

DEZ ANOS DE PRODUÇÃO DIEST: UM LEVANTAMENTO DAS AGENDAS TEMÁTICAS NA ANÁLISE DE POLÍTICAS PÚBLICAS¹

Natália Massaco Koga²

Pedro Palotti³

Rafael Lins⁴

Isabella Goellner⁵

1 INTRODUÇÃO

A relevância de estudos e pesquisas para subsidiar políticas públicas é um tema debatido desde a origem do campo de análise de políticas públicas (Laswell e Lerner, 1951; Weiss, 1979). Mais recentemente, a abordagem das Políticas Públicas Baseadas em Evidências (PPBE) retoma, por um lado, a convicção nos preceitos da racionalidade instrumental e da neutralidade científica como forma de embasamento das decisões de políticas públicas (Davies, Nutles e Smith, 2000) e, por outro, catalisa críticas provenientes de correntes analíticas mais recentes no campo, tais como a argumentativa ou pós-estruturalista, que dão base ao argumento da necessidade do alargamento do entendimento de evidências para além das evidências científicas (Pinheiro, 2020) e da compreensão do que ocorre na interface ou que a impede entre o campo acadêmico e da gestão pública (Caplan, 1979; Newman, Cherney e Head, 2016).

Entendemos que a atuação da Diretoria de Estudos e Políticas do Estado, das Instituições e da Democracia (Diest) encontra-se nessa interface e merece ser explorada a partir da perspectiva de análise de sua contribuição *sobre* e para a produção das políticas públicas. Vale lembrar, neste sentido, as competências institucionais da Diest, previstas no Estatuto do Ipea que podem ser sintetizadas em ao menos quatro papéis: i) assessoramento governamental; ii) produção de conhecimento; iii) disseminação de conhecimento; e iv) capacitação (entendida como uma forma de disseminação) “em questões ligadas à estrutura, organização e funcionamento do Estado brasileiro e de seus aparatos institucionais, bem como aos modos de relação entre o Estado e a sociedade nos processos de elaboração, implementação, monitoramento e avaliação de políticas públicas para o desenvolvimento do país” (Brasil, 2010).

Vale destacar que, nos casos da produção e disseminação de conhecimento, essas funções não estariam apenas dirigidas ao próprio governo, mas também para a sociedade em geral. Sustentamos que tais papéis tratam não apenas da produção de conhecimento, mas também se aproximam ao que estudos empíricos trazem acerca da função dos *knowledge brokers*, atores intermediários que atuam reduzindo o distanciamento entre as “duas comunidades” ao identificar demandas, traduzir,

1. DOI: <http://dx.doi.org/10.38116/bapi29art3>

2. Especialista em políticas pública e gestão governamental na Diretoria de Estudos e Políticas do Estado, das Instituições e da Democracia (Diest) do Ipea.

3. Especialista em políticas pública e gestão governamental na Diest/Ipea.

4. Pesquisador do Programa Nacional de Pesquisa Para o Desenvolvimento (PNPD) na Diest/Ipea.

5. Pesquisadora do PNPD na Diest/Ipea.

reinterpretar e difundir conhecimento em uma cadeia fluida e dinâmica de transmissão entre a Academia e a gestão (Ward, 2012; Newman, Cherney e Head, 2016).

Partindo desse escopo analítico, buscamos, neste artigo, realizar um estudo bibliométrico preliminar da produção da Diest ao longo de sua década de existência. Em que pese as limitações da base de dados hoje existente, entendemos que o estudo permite jogar luz sobre áreas temáticas elegidas na produção de conhecimento e, em alguma medida, escolhas de abordagens analíticas e estratégias de disseminação da produção da diretoria. A fim de interpretar os dados gerados, buscamos dialogar com alguns capítulos da publicação que detalham principais escolhas e caminhos ensejados e trilhados.

2 ESTRATÉGIA METODOLÓGICA

Para a elaboração do estudo descritivo proposto neste trabalho, iremos nos valer de uma base de dados constituída pela biblioteca do Ipea com as principais obras produzidas por autores provenientes da Diest no período de 2010 a 2019. Assumimos que os trabalhos produzidos por pesquisadores que fizeram parte da diretoria, em algum momento de sua trajetória profissional no Ipea, nesse período, ou produções de autores externos ao Ipea que participaram de publicações organizadas pela Diest foram, desse modo, resultados da contribuição da diretoria.

A base de dados utilizada neste artigo corresponde à produção bibliográfica encontrada no repositório do Ipea e nos livros editados entre 2010 e 2019. O repositório é uma base que armazena todas as publicações do Ipea em formato digital. Assim, caso a publicação não tenha uma versão digital, ela não estará disponível no repositório. No caso da Diest, no entanto, a produção não digitalizada é mínima, uma vez que a diretoria foi criada no momento de transição do Ipea para publicações sobretudo digitais.

As unidades bibliográficas consideradas na base compreendem todas as formas de publicação com circulação independente, o que inclui *Texto para Discussão* (TD), *Nota Técnica*, livros, capítulos de livros,⁶ relatórios de pesquisa, comunicações e artigos do *Boletim de Análise Político-Institucional* (Bapi). Não foram consideradas publicações externas, como as realizadas em congressos científicos e revistas indexadas.

O estudo adota o método estatístico Structural Topic Model (STM) e de análise de grafos. O STM é um método particularmente popular para ajustar um modelo de tópicos não supervisionado. Como um modelo de tópicos, ele propõe automaticamente uma classificação das palavras significantes para o pesquisador, em um conjunto de tópicos a partir dos documentos. Um grande atrativo é a possibilidade de descobrir tópicos e estimar sua relação com os metadados arbitrários sobre cada documento (Roberts *et al.*, 2019).

A análise de grafos permitiu a produção de uma rede compostas por vértices (palavras-chave), conectados por arestas (co-ocorrências). O posicionamento dos vértices foi obtido a partir de um algoritmo de posicionamento direcionado à força que atrai vértices com conectividade forte e repele aqueles com

6. Os capítulos de livro foram considerados quando parte de coletâneas, isto é, quando os organizadores não necessariamente são os únicos autores e os capítulos podem ser compreendidos por unidades independentes. Foram consideradas também partes do livro com o formato de apresentação, introdução ou prefácio. Vale destacar, no entanto, que nem todos os capítulos de livros constam do repositório, uma vez que, para serem desmembrados como unidades independentes para fins de disponibilização no relatório, demandam processo editorial de identificação separadamente do arquivo do livro do qual faz parte. Esse processo ainda não foi concluído para todos os livros.

conectividade baixa (Opsahl, Agneessens e Skvoretz., 2010). Assim, os vértices com menor conectividade na rede estão localizados na periferia do gráfico. Além disso, os nós com maior conectividade têm rótulos maiores, e os laços mais fortes entre os nós são mais escuros do que outras arestas. A conectividade entre as palavras-chave foi medida utilizando os pesos de Newman-Fowler (Fowler, 2006). O apêndice, por seu turno, trará algumas limitações da base de dados e o detalhamento do método de STM.

3 DISCUSSÃO DOS RESULTADOS

Apresentamos nesta seção os principais dados levantados por meio da análise de tipos de publicações e áreas temáticas da produção da Diest.

3.1 Tipos de publicações

A tabela 1 retrata o quantitativo dos principais tipos de publicações utilizados pela Diest no período de 2009 a 2019. Do total de 584 publicações, 215 foram artigos do Bapi, 120 TDs, 57 livros e 44 *Notas Técnicas*.⁷ A base disponível não permitiu identificar a quantidade de capítulos de livros que permitiria retratar com mais precisão o esforço total de produção da diretoria. De todo modo, os dados permitem identificar os principais veículos de disseminação adotados pela Diest.

TABELA 1
Principais tipos de publicação (2009-2019)

Tipo de publicação	Quantidade
Bapi	215
TD	120
Livros	57
Nota Técnica	44
<i>Brasil em Desenvolvimento</i> – capítulos	30
<i>Discussion Paper</i>	21
Comunicados do Ipea	21
<i>Desafios do Desenvolvimento</i> – artigos	18
Relatório de pesquisa	15
<i>Radar</i> – artigos	14
Outros	29
Total	584

Elaboração dos autores.

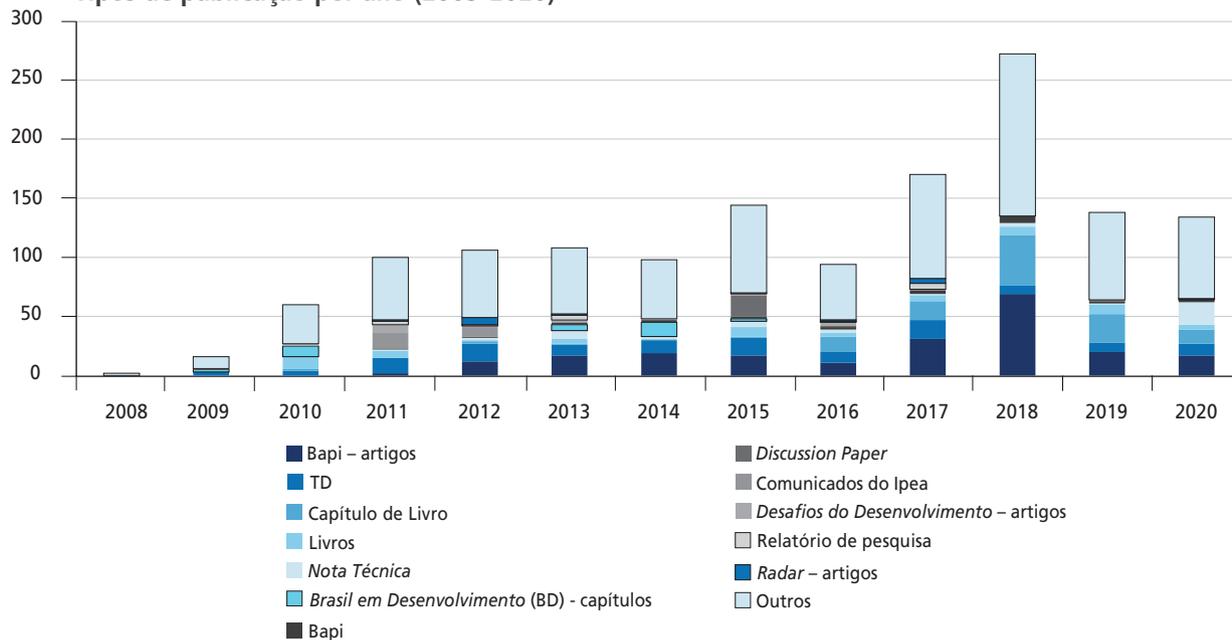
O gráfico 1 traz a quantidade de tipos de publicações por ano. Um primeiro destaque a ser feito diz respeito ao número significativo de artigos publicados por meio do Bapi, periódico criado pela Diest no final de 2011, com o intuito de ser um veículo informativo de formato leve e acessível para disseminar o trabalho dos técnicos da Diest junto a gestores públicos, pesquisadores e público interessado na agenda de pesquisa da diretoria. Este dado retrataria o esforço de disseminação relatado no segundo artigo deste boletim, de autoria de Flávia Schmidt e Janine Mello, com este tipo de veículo de publicação ágil e focado.

7. Na categoria outros encontram-se publicações como artigos para periódicos externos à Diest, como *Planejamento e Políticas Públicas* (PPP), *Boletim Regional, Urbano e Ambiental* (Brua) e *Boletim de Economia e Política Internacional* (Bepi), além de outros documentos como apresentações, slides e relatórios de atividades.

Os TDs e *Discussion Papers* (TDs traduzidos para a língua inglesa) são um veículo importante de divulgação do conhecimento produzido pela Diest, com uma média de doze publicações ao ano. Os livros também estão presentes ao longo de todo o período, tendo destaque para o ano de 2010, com dez livros publicados e, mais recentemente, em 2018, com sete, e 2019, com 8. As *Notas Técnicas* também podem ser encontradas, com variações, nos dez anos. Outros formatos, como *Comunicados do Ipea* e *Brasil em Desenvolvimento*, foram incentivados pela gestão em determinados períodos, sendo depois descontinuados.

Em linhas gerais, 2018 apresentou um número mais elevado de publicações, com 136 no total, mais que o dobro da média do período (58,4%), em função especialmente dos 69 artigos publicados no Bapi naquele ano. O gráfico 1 levanta a hipótese da caracterização de um modo próprio de produção e disseminação da Diest envolvendo seus três veículos mais permanentes, Bapi, TD e livro, em muitos casos de modo sequencial.

GRÁFICO 1
Tipos de publicação por ano (2009-2020)



Elaboração dos autores.

3.2 Áreas temáticas

Foram utilizadas duas fontes de dados para analisar as principais áreas temáticas às quais a Diest tem se dedicado. A primeira trata das palavras-chave, por meio da nuvem de palavras, e a segunda, os resumos identificados em cada publicação, por meio da análise de grafos e metodologia de STM.

A figura 1 retrata os termos de maior prevalência entre as principais palavras-chave utilizadas nas publicações da Diest, e, como tais, denotariam a forma como o estudo pretende ser localizado e a área de produção com a qual busca se engajar. Nesse sentido, apresentaria, em alguma medida, como a produção da Diest buscou ser identificada e disseminada junto aos potenciais leitores e buscadores.

A chave-analítica políticas públicas parece ser o objeto de análise comum de grande parte dos trabalhos, como esperado, dada a missão institucional do Ipea. Além dos termos gerais que denotam a atuação estatal, tais como Estado e administração pública, termos que caracterizam temáticas mais específicas como participação social; segurança pública; democracia; desenvolvimento econômico e social; planejamento governamental; entre outros, também se demonstraram proeminentes.

FIGURA 1
Diest: nuvem de palavras-chave das publicações



Elaboração dos autores.

Obs.: Figura cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Quando analisamos, ainda que de forma preliminar, os resumos das publicações por meio da análise de grafos representada na figura 2, percebemos convergências com os termos mais representados nas palavras-chave apresentados na figura 1. A figura 2 permite ainda identificar conexões presentes entre os distintos termos utilizados nos resumos. Novamente, podemos estabelecer um diálogo entre esses dados e as reflexões apresentadas por José Celso Cardoso Júnior (artigo 1) e Luseni Aquino e Maria Paula Santos (artigo 16) nesta publicação quanto ao nível de abordagem micro-meso e mais funcionalista, assumidos pela Diest para tratar as principais temáticas às quais o Ipea se dedica.

No que se refere ao debate de desenvolvimento, os dados apontam para um enfoque da Diest na atuação estatal, além de uma visão abrangente que abarca o desenvolvimento econômico e social. Em torno da agenda de segurança pública, também relacionada a políticas públicas e atuação estatal, forma-se uma sub-rede ligada a violência, política de drogas, homicídio e armas. Outras sub-redes facilmente identificáveis seriam as formadas em torno de temáticas específicas da participação social, Judiciário, saúde, empresas estatais, infraestrutura, agências reguladoras e mercado de trabalho, as quais se conectam em alguma medida. Entre elas, observamos também temas transversais, como inovação, planejamento e capacidades estatais.

do Estado. De fato, chama atenção a ausência do termo democracia ou expressões associadas tanto aos princípios republicanos de atuação do Estado como garantidor de direitos, liberdades e do bem comum, como aos princípios democráticos no exercício do poder público, como aventam as autoras no artigo 16 desta publicação. Enfim, uma análise mais aprofundada da rede temática da Diest, seus conglomerados e conexões mereceria ser conduzida a partir deste primeiro olhar.

Ainda buscando explorar as escolhas temáticas da produção da Diest, o quadro 1 traz uma tentativa de síntese temática dos vinte principais tópicos localizados, pela técnica do STM. Dois indicadores subsidiaram a síntese: o de maior probabilidade de um termo estar no tópico e de exclusividade de um termo ao tópico. A terceira e quarta coluna do quadro 1 trazem, assim, as dez principais palavras relacionadas aos dois indicadores, em ordem decrescente de relevância.

QUADRO 1
Tópicos e temas extraídos do STM

Tópico	Tema	Maior probabilidade de o termo estar no tópico	Exclusividade do termo ao tópico
Tópico 1	Presidência e núcleo de governo	Segurança, pública, agenda, governo, políticas, presidência, participação	Presidência, comunitários, segurança, janeiro, durante, entrevistas, república
Tópico 2	Bapi	Estudos, política, estado, democracia, diretoria, políticas, instituições	Diretoria, divulgar, democracia, temas, estudos, relevância, fomentar
Tópico 3	Saúde	Saúde, públicos, serviços, público, Brasil, discute, gastos	Renúncia, provisão, desigualdades, saúde, gasto, federativos, gastos
Tópico 4	Desenvolvimento	Desenvolvimento, países, China, internacional, parte, crise, questões	China, crise, chinês, unidos, chinesa, divisão, países
Tópico 5	Dados e informações: organização da sociedade civil (OSC), homicídios, letalidade, exportação, gênero	Brasileiro, OSCs, apresenta, evolução, informações, dados, análise	OSCs, letalidade, brasileiro, exportações, gênero, evolução, federalismo
Tópico 6	Legislativo, política e relações com parlamento	Política, reforma, representação, base, propostas, congresso, resultados	Vereador, Parlamento do Mercosul (Parlasul), eleições, local, reforma, eleitores, representação
Tópico 7	Governança e estatais	Governança, pública, estatais, políticas, administração, empresas, desenvolvimento	Governança, estatais, empresas, administração, perspectivas, conceito, regulatória
Tópico 8	Participação social	Participação, conferências, nacionais, social, participativas, públicas, conselhos	Conferências, participação, participativas, nacionais, conselhos, participativo, participativos
Tópico 9	Justiça e direitos	Pesquisa, geral, programas, direito, políticas, nacional, questões	Direito, conselheiros, geral, comunidades terapêuticas (CTs), empírica, pesquisa, programas
Tópico 10	Inovação e soluções de implementação	Inovação, objetivo, gestão, processos, implementação, programa, serviços	Soluções, dívida, Agência Nacional de Vigilância Sanitária (Anvisa), controle, identificação, interno, desafio
Tópico 11	Regulação e infraestrutura	Processo, Brasil, obras, infraestrutura, trata, principais, nacional	Transportes, mecanismo de desenvolvimento limpo (MDL), obras, transação, educação, judiciais, carbono
Tópico 12	Violência e exclusão social	Drogas, lei, armas, uso, políticas, social, Brasil	Drogas, armas, rua, fogo, uso, lei, crimes
Tópico 13	Burocracia e implementação	Políticas, públicas, desenvolvimento, burocracia, implementação, Estado, meio	Burocracia, cidade, manifestações, político-institucionais, implementação, volumes, mecanismos
Tópico 14	Judiciário, sistema e processos judiciais	Justiça, sistema, federal, principais, execução, fiscal, resultado	Justiça, entidades sem fins lucrativos (ESFLs), varas, fiscal, transferências, Conselho Nacional de Justiça (CNJ), orçamento
Tópico 15	Avaliação de políticas públicas	Avaliação, ouvidoria, políticas, comentários, desenvolvimento, direitos, destaca	Ouvidoria, comentários, avaliação, urbanização, destaca, tece, precários

(Continua)

(Continuação)

Tópico	Tema	Maior probabilidade de o termo estar no tópico	Exclusividade do termo ao tópico
Tópico 16	Social e direitos humanos	Social, estado, políticas, saúde, trabalho, análise, Brasil	Desenvolvimentista, vulnerabilidade, social, favelas, mulheres, seguridade, necessidades
Tópico 17	Planejamento e desenvolvimento	Planejamento, Brasil, desenvolvimento, gestão, trabalho, país, crescimento	Planejamento, investimento, Programa de Aceleração do Crescimento (PAC), cenários, Plano Plurianual (PPA), retomada, investimentos
Tópico 18	Inovação no setor público	Inovação, setor, público, organizações, objetivo, governo, inovações	Inovação, inovações, presidente, laboratórios, OSCs, setor, reino
Tópico 19	Dados e informações (homicídios, emprego, trabalho)	Dados, anos, pesquisa, emprego, informações, trabalho, homicídios	Superior, emprego, anual, cargos, Relação Anual de Informações Sociais (RAIS), homicídios, ocupações
Tópico 20	Mudanças em práticas governamentais	Mudanças, políticas, públicas, práticas, federal, casos, serviços	Gerenciamento, mudanças, práticas, casos, ênfase, apoio, passou

Elaboração dos autores

Tomemos, por exemplo, o tópico 8. O indicador de maior probabilidade de incidência de um termo no tópico aponta participação como a palavra com maior probabilidade, seguida de conferências, nacionais, sociais, assim por diante. Já o indicador de exclusividade aponta como palavras mais exclusivas a esse tópico: conferências, em primeiro lugar, e, em seguida, participação, participativas, nacionais, conselhos etc. Dessa forma, sugerimos que esse tópico possa ser identificado pela temática nomeada por participação social. Embora o exercício de identificação de temas gerais não seja preciso, sustentamos que o cotejamento com os dados apresentados nas figuras 1 e 2 demonstram convergência considerável.

Sobre os tópicos analisados, três pontos merecem ser destacados. O primeiro deles diz respeito ao fato de que foram identificadas correlações significativas em apenas dois pares de tópicos: 4 e 17, o que denota a atenção da diretoria para a discussão de desenvolvimento, qualificando-a tanto no sentido da contextualização internacional, como em sua relação com os elementos de planejamento e gestão, como observado no primeiro artigo desta publicação. E também entre os tópicos 8 e 13, o que talvez aponte para um enfoque da participação como forma e instrumento de implementação de políticas públicas e um interesse de análise da perspectiva da burocracia nos espaços participativos.

Além da questão das áreas temáticas, o tópico 2 permitiu aglutinar os Bapis como um tipo de produção específica da Diest. Por fim, um terceiro ponto que ilustra o relato produzido no segundo artigo deste boletim diz respeito aos tópicos 5 e 19, que evidenciam o esforço de produção de dados e informações que, além de subsidiar pesquisas, amplia as possibilidades de assessoramento e o fomento e a qualificação do debate público por meio da disseminação de bases de dados produzidas por projetos importantes da Diest, como o *Atlas da Violência*, *Mapa das OSCs* e *Atlas do Estado Brasileiro*.

4 CONSIDERAÇÕES FINAIS

Em linhas gerais, podemos sustentar que os dados apresentados convergem com reflexões apresentadas em outros artigos desta publicação quanto à atuação da Diest. Como sinalizado por José Celso Cardoso Júnior no primeiro artigo desta publicação, pôde-se observar avanços promovidos pela diretoria no sentido de introduzir abordagens meso e micro institucionais na produção, disseminação e capacitação em áreas e dimensões do conhecimento que compõem o núcleo duro das atividades estruturantes da administração pública brasileira. Da mesma forma, os dados retratam especializações

em áreas setoriais, como as de segurança pública e direitos humanos; estatais; social; regulação; e infraestrutura, mas em associação a temas e chaves-analíticas transversais, tais como burocracia, implementação, governança, capacidades e inovação. E, ainda, confirmam os achados do autor quanto à ampliação do escopo de análise do Executivo ao enxergá-lo e examiná-lo em sua relação com o Poder Judiciário, o Legislativo e a sociedade. Em síntese, podemos dizer que a Diest inova ao acrescentar uma abordagem micro-meso e relacional da atuação estatal nos grandes debates sobre desenvolvimento e democracia.

Contudo, os dados levantados também puderam evidenciar um aspecto importante trazido por Luseni Aquino e Maria Paula Santos no artigo 16 desta publicação quanto a uma ênfase na produção da Diest a uma perspectiva que poderíamos chamar de funcionalista ou operacional da análise das instituições do Estado. Temas que traduziriam preocupações relacionadas à efetividade das instituições para a garantia dos preceitos democráticos e republicanos não puderam ser claramente identificados nas análises apresentadas.

Podemos dizer que o estudo retratou, ainda, resultados do esforço e preocupação da Diest, detalhados por Flávia Schmidt e Janine Mello no segundo artigo desta publicação, em aprimorar e inovar tanto na geração e disponibilização de dados, como nos meios de divulgação da produção de seus técnicos. Por fim, em que pese o estágio preliminar na análise, acreditamos que o estudo tenha jogado luz sobre a contribuição da Diest tanto como produtora de conhecimento, como de *knowledge broker*, isto é, mediadora e disseminadora de conhecimento.

REFERÊNCIAS

BRASIL. Casa Civil. Decreto nº 7.142, de 29 de março de 2010. Aprova o Estatuto e o Quadro Demonstrativo dos Cargos em Comissão e das Funções Gratificadas do Instituto de Pesquisa Econômica Aplicada - IPEA, e dá outras Providências. **Diário Oficial da União**, Brasília, 30 mar. 2010.

CAPLAN, N. The two-communities theory and knowledge utilization. **American Behavioral Scientist**, n. 3, v 22, p. 459-470, 1979.

DAVIES, H. T. O.; NUTLEY, S.; SMITH, P. **What Works?** evidence-based policy and practice in public services. Bristol: The Policy Press, 2000.

FOWLER, J. H. Connecting the Congress: a study of Cosponsorship Networks. **Political Analysis**, n. 14, v. 4, p. 456-487, March 2006.

LASSWELL, H. D.; LERNER, D. (Org.), **The policy sciences**: recent developments in scope and method. Stanford: Stanford University Press, 1951.

NEWMAN, J.; CHERNEY, A.; HEAD, B. W. Do policy makers use academic research? Reexamining the “two communities”. theory of research utilization. **Public Admin Rev**, v. 76, p. 24-32, 2016.

OPSAHL, T., AGNEESSENS, F., SKVORETZ, J., Node centrality in weighted networks: generalizing degree and shortest paths. **Social Networks** n. 32, v. 3, p. 245-251, 2010.

PINHEIRO, M. M. S. **Políticas públicas baseadas em evidências (PPBEs)**: delimitando o problema conceitual. Rio de Janeiro: Ipea, 2020., (Texto para Discussão, n. 2554).

WARD, V. et al. Exploring knowledge exchange: a useful framework for practice and policy. **Soc Sci Med**, n. 74,v. 3, p. 297-304, 2012.

WEISS, C. The many meanings of research utilization. **Public Administration Review**, n. 5, v. 39, p.426-431, 1979.

APÊNDICE

METODOLÓGICO

BASE DE DADOS

Algumas ressalvas devem ser feitas quanto às limitações da base de dados utilizada, que podem implicar inclusões indevidas ou a não inclusão de produções devidas na base, relacionadas a seguir.

- a) Mudanças ou registros distintos dos nomes de autores – em razão de não haver um padrão no formato de citação dos nomes dos autores, ou de alguns sobrenomes eventualmente mudarem ao longo do tempo.
- b) Descompasso entre a publicação de alguma unidade bibliográfica e o período de atuação na diretoria – é comum que publicações levem algum tempo para serem concluídas, o que não necessariamente coincidirá com o tempo de trabalho na diretoria. Além disso, certos temas compõem a agenda de pesquisa de alguns pesquisadores, o que faz com que eles se mantenham ao longo do tempo, que independe da atuação na diretoria, inclusive com parcerias realizadas com autores de outras diretorias.
- c) Publicações institucionais, sem autoria definida – algumas publicações assumem um caráter institucional, sem uma definição explícita dos autores, embora os pesquisadores responsáveis sejam citados nas primeiras páginas da publicação.
- d) Dificuldade de precisar a data da publicação – Os *Textos para Discussão* e as *Notas Técnicas* apresentam o mês e ano em que foram publicados, mas os livros em geral apresentam somente o ano. Isso pode produzir um descompasso entre o período de atuação do pesquisador na Diest e o tipo de publicação realizada.

MÉTODO ESTATÍSTICO

O método estatístico do Structural Topic Model (STM) parte do princípio de que os documentos são escritos sobre tópicos, geralmente uma mistura desses, e cada tópico, está associado a um determinado conjunto (não excludente) de palavras de certo vocabulário, o qual, informalmente, representa um tema semântico subjacente. Isso permite que os documentos se “sobreponham” em termos de conteúdo de uma forma que espelha o uso típico da linguagem natural (Hall, Jurafsky e Manning, 2008; Jacobi, Atteveldt, Welbers e 2016).

No que tange ao tratamento dos dados, removemos algumas palavras que agregam pouco na análise, tais como advérbios, conjunções, preposições e pronomes. Nesse sentido, também foram removidas as localizadoras uniformes de recursos (Uniform Resource Locator – URLs) e números. Além disso, aplicamos o processo de lematização,¹ substituindo palavras flexionadas para a sua forma dicionarizada. Essa última etapa promove a redução do conjunto total de palavras para um conjunto reduzido igualmente representativo e menos redundante.

Visto que o modelo exige a especificação prévia do número de tópicos, faz-se necessária a utilização de métricas para orientar a melhor escolha. Nesse sentido, usamos algumas métricas comuns, tais como *heldout likelihood*, *semantic coherence* e *residual dispersion*, para avaliar como os modelos estão se comportam em vários números de tópicos (Roberts, Stewart e Tingley, 2019; Mimno *et al.*, 2011; Taddy, 2012).

1. A lematização consiste em representar as formas flexionadas de uma palavra para serem analisadas sob uma mesma forma canônica, o lema. O lema pode ser definido como o infinitivo do verbo e o masculino singular do substantivo ou adjetivo.

REFERÊNCIAS

- HALL, D.; JURAFSKY, D.; MANNING, C. D. Studying the history of ideas using topic models. *In*: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2008, Hawaii, USA. **Annals...** EMNLP: Hawaii, USA, 2008.
- JACOBI, C; ATTEVELD, W. van; WELBERS, K. Quantitative analysis of large amounts of journalistic texts using topic modelling, **Digital Journalism**, n. 4, v. 1, p. 89-106, 2016. DOI: <10.1080/21670811.2015.1093271>.
- MIMNO, D. *et al.* Optimizing semantic coherence in topic models. *In*: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2011, Stroudsburg. **Annals...** EMNLP: Stroudsburg: 2011.
- ROBERTS, M., STEWART, B.; TINGLEY, D. STM: R Package for Structural Topic Models. **Journal of Statistical Software**, n. 91, v. 2, p. 1-40, 2019.
- TADDY, M. A. On estimation and selection for topic models. *In*: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND STATISTICS, 15., 2012, La Paula, Canary Islands. **Annals...** La Paula, Canary Islands: AISTATS, 2012.