

O CADASTRO NACIONAL DE PESSOA JURÍDICA COMO RECURSO METODOLÓGICO PARA AS PESQUISAS SOCIOESPACIAIS: POTENCIALIDADES E LIMITAÇÕES DA BASE DE DADOS NACIONAL

Késia Anastácio Alves da Silva

Universidade Estadual Paulista, Departamento de Geografia, Presidente Prudente, SP, Brasil
kesia.anastacio@unesp.br

Natália Daniela Soares Sá Britto

Universidade Federal de Santa Catarina, Departamento de Geociências, Florianópolis, SC, Brasil
natalia.britto@ufsc.br

RESUMO

O objetivo deste trabalho é apresentar a base de dados do Cadastro Nacional de Pessoa Jurídica (CNPJ), disponibilizada para acesso público pela Receita Federal, como fonte e recurso metodológico para a produção de estudos espaciais, especialmente no âmbito da Geografia Urbana, Econômica e Regional. Para tanto, o artigo retoma alguns postulados sobre os aportes metodológicos na pesquisa socioespacial e, através de diversos exemplos, destaca as potencialidades de uso da base de dados do CNPJ, com seus desdobramentos escalares, setoriais e temporais. Além disso, o trabalho avalia alguns limites e cuidados metodológicos que devem ser considerados pelos pesquisadores ao utilizar e aplicar esta fonte.

Palavras-chave: Metodologia. Cadastro Nacional de Pessoa Jurídica. Pesquisa socioespacial. Fonte de dados.

THE NATIONAL REGISTRY OF LEGAL ENTITIES AS A METHODOLOGICAL RESOURCE FOR SOCIO-SPATIAL RESEARCH: POTENTIALS AND LIMITATIONS OF THE NATIONAL DATABASE

ABSTRACT

The main objective of this study is to present the database of the National Register of Legal Entities (CNPJ), which is made publicly available by the Federal Tax Administration, as a fundamental and methodological tool for conducting spatial analyses, particularly in the fields of urban, economic, and regional geography. The article deals with the main methodological principles of socio-spatial research and uses illustrative examples to show the many possible applications of the CNPJ database, including its impact on different scales, sectors, and temporal contexts. In addition, the study critically assesses certain methodological limitations and considerations that researchers must take into account when using this data source.

Keywords: Methodology. National Register of Legal Entities. Socio-spatial research. Data source.

INTRODUÇÃO

A interpretação dos processos e fenômenos socioespaciais, ou seja, a compreensão da produção social do espaço, tornou-se o centro dos estudos geográficos ao longo dos últimos quarenta anos. Os objetos que compõem as mais diversas pesquisas socioespaciais, através de métodos, técnicas, instrumentos de investigação científica, conceitos e teorias, buscam capturar as distintas articulações e manifestações dialéticas entre sociedade e espaço. Como destacado por Souza (2007, 2013), as pesquisas socioespaciais, para além da redução do adjetivo social, debruçam-se sobre a correlação entre os processos sociais e o espaço. A análise socioespacial, portanto, não se restringe apenas às estruturas espaciais, mas também examina os processos vivos e as interações que ocorrem nas mais diversas escalas que compõem o espaço geográfico.

A Geografia, enquanto ciência humana, detém um escopo teórico e conceitual próprio, exemplificado pelos conceitos de paisagem, região, lugar, território, rede e escala. Estes se configuram como

ferramentas intelectuais essenciais para a leitura e apreensão do espaço geográfico empiricizado (SANTOS, 1996; SUERTEGARAY, 2001). Do mesmo modo, a ciência geográfica faz uso de um conjunto amplo de instrumentos metodológicos e técnicas de pesquisa, sejam eles próprios deste campo científico ou não, nos quais são empregados para o entendimento dos elementos, processos e fenômenos que constituem seu objeto de estudo. Assim, no âmbito das pesquisas socioespaciais, tais instrumentos variam desde a observação sistemática da paisagem e sua representação até a compreensão de processos, estruturas, formas e funções que qualificam os sistemas de objetos e de ações (SANTOS, 1996), mediante a coleta, sistematização e uso de dados primários e/ou secundários.

Como em outras áreas da ciência, a clarividência do tema, do objeto e da pergunta a ser respondida, a construção teórica, e a sistematização dos conceitos e elementos que constituem a problemática analisada, permitem delinear os caminhos e os instrumentos metodológicos que orientarão a execução da investigação. Por essas razões, não apenas a escolha dos procedimentos, mas também a proposta de novas metodologias, ferramentas e novas fontes de dados ampliam a compreensão de temas que abrangem a pesquisa socioespacial, principalmente no contemporâneo - caracterizado pela rapidez das transformações socioespaciais e emergência de novos processos e fenômenos.

O caminho a ser delineado por meio de métodos e técnicas que caracterizam o rigor científico visa suscitar respostas e elementos analíticos que amplifiquem o entendimento da temática investigada. Neste contexto, torna-se fundamental compreender a importância da questão metodológica e conceitual, incluindo a adequação dos termos utilizados, em especial na pesquisa em Geografia, pois estes também são elementos de operacionalização dos objetos analisados. Dessa forma, os métodos e técnicas escolhidos pelo pesquisador estarão alinhados com os conceitos e termos anunciados na pesquisa, o que implica o detalhamento de conteúdos e teorias capazes de serem aplicados na pesquisa em Geografia.

Babbie (2003) cita as análises de conteúdo, a organização de banco de dados estatísticos, a construção de dados através de surveys, os estudos de casos e a observação participante como aportes metodológicos mais afeitos às ciências sociais e humanas. No caso da Geografia, para além desses procedimentos elencados, o trabalho de campo – importantíssimo para o geógrafo – auxilia o levantamento documental e fotográfico, na assimilação de processos socioeconômicos, culturais, históricos, entre outros. A aplicação de entrevistas qualitativas ou de questionários quantitativos permite o levantamento de informações sobre o objeto, visando compreender o conteúdo relatado, assim como a elaboração de modelos explicativos, a construção de gráficos, tabelas, mapas entre outras informações colhidas, construídas e trabalhadas para análise do objeto pesquisado.

Assim como o trabalho de campo e a produção de dados qualitativos/quantitativos construídos pelo próprio pesquisador, o uso de fontes de dados secundárias nas ciências sociais e na análise socioespacial é largamente difundido. No contexto das ciências sociais, um exemplo clássico do uso de fontes de dados secundárias foi o estudo do suicídio elaborado por Émile Durkheim em 1897 (BABBIE, 2003). Na Geografia - e nas diversas correntes que compõem a história do pensamento geográfico, da Geografia teórica quantitativa à Geografia crítica - o uso de fontes de dados secundárias, por vezes empregado com maior intensidade e outras menos, faz-se presente. O uso de dados secundários na pesquisa socioespacial é profícuo por apresentar vantagens econômicas, pois afasta do pesquisador a obrigatoriedade de elaborar seu próprio dado, não tendo custos com elaboração de questionários/ou roteiros de entrevistas, amostragem e recrutamento de sujeitos experimentais, aplicação dos questionários, codificações e construção do banco de dados. Por outro lado, como qualquer outra metodologia, possui desvantagens e limitações, principalmente porque a fonte utilizada não foi construída e pensada para o objeto de estudo analisado, e pode não apreender e representar o processo/fenômeno estudado em sua totalidade. Todavia, como afirma Babbie (2003 p.42), através de uma boa dose de engenhosidade o pesquisador pode contornar determinadas limitações.

Ao longo da história da ciência geográfica, a crítica às metodologias quantitativas se faz presente, e não podemos negar a pertinência das argumentações. A análise funcional do espaço geográfico, presente na Geografia teórica quantitativa, a diferenciação, classificação e hierarquização de áreas através de números e modelos matemáticos, em uma visão de sobrevoo que deixa esvaír as inúmeras nuances e processos que conformam a produção social do espaço (HARTSHORNE, 1978; SOUZA, 2007), compõem as críticas a esta corrente do pensamento geográfico e intensificam o embate entre metodologias qualitativas versus quantitativas.

Na pesquisa socioespacial, não podemos adotar apenas o olhar distante, pois com este desvanece o cotidiano, a percepção dos diversos agentes e atores que produzem o espaço, principalmente os contra

hegemônicos, escapa às singularidades do lugar e às resistências às verticalidades, bem como às nuances que a luta de classes, que também é espacial, assume no cotidiano. Por outro lado, o mergulho total no cotidiano, deixa escapar "a noção de conjunto, que faculta a apreensão das 'grandes linhas', aquela que permite trabalhar com as escalas da estratégia" (SOUZA, 2007, p.111).

Concordamos com Souza (2007); a problemática não está na adoção da visão de sobrevoo em si, mas sim na operacionalização e análise de processos socioespaciais apenas por uma perspectiva. Em outras palavras, a pesquisa socioespacial demanda, em muitos casos, a combinação de escalas, transitando entre a visão de sobrevoo ao mergulho no cotidiano. Além disso, é necessário que o pesquisador conheça as limitações e potencialidades das metodologias utilizadas, isto é, o alcance de determinadas ferramentas para estudo, elaboração de hipóteses e formulação de respostas perante os objetos analisados.

Neste contexto, o objetivo deste artigo não é apresentar uma nova metodologia, mas sim destacar a potencialidade de uma fonte de dados secundária para pesquisa socioespacial. Pretende-se elencar as possíveis aplicações do Cadastro Nacional da Pessoa Jurídica (CNPJ) nos campos da Geografia urbana, econômica, do comércio e regional. O trabalho, além de listar as características desta fonte de dados, suas variáveis e seus possíveis usos, detalha também as formas de construção do banco de CNPJs e apresenta exemplos analíticos. Para tanto, o artigo se subdivide em três seções, além desta introdução. A primeira descreve as características do Cadastro Nacional de Pessoa Jurídica, sua finalidade, criação, as informações contidas no cadastro. A segunda detalha a construção do banco de dados, sua manipulação através de um *software* estatístico e o processo de geocodificação. Por fim, a terceira parte detalha os possíveis usos dessas informações e as limitações que esta fonte apresenta no âmbito da pesquisa socioespacial. Destaca-se que os dados apresentados ao longo deste texto foram extraídos do site da Receita Federal no mês de julho de 2021.

O CNPJ COMO FONTE E RECURSO METODOLÓGICO NOS ESTUDOS ESPACIAIS

Uma breve apresentação da estrutura da fonte de dados

Institucionalizado pela Instrução Normativa SRF nº 27/1998, o Cadastro Nacional de Pessoa Jurídica (CNPJ) compreende a inscrição de entidades jurídicas de direito público ou privado¹ junto à Receita Federal do Brasil (RFB). Ele é utilizado para identificar as empresas e suas atividades econômicas, bem como para realizar operações financeiras e fiscais, representando uma evolução do extinto Cadastro Geral de Contribuintes² (CGC). O objetivo principal deste cadastro é reunir informações de interesse tributário para a administração pública, seja na esfera da União, dos estados, do Distrito Federal ou dos municípios.

De acordo com a Instrução Normativa RFB nº 1.863 de 27 de dezembro de 2018, que rege as disposições, conteúdos e competências do Cadastro Nacional da Pessoa Jurídica, são obrigadas a se inscrever no CNPJ todas as entidades, instituições, fundos e sociedades, públicas ou empresariais, domiciliadas no Brasil³, bem como cada um de seus estabelecimentos (localizados no Brasil ou no exterior), antes do início de suas atividades, conforme o Quadro 1 abaixo.

Quadro 1 - Entidades com Obrigatoriedade de Inscrição no CNPJ

1. Entidades empresariais domiciliadas no Brasil, bem como cada um de seus estabelecimentos, localizados no Brasil ou no exterior, antes do início de suas atividades
2. Órgãos públicos de qualquer dos poderes da União, dos estados, do Distrito Federal e dos municípios, desde que se constituam em unidades gestoras de orçamento
3. Condomínios edifícios, conceituados nos termos do art. 1.332 da Lei nº 10.406, de 10 de janeiro de 2002, e os setores condominiais na condição de filiais, desde que estes tenham sido instituídos por convenção de condomínio;

¹ Conforme o Art. 44 da Lei 10.406 de 2002 (Código Civil brasileiro), são pessoas jurídicas de direito privado: as associações, as sociedades, as fundações, as organizações religiosas, os partidos políticos e as empresas individuais de responsabilidade limitada.

² Instaurado e regulamentado pelas Leis 4.503/1964 e 5.614/1970.

³ Com exceção das organizações religiosas que não possuam autonomia administrativa ou que não sejam gestoras de orçamento, estas são dispensadas da obrigatoriedade de cadastro.

4. Grupos e consórcios de sociedades, constituídos, respectivamente, na forma prevista nos artigos 265 e 278 da Lei nº 6.404, de 15 de dezembro de 1976;
5. Consórcios de empregadores, constituídos na forma prevista no art. 25-A da Lei nº 8.212, de 24 de julho de 1991;
6. Clubes e fundos de investimento, constituídos segundo as normas da Comissão de Valores Mobiliários (CVM);
7. Representações diplomáticas estrangeiras no Brasil;
8. Representações diplomáticas do Estado brasileiro no exterior;
9. Representações permanentes de organizações internacionais ou de instituições extraterritoriais no Brasil;
10. Serviços notariais e de registro, de que trata a Lei nº 8.935, de 18 de novembro de 1994, inclusive aqueles que ainda não foram objeto de delegação do Poder Público;
11. Fundos públicos a que se refere o art. 71 da Lei nº 4.320, de 17 de março de 1964;
12. Fundos privados
13. Candidatos a cargo político eletivo e frentes plebiscitárias ou referendárias, nos termos de legislação específica;
14. Incorporações imobiliárias objeto de opção pelo Regime Especial de Tributação (RET), de que trata o art. 1º da Lei nº 10.931, de 2 de agosto de 2004, na condição de estabelecimento filial da incorporadora;
15. Comissões Polinacionais, criadas por ato internacional celebrado entre o Brasil e outros países;
16. Entidades domiciliadas no exterior que, no País: a) sejam titulares de direitos sobre bens móveis, imóveis e financeiros; b) que realizem: leasing, afretamento de equipamentos e embarcações, importação de bens sem cobertura cambial e consultoria de valores imobiliários
17. Instituições bancárias do exterior que realizem operações de compra e venda de moeda estrangeira com bancos no País
18. Sociedades em Conta de Participação (SCPs) vinculadas aos sócios ostensivos

Fonte: Instrução Normativa RFB nº 1.863/2018. Organização: autoras, 2023.

Desde dezembro de 2020, a RFB publica periodicamente um conjunto de informações cadastrais de todas as entidades registradas no país, fruto da Política Nacional de Dados Abertos, instituída em maio de 2016⁴. Em consonância com a Lei de Acesso à Informação, sancionada em novembro de 2011 (Lei 12.527/2011), a base de dados do CNPJ vem sendo publicada na forma de dados abertos no site da Receita Federal, com atualização mensal, apresentando informações do cadastro de mais de 21 milhões de entidades no país, bem como informações sobre o regime tributário das pessoas jurídicas, com atualização anual.

A base de dados do Cadastro Nacional de Pessoas Jurídicas está organizada em três grandes grupos: **a) Empresas; b) Sócios; c) Estabelecimentos**. Cada um desses grupos contém um conjunto de variáveis com informações cadastrais para cada um dos mais de 21 milhões de CNPJs do país, conforme especificado no quadro abaixo (Quadro 2). As variáveis contidas em cada um desses grupos apresentam potencialidades para a pesquisa socioespacial, principalmente pela capacidade de espacialização, periodização e qualificação do dado segundo atividades econômicas e outras dimensões. Neste contexto, nos próximos tópicos, serão detalhadas as variáveis cadastrais que compõem esses grupos de dados (Sócios, Empresas e Estabelecimentos), bem como as possibilidades analíticas que podem ser auferidas a partir dessas informações.

⁴ De acordo com o Decreto 8.777/2016, a Política Nacional de Dados Abertos tem, entre os seus objetivos, “promover a 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, aprimorar a cultura de transparência pública e franquear aos cidadãos o acesso, de forma aberta, aos dados produzidos ou acumulados pelo Poder Executivo federal, sobre os quais não recaia vedação expressa de acesso”.

Quadro 2 - Variáveis contidas no banco de dados do CNPJ segundo tipologia e grupo de dados

Empresas	Estabelecimentos	Sócios
Razão social ou nome empresarial da pessoa jurídica	Identificador do estabelecimento matriz/filial	Identificador de sócio (pessoa física, jurídica e/ou estrangeiro)
Natureza jurídica	Nome fantasia	Nome do sócio (pessoa física ou a razão social e/ou nome empresarial da pessoa jurídica e/ou nome do sócio/razão social do sócio estrangeiro)
Qualificação do responsável	Situação cadastral	CPF ou CNPJ do sócio
Capital social da empresa	Data situação cadastral	Código da qualificação do sócio
Porte da empresa	Motivo situação cadastral	Data de entrada na sociedade
Ente federativo responsável	Nome da cidade no exterior	Código país do sócio, no caso de sócio estrangeiro
Indicador da existência da opção pelo simples.	País	Número do CPF do representante legal
Data de opção e/ou exclusão pelo simples	Data de início atividade	Nome do representante legal
Indicador da existência da opção pelo MEI	Código da atividade econômica (CNAE) principal do estabelecimento	Código da qualificação do representante legal
Data de opção e/ou exclusão pelo MEI	CNAE fiscal secundária	Código correspondente à faixa etária do sócio
	Endereço completo	
	Telefone e outros contatos (e-mail, fax, site)	
	Situação especial da empresa	
	Data em que a empresa entrou em situação especial	

Fonte: RFB - Banco de dados de CNPJ. Organização: autores, 2023.

Empresas

O grupo de dados Empresas reúne informações sobre a natureza jurídica, o porte, o Quadro Sócio-Administrativo (QSA) e o capital social declarado pelas entidades jurídicas. Estes dados possibilitam a composição de um quadro sobre as empresas, os setores econômicos e os tipos jurídicos que predominam nesse universo. No caso dos dados sobre a natureza jurídica, é possível verificar a distribuição das empresas cadastradas entre os diferentes tipos de regimes jurídicos e societários que definem as normas e deveres para a realização das atividades empresariais, tais como Sociedades Anônimas de Capital Aberto (S.A), Sociedades em Conta de Participação (SCP), Microempreendedor Individual (MEI) e Empresas Individuais de Responsabilidade Limitada (EIRELI)⁵.

No Brasil, a classificação da natureza jurídica das empresas é estruturada em cinco grandes grupos, de acordo com a Comissão Nacional de Classificação (CONCLA, 2021), organismo responsável pelas classificações estatísticas do Brasil: a) Administração pública; b) Entidades empresariais; c) Entidades

⁵ O tipo jurídico EIRELI foi revogado tacitamente pela Lei nº 14.195/2021, além da revogação expressa pela recente Medida Provisória nº 1.085/2021. Dessa forma, as novas empresas que seriam abertas como EIRELIs devem ser registradas nas Juntas Comerciais como Sociedade Empresária Limitada.

sem fins lucrativos; d) Pessoas físicas; e) Organizações internacionais e outras instituições extraterritoriais (Quadro 3). Cada um desses grupos é composto por um amplo conjunto de naturezas jurídicas diferenciadas por códigos e descrições, o que permite classificar e caracterizar a composição das empresas. Esse é um dado inicial importante para compreender suas obrigações e forma de responsabilização relacionada ao seu instituto jurídico e societário.

Quadro 3 - Classificação da Natureza Jurídica segundo a Comissão Nacional de Classificação – CONCLA, 2021

Grupos	Tipos de Natureza Jurídica
Administração pública	1) Órgãos Públicos do Poder Executivo, legislativo ou judiciário, em diferentes níveis (Federal, Estadual, Distrito Federal ou municípios); 2) Autarquias em diferentes níveis (Federal, Estadual, Distrito Federal ou municípios); 3) Fundações Públicas de Direito Público ou Privado em diferentes níveis (Federal, Estadual, Distrito Federal ou municípios); 4) Órgãos Públicos Autônomos; 5) Comissões Polinacionais; 6) Consórcios Públicos de Direito Público (Associação Pública) ou de Direito Privado; 7) Fundos Públicos da Administração Direta ou Indireta de diferentes níveis (Federal, Estadual, Distrito Federal ou municípios);
Entidades empresariais	1) Empresa Pública; 2) Sociedade de Economia Mista; 3) Sociedade Anônima Aberta; 4) Sociedade Anônima Fechada; 5) Sociedade Empresária Limitada; 6) Sociedade Empresária em Nome Coletivo; 7) Sociedade Empresária em Comandita Simples; 8) Sociedade Empresária em Comandita por Ações; 9) Sociedade em Conta de Participação; 10) Empresário (Individual); 12) Cooperativa; 13) Consórcio de Sociedades; 14) Grupo de Sociedades; 15) Estabelecimento, no Brasil, de Sociedade Estrangeira; 16) Estabelecimento, no Brasil, de Empresa Binacional Argentino-Brasileira; 17) Empresa Domiciliada no Exterior; 18) Clube/Fundo de Investimento; 19) Sociedade Simples Pura; 20) Sociedade Simples Limitada; 21) Sociedade Simples em Nome Coletivo; 22) Sociedade Simples em Comandita Simples; 23) Empresa Binacional; 24) Consórcio de Empregadores; 25) Consórcio Simples; 26) Empresa Individual de Responsabilidade Limitada (de Natureza Empresária); 27) Empresa Individual de Responsabilidade Limitada (de Natureza Simples); 29) Sociedade Unipessoal de Advogados; 30) Cooperativas de Consumo; 31) Empresa Simples de Inovação - Inova Simples; 32) Investidor Não Residente.
Entidades sem fins lucrativos	1) Fundação Privada; 2) Serviço Social Autônomo; 3) Condomínio Edifício; 4) Comissão de Conciliação Prévia; 5) Entidade de Mediação e Arbitragem; 6) Entidade Sindical; 7) Estabelecimento, no Brasil, de Fundação ou Associação Estrangeiras; 8) Fundação ou Associação Domiciliada no Exterior; 9) Organização Religiosa; 10) Comunidade Indígena; 11) Fundo Privado; 12) Órgão de Direção Nacional de Partido Político; 13) Órgão de Direção Regional de Partido Político; 14) Órgão de Direção Local de Partido Político; 15) Comitê Financeiro de Partido Político; 16) Frente Plebiscitária ou Referendaria; 17) Organização Social (OS); 18) Demais Condomínios; 19) Plano de Benefícios de Previdência Complementar Fechada; 20) Associação Privada.
Pessoas físicas	1) Empresa Individual Imobiliária; 2) Segurado Especial; 3) Contribuinte individual; 4) Candidato a Cargo Político Eletivo; 4) Leiloeiro; 5) Produtor Rural (Pessoa Física)
Organizações internacionais e outras instituições extraterritoriais	1) Organização Internacional; 2) Representação Diplomática Estrangeira; 3) Outras Instituições Extraterritoriais.

Fonte: RFB – Banco de dados de CNPJ. Organização: autores, 2023.

As informações sobre o capital social, o quadro societário e o porte são declarados pelas entidades no momento do cadastro (e atualizadas anualmente), visto que é a partir dessas informações que será definido o seu regime tributário⁶. No caso das entidades empresariais, de acordo com a RFB (2018), o porte deve ser declarado com base no faturamento anual bruto⁷ da matriz e suas filiais (quando

⁶ A partir dessa classificação, é possível acessar uma série de benefícios fiscais e tributários, como os previstos na Lei Complementar 123/2006, conhecida como a Lei Geral da Micro e Pequena Empresa. No Brasil, os três tipos de regime tributário mais adotados são: Lucro Real, Lucro Presumido e Simples Nacional.

⁷ Além da classificação da Receita Federal, existem outras classificações de porte realizadas por diferentes instituições no Brasil, como a classificação do IBGE e do Sebrae, que utilizam o número de funcionários para a definição do porte. Outras instituições, como a Anvisa e o BNDES, consideram o faturamento anual das empresas,

existirem), podendo assim ser classificada de acordo com cinco grupos, conforme o quadro a seguir (Quadro 4).

Quadro 4 - Classificação do porte das empresas segundo a RFB, 2018

Tipo	Faturamento
Microempreendedor Individual (MEI)	Empresas cujo faturamento bruto anual não ultrapassa R\$ 81 mil reais ao ano
Microempresa (ME)	Empresas cujo faturamento bruto seja igual ou inferior a R\$ 360 mil reais anuais
Empresa de Pequeno Porte (EPP)	Empresas cuja receita bruta anual esteja entre R\$ 360 mil reais à R\$ 4,8 milhões de reais
Empresa de Médio Porte	Empresas em que o faturamento anual é maior de R\$ 4,8 milhões e menor que R\$ 300 milhões de reais
Grande Empresa	Empresas cuja receita bruta anual seja superior a R \$300 milhões de reais.

Fonte: RFB – Banco de dados de CNPJ. Organização: autores, 2023.

Já o capital social corresponde ao investimento inicial das empresas, visando apresentar os recursos iniciais para a realização de suas atividades, como reservas em mercadoria, matéria-prima, equipamentos, móveis e imóveis. Além disso, o capital social evidencia a divisão de responsabilidades e participação entre os sócios, por meio do quadro societário e dos administradores da pessoa jurídica (QSA), permitindo destacar a relação dos sócios com o patrimônio e suas funções dentro da empresa. No caso do Microempreendedor Individual (MEI), a declaração do capital social da empresa não é obrigatória, uma vez que se trata de um tipo de entidade jurídica sem a presença de sócios (QSA) e cujo faturamento anual é inferior a R\$ 81 mil. Por outro lado, para empresas que adotam outras naturezas jurídicas e se enquadram em diferentes portes, a declaração é obrigatória. O capital social declarado deve refletir um valor mínimo necessário para o funcionamento da empresa, podendo ser dividido proporcionalmente de acordo com o percentual de cada sócio (QSA), conforme estabelecido no contrato social da empresa.

Em conjunto, essas informações sobre o porte, o capital social e o QSA das empresas permitem compreender a estrutura jurídica e econômica predominante em cada setor, assim como a capilaridade societária e a abrangência geográfica dos agentes econômicos, como exemplificado no Quadro 5 a seguir.

Quadro 5 - Exemplo de dado empresarial: Capilaridade societária e geográfica da empresa MRV Engenharia e Participações S.A.⁸

Dados sobre a Empresa	Descrição do Dado
Nº de Inscrição CNPJ	08.343.492/0001-20
Razão Social	MRV Engenharia e Participações S. A
Natureza Jurídica	Sociedade Anônima Aberta

mas elencam outros grupos de classificação. No entanto, os dados do Cadastro Nacional de Pessoa Jurídica (CNPJ) classificam o porte apenas com base no faturamento anual bruto declarado pelas empresas, em conformidade com a Lei nº 123/2006.

⁸ As empresas destacadas nos Quadros 05, 07 e 09, assim como nas Figuras 03, 04, 05 e 06, foram escolhidas devido à sua representatividade nas atividades econômicas no território nacional. Priorizamos organizações-chave em setores relevantes, buscando ilustrar a abrangência da base de dados. A seleção foi orientada pela importância nacional e regional dessas empresas, destacando seu papel estratégico e evidenciando as múltiplas aplicações e potencialidades da fonte de dados em análise.

Porte	Grande
Capital Social declarado	R\$ 4.974.770.151,06 (Quatro bilhões, novecentos e setenta e quatro milhões, setecentos e setenta mil e cento e cinquenta e um reais e seis centavos)
QSA e Qualificação do responsável	Júnia Maria de Sousa Lima Galvão (Diretora); Rafael Nazareth Menin Teixeira De Souza (Presidente); Eduardo Fischer; Teixeira De Souza (Diretor); Raphael Rocha Lafeta; (Administrador); Ricardo Peixão Pinto Rodrigues (Diretor); Thiago Correa Ely (Administrador); Silvio Luiz Gava (Administrador)
Capilaridade societária - Total de empresas ativas com participação da MRV Engenharia e Participações em seu QSA	1.329 empresas
Capilaridade geográfica	MRV Engenharia e Participações é ou já foi sócio de 7 empresas no estado de Paraná, 10 em Santa Catarina, 280 em Minas Gerais, 1 em Maranhão, 4 em Bahia, 16 em Rio Grande Do Sul, 11 em Distrito Federal, 227 em São Paulo, 1 em Alagoas, 34 em Ceará, 55 em Mato Grosso, 4 em Tocantins, 22 em Espírito Santo, 28 em Mato Grosso do Sul, 36 em Pernambuco, 44 em Goiás, 8 em Rio Grande Do Norte e 134 em Rio De Janeiro.

Fonte: RFB – Banco de dados de CNPJ. Organização: autores, 2023.

Do ponto de vista analítico, as informações apresentadas no quadro têm o potencial de contribuir significativamente para a compreensão de diversos aspectos relevantes no contexto da pesquisa socioespacial. Isso inclui a identificação da seletividade dos lugares em diferentes escalas geográficas, resultando em desigualdades e hierarquias espaciais que moldam as decisões e estratégias de localização das empresas. Além disso, permite examinar a extensão da participação dessas empresas em outros ramos e setores, revelando processos como centralização, concentração e formação de grupos econômicos por meio da associação de empresas distintas sob personalidades jurídicas diversas.

Sócios

Complementar às informações do grupo Empresas, em especial no que diz respeito ao QSA (Quadro de Sócios e Administradores), estão concentradas as informações a respeito dos Sócios. Neste grupo, é possível obter informações de cada membro do QSA da empresa, independentemente deste membro configurar uma pessoa física ou jurídica, ou ainda, ser domiciliado no país ou no estrangeiro.

Dentre as variáveis que aparecem neste grupo, estão as informações sobre a data de entrada de cada pessoa (física ou jurídica) na sociedade, suas qualificações (presidente, administrador, diretor, entre outros), bem como o país do sócio (no caso de sócio estrangeiro), sua faixa etária e a identificação do seu representante legal (no caso de os sócios serem menores de 18 anos). É importante destacar que, como vimos anteriormente, existem muitos tipos de natureza jurídica para a classificação das empresas. No entanto, será o tipo societário⁹ escolhido na hora do cadastro da entidade jurídica que determinará a forma como a empresa será estruturada em termos de natureza, seu objeto social e participação dos sócios. De acordo com os dados do Boletim do 1º Quadrimestre de 2022 do Mapa de Empresas (Quadro 6), publicado pelo Ministério da Economia, o tipo societário com maior adesão no Brasil são as sociedades do tipo limitada (LTDA), com mais de 4,6 milhões de CNPJS deste tipo ativo no país.

⁹ No caso de empresas do tipo Empresário Individual e MEI (Microempreendedor Individual), não há a composição de sócios, uma vez que se trata de empresas de responsabilidade individual.

Quadro 6 - Movimento de abertura segundo tipo societária, primeiro quadrimestre de 2022

Tipo societário	Empresas ativas	Empresas abertas no 1º quadrimestre de 2022
Sociedade Empresária Limitada	4.667.178	226.549
Sociedade Anônima	177.898	3.749
Cooperativa	35.169	880
Demais tipos de empresas societárias ¹⁰	66.832	1.742

Fonte: RFB – Banco de dados de CNPJ. Organização: autores, 2023.

Do ponto de vista analítico, as informações sobre os sócios permitem, além do detalhamento de informações como faixa etária e país de residência, verificar a atuação dos sócios em outras empresas e setores econômicos diversos, contribuindo assim para análises que se debruçam sobre os agentes econômicos e suas atuações ramificadas (Quadro 7).

Quadro 7 - Exemplo de dado societário: capilaridade empresarial e geográfica de um dos sócios da MRV Engenharia e Participações S.A.

Dado societário	Descrição
Identificador de sócio (pessoa física, jurídica e/ou estrangeiro)	Pessoa física
Nome do sócio (pessoa física ou a razão social e/ou nome empresarial da pessoa jurídica e/ou nome do sócio/razão social do sócio estrangeiro)	Júnia Maria De Sousa Lima Galvão
Participação em outras empresas	Já participou de 404 empresas matrizes e 1395 filiais
Total de empresas ativas com participação do sócio:	1742
Capilaridade geográfica do sócio:	É ou já foi sócio de 7 empresas no estado de Paraná, 109 em Minas Gerais, 1 em Bahia, 13 em Rio Grande Do Sul, 2 em Distrito Federal, 156 em São Paulo, 1 em Alagoas, 16 em Ceará, 1 em Paraíba, 10 em Santa Catarina, 7 em Mato Grosso, 16 em Espírito Santo, 2 em Mato Grosso do Sul, 27 em Pernambuco, 6 em Goiás, 6 em Rio Grande Do Norte e 24 em Rio De Janeiro.
Código país do sócio, no caso de sócio estrangeiro	Não possui
Nome do representante legal	Não possui

Fonte: RFB – Banco de dados de CNPJ. Organização: autores, 2023.

Estabelecimentos

Por fim, o terceiro grupo de informações do banco de dados do CNPJ reúne variáveis sobre os estabelecimentos que, de acordo com a Instrução Normativa RFB nº 1.863/2018, são compreendidos como o "local privado ou público, edificado ou não, móvel ou imóvel, próprio ou de terceiros, onde a entidade exerce suas atividades em caráter temporário ou permanente ou onde se encontram armazenadas mercadorias, incluindo as unidades auxiliares". O conjunto de variáveis divulgadas neste

¹⁰ Consórcio de Sociedades; Consórcio Simples; Empresa Pública; Sociedade Estrangeira e Empresas Binacionais; Grupo de Sociedades; Sociedade de Economia Mista; Sociedade em Comandita por Ações; Sociedade em Comandita Simples; Sociedade em Conta de Participação; e Sociedade em Nome Coletivo. Fonte: Ministério da Economia, 2022.

grupo permite obter informações como o nome fantasia da empresa, a identificação do estabelecimento (se matriz ou filial), o código CNAE principal e secundário das atividades realizadas pela empresa, o endereço completo do estabelecimento (com logradouro, número, CEP, cidade, UF e, em alguns casos, e-mail de contato), a data de início da atividade e a data do evento da situação cadastral, fornecendo indicações sobre a sua situação cadastral a partir da última atualização do estabelecimento junto à RF (ativa, baixada, inapta, suspensa ou nula).

No caso do nome fantasia, a variável permite identificar o nome comercial do estabelecimento, uma vez que, em muitos casos¹¹, este pode ser diferente da razão social da empresa. A identificação do estabelecimento através de códigos de filial ou matriz possibilita compreender a capilaridade empresarial da entidade, evidenciando a extensão organizacional estabelecida pela empresa através de seus estabelecimentos. Neste caso, a matriz refere-se ao estabelecimento responsável legalmente pela gestão da empresa e de seus custos operacionais e tributários, enquanto as filiais correspondem aos estabelecimentos subordinados à matriz na forma de uma extensão da empresa já existente. Nesse sentido, pode-se conjecturar que a existência de empresas filiais corresponde, geralmente, a uma estratégia de expansão de negócios de uma entidade empresarial e atividade econômica.

Outra informação constante no grupo estabelecimento é a identificação do chamado código da CNAE das atividades, ou seja, o código identificador do estabelecimento segundo a Classificação Nacional de Atividades Econômicas (CNAE). De acordo com a Comissão Nacional de Classificação (CONCLA), a CNAE é um instrumento de padronização nacional dos códigos de atividade econômica e dos critérios de enquadramento utilizados pelos diversos órgãos da Administração Pública e Tributária do Brasil (cadastros e registros administrativos nas três esferas de poder). Sua aplicação é regulada pela Resolução Concla 01/2006 publicada no Diário Oficial em 05/09/2006, sendo utilizada para determinar quais atividades são exercidas por uma empresa. De acordo com essa última resolução, a CNAE 2.0 (versão atual) se estrutura em cinco níveis de atividades, subdivididos em 21 seções, 87 divisões, 285 grupos, 637 classes e 1301 subclasses de atividades econômicas.

No momento do cadastro do CNPJ junto à RFB, toda a empresa deve indicar o código correspondente às atividades econômicas que serão exercidas, podendo apresentar mais de uma CNAE (principal e secundária), desde que tenham naturezas relacionadas. Essa informação, do ponto de vista analítico, permite operacionalizar análises setoriais a partir dos estabelecimentos e empresas, fornecendo assim importantes elementos para pesquisas econômicas estruturais e conjunturais (Quadro 8), e a respeito da capilaridade empresarial e da diversidade de atividades exercidas pelos agentes econômicos (Quadro 8).

Quadro 8 - Exemplo de dado setorial: Brasil, empresas de Incorporação Imobiliária (CNAE 41107) abertas entre maio e junho de 2022 segundo a natureza jurídica

Natureza Jurídica	CNPJs abertos no período
Sociedade Anônima Aberta	1
Sociedade Anônima Fechada	6
Sociedade Empresária Limitada	676
Empresário (Individual)	2
Consórcio de Sociedades	1
Empresa Individual de Responsabilidade Limitada (de Natureza Empresária)	1
Total geral	687

Fonte: RFB – Banco de dados de CNPJ. Organização: autores, 2023.

¹¹ À exemplo do que comumente acontece em lojas comerciais em que o nome fantasia não revela de imediato a razão social e/ou o grupo econômico ao qual o estabelecimento pertence.

A data de início das atividades, a indicação da situação cadastral e a data da última atualização do cadastro, são importantes variáveis deste grupo. A base de dados disponibilizada pela RFB contempla uma longa série histórica de CNPJs, sendo possível identificar empresas cuja abertura remonta à década de 1960, como no caso do exemplo anterior (Quadro 9). Essa informação auxilia na produção de análises temporais sobre as empresas e os setores de atividade, observando movimentos de abertura e fechamento dos estabelecimentos ao longo dos anos.

Quadro 9 - Exemplo de dado do estabelecimento: capilaridade empresarial da Matriz e diversidade de atividades econômicas da empresa Magazine Luiza S.A.

Estabelecimento	Descrição da Informação
Nº de Inscrição CNPJ	47.960.950/0001-21
Razão Social	Magazine Luiza S/A
Nome Fantasia	Magazine Luiza
Natureza Jurídica	Sociedade Anônima Aberta
Identificador do estabelecimento	Matriz
Nº de filiais	1191 filiais
Atividade econômica principal (Código CNAE e Descrição da Atividade):	4713004 - Lojas de departamentos ou magazines, exceto lojas francas (<i>duty free</i>).
Atividades econômicas secundárias (Código CNAE + Descrição):	3329501 - Serviços de montagem de móveis de qualquer material; 4693100 - comércio atacadista de mercadorias em geral, sem predominância de alimentos ou de insumos agropecuários; 4711302 - comércio varejista de mercadorias em geral, com predominância de produtos alimentícios – supermercados; 4723700 - comércio varejista de bebidas; 4729699 - comércio varejista de produtos alimentícios em geral ou especializado em produtos alimentícios não especificados anteriormente; 4753900 - comércio varejista especializado de eletrodomésticos e equipamentos de áudio e vídeo; 4771701 - Comércio varejista de produtos farmacêuticos, sem manipulação de fórmulas; 4771704 - comércio varejista de medicamentos veterinários; 4772500 - Comércio varejista de cosméticos, produtos de perfumaria e de higiene pessoal; 4774100 - comércio varejista de artigos de óptica; 4789099 - comércio varejista de outros produtos não especificados anteriormente; 6619302 - Correspondentes de instituições financeiras 7490104 - Atividades de intermediação e agenciamento de serviços e negócios em geral, exceto imobiliários; 7911200 - agências de viagens; 7912100 - operadores turísticos 7990200 - Serviços de reservas e outros serviços de turismo não especificados anteriormente.
Data de abertura do CNPJ (Matriz)	24/10/1966
Situação Cadastral	Ativa
Data da Situação Cadastral	03/11/2005

Fonte: RFB – Banco de dados de CNPJ. Organização: autores, 2023.

Por fim, o banco fornece o endereço completo de cada estabelecimento registrado no Cadastro Nacional, com informações como logradouro, número, CEP, bairro, cidade e UF, o que permite a geocodificação dessas entidades empresariais. A possibilidade de espacializar o conjunto de informações contidas no banco de CNPJ permite, por exemplo, realizar inferências sobre as transformações no conteúdo e função de determinadas áreas de uma cidade ou região através da emergência de determinados tipos de atividades econômicas. Neste contexto, as próximas seções, além de detalharem as formas de operacionalização dessa fonte de dados, pontuam as potencialidades e limites do Cadastro Nacional de Pessoas Jurídicas no âmbito das pesquisas socioespaciais.

QUESTÕES OPERACIONAIS: CONSTRUÇÃO DO BANCO DE DADOS DE CNPJS, GEOCODIFICAÇÃO E ESPACIALIZAÇÃO

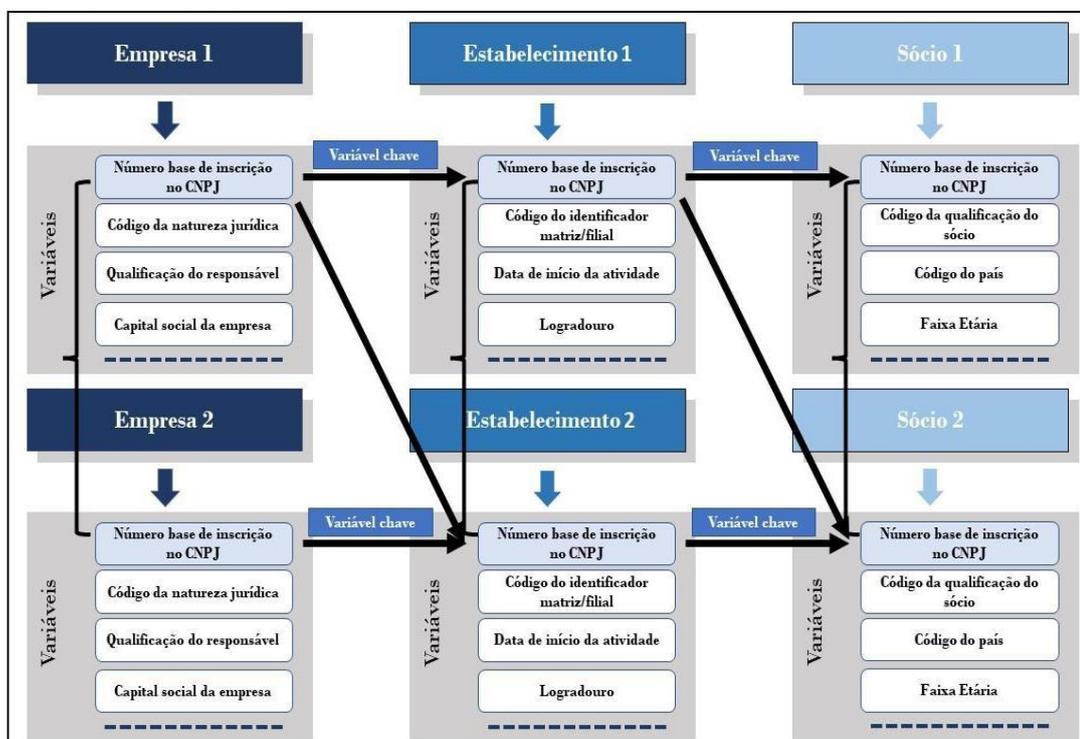
A organização de um banco de dados relacional

Considerando os tipos de variáveis contidas no banco de dados de CNPJs e a potencialidade desta fonte de informação para as pesquisas socioespaciais, a presente seção tem como objetivo detalhar brevemente as formas de construção e manipulação desta fonte de dados. O formato do arquivo do banco de dados de CNPJs disponibilizado pela Receita Federal é do tipo Banco de Dados Relacional (RDBMS – *Relational Database Management Systems*) e, para sua manipulação, gerenciamento e utilização, são necessários programas específicos.

De forma simplificada, pode-se dizer que um Banco de Dados Relacional (RBD) compreende uma tipologia de banco de dados que agrupa, conecta e organiza as informações de maneira relacional. Os RBDs são utilizados para estruturar e organizar grandes bases de dados, que possuem diferentes tabelas, colunas e linhas, através de vinculações (relacionamentos) entre as informações. O que permite uma conexão lógica entre as diferentes tabelas que compõem o banco, bem como a realização de insights e cruzamentos entre as colunas e linhas. Portanto, essa tipologia de banco de dados possibilita que um conjunto de tabulações seja vinculado e manipulado através de um atributo comum contido nas mesmas.

No caso do banco de CNPJs, como mencionado anteriormente, a Receita Federal organiza esses dados em três grandes grupos: **Empresas, Estabelecimentos e Sócios**, e cada um desses grupos é composto por diferentes tabelas. Por exemplo, o grupo Empresa é composto por nove tabelas que contêm informações sobre as empresas em todo o país. O banco de dados relacional concatena esses três grupos e as diferentes tabelas que os compõem. A Figura 1 resume a estrutura relacional entre os grupos e tabelas que compõem o banco de CNPJs. Observa-se neste esquema que a relação entre as tabelas que compõem o mesmo grupo, como Empresa 1 e Empresa 2, e os diferentes grupos é feita através da variável Número Base de Inscrição no CNPJ, que consiste nos oito primeiros dígitos do número atribuído a cada pessoa jurídica.

Figura 1 - Esquema de estruturação relacional do banco de dados do CNPJ



Organização: autoras, 2023.

Para estruturar os inúmeros arquivos que compõem a base de CNPJs em um banco integrado, foi necessário realizar o *download* de todos os arquivos listados no site da receita federal e criar uma base relacional. Esse processo de agrupamento foi realizado no *software* MySQL¹², que é um sistema de gerenciamento de banco de dados relacional que utiliza a linguagem SQL como interface. Após o agrupamento dessas tabelas em uma única base de dados, optou-se por exportar esse banco para um segundo programa denominado SPSS (*Statistical Package for the Social Sciences*). O SPSS é um programa estatístico amplamente utilizado nas ciências sociais e aplicadas, que possui uma interface fácil de usar¹³. Neste programa, foi realizada a codificação das variáveis¹⁴, ou seja, a renomeação dos códigos em valores nominais para determinadas variáveis, bem como a criação de novas variáveis com base em outras já existentes. A manipulação dessas extensas bases de dados em programas estatísticos permite uma ampla variedade de cruzamentos entre as variáveis que compõem o banco, o que possibilita avançar analiticamente em determinados objetos de estudo.

A espacialização dos dados do banco de dados do CNPJ

Além da estruturação, manipulação e cruzamento das variáveis contidas na base de CNPJs por meio de *softwares* estatísticos, há a possibilidade de espacializar o conjunto de informações. No entanto, a espacialização demanda certos procedimentos computacionais, os quais serão detalhados nos próximos parágrafos.

Uma vez que o banco de CNPJs contém informações de endereço, tais como nome da rua, número da residência, município, unidade da federação e CEP, torna-se possível geocodificar esses dados. Esse procedimento é indispensável, uma vez que o banco não possui coordenadas espaciais. Portanto, para delegar atributos espaciais a esses endereços, como informações de latitude e longitude, foram necessários alguns procedimentos.

O primeiro passo foi a seleção dos dados relevantes para o objeto de estudo e a subsequente exportação desses dados para o Excel. Por exemplo, tivemos o interesse em analisar a concentração de comércios e serviços em áreas urbanas de cidades médias. Para isso, selecionamos as cidades de foco do estudo e os códigos das atividades econômicas relevantes para a análise. Após a seleção e exportação dos dados para um arquivo Excel, foi necessário organizar as informações sobre o endereço. Devido ao fato de as informações sobre a tipologia de logradouro, nome e número do estabelecimento estarem separadas na base da Receita Federal, tornou-se necessário unir esses elementos em uma única coluna. Isso ocorre porque a maioria dos *softwares* utilizados para geocodificar endereços não permite que a tipologia do logradouro, nome e número do local a serem geocodificados estejam separados. Para reduzir as inconsistências no processo de geocodificação, também foi necessário corrigir endereços com erros de digitação e excluir aqueles que não puderam ser corrigidos. Esse procedimento é crucial para evitar imprecisões na localização dos pontos geocodificados.

A geocodificação dos endereços foi realizada no *software* ArcGIS Pro, o qual transformou os endereços em pontos no formato *shapefile*. De fato, durante a execução desses procedimentos, ocorrem perdas de endereços; no entanto, essas perdas não são estatisticamente significativas¹⁵.

¹² Além desse *software*, existem outros que são utilizados para criar e gerenciar um banco de dados relacional, tais como: *PostgreSQL*, *Microsoft SQL Server* e *Oracle*.

¹³ A exportação do banco para um segundo *software* não se faz necessária; contudo, optou-se por manipular o banco no SPSS em um segundo momento, devido à facilidade que este programa proporciona para usuários que não possuem base em linguagem de programação.

¹⁴ Em qualquer programa de manipulação de banco de dados, é possível atribuir rótulos de valores a variáveis que não são nomeadas nominalmente, como é o caso da variável cidade contida no banco de dados de CNPJ. Nessa variável, a identificação não é feita pelo nome do município, mas sim por valores numéricos. Assim, por exemplo, por meio de um dicionário de variáveis disponibilizado pela Receita Federal, foram definidos rótulos nominais, ou seja, o nome do município para os números correspondentes a cada um deles.

¹⁵ Na aplicação do procedimento para a cidade de Presidente Prudente, dos 52.876 CNPJs baixados e ativos até o ano de 2021 e que se enquadraram nos CNAEs pré-estabelecidos pelos pesquisadores em outro estudo, 51.484 foram geocodificados. A perda de aproximadamente 1.300 endereços ocorreu durante os procedimentos de correção dos endereços e geocodificação. Em exercícios realizados para outras cidades, o número de endereços perdidos não ultrapassou 5% do total, o que não compromete o conjunto total de dados

POTENCIALIDADES E LIMITES DA FONTE

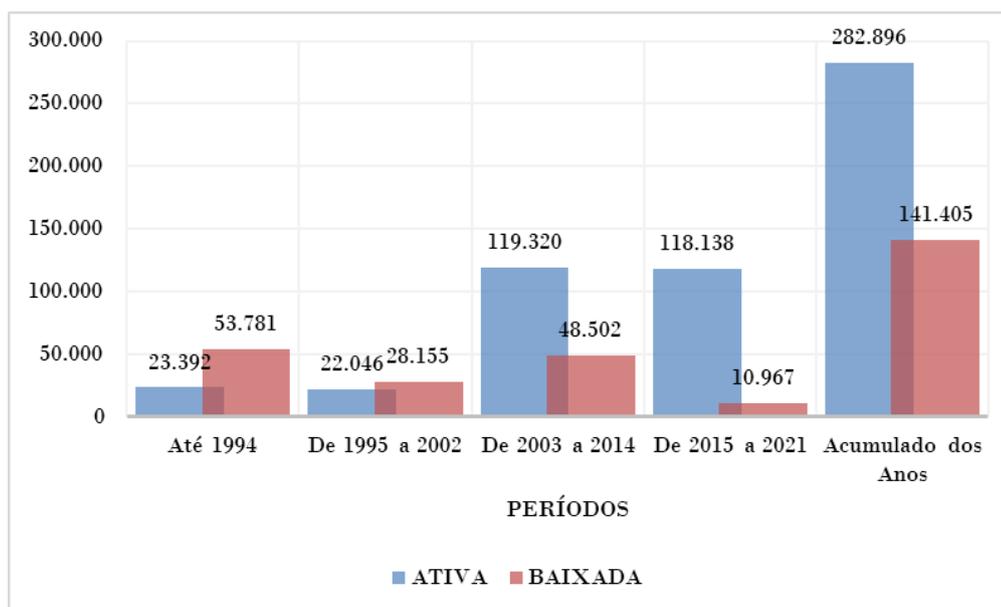
A descrição das variáveis contidas no banco de dados de CNPJs, realizada nos tópicos anteriores, ressaltou diversas contribuições significativas desta fonte de informação. Foram ressaltadas análises como a distribuição das empresas conforme seu regime jurídico, a capilaridade das empresas e sua extensão organizacional, reveladas pelas informações sobre matriz, filial e Quadro Societário (QSA), assim como análises setoriais com base no tipo de atividade, entre outras aplicações. De fato, a potencialidade de uma fonte de dados está intrinsecamente ligada ao objeto de estudo e à habilidade do pesquisador em explorá-la, reconhecendo também suas limitações. Nesse sentido, os próximos parágrafos aprofundarão e exemplificarão um pouco mais as aplicações, potencialidades e limites do Cadastro Nacional de Pessoas Jurídicas para a pesquisa socioespacial.

Uma das principais vantagens da base de CNPJs, como mencionado anteriormente, é a sua constante atualização. A inclusão regular de novos cadastros e as atualizações sobre a situação cadastral, incluindo o encerramento de empresas, permitem o monitoramento contínuo das entidades com personalidade jurídica. Poucas fontes de dados no Brasil possuem essa frequência de atualização. Além disso, esse banco de dados possui uma vasta série histórica que abrange informações sobre entidades jurídicas com registros que datam desde a década de 1960, proporcionando importantes inferências temporais para análises socioespaciais.

Através de variáveis como a data de início da atividade e a data em que a empresa entrou em situação especial, torna-se possível analisar a abertura e o fechamento de empresas/estabelecimentos, bem como sua localização, deslocamento ou desaparecimento territorial em diferentes escalas, desde o intraurbano até o regional e nacional, ao longo de diversos períodos. Os dados cadastrais também possibilitam uma análise detalhada de atividades econômicas específicas, graças à classificação dos estabelecimentos conforme a Classificação Nacional de Atividades Econômicas (CNAE). Essa abordagem permite esclarecer movimentos macroeconômicos e conjunturais, identificando crises cíclicas, ajustes na economia, entre outros processos relevantes.

Um exemplo que ilustra essa capacidade é apresentado na Figura 2, que exhibe a evolução da abertura e fechamento de empresas no setor da construção civil¹⁶ no Brasil, com dados históricos a partir de 1960. Observa-se nesse gráfico um crescimento contínuo no número de empresas ativas (matrizes) ao longo dos anos. Essas informações fornecem uma base sólida para análises relacionadas ao setor, possibilitando uma investigação mais aprofundada de seus ciclos e desempenho. Além disso, facilita a compreensão da influência das políticas macroeconômicas e anticíclicas.

Figura 2 - Brasil: Evolução do setor da construção civil nos últimos 60 anos, 1960 e 2021



Fonte: RFB – Banco de dados de CNPJ. Organização: autoras, 2023.

¹⁶ CNAEs 4110700 e 4120400, referente às atividades econômicas ligadas à construção de edifícios e Incorporação de empreendimentos imobiliários.

A análise espacial dos setores produtivos também proporciona uma compreensão sobre a especialização produtiva dos diferentes lugares (HARVEY, 2013) e os efeitos dessa especialização sobre outras atividades econômicas. Essa abordagem pode ser ainda mais aprofundada ao combinar outras variáveis do banco de CNPJ, como os dados societários.

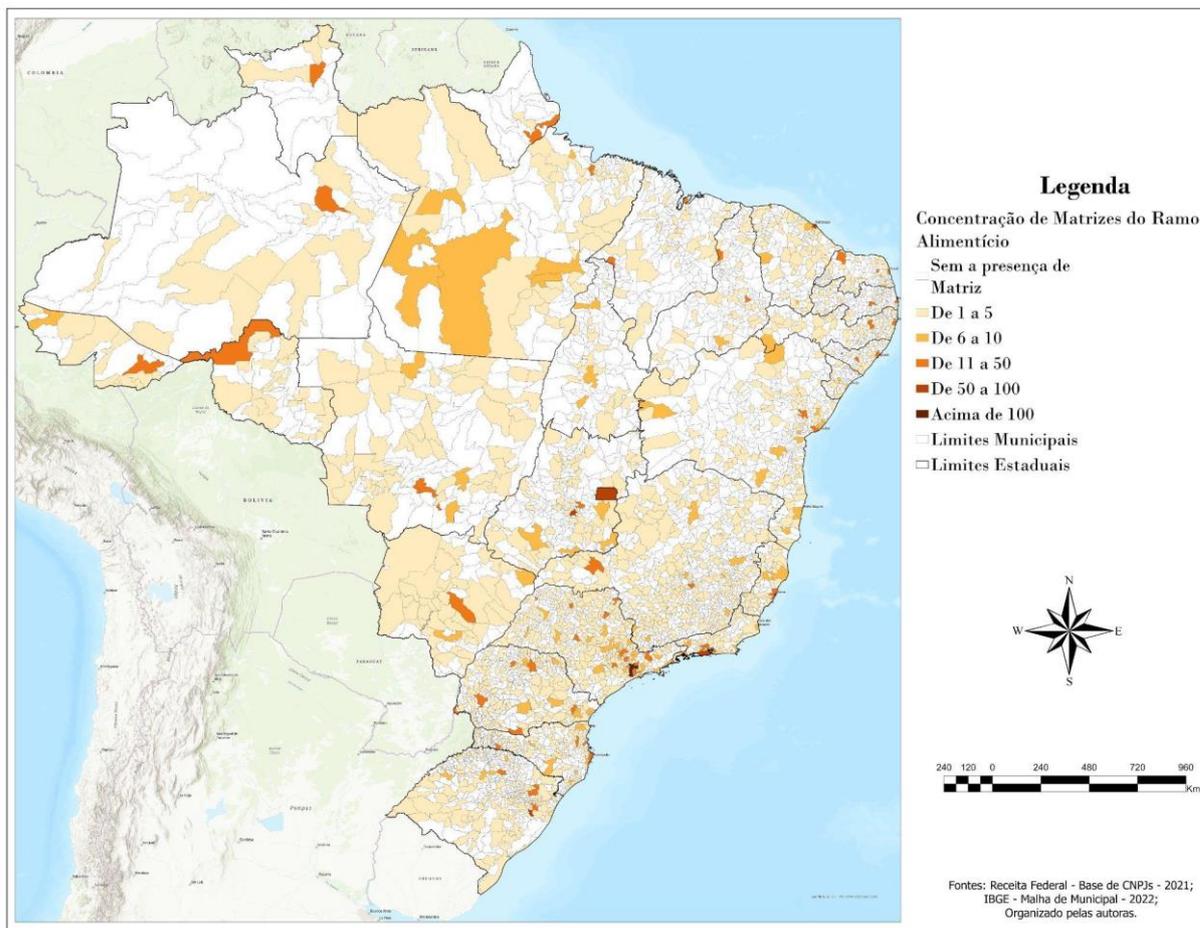
Dessa forma, a análise integrada dos dados proporciona uma visão abrangente e detalhada da dinâmica econômica, permitindo identificar tendências, padrões e influências conjunturais e estruturais em diversos setores da economia, bem como na produção do espaço, através da identificação das lógicas, estratégias e escalas de atuação/localização de diversas atividades econômicas, empresas ou agentes, incluindo diferentes níveis hierárquicos, mesmo na escala intraurbana.

Um exemplo ilustrativo pode ser observado nas Figuras 3 e 4, que mostram a espacialização¹⁷ das matrizes e filiais do setor alimentício nos municípios brasileiros em 2021¹⁸. Ao analisar a concentração de matrizes por estados (Figura 3), percebemos a presença de vazios no território brasileiro, seguidos por uma vasta região concentrada, especialmente liderada pelo estado de São Paulo, que se destaca como o único do Brasil com um município, no caso a capital paulista, concentrando mais de 100 matrizes. Essa concentração relativa indica ainda uma centralização do capital comercial nessa região, revelando desigualdades regionais e evidenciando uma tendência locacional específica do setor.

¹⁷ Em busca de uma espacialização ainda mais enriquecedora, podemos combinar diversas variáveis presentes no banco de CNPJ. Isso inclui, por exemplo, a data de início das atividades das matrizes, o ano em que as filiais se expandiram e as regiões em que essa expansão ocorreu. Além disso, é importante analisar as redes de mercados e hipermercados que predominam em cada nível escalar, bem como outros cruzamentos de dados relevantes.

¹⁸ Nas presentes ilustrações (Figuras 3 e 4), consideram-se apenas as matrizes do setor alimentício que tinham filiais; aquelas que não apresentaram desdobramentos de suas matrizes foram desconsideradas. Os seguintes CNAES foram considerados nas representações: 4639701, 4691500, 4711301 e 4711302, que correspondem, respectivamente, a: Comércio atacadista de produtos alimentícios em geral; Comércio atacadista de mercadorias em geral, com predominância de produtos alimentícios; Comércio varejista de mercadorias em geral, com predominância de produtos alimentícios – hipermercados; Comércio varejista de mercadorias em geral, com predominância de produtos alimentícios – supermercados.

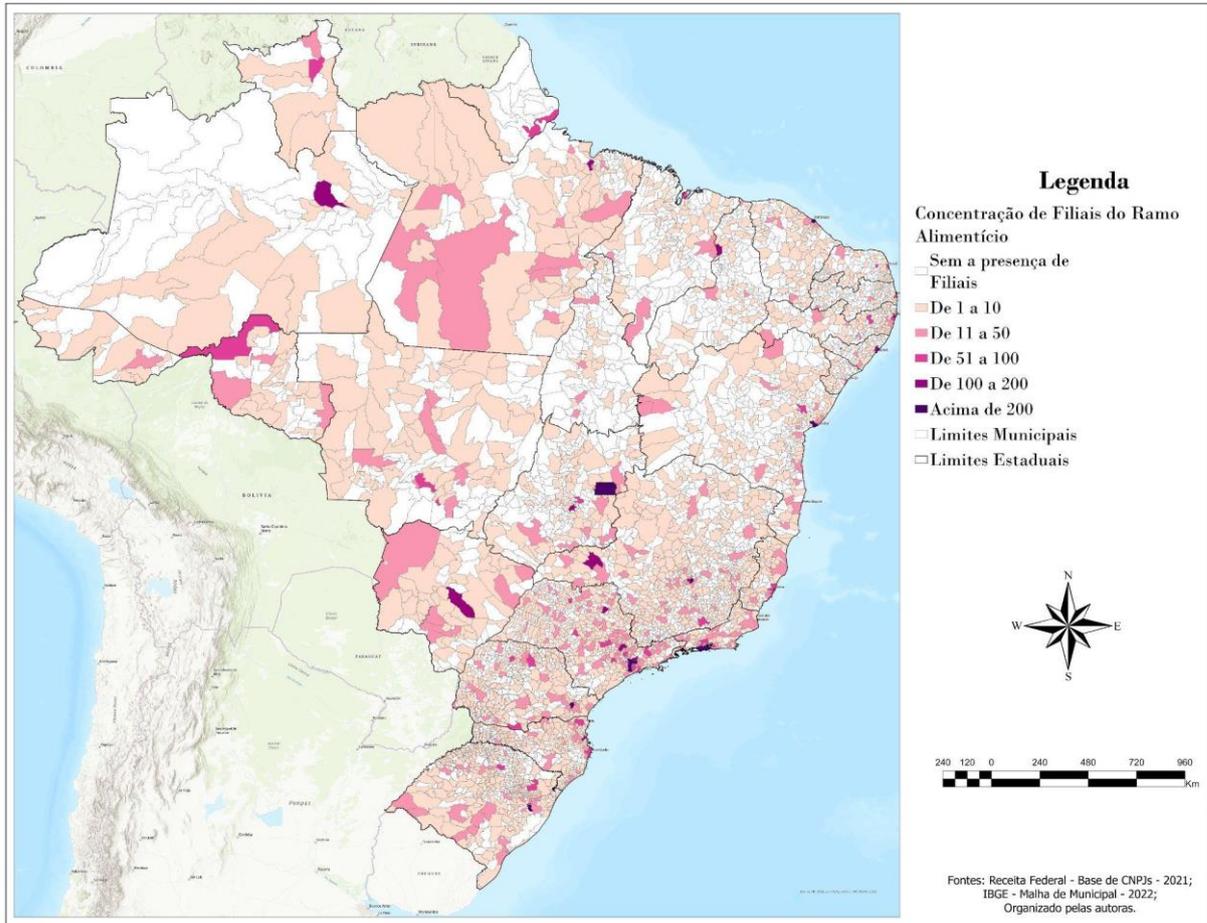
Figura 3 - Brasil: Concentração de matrizes que possuem filiais do ramo alimentício (atacado e varejo incluindo supermercados e hipermercados), 2021



Fonte: RFB – Banco de dados de CNPJ. Organização: autoras, 2023.

Da mesma forma, a presença de filiais (Figura 4) também revela vazios no território, embora em menor intensidade comparada às matrizes (Figura 3). Os vazios concentram-se predominantemente nos estados das regiões Norte e Nordeste. Por outro lado, verifica-se uma notável densidade de filiais nos estados de São Paulo, Minas Gerais, Rio de Janeiro, Rio Grande do Sul, entre outros, com a presença de municípios que abrigam mais de 200 filiais. Este tipo de análise é importante para compreensão das dinâmicas econômicas e das estratégias adotadas pelas empresas do ramo alimentício ao expandirem suas operações por diferentes regiões do país. Assim como contribui para identificar possíveis influências de infraestrutura, mercado consumidor, logística e políticas governamentais no processo de concentração, promovendo um entendimento sobre as desigualdades regionais.

Figura 4 - Brasil: Concentração de filiais que possuem filiais do ramo alimentício (atacado e varejo incluindo supermercados e hipermercados), 2021



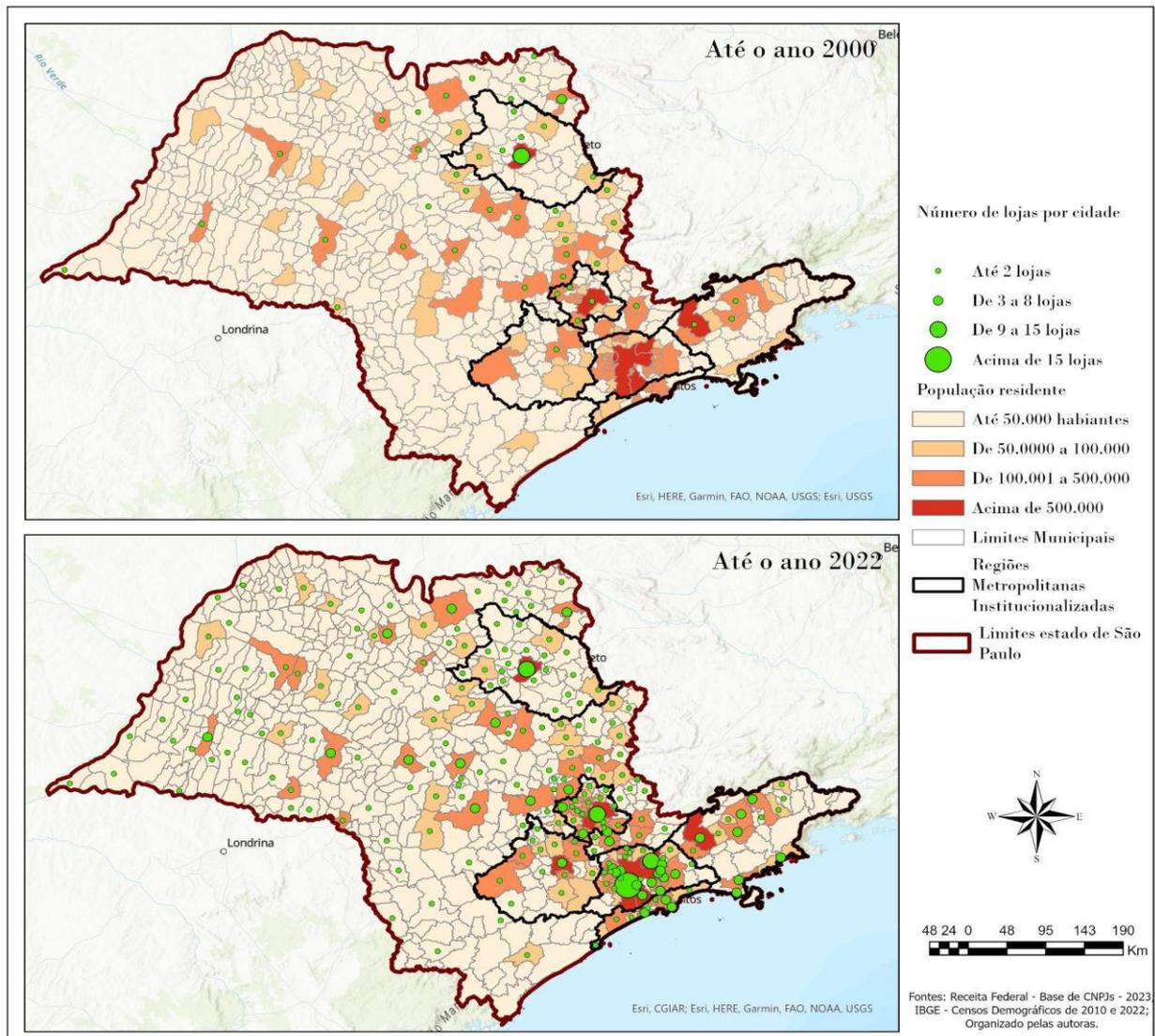
Fonte: RFB – Banco de dados de CNPJ. Organização: autoras, 2023

Conforme evidenciado nos exemplos anteriores, a análise setorial, fundamentada na identificação das atividades econômicas, proporciona uma compreensão abrangente de diversos processos socioespaciais, como a produção das desigualdades regionais, a especialização dos lugares e os ajustes espaciais. No entanto, é relevante destacar que esses objetos analíticos podem ser explorados de maneira ainda mais detalhada por meio da variável Razão Social, presente no Cadastro Nacional de Pessoas Jurídicas (CNPJ). Ao contrário da variável de atividade econômica, a razão social identifica tanto a empresa ou grupo econômico quanto a atividade econômica principal exercida por essa empresa ou grupo.

Dessa forma, a inclusão da variável Razão Social oferece uma visão mais aprofundada das dinâmicas econômicas e das estratégias adotadas por cada entidade jurídica, permitindo compreender de que maneira suas ações se materializam no espaço, em diferentes escalas territoriais, incluindo a reconfiguração da estrutura socioespacial tanto em níveis interurbanos quanto intraurbanos (Spósito e Spósito, 2017). Esse enfoque expandido representa uma contribuição para a pesquisa socioespacial, possibilitando análises mais precisas e detalhadas das influências das atividades econômicas na organização do espaço. Com essa abordagem, os pesquisadores podem explorar mais profundamente as relações entre as atividades econômicas, os lugares onde estão localizadas e as interações entre os diversos setores da economia. Essa compreensão mais detalhada contribui para o entendimento dos processos que moldam a geografia econômica e social, auxiliando na formulação de políticas e estratégias mais adequadas para o desenvolvimento regional e urbano. Para ilustrar esse ponto, a Figura 5 retrata a localização dos estabelecimentos do Grupo Magazine Luiza no estado de São Paulo em dois momentos: desde sua fundação até o ano 2000 e após o ano 2000 até 2021 - segundo classes

de tamanho de municípios e tipologias de aglomerado¹⁹. A representação cartográfica apresentada ilustra a expansão desta empresa, evidenciando sua estratégia de localização em direção à área metropolitana de São Paulo, com a capital paulistana concentrando a maioria dos estabelecimentos da rede. Além disso, a preferência por cidades com mais de 100 mil habitantes ou próximas a essas cidades pode estar relacionada à densidade do mercado consumidor. Essas informações são essenciais para compreender os efeitos das ações das empresas na transformação do espaço em diferentes escalas.

Figura 5 - São Paulo, Distribuição da rede de lojas das lojas do grupo Magazine Luiza por municípios, 1990 e 2022

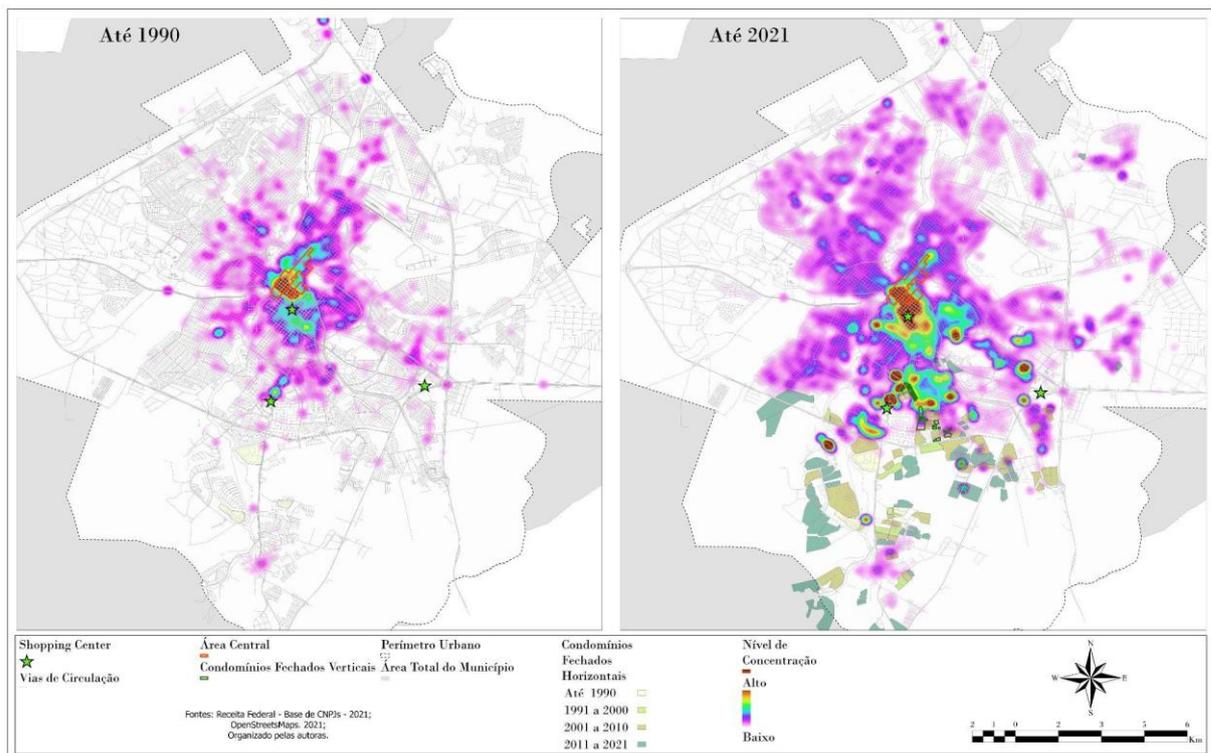


Fonte: RFB – Banco de dados de CNPJ. Organização: autores, 2023.

¹⁹ A rede Magazine Luiza foi fundada em 1957, na cidade de Franca. Até o ano de 2000, a empresa possuía apenas 39 lojas localizadas no estado de São Paulo, concentradas principalmente nas regiões e municípios próximos a Franca, como o aglomerado de Ribeirão Preto, além de outras cidades fora da Região Metropolitana de São Paulo. No entanto, a partir do ano 2000, a empresa iniciou um notável processo de expansão, registrando um crescimento expressivo. Em 2021, a rede Magazine Luiza alcançou a marca de 249 estabelecimentos apenas no estado de São Paulo. A nível nacional, sua presença também se tornou relevante, contando com um total de 508 estabelecimentos espalhados por diversas regiões do país. Esse crescimento significativo demonstra o sucesso e a influência da empresa no cenário varejista brasileiro

Por fim, outra aplicabilidade do Banco de Dados de CNPJ é o estudo das mudanças na estrutura intraurbana, como evidenciado na Figura 6²⁰, que mostra a concentração de atividades de comércio e serviços na cidade de Ribeirão Preto nos anos de 1990 e 2021. Através dessa representação cartográfica, é possível observar a expansão da concentração de comércios e serviços para além da área central principal, em direção à zona sul da cidade, indicando a emergência de subcentros e novas áreas centrais ao longo do tempo.

Figura 6 - Ribeirão Preto: Concentração de comércios e serviços, 1990 e 2021



Fonte: RFB – Banco de dados de CNPJ. Organização: autores, 2023.

Destaca-se a importância da qualificação das concentrações de comércio e serviços, considerando fatores como o tipo de atividade (matriz ou filial), os grupos econômicos envolvidos e o porte das empresas. Essas dimensões, juntamente com informações sobre o perfil dos frequentadores dessas áreas (renda, escolaridade etc.) e a qualidade do entorno onde estão situadas (renda, tipos de habitações etc.), provenientes também de outras fontes de dados e metodologias, desempenha um papel fundamental na compreensão do processo de produção desigual das cidades, como a emergência de multi(poli)centralidade e outros fenômenos relevantes.

Em síntese, as variáveis disponíveis no banco de CNPJ apresentam um potencial significativo para as pesquisas socioespaciais, permitindo análises detalhadas, como o surgimento e evolução de centralidades urbanas, especialmente quando combinadas com outras fontes de dados e métodos de pesquisa. Essa abordagem enriquece a compreensão dos processos socioespaciais e revela-se como uma ferramenta para estudos setoriais, possibilitando uma investigação abrangente e minuciosa da dinâmica econômica, estratégias de atuação e localização dos setores econômicos em diversas escalas territoriais. Entretanto, é essencial reconhecer que, como em qualquer fonte de dados, o CNPJ

²⁰ Na Figura 6, buscamos representar também a evolução dos espaços residenciais fechados na cidade de Ribeirão Preto. Ao analisar essas duas informações, percebe-se que o surgimento das novas áreas de concentração de comércio e serviços está próximo às regiões onde estão localizados os condomínios fechados. Esse tipo de cruzamento de dados pode fornecer *insights* sobre a configuração de centralidades voltadas para determinados tipos de capitais e grupos de pessoas.

possui suas limitações, e os pesquisadores devem estar cientes delas ao conduzir suas análises. Dentre as principais limitações da base de dados, duas merecem destaque.

Primeiramente, evidencia-se a sobreenumeração de registros, que ocorre quando uma mesma empresa realiza múltiplas aberturas de CNPJs sem realizar a devida baixa da inscrição de atividades encerradas, resultando na duplicação das informações e superestimação do dado. Da mesma forma, a presença dos chamados CNPJs fantasmas²¹ também se mostra como uma limitação que exige cautela do pesquisador, especialmente quando o dado é utilizado para análises espaciais, apontando inconsistências que podem ser identificadas através da produção de mapas, como por exemplo, na identificação da concentração elevada de registros de atividades econômicas em determinados endereços cujo uso é predominantemente residencial. No entanto, apesar destes entraves, desde 2018 a Receita Federal do Brasil (RFB) passou a inativar automaticamente o CNPJ de empresas que não atualizam suas declarações no prazo de dois anos, o que tem mitigado esse problema em certa medida. A segunda limitação refere-se à subnotificação do capital social das empresas. O capital social representa o patrimônio inicial investido na abertura de uma empresa. Geralmente, esse valor é subestimado para evitar maiores taxações e tributações, o que limita a precisão das informações e pode distorcer o porte real das empresas.

Por fim, é importante destacar que as informações do CNPJ apresentam variações em determinados períodos do ano, correspondendo aos prazos legais de declaração dos cadastros, como no caso dos prazos do Simples Nacional, que ocorre todo mês de fevereiro de cada ano. Portanto, as análises temporais e setoriais devem levar em consideração essas variações, que podem influenciar em picos de baixa e/ou abertura de novas empresas.

CONSIDERAÇÕES FINAIS

Em um período marcado pela aceleração contemporânea (SANTOS, 1996), por mudanças sociais, econômicas, culturais e espaciais profundas, é preciso novos aportes teóricos-metodológicos que auxiliem a compreensão das transformações e da complexidade que caracterizam os processos socioespaciais no contemporâneo, marcados pela fluidez e multiescalaridade. A adoção de novas metodologias, ou o uso conjunto de diversas metodologias, como qualitativas e quantitativas, pode ser visto como um caminho para entender a complexidade da realidade contemporânea em suas múltiplas escalas. Em síntese, as metodologias podem ser interpretadas como o processo de transformação da realidade em dados apreensíveis e cognoscíveis, que busca tornar inteligível um objeto de estudo.

É neste contexto que o banco de dados do Cadastro Nacional de Pessoas Jurídicas (CNPJ), fornecido mensalmente pela Receita Federal brasileira, destaca-se como ferramenta para pesquisas socioespaciais. Apesar de suas limitações, como erros de imputação do capital social e sobreenumeração de CNPJs, suas potencialidades, como a atualização periódica e a presença de variáveis temporais, o tornam uma fonte secundária relevante. No contexto das pesquisas socioespaciais, a possibilidade de espacialização do dado e capacidade de trabalhar com esta fonte em diversos níveis escalares, também se coloca como potência no contexto dos estudos socioespaciais. Ao longo do texto, demonstramos alguns exemplos de aplicabilidade desta fonte secundária, em distintos níveis escalares e cruzando com outras fontes de dados. Compreender processos socioespaciais como desigualdades regionais e intraurbana, especialização dos lugares, emergência de centralidades, e as relações de mudanças macroeconômicas em distintas escalas, foram algumas das aplicabilidades do CNPJ. De fato, as aplicabilidades, cruzamentos de variáveis, e a combinação desta análise quantitativa através desta fonte de dados com outras metodologias, são muito maiores do que os exemplos pontuados no texto, e se relaciona com a expertise do pesquisador e com objeto investigado como já reiterado.

Capilaridade de empresas e de agentes, estratégias e lógicas de localização e de expansão de atividades econômicas, agentes e empresas são também recursos analíticos possíveis de serem

²¹ O termo CNPJ fantasma faz referência a uma entidade jurídica que não possui atividade real no mercado, mas que ainda assim possui um registro formal. Essa entidade pode apresentar informações como razão social, quadro societário (QSA) e endereços fictícios ou associados a residências, muitas vezes sem o conhecimento dos verdadeiros moradores. No contexto empresarial, essas entidades também são conhecidas como empresas de fachada, empresas de prateleira (shelf company) ou empresas *ready-made*. Tais entidades são frequentemente criadas com intenções criminosas, sendo utilizadas para cometer delitos relacionados à Ordem Tributária, à Administração Pública (como contrabando e descaminho) e ao Sistema Financeiro Nacional (como lavagem de dinheiro).

operacionalizados a partir do banco de dados de CNPJ, e utilizados em investigações que buscam compreender, por exemplo, mudanças na rede urbana a partir da expansão de atividades e agentes para distintos níveis hierárquicos de cidades. Portanto, multiescalaridade, periodização e análise do conteúdo e função, são algumas das empirias que o banco permite.

AGRADECIMENTOS

Este texto está associado ao desenvolvimento da pesquisa financiada pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), na modalidade projeto temático (processo 2018/07701-8), intitulada “Fragmentação socioespacial e urbanização brasileira: escalas, vetores, ritmos, formas e conteúdo”. O texto também está associado a outra pesquisa financiada pela FAPESP (vinculada ao projeto citado acima, na modalidade Bolsas de pós-doutorado processo 2019/04392-7 e processo 2019/05225-7)

REFERÊNCIAS

BABBIE, E. **Métodos de Pesquisas de Surveys**. Belo Horizonte: Editora UFMG, 2001, 519 p.

BRASIL. Lei nº 4.503, de 30 de novembro de 1964. Institui, no Ministério da Fazenda, o cadastro geral de pessoas jurídicas, cria o Departamento de Arrecadação e dá outras providências **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 30 nov. 1964. Seção 1 - Suplemento, p. 48.

BRASIL. Lei nº 5.614, de 5 de outubro de 1970. Dispõe sobre o Cadastro Geral de Contribuintes (C.G.C.). **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 5 out. 1970.

BRASIL. Lei nº 6.404, de 15 de dezembro de 1976. Dispõe sobre as Sociedades por Ações. **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 17 dez. 1976.

BRASIL. Lei nº 8.212, de 24 de julho de 1991. Dispõe sobre a organização da Seguridade Social. **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 25 jul. 1991.

BRASIL. Lei nº 8.935, de 18 de novembro de 1994. Regulamenta o art. 236 da Constituição Federal, dispondo sobre serviços notariais e de registro (Lei dos cartórios). **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 21 nov. 1994

BRASIL. Lei nº 10.406, de 10 de janeiro de 2002. Institui o Código Civil. **Diário Oficial da União: seção 1, Diário oficial [da] República Federativa do Brasil**, Brasília, DF, ano 139, n. 8, p. 1-74, 11 jan. 2002.

BRASIL. Lei nº 10.931, de 2 de agosto de 2004. Dispõe sobre o patrimônio de afetação de incorporações imobiliárias, Letra de Crédito Imobiliário, Cédula de Crédito Imobiliário, Cédula de Crédito Bancário, altera o Decreto-Lei nº 911, de 1º de outubro de 1969, as Leis nº 4.591, de 16 de dezembro de 1964, nº 4.728, de 14 de julho de 1965, e nº 10.406, de 10 de janeiro de 2002. **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 3 ago. 2004.

BRASIL. Lei Complementar nº 123, de 14 de dezembro de 2006. Institui o Estatuto Nacional da Microempresa e da Empresa de Pequeno Porte. **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 14 dez. 2006. Seção 1, p. 1.

BRASIL. Lei nº 12.527, de 18 de novembro de 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. **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 18 nov. 2011.

BRASIL. Lei nº 12.527, de 18 de novembro de 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. **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 18 nov. 2011. Seção 1, p. 1. Edição Extra.

BRASIL. Decreto nº 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo Federal. **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 12 maio 2016, p. 21, col. 1.

BRASIL. Lei nº 14.195, de 26 de agosto de 2021. Dispõe sobre a facilitação para abertura de empresas. **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 26 ago. 2021. Seção 1, p. 1

BRASIL. Medida Provisória nº 1.085, de 27 de dezembro de 2021. Dispõe sobre o Sistema Eletrônico dos Registros Públicos - SERP. **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 27 dez. 2021. Seção 1, p. 1.

BRASIL. Secretaria da Receita Federal. Instrução Normativa SRF nº 27, de 05 de março de 1998. Dispõe sobre o Cadastro Nacional da Pessoa Jurídica - CNPJ **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 09 mar. 1998. Seção 1, p. 18.

BRASIL. Receita Federal do Brasil. Instrução Normativa RFB nº 1863, de 27 de dezembro de 2018. Dispõe sobre o Cadastro Nacional da Pessoa Jurídica (CNPJ). **Diário oficial [da] República Federativa do Brasil**, Brasília, DF, 28 dez. 2018. Seção 1, p. 354.

HARTSHORNE, R. **Propósitos e natureza da Geografia**. 2. ed. São Paulo: Hucitec, 1978.

HARVEY, D. **Os limites do capital**. São Paulo: Boitempo, 2013^a

Ministério da Economia (Brasil). **Cadastro Nacional da Pessoa Jurídica (CNPJ)**. Disponível em: <https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj>. Acesso em: 16 de julho de 2023.

SANTOS, M. **A natureza do espaço** – Técnica e tempo. Razão e emoção. São Paulo: Hucitec, 1996.

SOUZA, M.L. Da “diferenciação de áreas” à “diferenciação socio-espacial: “visão (apenas) de sobrevoos” como uma tradição epistemológica e metodológica limitante. **Revista Cidades**. Presidente Prudente, v 4 n 6, 2007.

SOUZA, M.L. **Os conceitos fundamentais da pesquisa sócio-espacial**. 1. ed. Rio de Janeiro: Bertrand Brasil, 2013. v. 1. 319p <https://doi.org/10.36661/2448-1092.2007v4n6.12798>

SPOSITO, M.E.B.; SPOSITO, E.S. Articulação entre múltiplas escalas geográficas: lógicas e estratégias espaciais de empresas. **GEOUSP Espaço e Tempo (Online)**, v. 21, n. 2, p. 462-479, 2017. <https://doi.org/10.11606/issn.2179-0892.geousp.2017.131655>

SUERTEGARAY, D.M. **España geográfico uno e múltiplo**. **Scripta Nova**: revista electrónica de geografía y ciencias sociales, 2001.

Recebido em: 26/07/2023

Aceito para publicação em: 06/02/2024