Proposta de Reformulação da BADEPI para análise de dados de Marcas Coletivas

Raul Tavares Cecatto[1,2], Fabrício Geraldo dos Santos Rodrigues[2], Thiago Maia Sayão de Moraes[3]

[1]Instituto Federal de Educação Ciência e Tecnologia de Mato Grosso – *Campus* Campo Verde
Avenida Isidoro Luiz Gentilin, 585 – 78841-006 – Campo Verde – MT – Brasil

[2]Instituto Federal de Educação Ciência e Tecnologia de Mato Grosso – *Campus* Primavera do Leste

[3]SECITECI – Escola Técnica Estadual de Primavera do Leste

{raul,fabricio}@ifmt.edu.br, thiagomsayao@gmail.com

Abstract. *The article analyzes the Collective Trademark as an instrument of recognition and productive inclusion, highlighting the lack of quantitative studies on its use by traditional peoples and communities. Given the limitations of INPI's databases, especially BADEPI and PePI, the study proposes a three-step method for data restructuring and enrichment: importing into a relational database (SQL), extraction via web scraping with Python, and supplementation through the Brasil API. After textual cleaning and standardization, the consolidated database gathered legal, sectoral, and geographic information, made publicly available. The research demonstrates the relevance of computational organization of public databases to democratize access to information and foster studies on industrial property and social innovation.*

Resumo. *O artigo analisa a Marca Coletiva como instrumento de valorização e inclusão produtiva, destacando a falta de estudos quantitativos sobre seu uso. Diante das limitações das bases do INPI, especialmente da BADEPI e do PePI, o trabalho propõe um método de reestruturação e enriquecimento de dados em três etapas: importação em banco relacional (SQL), extração via web scraping com Python e complementação por meio da Brasil API. Após limpeza e padronização textual, a base consolidada reuniu informações jurídicas, setoriais e geográficas, disponibilizadas publicamente. A pesquisa demonstra a relevância da organização computacional de bases públicas para democratizar o acesso à informação e fomentar estudos sobre propriedade industrial e inovação social.*

1. Introdução

A Marca Coletiva é um instrumento jurídico previsto na Lei de Propriedade Industrial (Lei nº 9.279/1996) que serve para a valorização de identidades e o fortalecimento da inclusão produtiva de grupos organizados. Contudo, a literatura científica carece de estudos quantitativos sistemáticos que revelem a dimensão da apropriação desse instituto , por exemplo, por povos e comunidades tradicionais (PCTs). A principal razão para a ausência desse diagnóstico é, em grande parte, a assimetria informacional e as barreiras de acesso e organização dos dados brutos disponibilizados pelo órgão regulador, o INPI (Instituto Nacional da Propriedade Industrial). As limitações das bases oficiais, como a Base de Dados de Propriedade Industrial (BADEPI) e o portal Pesquisa

em Propriedade Industrial (PePI) , tornam essencial a consolidação metodológica e computacional para superar as deficiências das fontes originais.

O desafio central para a análise estatística de marcas coletivas reside nas limitações das bases oficiais do Instituto Nacional da Propriedade Industrial (INPI). A Base de Dados de Propriedade Industrial (BADEPI) fornece um grande volume de informações em formato CSV, mas essas informações não são completas e padronizadas para a reconstrução histórica dos registros. Por sua vez, o portal Pesquisa em Propriedade Industrial (PePI) não permite buscas específicas por modalidade de marca (como a coletiva), restringindo a pesquisa a apenas alguns campos textuais.

Diante disso, a consolidação metodológica e computacional da base de dados torna-se uma etapa essencial, pois supera as deficiências das fontes originais. Este artigo se concentra em demonstrar a importância da organização e reconstrução dos dados para viabilizar análises variadas.

2. A reformulação da BADEPI

O método adotado neste trabalho insere-se no campo da pesquisa documental e exploratória e baseia-se na reestruturação e enriquecimento de bases de dados públicas para superar as barreiras de assimetria informacional. A metodologia empregada seguiu as diretrizes propostas por Cecatto, Rodrigues e Moraes (2025) para a extração e tratamento de dados de marcas coletivas.

A reestruturação da base de dados envolveu três etapas computacionais principais: (1) a importação e organização em um banco de dados relacional; (2) a extração delimitada e o *web scraping* para complementação de dados; e (3) o enriquecimento via API para coleta de informações sobre os titulares. Estas etapas foram consolidadas na Figura 1.

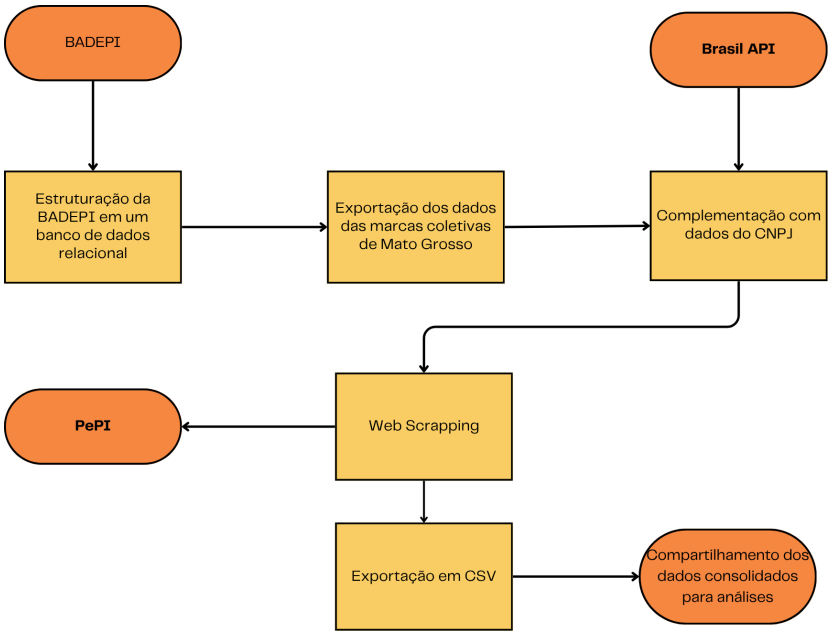

Figura 1 - Fluxo metodológico da extração e enriquecimento de dados (BADEPI + BrasilAPI)

2.1 Importação e Organização dos Dados Brutos

O ponto de partida foi a Base de Dados de Propriedade Industrial (BADEPI) do INPI, que fornece um grande volume de dados em formato CSV. Para permitir uma manipulação eficiente e consultas complexas, as quatro planilhas principais da BADEPI (classes de Nice, depositantes, despachos e pedidos) foram importadas em um banco de dados relacional (SQL).

Essa abordagem foi crucial, pois a estrutura relacional do SQL, pela sua proximidade com a forma como a BADEPI é conceitualmente organizada, permitiu consolidar e relacionar milhões de registros, superando a limitação do portal oficial PePI, que não permite buscas específicas por modalidade de marca (Marca Coletiva).

2.2 Extração Delimitada e *Web Scraping*

Após a organização em SQL, foi possível delimitar o escopo da busca antes da extração final, algo restrito no portal PePI. A consulta foi restringida a processos de depósito de marcas coletivas concedidas a pessoas jurídicas brasileira.

Em seguida, o número do processo (NO_PROCESS) extraído da BADEPI foi utilizado em um processo de *web scraping*. Foi desenvolvido um *script* em Python utilizando a biblioteca Selenium para automatizar a coleta de informações diretamente do portal Pesquisa em Propriedade Industrial (PePI) do INPI. O objetivo principal do *web scraping* foi capturar a situação completa do registro, que não é totalmente descrita na entidade que concentra os despachos da BADEPI.

2.3 Enriquecimento e Consolidação da Base

Para um melhor entendimento das marcas concedidas e seus titulares, a base de dados foi enriquecida com informações adicionais: As organizações vinculadas aos pedidos (titulares) foram consultadas por meio da Brasil API. Essa consulta permitiu complementar os dados com informações do Cadastro Nacional da Pessoa Jurídica (CNPJ) das entidades.

O procedimento resultou em uma tabela final consolidada, a qual não foi normalizada para as análises, essa decisão metodológica foi tomada para facilitar as análises , garantindo que todos os atributos cruciais—como natureza jurídica do titular, CNAE, Classe de Nice, e localização geográfica —estivessem prontos para uso. Dessa forma, os dados, já tratados, enriquecidos e classificados, são disponibilizados publicamente para que pesquisadores e o público em geral, incluindo não especialistas em grandes volumes de dados, possam utilizá-los. Essa tabela reuniu atributos cruciais para a inferência, incluindo: Natureza jurídica do titular; CNAE; Classe de Nice; Localização geográfica; Denominação social e nome da marca.

Essa base de dados consolidada e enriquecida, acessível publicamente (https://zenodo.org/records/16995708), forneceu o alicerce para a aplicação posterior do procedimento de inferência por proxies (legais, textuais, setoriais e territoriais), que visou estimar a probabilidade de vínculo dos registros com povos e comunidades tradicionais.

3. Análises possíveis

Embora não seja o foco deste trabalho, os dados organizados desta forma permitem que uma série de análises e gráficos sejam elaborados. A contagem de registros por estado (Figura 2a), já era possível executar com a base bruta, embora envolvesse conhecimentos de junções das tabelas originais, além de um compreendimento avançado na forma com que os processos eram tramitados no INPI, para exclusão das marcas que não estejam ativas. O mesmo ocorre para a contagem das marcas por classe de Nice (Figura 2b), que nos permite entender as categorias dos produtos e serviços registrados. Já a distribuição dos registros por CNAE (Figura 2c) é uma facilidade trazida por este estudo, ao relacionar os dados da receita federal aos da BADEPI, e permite complementar a análise por Classe de Nice e entender as atividades que a organização exerce primariamente. Além disso, é possível analisar:

- Trajetória de Continuidade (Longevidade): É possível delinear a trajetória de continuidade das marcas, por exemplo, verificando quantas marcas alcançaram o segundo decênio de vigência.
- Situação do Registro: A base contém campos que detalham a situação do registro (ativo/em vigor, extinto, nulo), permitindo analisar os motivos de extinção (por exemplo, extinção por encerramento/liquidação voluntária, como visto em alguns registros).

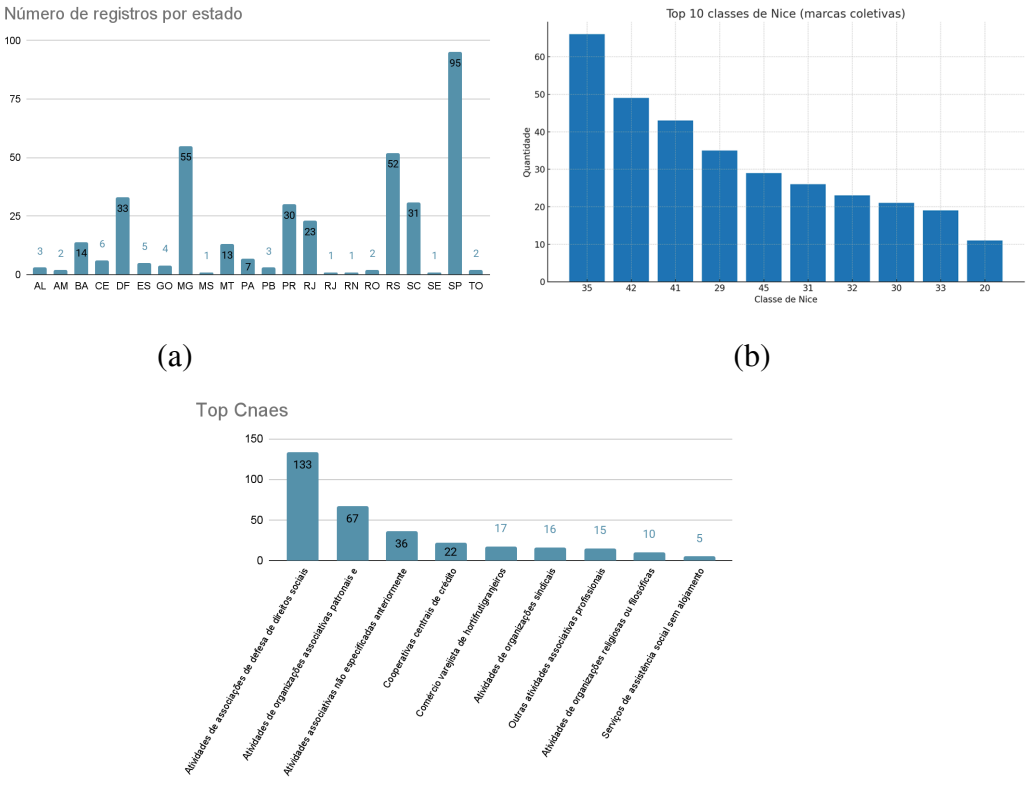

Figura 2 - Análises possíveis (a) Registros por estado (b) Classes de Nice (c) Distribuição por CNAE

4. Conclusão

A reestruturação e organização computacional rigorosa da base de dados representa uma das principais contribuições metodológicas deste trabalho, demonstrando a aplicabilidade de metodologias inovadoras de tratamento e inferência de dados em bases públicas de propriedade industrial. Esse processo foi fundamental para superar as barreiras de acesso e a complexidade da manipulação do grande volume de dados brutos da BADEPI (Base de Dados de Propriedade Industrial).

A base de dados final, que não foi normalizada para facilitar as análises, reuniu todos os atributos cruciais, desde a natureza jurídica do titular e o CNAE até a pontuação de inferência por proxies, em um formato pronto para uso. Isso transformou milhões de registros complexos e dispersos, que inicialmente exigiam a importação para um banco SQL e o uso de *web scraping* em Python com Selenium para serem complementados, em um conjunto de evidências quantitativas inéditas e transparentes.

Essa disponibilização pública dos dados, já tratados, enriquecidos e classificados, é essencial para que outros pesquisadores e o público em geral, incluindo não especialistas em grandes volumes de dados, possam utilizar a base consolidada. Isso permite que a comunidade acadêmica e a sociedade civil conduzam reanálises com diferentes parametrizações, democratizando o acesso às informações e avançando na agenda de pesquisa que articula direito e inovação social.

References

BRASIL. Instituto Nacional da Propriedade Industrial (INPI). Base de Dados de Propriedade Industrial – BADEPI. Rio de Janeiro: INPI, [2025]. Disponível em: https://inpidrive.inpi.gov.br/index.php/s/pTASUok7glOjSeO. Acesso em: 22 ago. 2025.

BRASIL. Instituto Nacional da Propriedade Industrial (INPI). Pesquisa em Propriedade Industrial – PePI. Rio de Janeiro: INPI, [2025]. Disponível em: https://busca.inpi.gov.br/pePI/. Acesso em: 23 ago. 2025.

BRASIL. Lei nº 9.279, de 14 de maio de 1996. Regula direitos e obrigações relativos à propriedade industrial. Diário Oficial da União: seção 1, Brasília, DF, 15 maio 1996.

BRASIL. Receita Federal do Brasil (RFB). Brasil API – Cadastro Nacional da Pessoa Jurídica (CNPJ). Brasília: RFB, [2025]. Disponível em: https://brasilapi.com.br. Acesso em: 23 ago. 2025.

CECATTO, Raul Tavares; RODRIGUES, Fabrício Geraldo dos Santos; MORAES, Thiago Maia Sayão de. Método para extração de dados das marcas coletivas. In: PROSPECT&I, 15.; CONGRESSO INTERNACIONAL PROFNIT, 9., 2025, Brasília, DF. Anais… [S.l.: s.n.], 2025.