano de Dados Abertos*. PDA-Ebserh. 2017/2018. https://www.gov.br/ebserh/pt-br/hospitais-universitarios/regiao-sudeste/hu-uffg/governanca/superintendencia/unidade-de-planejamento/dados-abertos/Plano_Dados_Abertos_20172018_EBSERH.pdf
- EBSERH. (2020). Dados Abertos [Portal gov.br]. Empresa Brasileira de Serviços Hospitalares. <https://www.gov.br/ebserh/pt-br/aceso-a-informacao/dados-abertos>
- Filgueiras, F., Palotti, P. L. M., & Nascimento, M. I. B. (2022). Policy design e uso de evidências: O caso da plataforma Gov.br. Em *Políticas públicas e usos de evidências no Brasil* (1a). Instituto de Pesquisa Econômica Aplicada (Ipea). https://www.ipea.gov.br/portal/images/stories/PDFs/livros/livros/220412_lv_o_que_informa_miolo_cap16.pdf
- FOSTER. (2015). *Open Government Data*. FOSTER Portal. <https://www.fosteropenscience.eu/taxonomy/term/115>
- Heijlen, R., & Cromptvoets, J. (2021). Open health data: Mapping the ecosystem. *Digital Health*, 7, 205520762110501. <https://doi.org/10.1177/20552076211050167>
- Langenegger, N. (2021). Abertura de dados à Lei Geral de Proteção de Dados. Em *Publicadores de dados [recurso eletrônico]: Da gestão estratégica à abertura* (p. 39–45). Open Knowledge Brasil. <https://www.ok.org.br/wp-content/uploads/2021/05/PublicadoresDeDados2.pdf>
- Lei n. 12.527, de 18 de novembro de 2011. (2011). Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Recuperado de http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/l12527.htm
- Lei n. 12.550, de 15 de dezembro de 2011. (2011). Autoriza o Poder Executivo a criar a empresa pública denominada Empresa Brasileira de Serviços Hospitalares—EBSERH; acrescenta dispositivos ao Decreto-Lei no 2.848, de 7 de dezembro de 1940—Código Penal; e dá outras providências. Recuperado de http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/Lei/L12550.htm
- Meijer, A. J., Curtin, D., & Hillebrandt, M. (2012). Open government: Connecting vision and voice. *International Review of Administrative Sciences*, 78(1), 10–29. <https://doi.org/10.1177/0020852311429533>
- Nascimento-Silva, P., & Kerr-Pinheiro, M. M. (2019). Métrica alternativa para dados governamentais abertos na América Latina. *Transinformação*, 31, e190009. <https://doi.org/10.1590/2318-0889201931e190009>
- OGP. (2022). *About Open Government Partnership*. *Open Government Partnership*. <https://www.opengovpartnership.org/about/>

Open Knowledge Brasil. (2021a). *Emergência dos dados [recurso eletrônico]: Como o Índice de Transparência da Covid-19 impulsionou a abertura de dados da pandemia no Brasil*. Open Knowledge Brasil. https://ok.org.br/wp-content/uploads/2021/11/Ebook_EmergenciaDados_OKBR.pdf

Open Knowledge Brasil. (2021b). *Publicação de microdados de casos de Covid-19*. Open Knowledge Brasil. https://transparenciacovid19.ok.org.br/files/Toolkit_1_microdados_basicosV2.pdf

Open Knowledge Foundation. ([s.d.]). *What is open?* Recuperado 28 maio, 2022, de <https://okfn.org>

Open Knowledge Foundation. (2015). *The Open Definition—Open Definition—Defining Open in Open Data, Open Content and Open Knowledge*. <https://opendefinition.org/>

O'Reilly, T., Malamud, C., Sifry, M., Holovaty, A., & O'Neil, D. X. (2007). *The 8 Principles of Open Government Data*. OpenGovData.org. <https://opengovdata.org/>

Pires, M. T. (2015). Guia de Dados Abertos. Governo do Estado de São Paulo, Governo do Reino Unido, Núcleo de Informação e Coordenação do Ponto Br (NIC.Br). https://nic.br/media/docs/publicacoes/13/Guia_Dados_Abertos.pdf

Pires, K. M. (2019). Dados abertos nas universidades federais: Envolvimento interno e divulgação para a sociedade. *Revista Brasileira de Biblioteconomia e Documentação*, 15(2), 58–80.

Santos, P. X. dos, Almeida, B. de A., & Henning, P. (Orgs.). (2017). *Livro Verde—Ciência aberta e dados abertos: Mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional*. Fiocruz. <https://www.arca.fiocruz.br/bitstream/icict/24117/2/Livro-Verde-07-06-2018.pdf>

Silva, M. P. A. e, & Dominguez, M. (2017). *E-transparência e dados abertos governamentais: O ambiente digital como locus dos avanços do governo brasileiro nas políticas, leis e parceiras internacionais*. Compolítica. http://compolitica.org/novo/anais/2017_Almada_Dominguez.pdf

Silva, P. N. (2018). *Dados governamentais abertos: Métricas e indicadores de reúso* [Tese (Doutorado), Universidade Federal de Minas Gerais]. <https://repositorio.ufmg.br/handle/1843/BUBD-AYNG4U>

Tribunal de Contas da União (2015). *5 motivos para abertura de dados na Administração Pública*. https://portal.tcu.gov.br/data/files/81/55/71/DB/A592C710D79E7EB7F18818A8/5_motivos_abertura_dados_administracao_publica.PDF



Gabriela de Oliveira Silva

<http://orcid.org/0000-0003-1809-3789>

Escola Nacional de Administração Pública, Empresa Brasileira de Serviços Hospitalares

gabioliveira18@gmail.com

Farmacêutica na Empresa Brasileira de Serviços Hospitalares (Ebserh). Doutoranda em Políticas Públicas pela Escola Nacional de Administração Pública (Enap), Mestra em Políticas Públicas em Saúde pela Escola Fiocruz de Governo (2017), Especialista em Vigilância Sanitária pela PUC-GO (2016), graduada em Farmácia (2010) com Habilitação em Indústria pela Universidade Federal de Minas Gerais (2011).

**Renata Mendes de Araujo**

<https://orcid.org/0000-0002-8674-1728>

Universidade Presbiteriana Mackenzie (principal), Escola Nacional de Administração Pública e Universidade de São Paulo

renata.araujo@mackenzie.br

Professora na Faculdade de Computação e Informática e pesquisadora do Programa de Pós-Graduação em Computação Aplicada da Universidade Presbiteriana Mackenzie. Pesquisadora do Programa de Pós-Graduação em Sistemas de Informação da EACH-USP. Pesquisadora do Doutorado Profissional em Administração Pública da Enap. Pesquisadora convidada do LUDS-Laboratório de Ludologia, Engenharia e Simulação da COPPE/UFRJ. Bolsista de Produtividade em Desenvolvimento Tecnológico e Extensão Inovadora pelo CNPq (#313210/2019-5). Possui graduação em Informática pela Universidade Federal do Rio de Janeiro (1992), mestrado (1994) e doutorado (2000) em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro.

**Fernando Filgueiras**

<https://orcid.org/0000-0001-9570-8113>

Universidade Federal de Goiás e Escola Nacional de Administração Pública

fernandofilgueiras@hotmail.com

Professor Associado da Faculdade de Ciências Sociais da Universidade Federal de Goiás (UFG). Pesquisador Associado, Ostrom Workshop in Political Theory and Policy Analysis - Indiana University. Pesquisador do Instituto Nacional de Ciência e Tecnologia - Democracia Digital (INCT-DD), Universidade Federal da Bahia (UFBA). Bolsista de Produtividade em Pesquisa pelo CNPq (#303273/2020-8). Foi Diretor de Pesquisa e Pós-Graduação da Escola Nacional de Administração Pública (Enap). Membro efetivo do College da International Public Policy Association (IPPA). Doutor em Ciência Política pelo Instituto Universitário de Pesquisas do Rio de Janeiro (IUPERJ).

A trajetória mexicana na consolidação do direito de acesso à informação pública: histórico e desafios^{1,2}

The Mexican path in consolidating the right of access to public information: historic and challenges

La trayectoria mexicana en la consolidación del derecho de acceso a la información pública: historia y retos

Luma Poletti Dutra

<https://doi.org/10.36428/revistadacgu.v14i26.525>

Resumo: Este artigo apresenta a trajetória de desenvolvimento de políticas de acesso à informação pública no México, em especial a aprovação da Lei Federal de Acesso à Informação Pública, em 2002, e suas posteriores reformas. Partindo de um breve olhar sobre aspectos que marcam o contexto histórico político do país, busca-se compreender como se deu essa trajetória, quais foram os atores envolvidos e os desafios que persistem no cenário mexicano. A partir de revisão de literatura e entrevistas realizadas com pesquisadores e ativistas mexicanos, conclui-se que o país foi uma referência para a região da América Latina em razão de seu pioneirismo na regulamentação do direito de acesso à informação. No entanto, a despeito dos avanços conquistados neste campo, o México não deixa de enfrentar desafios relacionados ao tema diante de novos cenários políticos que impactam desde o financiamento do INAI até o atendimento aos pedidos de acesso à informação.

Palavras-chave: México; Lei de Acesso à Informação; *Accountability*; transparência.

Abstract: This article presents the development trajectory of policies for access to public information in Mexico, in particular the approval of the Access to Information Law, in 2002, and its subsequent reforms. Starting from a brief look at aspects that mark the historical political context of the country, we seek to understand how the right of access to public information has been consolidated and what challenges persist in the Mexican scenario. Based on a literature review and interviews carried out with Mexican researchers and activists, it is concluded that the country – which has some similarities with Brazil at a historical and political level – was a reference for the Latin American region due to its pioneering spirit in the regulation of the right of access to information. However, despite the advances achieved in this field, the country does not fail to face challenges related to the theme in the face of new political scenarios that affect INAI's funding and the fulfillment of requests for access to information.

Keywords: Mexico; Access to Information Law; *Accountability*; transparency

Resumen: Este artículo presenta la trayectoria de desarrollo de las políticas de acceso a la información pública en México, en particular la aprobación de la Ley de Acceso a la Información, en 2002, y sus reformas posteriores. A partir de una breve mirada a aspectos que marcan el contexto político histórico del país, buscamos comprender cómo se ha consolidado el derecho de acceso a la información pública y qué desafíos persisten en el escenario mexicano. A partir de una revisión de la literatura y entrevistas realizadas con investigadores y activistas mexicanos,

1. Artigo submetido em 23/06/2022 e aceito em 17/11/2022.

2. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

se concluye que el país, que tiene algunas similitudes con Brasil a nivel histórico y político, fue un referente para la región latinoamericana por su pionero. espíritu en la regulación del derecho de acceso a la información. Sin embargo, a pesar de los avances logrados en este campo, el país no deja de enfrentar desafíos relacionados con el tema ante los nuevos escenarios políticos que impactan desde el financiamiento del INAI hasta el cumplimiento de las solicitudes de acceso a la información.

Palabras clave: México; Ley de Acceso a la Información; rendición de cuentas; transparência.

INTRODUÇÃO

Brasil e México possuem algumas semelhanças em suas trajetórias políticas. Ambos países são marcados por uma tradição de opacidade na administração pública, característica do patrimonialismo, sistema marcado pela difícil distinção entre público e privado (Costa; Dutra; Paulino, 2021). Neste sentido, este artigo lança um olhar sobre a consolidação do direito de acesso à informação no cenário mexicano, com especial interesse por sua regulamentação. O objetivo é compreender como se deu essa trajetória, quais foram os atores envolvidos e os desafios que se apresentam, vinte anos depois da aprovação da Lei Federal de Acesso à Informação. Para isso, reúne informações obtidas por meio de revisão de literatura e apresenta trechos de entrevistas semiestruturadas com pesquisadores e ativistas, realizadas na Cidade do México entre 2019 e 2020, durante período de pesquisa no país.

A Lei Federal de Transparência e Acesso à Informação Pública do Governo do México foi sancionada em junho de 2002, nove anos antes da LAI brasileira. A regra mexicana serviu de parâmetro (juntamente com outras experiências internacionais à época) para a elaboração da norma brasileira (Controladoria-Geral da União [CGU], 2008).

Considerada uma referência a nível internacional (Michener, 2011), a lei mexicana sofreu algumas reformas, de modo que a legislação em vigor atualmente, aprovada em 2015, apresenta mudanças consideráveis em relação ao texto sancionado em 2002, como se verá neste artigo. Os quase vinte anos de trajetória desde a entrada em vigor da primeira lei no país ajudaram a consolidar o direito de acesso à informação e trouxeram novos desafios (Delgado, 2020). Assim, olhar para a experiência mexicana pode oportunizar um valioso aprendizado para o Brasil, que em 2021 completou dez anos de sanção de sua Lei de Acesso à Informação.

Pontos de conexão entre Brasil e México

Se ao chegarem ao Brasil os portugueses buscaram criar um mundo à imagem da Coroa (Faoro, 2011), com a colonização espanhola no México não foi diferente. As culturas que ali existiam tornaram-se subsidiárias do império espanhol. Na chamada Nova Espanha, nada era elaborado ou criado, mas sim replicado e adaptado. Com fronteiras difusas entre as esferas pública e privada, entre família e Estado, o patrimonialismo também é uma marca da cultura política mexicana (Paz, 2014).

Nas palavras do Nobel de Literatura Octavio Paz: “a corrupção governamental é uma característica pré-moderna de nossa sociedade que vem do patrimonialismo da Nova Espanha; a persistência do patrimonialismo, por sua vez, se explica por séculos de simbiose entre o poder e a tradicional família mexicana” (Paz, 2014, p. 14, tradução nossa). No ensaio *El ogro filantrópico*, publicado originalmente em 1979, Octavio Paz faz uma análise política e cultural do país, e destaca o perfil de governante que predominou no México ao longo dos anos: “Salvo durante os intervalos de anarquia e guerra civil, os mexicanos temos vivido à sombra de governos alternativamente despóticos ou paternos, mas sempre fortes: o rei-sacerdote asteca, o vice-rei, o ditador, o senhor presidente” (Paz, 2014, p.317, tradução nossa).

Assim, além do patrimonialismo, a excessiva valorização do Poder Executivo é outro traço político comum entre Brasil e México. Segundo Carvalho (2012), esta é uma herança compartilhada por países de tradição ibérica, que colocam em segundo plano os demais poderes. “O Estado é sempre visto como todo-poderoso, na pior hipótese como repressor e cobrador de impostos; na melhor, como um distribuidor paternalista de empregos e favores” (Carvalho, 2012, p.220).

Paz (2014) descreve o Estado mexicano como pertencente à duas alas opostas, mas que convivem: a tecnocracia administrativa (a chamada burocracia profissional) e a casta política. Essa característica revela

uma enorme contradição entre um corpo de funcionários com valores modernos e dispostos a modernizar o país, e uma sociedade da Corte, formada por amigos, parentes e demais favorecidos que representam a manutenção do patrimonialismo, segundo o autor.

A contradição entre a sociedade da Corte e a burocracia tecnocrata não imobiliza o Estado, mas torna sua marcha difícil e sinuosa. Não existem duas políticas dentro do Estado: existem duas maneiras de entender a política, dois tipos de sensibilidade e de moral. (Paz, 2014, p. 323, tradução nossa).

Diante deste contexto político e cultural, fica claro que a aprovação da primeira Lei de Acesso à Informação do país, em 2002, possui um peso simbólico que ultrapassa o aspecto jurídico.

O acesso à informação como ruptura com o passado

Para tratar do desenvolvimento de políticas de acesso à informação no contexto mexicano, toma-se como ponto de partida o ano de 1977, e o papel desempenhado pela Suprema Corte de Justiça da Nação (SCJN) à época. Naquele ano, uma reforma constitucional (decorrente de uma reforma política eleitoral, promovida pelo então presidente José López Portillo, do PRI³, acrescentou ao texto do artigo 6º da carta magna que o direito à informação seria assegurado pelo Estado. Em 1985, a SCJN foi questionada sobre o alcance desta garantia. A corte estabeleceu que o direito de acesso à informação se traduzia como um direito social⁴, que assegurava espaços midiáticos a partidos políticos para a manifestação da diversidade de opiniões e posicionamentos ideológicos.

De acordo com o entendimento da corte, a alteração do texto constitucional em 1977 não pretendia “estabelecer uma garantia individual consistindo em que qualquer governado, no momento em que julgue apropriado, solicite e obtenha de órgãos do Estado de-

terminada informação” (Suprema Corte de Justiça da Nação [SCJN], 1992, p. 44, tradução nossa). Neste sentido, conforme analisou Quintanilla-Madero (1988), “a atenção dada às preocupações da mídia resultou no adiamento das discussões sobre outros aspectos substantivos do direito de acesso à informação governamental” (Quintanilla-Madero, 1988, p.100, tradução nossa).

O entendimento mudou em 1996, após o episódio conhecido como *Masacre de Aguas Blancas*⁵, a corte ampliou o alcance do direito de acesso à informação ao reconhecer sua estreita ligação com o direito à verdade⁶. Na nova interpretação, a Suprema Corte se manifestou da seguinte forma:

Tal direito é, portanto, básico para a melhoria de uma consciência cidadã que contribui para torná-la mais consciente, essencial para o progresso de nossa sociedade. Se as autoridades públicas, eleitas ou designadas para servir e defender a sociedade, assumem diante desta atitudes que lhes permitem atribuir comportamentos antiéticos, fornecendo à comunidade informações incompletas e manipuladas, condicionadas aos interesses de grupos ou indivíduos, o que a proíbe de conhecer a verdade para participar livremente na formação da vontade geral, incorre em grave violação de garantias individuais nos termos do artigo constitucional, segundo parágrafo, porque seu processo leva a considerar que existe uma propensão a incorporar em nossa vida política o que poderíamos chamar de cultura do engano, da maquinação e da ocultação, ao invés de enfrentar a verdade e tomar ações rápidas e eficazes para alcançá-la e torná-la conhecida aos governados. (SCJN, 1996, p. 513, tradução nossa).

3. Partido Revolucionário Institucional. Seguindo com as reflexões de Octavio Paz, o autor dizia que junção dos termos “Revolucionário” e “Institucional” em uma mesma sigla representa “uma curiosa ilustração dos paradoxos da política mais do que da lógica” (PAZ, 2014, p. 262, tradução nossa).

4. Marshall (1967) define que direitos sociais vão desde o mínimo de bem-estar econômico e segurança, até o “direito de participar por completo na herança social e levar a vida de um ser civilizado de acordo com os padrões que prevalecem na sociedade” (Marshall, 1967, p.64). Ao discorrer sobre os fundamentos dos direitos do homem, Bobbio (2004) apresenta os direitos sociais como *poderes*, em contraposição aos direitos individuais tradicionais, que consistem em *liberdades*. Para o autor italiano, os direitos sociais são mais difíceis de se proteger e requerem uma intervenção direta do Estado.

5. Episódio de violência ocorrido em 1995, em que 17 camponeses do estado de Guerrero foram assassinados pela polícia. A versão oficial das autoridades dava conta de que os militares se defenderam de um ataque armado – o que se provou falso.

6. O conceito é trabalhado por Osmo (2014), que traça um panorama sobre sua inserção nos debates sobre justiça de transição e seu reconhecimento, no âmbito do direito internacional, diante de práticas de desaparecimentos forçados. A autora afirma que o direito à verdade ampliou seu significado com a ascensão de regimes democráticos (instituídos após períodos de violação de direitos humanos), e identifica seus três objetivos principais: verdade fática, veracidade e justiça.

Dessa forma, o direito de acesso à informação, antes entendido como um direito social, se tornou uma garantia individual. Essa interpretação foi reforçada pelo tribunal em julgamentos subsequentes. Por meio de outras decisões, “a Suprema Corte ampliou a compreensão desse direito, entendendo-o, também, como garantia individual limitada, pelos interesses nacionais e da sociedade, assim como pelo respeito aos direitos de terceiros” (SCJN, 2000, p. 72, tradução nossa). Esses acontecimentos precederam a aprovação da Lei Federal de Acesso à Informação Pública, e fazem parte do contexto que permitiu que as discussões sobre a norma avançassem no país.

Além da jurisprudência da Suprema Corte, a eleição de 2000 também representou um marco importante para as políticas de acesso à informação no México, pois encerrou a hegemonia do PRI no poder após 72 anos de domínio político. O monopólio começou a ser rompido em 1997, com o aumento da diversidade de partidos com representantes na Câmara dos Deputados, e se consolidou com a vitória de Vicente Fox, do PAN (Partido da Ação Nacional), para a Presidência. O mesmo pleito elegeu uma maioria de congressistas opositores a Fox, em uma configuração política que favorece o desejo de intensificar a fiscalização sobre o novo governo.

Para Sergio López Ayllón (2002), a mudança na coalizão governamental foi um dos principais fatores que permitiram a introdução do debate sobre o direito de acesso à informação pública na agenda nacional.

De fato, uma vez que a equipe de transição do então presidente eleito Fox foi formada, o grupo jurídico identificou a necessidade de

elaborar uma iniciativa de lei sobre acesso à informação governamental como um dos aspectos jurídicos prioritários para consolidar e completar a transição política no México. Outros integrantes da mesma equipe de transição concordaram com essa opinião e o tema foi instituído na agenda do novo governo. (López-Ayllón, 2002, p. 262, tradução nossa).

O autor também afirma que, com a mudança de governo, o PRI passou a ocupar o papel de oposição, o que fez com que a legenda tivesse “incentivos para mudar sua posição histórica de apoio implícito ao sigilo administrativo para apoiar uma iniciativa que abriria enormemente o acesso às informações do governo” (López-Ayllón, 2005, p.146, tradução nossa). Além disso, o envio de um projeto de lei de acesso à informação ao Congresso estava entre as promessas de campanha de Fox (Pla, 2009).

No mesmo ano da eleição, em 2000, foi aprovada a Lei de Aquisições, Arrendamentos e Serviços do Setor Público, e estabeleceu que compras e contratações no setor público deveriam ser efetuadas de maneira transparente. Em 2002, ano em que a lei de acesso foi sancionada, também foi aprovada a Lei de Transparência e Promoção da Concorrência em Crédito Garantido, para dar transparência aos processos de concessão de crédito por instituições financeiras. Essas e outras legislações subsequentes sinalizam um empenho no processo de ruptura com o passado político recente do país (Robledo, 2014).

QUADRO1: INSTRUMENTOS NORMATIVOS ANTERIORES E POSTERIORES À LEI FEDERAL DE TRANSPARÊNCIA E ACESSO À INFORMAÇÃO PÚBLICA DO GOVERNO, QUE TRATAM DA GESTÃO DOCUMENTAL E TRANSPARÊNCIA NA ADMINISTRAÇÃO PÚBLICA

INSTRUMENTOS NORMATIVOS SOBRE ACESSO À INFORMAÇÃO PÚBLICA	
INSTRUMENTO NORMATIVO	CONTEÚDO
Lei de Aquisições, Arrendamentos e Serviços do Setor Público, de janeiro de 2000	Regulamenta as compras e contratos firmados pelo governo federal para que sejam cumpridos de acordo com os princípios da transparência e eficiência.
Lei de Fiscalização Superior da Federação, de dezembro de 2000	Base legal para a criação da Auditoria Superior da Federação, responsável por fiscalizar as contas públicas por meio de auditorias realizadas nos três Poderes da União, nos órgãos constitucionalmente autônomos, nos estados e municípios do país, bem como em qualquer entidade que exerça recursos públicos federais, inclusive pessoas físicas.
Lei de Transparência e Promoção da Concorrência em Crédito Garantido, de dezembro de 2002	Regula os serviços financeiros de concessão de Crédito Garantido, para assegurar a transparência na operação.
Lei de Transparência e Regulamentação de Serviços Financeiros, de janeiro de 2004	Estabelece regras de transparência que obrigam o Banco do México a publicar informação relativa à cobrança de comissões sobre os diversos serviços financeiros.
Lei de Fiscalização e Prestação de Contas da Federação, de maio de 2009	Regulamenta o exercício do controle das contas públicas por meio da revisão das receitas, despesas, entre outros, bem como o supervisão e avaliação de todas as informações financeiras das entidades auditadas.
Lei Federal de Proteção de Dados Pessoais Mantidos por Pessoas Físicas, de julho de 2010	Protege os dados pessoais em posse de particulares com o objetivo de regular o seu tratamento legítimo, controlado e informado, de forma a garantir a privacidade e o direito à autodeterminação informativa das pessoas.
Lei Federal de Arquivos, de janeiro de 2012	Tornou necessário documentar e cuidar das informações que devem ser integradas aos arquivos dos Três Poderes federais e dos órgãos autônomos do Estado.
Lei Federal Anticorrupção nas Contratações Públicas, de junho de 2012	Estabelece as bases para a punição de pessoas físicas por infrações de natureza contratual com o Estado em matéria de compras públicas e por infrações cometidas na participação em transações comerciais internacionais e em contratos federais.

Fonte: Robledo, 2014, tradução nossa.

Soma-se a este cenário o contexto internacional, com acordos comerciais firmados cuja contrapartida consistia na adoção de medidas de transparência, impulsionadas principalmente pela Organização Mundial do Comércio (Pla, 2009).

A aprovação da Lei de Acesso à Informação mexicana ainda contou com importante atuação de organizações da sociedade civil, com destaque para o Grupo Oaxaca (Costa et al., 2021), que conseguiu inserir o tema na agenda de debates e participou da negociação do texto final junto com o governo. Finalmente, em 30 de abril de 2002 foi aprovada a Lei Federal de Transparência e Acesso à Informação Pública Governamental.

A lei determinou a criação de *Unidades de Enlace* em todos os órgãos públicos, setores responsáveis por receber as solicitações e dar andamento aos pedidos internamente, bem como auxiliar os requerentes (Art. 28). Não foi designado orçamento ou mão de obra extra para as unidades, o que, a princípio, gerou uma expectativa de que isso pudesse levar a um “ressenti-

mento burocrático e resistência das entidades” (Doyle, 2002, p. 4, tradução nossa). Um dos pontos de maior destaque da lei foi a criação do IFAI, instituído para atuar como órgão fiscalizador, instância recursal no âmbito do Executivo federal e promover treinamento para os servidores públicos atuarem de acordo com os preceitos da nova regra (Sobel et al., 2006).

As reformas sofridas pela legislação

Depois da aprovação da lei, ocorreram duas reformas constitucionais no âmbito do art. 6º, o mesmo que fora reformado em 1977 para a incluir a garantia do direito à informação por parte do Estado (México, 2007; 2014).

A primeira foi publicada em julho de 2007, com o objetivo de detalhar os princípios e bases que deveriam balizar o direito de acesso à informação em nível federal e estadual, em razão da diversidade de regulamentações adotadas pelos estados, que variavam em relação à especificação de entidades submetidas à lei,

prazos e procedimentos para o acesso à informação (López-Ayllón; Laborde, 2007). Do mesmo modo, a atuação de órgãos autônomos para garantir o cumprimento da lei também variava de acordo com o estado (López-Ayllón; Laborde, 2007).

Sobre a disparidade de regulamentações adotadas pelos estados, vale mencionar a pesquisa conduzida por Daniel Berliner e Aaron Erlich (2015), que analisou as normas adotadas por 31 estados mexicanos e pelo Distrito Federal. Os pesquisadores observaram que disputas políticas locais funcionam como um grande catalizador deste tipo de legislação. Eles identificaram dois mecanismos que aceleram aprovação de leis de acesso em cenários de competição política. Um deles funciona como uma espécie de “seguro” (*insurance*), em que o grupo político dominante antevê a iminente alternância de poder e apoia a criação da lei para se certificar de que poderá monitorar o grupo rival no futuro. O outro mecanismo gira em torno da perspectiva de reeleição do grupo político dominante, em que a aprovação de uma lei de acesso é vista como um ativo na busca de votos, por demonstrar seu comprometimento com a transparência.

Em fevereiro de 2014, foi publicada no Diário Oficial outra reforma constitucional em matéria de transparência (México, 2014). O decreto implementou o sistema integral de acesso à informação pública e estabeleceu que o Instituto Federal de Acesso à Informação seria transformado em Instituto Nacional de Transparência, Acesso à Informação e Proteção de Dados Pessoais (INAI), ampliando seu escopo de atuação para os três Poderes. Antes, a instituição supervisionava 247 órgãos públicos, após a mudança esse número passou para 872 (Mora, 2016).

A reforma também determinou a criação de uma lei geral sobre o tema. Assim, em maio de 2015 foi publicada a Lei Geral de Transparência e Acesso à Informação Pública. A nova regra ampliou de dezesseis para quarenta e oito o número de informações que devem ser publicadas obrigatoriamente por meio de transparência ativa por todos os órgãos públicos do país – além de outras especificações para determinadas instituições de acordo com sua função, como o Poder Judiciário (ao qual se referem os artigos 70 e 73). Enquanto a primeira lei, de 2002, possibilitava que os demais poderes e órgãos autônomos regulamentassem a norma em seu próprio âmbito, a Lei Geral se aplica

a todos. Outra mudança importante trazida pela nova lei, foi a inclusão de partidos políticos e sindicatos no rol de organizações às quais a regra se aplica.

Novos e velhos desafios

Feito o panorama da trajetória que resultou na aprovação da Lei Federal de Transparência e Acesso à Informação Pública, em 2002, e suas alterações posteriores, neste momento serão apresentados dados e relatos coletados em entrevistas realizadas na Cidade do México, entre outubro de 2019 e março de 2020. Ao consultar pesquisadores, jornalistas e ativistas, o objetivo é identificar novos e velhos desafios que o direito de acesso à informação enfrenta no país, por meio de seus relatos são fornecidos elementos que auxiliam na compreensão do contexto estudado (Duarte, 2012).

Um dos desafios que permanece diz respeito ao perfil de usuários dos mecanismos instituídos pela Lei Geral de Transparência. De acordo com a Pesquisa Nacional de Acesso à Informação Pública e Proteção de Dados Pessoais de 2019 (ENAIID, sigla em espanhol), 48,4% da população considera que obter informação governamental é difícil ou muito difícil. Além disso, 54,8% disse que conhecia ou havia escutado sobre a existência de uma instituição de governo que garante o direito de acesso à informação pública (na pesquisa anterior, realizada em 2015, este índice era de 50,6%). Observa-se, portanto, um avanço gradual.

No entanto, a pesquisa ainda revela que apenas 3,6% da população já fez alguma solicitação de acesso à informação. Deste grupo, 47,9% formalizou o pedido pessoalmente, enquanto 43,1% o fez pela internet. Entre aqueles que não fizeram solicitações, 77,8% afirmou não ter interesse em fazê-lo, enquanto 21,1% manifestou interesse. Ou seja, não se trata somente de orientar sobre os procedimentos para solicitar informações públicas, mas conscientizar sobre o porquê fazê-lo – uma tarefa muito mais desafiadora.

De acordo com o relatório de atividades de 2020 do INAI, o perfil de solicitante frequente (dentre aqueles que informaram os dados) no último ano se caracteriza por ser jovem (menores de 18 anos correspondiam a 17,2% dos solicitantes, seguido pela faixa etária de 35 a 39 anos, com 15,8% e 30 a 34 anos, com 14%) e do sexo masculino (62,9%). Quanto à escolaridade, 46,9% têm ensino superior completo e 32,4%, pós-graduação. Em

7. Disponível em: <https://micrositios.inai.org.mx/infomesinai/?page_id=519> Acesso em 11 dez. 2021.

relação à ocupação, a maioria atua em empresas privadas, em seguida estão os acadêmicos e, em terceiro lugar, os meios de comunicação.

Em entrevista para esta pesquisa, o presidente da *Asociación Mexicana de Derecho a la Información* (Amedi), Jorge Bravo⁸, afirma que difundir o uso da Lei Geral de Transparência entre os jornalistas ainda é uma tarefa pendente.

Quem pede mais? Acadêmicos e empresas. Empresas porque querem saber o que o seu concorrente está fazendo, pedem contratos obtidos. Acadêmicos porque queremos escrever sobre os temas. Mas o jornalista é a grande questão pendente. Ainda não é a ferramenta ideal para quem vive no dia a dia, da declaração na rua. Você não quer esperar um mês pelas informações ou não tem tempo. Se o fizessem, realmente melhoraria muito jornalismo no México. (J. Bravo, comunicação pessoal, 6 de março de 2020, tradução nossa).

Em um evento promovido pelo INAI em 2020, o pesquisador Brian Palmer-Rubin, da Universidade de Marquette, exibiu alguns resultados preliminares de um trabalho coletivo que analisou mais de um milhão de pedidos de acesso à informação protocolados no México, entre 2004 e 2015, utilizando métodos de big data⁹. Diante dos dados levantados pela ENAID e pelo mais recente relatório de atividades do INAI, é importante agregar as conclusões apresentadas por Palmer-Rubin na ocasião:

Observamos um processo de aprendizagem onde vemos que os cidadãos se sofisticam cada vez mais na forma de solicitar informações, na linguagem que têm que usar nas suas solicitações e cada vez mais os órgãos públicos respondem melhor os pedidos mais sofisticados. Ou seja, nos dados que temos [...] observamos que a partir de 2008 há uma crescente especialização do sistema. Torna-se mais fácil para um especialista utilizar o sistema de acesso à informação e mais difícil para uma pessoa sem

tanta sofisticação, sem tanto conhecimento interno da administração pública mexicana. (B. Palmer-Rubin, comunicação verbal, 25 de fevereiro de 2020, tradução nossa).

A análise conduzida pelo pesquisador e sua equipe revela que a qualidade da resposta recebida varia de acordo com a linguagem utilizada na solicitação. A partir de 2008, observou-se um aumento de pedidos com linguagem formal, especializada e com referências legais – justamente estas solicitações levaram menos tempo para serem atendidas, o que não acontecia antes (Berliner et al., 2022). O cenário evidencia uma crescente importância do domínio de conhecimentos sobre a burocracia estatal e de termos específicos da legislação para fazer valer o direito de acesso à informação pública (Berliner et al., 2022).

A principal pergunta é: o quanto o México precisa de um sistema de acesso amplo para todo o público, e o quanto o México precisa de um sistema que se preste a um grupo de elites que representam um grupo de intermediários, que são muito sofisticados para solicitar informação, e então eles, esses intermediários, podem filtrar a informação ao público? (B. Palmer-Rubin, comunicação verbal, 25 de fevereiro de 2020, tradução nossa).

Vale ressaltar que o fenômeno da discriminação no atendimento a pedidos de informação de acordo com a identificação do solicitante é observado quando se tratam de jornalistas (Fox & Haight, 2011). Este problema também já foi verificado no Brasil (Artigo 19, 2018; Michener et al., 2018).

Em um trabalho anterior, os pesquisadores Berliner et al. (2018) utilizaram a mesma amostra de um milhão de pedidos de acesso à informação ao governo mexicano para mapear os principais temas que despertavam interesse entre os usuários da lei. O objetivo era compreender se os assuntos seriam pertinentes sob o viés da *accountability* ou se diziam respeito a interesses particulares¹⁰.

8. Entrevista realizada na Cidade do México, em 6 mar. 2020.

9. O painel Acceso a la Información: Perspectivas y Retos para una Sociedad más Justa, que contou com a apresentação de Brian Palmer-Rubin, está disponível em: <https://www.youtube.com/watch?v=T1jXbT2uWnQ> Acesso em 12 de mar. 2021.

10. Os autores deixam claro que a intenção não é desmerecer pedidos de informação de cunho particular: “Nós não pretendemos denegrir os usos privados. A capacidade de buscar informações para objetivos pessoais pode ajudar a equalizar o acesso a serviços que de outra forma pode ser tendencioso. Oferecer uma plataforma para interação entre cidadão e Estado pode demonstrar a responsividade do governo independente da informação solicitada. No entanto, esse tipo de uso é inconsistente com aqueles que motivaram a criação de sistemas de acesso à informação e podem ser melhor atendidos por meio de um sistema de dados abertos, ao invés de um sistema de solicitação que demanda tempo intensivo” (Berliner et al., 2018, p.227, tradução nossa)

O levantamento revelou que questões ambientais (relatórios de impacto, informações sobre posse e uso da terra, etc.) foram os temas que mais se mostraram frequentes entre os pedidos. Em seguida aparecem as solicitações agrupadas como “necessidades individuais”, que consistem em pedidos de informação sobre como se inscrever em determinado programa de microcrédito do governo, por exemplo, ou dúvidas sobre procedimentos para ter acesso a benefícios sociais. O terceiro tema mais buscado é referente a orçamentos e despesas do governo federal (em especial os gastos com publicidade).

Informações comerciais são o quarto assunto mais demandado pelos usuários da lei de acesso mexicana, como registros de empresas, processos administrativos envolvendo pessoas jurídicas e registros de marcas. O levantamento mostra que esse tipo de solicitação vem aumentando com o passar do tempo, o que demonstra que o ambiente empresarial tem se apropriado cada vez mais da Lei Geral de Transparência como uma ferramenta para os negócios.

Por fim, o quinto tema mais buscado pelos usuários é referente à segurança pública, crimes e polícia. Os pesquisadores observaram que o volume de demandas sobre este assunto se mantém estável ao longo dos anos, mas alcança picos em momentos de crise de segurança, como em episódios de guerra contra o narcotráfico. Em entrevista aos pesquisadores, uma autoridade do governo federal resumiu em uma frase a ligação entre os eventos de interesse público e as demandas que chegam pela lei: “A manchete de hoje é o pedido de informação de amanhã” (Berliner et al., 2018, p. 231, tradução nossa).

Os autores concluem que os cidadãos mexicanos de fato solicitam informações consideradas relevantes sob o viés da *accountability*:

Assim, oferecemos uma resposta afirmativa retumbante à pergunta ‘Se você construir, eles virão?’. Isso sugere que, dado um sistema de acesso à informação razoavelmente funcional, com poucas barreiras de uso, a demanda por informação não é o obstáculo-chave para o sucesso da *accountability* baseada em informações. (Berliner et al., 2018, p.233, tradução nossa).

A despeito das boas avaliações da lei mexicana, o pesquisador Guillermo Cueto (2017) questiona porquê o país continua figurando em rankings internacionais com altos índices de percepção de corrupção¹¹. A resposta, segundo o autor, é simples: mais transparência não significa, necessariamente, mais *accountability*. É necessário investir em fiscalização e processos de prevenção à corrupção. “Propor um sistema de prestação de contas em que a única figura concreta seja a política de transparência ou o direito de acesso à informação é deixar esse mesmo sistema coxo” (Cueto, 2017, p.97, tradução nossa).

O autor também destaca a necessidade de difundir os mecanismos de acesso à informação para todos os estratos da sociedade, assim como traduzir as informações disponibilizadas, de forma a lograr um impacto social maior. “Tornar a informação transparente não adianta nada se não conseguimos transformar essa informação em conhecimento” (Cueto, 2017, p.97, tradução nossa).

Na mesma linha, ao analisar os espaços de participação cidadã no país, Mariñez Navarro (2021) inclui a Lei Geral de Transparência no conjunto de marcos normativos que formam o ambiente propício para a prestação de contas e fiscalização. Também integram o grupo, a Lei Geral de Arquivos, a Lei Geral do Sistema Nacional Anticorrupção, a Lei de Fiscalização e Prestação de Contas da Federação, entre outras. Há, portanto, a construção de um ambiente normativo no qual a transparência é apenas um dos elementos de *accountability*.

Além disso, pesquisadores e ativistas entrevistados no âmbito do presente estudo fazem críticas à Lei Geral de Transparência, que, segundo eles, teria se tornado demasiadamente “burocrática” em razão da quantidade de itens obrigatórios a serem informados por meio da transparência ativa por todos os órgãos submetidos à norma. É o que afirma a professora da Universidade Nacional Autônoma do México Jacqueline Peschard¹², que desempenhou a função de comissionária-presidente do então IFAI entre 2009 e 2013:

O problema é que se trata de uma lei muito rigorosa, mas também implica uma enorme responsabilidade do órgão fiscalizador. É um órgão que não chega a vigiar a todos porque lhe deram tantas competências que o que

11. Na edição de 2020 do Índice de Percepção da Corrupção, desenvolvido pela ONG Transparência Internacional, o México somou 31 pontos, ocupando a 124ª posição entre 180 países. O Brasil obteve 38 pontos e ficou em 94º. Disponível em: <https://www.transparency.org/en/cpi/2020/index/bra> Acesso em 9 dez. 2021.

12. Entrevista realizada na Cidade do México, em 30 de out. 2019.

faz então é uma vigilância muito formalista, muito baseada em: preencha estes formulários e com o fato de preencher estes formulários, eu já te fiscalizei. Mas a verdade é que não. (J. Peschard, comunicação pessoal, 30 de outubro de 2019, tradução nossa).

Em entrevista para a presente pesquisa, o diretor do Programa de Direito de Acesso à Informação da ONG Artigo 19 no México, Carlos Dorantes¹³, concorda que a grande quantidade de informações a serem divulgadas em transparência ativa por todos os órgãos públicos é algo que deveria ser revisto.

É uma lei bem avançada, eu digo que tem seus entraves, seus problemas, mas mesmo assim, se essa lei fosse cumprida, seria o ideal. O que acontece é que não se cumpre, então finalmente não importa tanto se está num ranking internacional de leis de transparência. Entre as dificuldades que considero que esta lei tem, uma é a transparência judicial, outra é a quantidade de obrigações de transparência que tem tornado muito difícil para as instituições cumpri-las (C. Dorantes, comunicação pessoal, 10 de março de 2020, tradução nossa)

Também em entrevista à autora, o pesquisador Gregory Michener¹⁴ endossa a crítica:

As obrigações de transparência ativa são tão onerosas e extensas, é uma lista tão grande que ninguém consegue. Você passa o tempo todo atualizando os dados e tentando cumprir com a lei. Eu já falei com muitas pessoas

no Executivo Federal e eles falam ‘a gente não está cumprindo porque a gente não consegue’. Então eles foram além do limite. (G. Michener, comunicação pessoal, 14 de novembro de 2019).

Além das dificuldades pontuadas por pesquisadores, jornalistas e ativistas, o órgão responsável pela fiscalização do cumprimento das políticas de acesso à informação e proteção de dados pessoais tem enfrentado outro desafio, de cunho político.

Eleito em 2018, o presidente Andrés Manuel López Obrador, do partido Morena, já questionou a eficácia do INAI em diversas coletivas de imprensa. “Custa um bilhão de pesos manter este órgão, ele foi criado e em que contribuiu para reduzir a corrupção? Ao contrário, a corrupção cresceu como nunca ao mesmo tempo em que se criou este órgão” (Redación AN/GV, 2018), disse em dezembro de 2018. Dois anos depois, em dezembro de 2020, reafirmou o argumento em outra coletiva de imprensa: “Para que servem estes órgãos? São decorativos. E, antes, para simular que, desta maneira, o governo se preocupava para que não houvesse corrupção” (Caña, 2020). Além das críticas públicas ao INAI, levantamentos revelam que o número de recursos questionando negativas de acesso à informação aumentou desde o início da gestão do novo presidente (Langner, 2020). Como efeito deste contexto político, o orçamento do INAI vem enfrentando reduções nos últimos anos, consequência da perspectiva do governo sobre a importância das políticas de acesso à informação pública, como mostra a tabela abaixo:

TABELA 1 – EVOLUÇÃO DO ORÇAMENTO DO INAI ENTRE 2017 E 2020. VALORES EM PESOS MEXICANOS

ANO	ORÇAMENTO - INAI
2017	\$ 896.931.201,11
2018	\$ 1.041.623.750,75
2019	\$ 164.715.547,60
2020	\$ 135.847.505,55

Fonte: *Elaboração própria a partir de base de dados INAI, 2020¹⁵.*

13. Entrevista realizada na Cidade do México, em 10 mar. 2020.

14. Entrevista realizada na Cidade do México, em 14 nov. 2019.

15. Informações orçamentárias disponíveis em: <<https://micrositios.inai.org.mx/planeacion/>> Acesso em 12 jun. 2021.

A percepção de que os escândalos de corrupção aumentaram se deve justamente ao fato de que, com efetivo acesso à informação, é possível trazer à tona irregularidades (Heald, 2006). É o caso de investigações jornalísticas que utilizaram ferramentas de transparência que revelaram o escândalo conhecido como *La Casa Blanca*¹⁶ em 2014, na gestão do presidente Enrique Peña Nieto, por exemplo. Ou a descoberta de milhares de fossas clandestinas pelo país, resultantes da política de guerra contra o narcotráfico entre 2006 e 2016¹⁷.

Legislações de acesso à informação tornam-se instrumentos norteadores de políticas públicas cujos efeitos são a longo prazo e difusos, difíceis de serem quantificados (Dutra, 2021). Essas características fazem com que estas normas possam se tornar alvos de governantes que não veem vantagens imediatas que justifiquem o investimento em políticas públicas voltadas ao direito de acesso a informações – que podem inclusive gerar revezes ou custos reputacionais ao revelar malfeitos, falhas ou omissões por parte da administração pública (Dutra, 2021).

CONCLUSÕES

Brasil e México são países latino-americanos marcados por heranças ibéricas similares. Tanto Portugal quanto Espanha buscaram aplicar em suas colônias regras e modelos administrativos pré-fabricados em seus respectivos reinos. A despeito das singularidades da trajetória histórica de cada país, temos Estados marcadamente patrimonialistas, que nasceram fechados, distantes da população, e um sistema político cujas atenções se voltam majoritariamente para o chefe do Executivo. Todas essas particularidades conferem aos dois países muitos pontos de congruência no que tange às políticas de acesso à informação pública e proporcionam oportunidades de aprendizado.

No México, adota-se como marco inicial desta trajetória o ano de 1977, com a reforma constitucional do artigo 6º, que incluiu o direito na carta magna do país. No entanto, ele se restringiu, em um primeiro momento, à garantia de espaços midiáticos a partidos políticos. Em 1996, a partir de uma mudança de posi-

cionamento da SCJN, este entendimento é ampliado e o direito à informação passa a se vincular ao direito à verdade.

A ascensão de Vicente Fox (PAN) ao poder na virada dos anos 2000 marcou o fim de uma hegemonia de 72 anos do PRI. Neste momento, teve início um processo de transição democrática em que um projeto de lei de acesso à informação começou a ser discutido por organizações da sociedade civil e pelo governo. Concomitantemente, foram aprovados outros instrumentos normativos com o objetivo de aumentar a transparência na administração pública. Finalmente, em 2002, a Lei Federal de Transparência e Acesso à Informação Pública Governamental foi enfim aprovada.

A regra sofreu alterações, e em 2015 foi sancionada a Lei Geral de Transparência e Acesso à Informação Pública, em vigor atualmente. Diante de todo este caminho percorrido, foi possível identificar atores e instituições que desempenharam importante papel, com destaque para a atuação da sociedade civil, na figura do Grupo Oaxaca.

No entanto, vinte anos depois da sanção da primeira lei de acesso, é possível indicar alguns desafios que se apresentam no contexto mexicano. O primeiro deles diz respeito à popularização dos mecanismos de acesso à informação, que precisam ser amplamente difundidos com o objetivo de diversificar o perfil de solicitantes. O segundo foi apontado em entrevistas durante a presente pesquisa: a excessiva burocratização da lei, que estabelece um extenso rol de requisitos a todos os órgãos da administração pública. O que pode ter partido de uma intenção de aumentar o volume de informações disponíveis pode, ao final, não ser passível de execução e fiscalização.

Finalmente, o terceiro desafio decorre de mudanças no cenário político, que impactam desde as atividades de órgãos fiscalizadores como INAI (colocando em dúvida sua eficácia e reduzindo seu orçamento) até a qualidade no atendimento aos pedidos de acesso à informação enviados ao governo. A experiência mexicana demonstra que a defesa do direito de acesso à informação vai além da aprovação de marcos normativos. Carece de constante vigilância para evitar retrocessos e garantir avanços.

16. O escândalo conhecido como "La Casa Blanca" foi revelado em 2014 pela equipe da jornalista Carmen Aristegui. Trata-se da aquisição de uma mansão pelo valor de 7 milhões de dólares pelo ex-presidente mexicano Enrique Peña Nieto, que comprou o imóvel de um empreiteiro com negócios no governo. Disponível em: <<https://aristeguinoicias.com/0911/mexico/la-casa-blanca-de-enrique-pena-nieto/>> Acesso em 6 maio 2021.

17. Em reportagem publicada em 2018, jornalistas revelaram a descoberta de quase duas mil fossas, onde eram enterradas vítimas de criminosos entre 2006 e 2016. A publicação rendeu diversos prêmios jornalísticos internacionais. Disponível em: <<https://adondevanlosdesaparecidos.org/2018/11/12/2-mil-fosas-en-mexico/>> Acesso em 6 maio 2021.

REFERÊNCIAS

- Artigo 19 Brasil. (2018). *Identidade revelada: entaves na busca por informação pública no Brasil*. Artigo 19.
- Berliner, D.; Palmer-Rubin, B.; Reyes, J. E. T.; Bagozzi, B., & Erlich, A. (2022). *Big data y acceso a la información en México*. OSF Preprints.
- Berliner, D., Bagozzi, B. E., & Palmer-Rubin, B. (2018). What information do citizens want? Evidence from one million information requests in Mexico. *World Development*, 109, 222-235.
- Berliner, D., & Erlich, A. (2015). Competing for transparency: political competition and institutional reform in Mexican states. *American Political Science Review*, 109(1), 110-128.
- Bobbio, N. (2004). *A era dos direitos*. Elsevier.
- Caña, P. V. y. (2020, 7 de dezembro) Organismos de transparencia, como el INAI, son floreros: AMLO. *El Universal*. Disponível em: <https://bityli.com/R7Xbw7> Acesso em 6 maio 2021.
- Carvalho, J. M. de. (2012). *Cidadania no Brasil: o longo caminho*. Civilização Brasileira.
- Controladoria-Geral da União. (2008). *Ata da 9ª reunião do Conselho da Transparência Pública e Combate à Corrupção*. Brasília, 9 set., 6p. Disponível em: <https://www.gov.br/cgu/pt-br/assuntos/transparencia-publica/conselho-da-transparencia/documentos-de-reunioes/atas/ata-da-9a-reuniao-set-2008.pdf> Acesso em 5 ago. 2021.
- Costa, A. B. L., Dutra, L. P., & Paulino, F. O. (2021). Derecho de acceso a la información en Brasil y México: entre cabestros, jefes y sociedad civil. *Millcayac - Revista Digital De Ciencias Sociales*, 15(8), 117-142.
- Cueto, G. A. T. (2017). El Derecho de acceso a la información en Iberoamérica y su concreción como garantía constitucional: El caso mexicano. *Gestión y Análisis de Políticas Públicas*, (17), 79-99.
- Delgado, J. F. E. (Ed.) (2020). *La agenda democrática del Grupo Oaxaca: balance y futuro*. INAI.
- Doyle, K. (2002). Mexico's new Freedom of Information Law. *The National Security Archive*. Cidade do México, 10 jun.
- Duarte, J. (2012). Entrevista em profundidade. In: J. Duarte, & A. Barros (Eds.), *Métodos e técnicas de pesquisa em comunicação*. Atlas.
- Dutra, L. P. (2021). Acesso à informação e comunicação pública: uma análise das Supremas Cortes do Brasil e do México. 224 p. Tese (Doutorado em Comunicação), Faculdade de Comunicação, Universidade de Brasília, Brasília.
- Faoro, Raymundo. (2001). *Os donos do poder*. Globo.
- Fox, J., & Haight, L. (2011). Mexico's transparency reforms: theory and practice. *Government Secrecy*, 19, 353–379.
- Heald, D. (2006). Varieties of transparency. In: C. Hood, & D. Heald (Eds.), *Transparency: The key to better governance?* Oxford University Press.
- Instituto Nacional de Estadística y Geografía. (2020). *Encuesta Nacional de Acceso a la Información Pública y Protección de Datos Personales (ENAIID) 2019*. México.
- Langner, A. (2020, 22 de janeiro). Aumentan declaraciones de inexistencia de información: Inai. *La Jornada*. Disponível em: <https://bityli.com/chAmgA> Acesso em 6 jun. 2020.
- López-Ayllón, S.; Laborde, M. M. (Eds.) (2007). *La transparencia en la república: un recuento de buenas prácticas*. IFAI, CIDE.
- López-Ayllón, S. (2005). Democracia y rendición de cuentas: La Ley mexicana de transparencia y acceso a la información. *Documentación Administrativa*, (273).

- _____. (2002). Comentario a la iniciativa de ley federal de transparencia y acceso a la información presentada por el Ejecutivo Federal al Congreso de la Unión. *Cuestiones Constitucionales*, 7, 261-274.
- Mariñez Navarro, F. (2021). Participación Ciudadana Colaborativa, Rendición de Cuentas y Fiscalización. El caso de México. *Revista da CGU*, 13(23), 71-85.
- Marshall, T. H. (1967). *Cidadania, classe social e status*. Zahar.
- México (2018). *Ley General de Archivos*. Diario Oficial. Disponible em: <<https://bitly.com/iZLKgF>>
- _____. (2016). *Ley de Fiscalización y Rendición de Cuentas de la Federación*. Diario Oficial. Disponible em: <<https://bitly.com/tvXHax>>
- _____. (2016). *Ley General del Sistema Nacional Anticorrupción*. Diario Oficial. Disponible em: <<https://bitly.com/JtXgfG>>
- _____. (2015). *Ley General de Transparencia y Acceso a la Información Pública*. Diario Oficial. Disponible em: <<https://bitly.com/slBkMe>>
- _____. (2014). *Decreto por el que se reforman y adicionan diversas disposiciones de la Constitución Política de los Estados Unidos Mexicanos, en materia de transparencia*. Diario Oficial. Disponible em: <<https://bitly.com/mtUnOo>>
- _____. (2007). *Decreto por el que se adiciona un segundo párrafo con siete fracciones al Artículo 6o. de la Constitución Política de los Estados Unidos Mexicanos*. Diario Oficial. Disponible em: <<https://bitly.com/gELstO>>
- _____. (2002). *Ley federal de transparencia y acceso a la información pública gubernamental*. Diario Oficial. Disponible em: <<https://bitly.com/epVZmp>>
- Michener, G., Contreras, E., & Niskier, I. (2018). Da opacidade à transparência? Avaliando a Lei de Acesso à Informação no Brasil cinco anos depois. *Revista de Administração Pública*, 52(4), 610-629.
- Michener, G. (2011). FOI Laws around the world. *Journal of Democracy*, 22(2), 145-159.
- Mora, X. P. de la. (2015, 11 de novembro). *Implementación de la Reforma en Materia de Transparencia: Retos y Perspectivas para el Poder Judicial de la Federación*. [Apresentação de Painel]. Seminario Internacional de Transparencia Judicial, 'El Nuevo Modelo de Acceso a la Información y Justicia Abierta', Cidade do México. Disponible em: <https://bitly.com/yQSrde>
- Osmo, C. (2014). *Direito à verdade: origens da conceituação e suas condições teóricas de possibilidade com base em reflexões de Hannah Arendt*. 299 p. Tese (Doutorado em Direito), Faculdade de Direito, Universidade de São Paulo, São Paulo.
- Paz, O. (2014). El ogro filantrópico. In O. Paz, *El peregrino en su patria: historia y política de México*. (2 ed., pp. 316-331). FCE.
- Pla, I. L. (2009). *Movimiento social del derecho de acceso a la información en México*. Universidad Nacional Autónoma de México.
- Quintanilla-Madero, C. (1988). Toward a law of public access to government information in Mexico. *U. Miami Inter-Am. L. Rev.*, 20, 93-113.
- Redación AN/GV. (2018, 7 de dezembro). AMLO critica sueldos del Inai; "corrupción creció a la par" de que se creó este organismo, dijo. *Aristegui Noticias*. Disponible em: <<https://bitly.com/IjCVjK>> Acceso em 8 jun. 2020.
- Robledo, M. M. (2014). Acceso a la información, transparencia y rendición de cuentas en México: 37 años de evolución. *El Cotidiano*. 203-215.

Sobel, D., Davis Noll, B., Fernandez Bogado, B., Price, M., & TCC Group. (2006). *The Federal Institute for Access to Information in Mexico and a Culture of Transparency*. Other

Publications from the Center for Global Communication Studies.

Suprema Corte de Justicia de la Nación. (2000). *Semanario Judicial de la Federación y su Gaceta, novena época, Pleno, tomo XI, Abril, tesis P. XLV/2000*, p. 72. Registro: 191981. Disponível em: <https://sjf2.scjn.gob.mx/detalle/tesis/191981>

_____. (1996). *Semanario Judicial de la Federación y su Gaceta, novena época, Pleno, tomo III, junio, tesis P. LXXXIX/96*, p. 513. Registro: 200111. Disponível em: <https://sjf2.scjn.gob.mx/detalle/tesis/200111>

_____. (1992). *Semanario Judicial de la Federación, Octava Época, Segunda Sala, tomo X, agosto, Tesis 2a. I/92*, p. 44. Registro: 206435. Disponível em: <https://sjf2.scjn.gob.mx/detalle/tesis/206435>



Luma Poletti Dutra

lumadutra@gmail.com

<https://orcid.org/0000-0003-0834-6359>

Jornalista, doutora e mestra em Comunicação pela Universidade de Brasília, extensionista da Universidade Federal do Mato Grosso.

 cguonline  cguonline  cguoficial  cguonline  cguoficial

CONTROLADORIA-GERAL
DA UNIÃO

