

**REVISTA SEMESTRAL DE  
DIREITO EMPRESARIAL**

**Nº 26**

Publicação do Departamento de Direito Comercial e do Trabalho  
da Faculdade de Direito da Universidade do Estado do Rio de Janeiro

Rio de Janeiro  
**Janeiro / Junho de 2020**



Publicação do Departamento de Direito Comercial e do Trabalho da Faculdade de Direito da Universidade do Estado do Rio de Janeiro (Prof. Alexandre Ferreira de Assumpção Alves, Prof. Eduardo Henrique Raymundo Von Adamovich, Prof. Enzo Baiocchi, Prof. Ivan Garcia, Prof. João Batista Berthier Leite Soares, Prof. José Carlos Vaz e Dias, Prof. José Gabriel Assis de Almeida, Prof. Leonardo da Silva Sant'Anna, Prof. Marcelo Leonardo Tavares, Prof. Mauricio Moreira Menezes, Prof. Rodrigo Lychowski e Prof. Sérgio Campinho).

**EDITORES:** Sérgio Campinho e Mauricio Moreira Menezes.

**CONSELHO EDITORIAL:** Alexandre Ferreira de Assumpção Alves (UERJ), Ana Frazão (UNB), Antônio José Avelãs Nunes (Universidade de Coimbra), Carmen Tiburcio (UERJ), Fábio Ulhoa Coelho (PUC-SP), Jean E. Kalicki (Georgetown University Law School), John H. Rooney Jr. (University of Miami Law School), Jorge Manuel Coutinho de Abreu (Universidade de Coimbra), José de Oliveira Ascensão (Universidade Clássica de Lisboa), Luiz Edson Fachin (UFPR), Marie-Hélène Monsérié-Bon (Université des Sciences Sociales de Toulouse), Paulo Fernando Campos Salles de Toledo (USP), Peter-Christian Müller-Graff (Ruprecht-Karls-Universität Heidelberg) e Werner Ebke (Ruprecht-Karls-Universität Heidelberg).

**CONSELHO EXECUTIVO:** Carlos Martins Neto e Mariana Pinto (coordenadores). Guilherme Vinseiro Martins, Leonardo da Silva Sant'Anna, Livia Ximenes Damasceno, Mariana Campinho, Mariana Pereira, Mauro Teixeira de Faria, Nicholas Furlan Di Biase e Rodrigo Cavalcante Moreira.

**PARECERISTAS DESTA NÚMERO:** Bruno Valladão Guimarães Ferreira (PUC-Rio), Caroline da Rosa Pinheiro (UFJF), Gerson Branco (UFRGS), José Gabriel Assis de Almeida (UERJ), Fabrício de Souza Oliveira (UFJF), Fernanda Valle Versiani (UFMG), Marcelo Féres (UFMG), Marcelo Lauar Leite (UFERSA), Milena Donato Oliva (UERJ), Pedro Wehrs do Vale Fernandes (UERJ), Samuel Max Gabbay (IFRJ), Sergio Negri (UFJF) e Uinie Caminha (UNIFOR).

**PATROCINADORES:**



ISSN 1983-5264

CIP-Brasil. Catalogação-na-fonte  
Sindicato Nacional dos Editores de Livros, RJ.

---

Revista semestral de direito empresarial. — n° 26 (janeiro/junho 2020)  
. — Rio de Janeiro: Renovar, 2007-.

v.

UERJ  
Campinho Advogados  
Moreira Menezes, Martins Advogados

Semestral

1. Direito — Periódicos brasileiros e estrangeiros.

94-1416.

CDU — 236(104)

---

\* Publicado no segundo semestre de 2021.

# **VALIDADE E CONFIABILIDADE DAS INFORMAÇÕES TECNOLÓGICAS PARA PESQUISAS EMPÍRICAS: O CASO DOS REGISTROS DE PATENTE NO INSTITUTO NACIONAL DE PROPRIEDADE INDUSTRIAL<sup>1</sup>**

## **VALIDITY AND RELIABILITY OF TECHNOLOGICAL INFORMATION FOR THE EMPIRICAL RESEARCH: THE CASE OF PATENT APPLICATIONS IN THE NATIONAL INSTITUTE OF INDUSTRIAL PROPERTY**

*Elisa Mara Coimbra\**  
*Marcos Vinício Chein Feres\*\**

---

<sup>1</sup> Artigo recebido em 16.07.2021 e aceito em 28.09.2021.

\* Doutoranda em Empresa e Atividades Econômicas pela Universidade do Estado do Rio de Janeiro – UERJ. Mestre em Direito Constitucional e Teoria do Estado pela Pontifícia Católica do Rio de Janeiro – PUC-Rio. Graduada em Direito pela Universidade Federal de Juiz de Fora – UFJF. Advogada da Financiadora de Estudos e Projetos – Finep. Integrante do Grupo de Pesquisa “Argumentação, direito e inovação” da Universidade Federal de Juiz de Fora – UFJF. Integrante do Grupo de Pesquisa, Ensino e Extensão em Direito Administrativo Contemporâneo Estudos – GDAC. Colaboradora do Núcleo Jurídico OIC-IEA/USP. E-mail: elisacoimbra775@yahoo.com.br

\*\* Professor Associado da Universidade Federal de Juiz de Fora – UFJF. Professor do Corpo Permanente do Programa de Pós-Graduação Estrito Senso em Direito e Inovação da Faculdade de Direito da Universidade Federal de Juiz de Fora – UFJF. Doutor em Direito pela Universidade Federal de Minas Gerais – UFMG. Mestre em Direito pela Universidade Federal de Minas Gerais – UFMG. Graduado em Direito pela Universidade Federal de Juiz de Fora – UFJF. Bolsista de Produtividade do CNPq e Professor Colaborador do Programa em Pós-Graduação em Direito da Universidade do Estado do Rio de Janeiro – UERJ. Desempenhou a função de Diretor da Faculdade de Direito da UFJF (2006-2014). Desempenhou a função de Vice-Reitor da UFJF (2014-2016), assumindo o exercício da Reitoria de novembro de 2015 a abril de 2016. Compõe, como participante, a *Collaborative Research Network* (047) da *Law and Society Association*, sobre “economic and social rights”. Possui projetos na área de Propriedade Intelectual e Teoria do Direito Aplicada financiados pela FAPEMIG e pelo CNPq. Tem experiência na área de Direito, com ênfase em pesquisa empírica em direito aplicada ao Direito Econômico, atuando principalmente nos seguintes temas: Pesquisa Empírica em Direito, Argumentação, Direito e Inovações Tecnológicas, Direito Econômico, Direito de Propriedade Intelectual (Marcas, Patentes e Inovação e Transferência de Tecnologia) e Metodologia da Pesquisa e do Ensino Jurídico. E-mail: mvchein@gmail.com

*Resumo:* Este artigo visa estudar a qualidade dos dados de propriedade intelectual disponibilizados pelo Instituto Nacional de Propriedade Industrial (INPI), considerando os conceitos de validade e de confiabilidade definidas por Lee Epstein e Gary King. O problema inicialmente identificado foi o de como coletar dados existentes nos bancos de dados de institutos de propriedade intelectual, comparando duas fontes possíveis, ambas do próprio INPI – uma fonte oficial formalizada, dotada de confiabilidade por lei, e uma fonte oficial, porém não legalmente formalizada (banco de busca de dados eletrônico), mais amigável do ponto de vista tecnológico. A importância deste trabalho é validar uma metodologia a ser posteriormente inseridas em pesquisas empíricas que se utilizem de tais informações tecnológicas, especialmente para trabalhos acadêmicos ou mercadológicos, assim como para formulação de políticas públicas. Também foi utilizado o *software* Anaconda, uma distribuição Python de código livre, licença New BSD, disponível gratuitamente, para viabilizar a comparação entre as fontes anteriormente indicadas.

*Palavras-chave:* Pesquisa empírica em Direito. Técnica de Pesquisa. Coleta de dados. Propriedade Intelectual.

*Abstract:* This article aims to study the quality of intellectual property data provided by the National Institute of Industrial Property, taking into account the concepts of validity and reliability defined by Lee Epstein and Gary King. The problem initially identified was how to collect data concerning the database of intellectual property institutes, comparing two possible sources, both from the INPI – an official formalized source, endowed with reliability by law, and an official source but not legally formalized (electronic database), more tech-friendly. The importance of this work is to validate a methodology approach in order to be later inserted at empirical research which may use this kind of data, especially not only for academic and market-oriented works but also for public policy formulation. In addition, the Anaconda software, a free source Python distribution, New BSD License, available for free, had been utilized so as to enable a comparison between the sources previously indicated.

*Keywords:* Empirical legal research. Research Technique. Data Collection. Intellectual Property.

*Sumário:* Introdução. 1. Recorte metodológico. 2. Limites comparativos das fontes de dados relacionados à propriedade intelectual. 3. Discussão de Resultados. Conclusão.

## **Introdução.**

A investigação a respeito da propriedade intelectual tem se revelado tema de destaque pelo mundo a fora, especialmente pela sua relação com a inovação, esta considerada como um dos *drivers* do crescimento econômico.<sup>2</sup> No entanto, a relação entre propriedade intelectual e inovação não é uma relação estática e definitiva. Tanto é assim que existe uma cizânia literária entre, de um lado, os estudos econômicos que demonstram que a propriedade industrial fomenta a inovação de maneira determinante<sup>3</sup> e, por outro lado, estudos que demonstram que propriedade industrial não fomenta significativamente a inovação.<sup>4</sup> Além disso, segundo Hudson e Minea,<sup>5</sup> a propriedade industrial afeta a inovação de maneira não linear, de modo que um sistema de proteção de patente forte nem sempre colabora para o processo de inovação.

Neste cenário, é a análise empírica que permite o estudo dos dados primários,<sup>6</sup> especialmente neste tema, cujas numerosas variáveis impactam como a propriedade intelectual influencia a inovação,

---

2 DAHLSTRAND, A. L.; STEVENSON, L. Innovative entrepreneurship policy: linking innovation and entrepreneurship in a European context. *Annals of Innovation & Entrepreneurship*, v. 1, n. 1, 2010, p. 5602; OECD/Eurostat. Oslo Manual: Guidelines for Collecting, Reporting and Using Data on Innovation. 4. ed. Paris/Eurostat, Luxembourg: OECD Publishing, 2018.

3 PAPAGEORGIADIS, N.; SHARMA, A. Intellectual property rights and innovation: A panel analysis. *Economics Letters*, 141: 70-72, 2016.

4 QIAN, Y. Do national patent laws stimulate domestic innovation in a global patenting environment? A cross-country analysis of pharmaceutical patent protection, 1978–2002. *The Review of Economics and Statistics*, v. 89, n. 3, p. 436-453, 2007.

5 HUDSON, J.; MINEA, A. Innovation, intellectual property rights, and economic development: a unified empirical investigation. *World Development*, Amsterdam, n. 46, p. 66-78, 2013.

6 MACHADO, M. R. Pesquisar empiricamente o direito. São Paulo: *Rede de Estudos Empíricos em Direito*, p. 119-160, 2017.

observadas por meio de regras de inferência.<sup>7</sup> Isso porque a propriedade intelectual é um instrumento jurídico de apropriação de ativos intangíveis por meio do direito de propriedade conferido a todos aqueles que preencham determinados requisitos legais.<sup>8</sup> Portanto, estudos empíricos são mais que recomendáveis, são imprescindíveis não apenas para uma compreensão adequada do tema como para a elaboração de políticas públicas industriais, de inovação e/ou de desenvolvimento.<sup>9</sup>

Neste contexto, estuda-se, aqui, a qualidade dos dados de propriedade intelectual disponibilizados pelo Instituto Nacional de Propriedade Industrial, sob uma perspectiva metodológica. O problema inicialmente identificado foi o de como coletar dados confiáveis concernentes aos institutos de propriedade intelectual de maneira adequada e a mais facilitada possível, considerando que o veículo oficial de publicação de atos, decisões, despachos e matérias relacionadas aos serviços do INPI (Instituto Nacional de Propriedade Industrial), a Revista da Propriedade Industrial (RPI), não está totalmente disponível, desde a criação do órgão (1971), bem como considerando a dificuldade de consulta das informações, quando disponível.

Para tal, toma-se como estratégia metodológica deste trabalho comparar as informações obtidas de outra fonte que não a RPI, um banco de dados desenvolvido pelo próprio INPI (“consulta à base de dados do INPI”), com os dados das RPI disponíveis. Neste sentido, foram consideradas as limitações deste banco indicadas pelo próprio órgão em relação ao recorte temporal a partir do qual os dados passaram a ser inseridos. O objetivo é o de extrair o máximo de informações confiáveis deste banco devido às possibilidades tecnológicas disponíveis na plataforma que o compreende. Afinal, a consulta a um

---

7 EPSTEIN, L.; KING, G. *Pesquisa empírica em direito: as regras de inferência*. Vários Tradutores. São Paulo: Direito GV, 2013.

8 BARBOSA, D. B. 2010. *Uma Introdução à Propriedade Intelectual*. Ed. Lumen Juris: Rio de Janeiro, 2010.

9 FERREIRA, A. A.; GUIMARÃES, E. R.; CONTADOR, J. C. Patente como instrumento competitivo e como fonte de informação tecnológica. *Gestão & Produção*, v. 16, n. 2, p. 209-221, 2009; OCDE. *Manual de Oslo: Diretrizes para coleta e interpretação de dados sobre inovação*. 3. ed. Trad. Rio de Janeiro: Finep, 2005.



sistema de dados eletrônico ainda que incompleto facilita muito o trabalho do pesquisador que pode se dedicar mais especialmente ao seu problema original, poupando tempo e recursos.

Para tal, utiliza-se como marco teórico as considerações de Lee Epstein e Gary King<sup>10</sup> sobre a importância dos dados, dos processos de como os dados são observados, bem como do processo pelo qual as observações potenciais são geradas, a fim de subsidiar a subsequente investigação por inferência nos estudos empíricos relacionados à propriedade intelectual.

Ressalta-se ainda que a questão apresentada surgiu no contexto de uma pesquisa de doutorado que visa estudar como a proteção jurídica da propriedade intelectual configura-se como um mecanismo para alavancar o processo de inovação, no sentido de auxiliar a sociedade empresária, no caso a Embraer, na criação de uma vantagem competitiva no seu setor econômico específico.

## **1. Recorte metodológico.**

De acordo com Lee Epstein e Gary King,<sup>11</sup> a inferência, definida pelos autores como o aprendizado de fatos que não conhecemos pelo uso de fatos conhecidos, requer dados. Neste sentido, faz-se necessária a coleta de dados, bem como de críticas a esses dados, de modo a se tirar o máximo proveito em prol de mais qualidade às possíveis conclusões.

Ainda segundo Lee Epstein e Gary King,<sup>12</sup> a recomendação por coletar a maior quantidade de dados não é incompatível com a possibilidade de serem utilizados métodos mais facilitados de fazê-la, haja vista que a tarefa mais especializada de um pesquisador empírico está relacionada ao processo de fazer as inferências.

Neste contexto, a primeira dúvida que surgiu durante a pes-

---

10 EPSTEIN, L.; KING, G. *Pesquisa empírica em direito: as regras de inferência*. Vários Tradutores. São Paulo: Direito GV, 2013.

11 *Ibidem*, p. 11.

12 *Idem*.

quisa para a tese de doutorado que envolve dados de propriedade intelectual estava relacionada com qual método de coleta seria o mais adequado e facilitado para o pesquisador, pois existiriam duas fontes de dados possíveis, ambas do próprio INPI, uma fonte oficial formalizada (RPI) e uma fonte oficial, porém não legalmente formalizada (banco de busca de dados eletrônico). Embora a fonte oficial formalizada seja a dotada de confiabilidade por lei, a fonte não legalmente formalizada seria mais amigável do ponto de vista tecnológico por oferecer ferramentas para a construção do banco de dados personalizado para as possíveis pesquisas, inclusive a que se intenta realizar em nível de doutorado.

Desse modo, de um lado, está o repositório das RPI, veículo oficial de publicação do INPI, destinado a publicar seus atos, despachos, decisões e matérias relacionadas aos seus serviços, originalmente previsto no artigo 9º, p.u., da Lei nº 5.648/1970. Por outro lado, está a base de busca também do próprio INPI, que funciona a partir da utilização de palavras-chave e filtros que parametrizam os resultados da pesquisa, mas não é o veículo oficial de publicação, além de conter uma restrição por limitação temporal indicada pelo próprio órgão que varia a depender da categoria do serviço.

Mais especificamente, o problema foi investigar se a base de dados não formalizada poderia ser considerada confiável do ponto de vista de conter as informações utilizadas na pesquisa de doutorado e quais seriam os seus limites, a fim de se otimizar os recursos disponíveis sem comprometer os resultados finais e, eventualmente, validando ou invalidando outras pesquisas que se utilizam de uma fonte ou outra.

Portanto, antes de explicitar a pesquisa metodológica prévia à coleta de dados em si, faz-se necessário esclarecer, de forma sintética, o objeto da tese de doutorado, a fim de justificar as escolhas feitas na próxima seção.

O problema originalmente apresentado foi o de descrever como a proteção da propriedade intelectual relaciona-se com o processo de inovação, considerando, em primeiro lugar, que a propriedade intelectual é um mecanismo jurídico de apropriação do conhecimento, hábil a incrementar o avanço das fronteiras tecnológicas e, em segundo lugar, que a propriedade intelectual pode potencializar

a vantagem competitiva de uma sociedade empresária em relação aos concorrentes que temporariamente não poderão se apropriar da tecnologia em questão.

Em outras palavras, propôs-se a seguinte pergunta de pesquisa: como a proteção jurídica da propriedade intelectual, no caso Embraer, pode se configurar como um mecanismo para alavancar o processo de inovação, no sentido de auxiliá-la a potencializar uma vantagem competitiva no seu setor econômico específico?

Assim sendo, a hipótese preliminarmente indicada foi a de que os direitos de propriedade intelectual, seja qual modalidade for, marca, patente, desenho industrial, direitos decorrentes de transferência de tecnologia, entre outros, podem ser mecanismos de apropriação do conhecimento utilizados pela Embraer, no sentido de potencializar uma vantagem competitiva no setor econômico no qual está inserida.

A proposta foi a de analisar de forma estruturada em um caso concreto como a Embraer tem protegido o seu conhecimento, ao longo dos anos, e analisar a repercussão destas políticas na história da sociedade empresária para criar uma vantagem competitiva. O objetivo seria o de conferir como e quando a proteção da propriedade intelectual foi administrada, contextualizando no processo de inovação as suas tendências e os seus benefícios em relação às estratégias de crescimento da empresa.

Sabe-se pela bibliografia revisada e indicada acima que, até o momento, não existe um modelo teórico consolidado para explicar se a propriedade intelectual fomenta a inovação e como dar-se-ia esta relação. Por isso, optou-se por estudar um caso concreto em que fosse evidente a ocorrência do processo de inovação, a fim de identificar como a propriedade intelectual interagiria com a inovação neste contexto. Por conseguinte, a estratégia pela pesquisa empírica foi a melhor opção. No entanto, para tal faz-se necessário verificar adequadamente a confiabilidade dos dados utilizados para que as conclusões possam ser as mais acertadas possível.

De um lado, sobre a confiabilidade, Lee Epstein e Gary King<sup>13</sup> ensinam que “é a extensão à qual se pode replicar uma medida, reproduzindo o mesmo valor (indiferente de ser este o valor correto ou não) no mesmo padrão para o mesmo tópico a um mesmo tempo”. O exemplo trazido pelos autores é: se alguma pessoa subir consecutivamente em uma balança, ela registrará sempre o mesmo valor. Por outro lado, a validade está relacionada com a correção do registro.<sup>14</sup>

Portanto, simplesmente escolher aleatoriamente um modelo teórico ainda não consolidado e, a partir dele, dedutivamente propor hipóteses para, em seguida, extrair conclusões, não contribuiria para o avanço da questão, dada a provável superação deste modelo inicial, deixando igualmente para trás as conclusões dele extraídas.

Ocorre que a coleta de tais dados não se revelou uma iniciativa simples como inicialmente planejado, devido às dificuldades apresentadas na próxima seção.

## **2. Limites comparativos das fontes de dados relacionados à propriedade intelectual.**

A verificação de como a Embraer utiliza o sistema de proteção da propriedade intelectual fez com que fosse necessário compreender como funciona o registro das propriedades intelectuais e direitos correlatos, bem como os mecanismos de dar publicidade a estas informações, a fim de que os dados pudessem ser posteriormente coletados. Por isso, houve necessidade de procurar o INPI, já que é o escritório de patentes nacional, órgão responsável pelo processamento de pedidos de patentes e outros direitos de propriedade industrial

---

13 *Ibidem*, p. 105.

14 *Ibidem*, p. 111; para uma compreensão sobre a diferença entre validade e confiabilidade aplicadas, ver: FERES, M. V. C.; SILVA, A. R.; MORAIS, A. R.; SOUZA, A. M. de. A medida da inovação farmacêutica e os pedidos de patente: o caso da doença de chagas. *Revista de Estudos Empíricos em Direito*, São Paulo, v. 5, n. 3, p. 118–135, 2018.

no Brasil (Lei nº 5.648/1970), considerando que a proteção da propriedade intelectual é territorial, ou seja, uma patente protegida no Brasil tem proteção em todo o âmbito nacional apenas (Lei nº 9.279/1996).

Portanto, a despeito de considerar por lei a RPI como fonte oficial das publicações do INPI, identificou-se a ferramenta de “consulta à base de dados do INPI” (<https://gru.inpi.gov.br/pePI/servlet/LoginController?action=login>). Ao clicar na consulta à base de dados identifica-se a seguinte tela:

Figura 1 - Tela inicial da consulta sugerida pelo site do INPI à base de dados de propriedade Industrial



Fonte: Site do INPI. Disponível em: <http://revistas.inpi.gov.br/rpi/>.

Nesta modalidade de consulta, a base de dados é classificada em categorias específicas: marca, patente, desenho industrial, indicação geográfica, programa de computador, topografia de circuito integrado, transferência de tecnologia, informação tecnológica de paten-

tes. Assim, para usufruir da consulta, faz-se necessária a escolha prévia da categoria de propriedade intelectual que se pretende pesquisar.

Em seguida, são abertas duas possibilidades de pesquisa: a pesquisa avançada e a pesquisa básica. Para utilizar os filtros da pesquisa avançada, é necessário conhecer previamente alguns parâmetros, os quais, na base de dados de patentes, são classificados como: números, datas, classificação, palavras-chave, depositante/titular/inventor. Abaixo:

Figura 2 - Tela da pesquisa avançada após escolher a pesquisa à base de dados de patentes.

**PESQUISA AVANÇADA**  
*Forneça abaixo as chaves de pesquisa desejadas. Evite o uso de frases ou palavras genéricas.*

<b>Números</b>	
(21) Nº do Pedido:	<input type="text"/> ?
<input type="radio"/> Calendário de Patentes expiradas/a expirar ? <input type="radio"/> Patente Concedida ?	
(33)/(31) País/Nº da Prioridade:	<input type="text"/> ?
(86) Nº do Depósito (PCT):	<input type="text"/> ?
<b>Datas</b>	
(22) Data Depósito:	<input type="text"/> a <input type="text"/> ?
(32) Data da Prioridade:	<input type="text"/> a <input type="text"/> ?
(86) Data do Depósito (PCT):	<input type="text"/> a <input type="text"/> ?
(87) Data da Publicação (PCT):	<input type="text"/> a <input type="text"/> ?
<b>Classificação</b>	
(51) Classificação IPC:	<input type="text"/> ?
Palavra-chave no classificador IPC:	<input type="text"/> ?
<b>Palavra Chave</b>	
(54) Título:	<input type="text"/> ?
(57) Resumo:	<input type="text"/> ?
<b>Depositante/Titular/Inventor</b>	
(71/73) Nome do Depositante/Titular:	<input type="text"/> ?
CPF/CNPJ do Depositante:	<input type="text"/> ?
(72) Nome Inventor:	<input type="text"/> ?

Fonte: Site do INPI

Interessante registrar que, a depender da categoria de propriedade intelectual escolhida, os filtros da pesquisa avançada variam de acordo com as suas peculiaridades. De qualquer modo, para continuar a pesquisa, algum tipo de dado prévio a respeito da propriedade

intelectual escolhida faz-se necessário. Por isso, tal tipo de busca não é adequado para o tipo de análise que se pretende neste trabalho, pois se desconhece, a princípio, como a Embraer utilizou o sistema de proteção da propriedade intelectual ao longo do tempo, ou seja, previamente se desconhece quaisquer características dos direitos de propriedade intelectual e direitos correlatos requeridos e concedidos ao longo do tempo (marco zero da pesquisa).

Ocorre que, ao se retornar ao ponto de partida, dando a devida atenção aos limites da base de dados utilizada, identificou-se a restrição contida na Figura 1 (parte inferior) no sentido de que o acervo da base de dados está restrito aos documentos publicados a partir de 2000, exceto para os contratos de tecnologia, cujo acervo da base de dados está restrito aos documentos publicados a partir de 2009. Assim, pode-se supor que tal modalidade de pesquisa tem uma alta probabilidade de resultar em “falso negativo”, de modo que não é possível, por ora, atestar a confiabilidade do resultado observado na Figura 4.

Diante dessa restrição temporal na busca por produtos anteriores ao ano 2000, recorre-se à Revista de Propriedade Industrial (doravante RPI), publicada pelo INPI, considerada o canal oficial de publicações relativas às atividades do INPI, com o intuito de se realizarem as buscas por dados nela.

Ocorre que, embora atualmente as últimas edições estejam em formato eletrônico, nem sempre foi assim. Na verdade, a maior parte delas foi publicada fisicamente. Por isso, foi necessária uma visita à biblioteca do INPI para se constatar que a primeira RPI foi publicada fisicamente em 1972, mais precisamente em 04/04/1972 e que as RPI físicas estavam sendo digitalizadas bem como recebendo uma camada de OCR (*Optical Character Recognition*), ferramenta que permitiria a localização de palavras-chave. Verificou-se também que as edições de RPI mais recentes, a partir de 2005, foram publicadas já em formato eletrônico, o que não exigiria o processo de digitalização, pois já estavam disponíveis.<sup>15</sup>

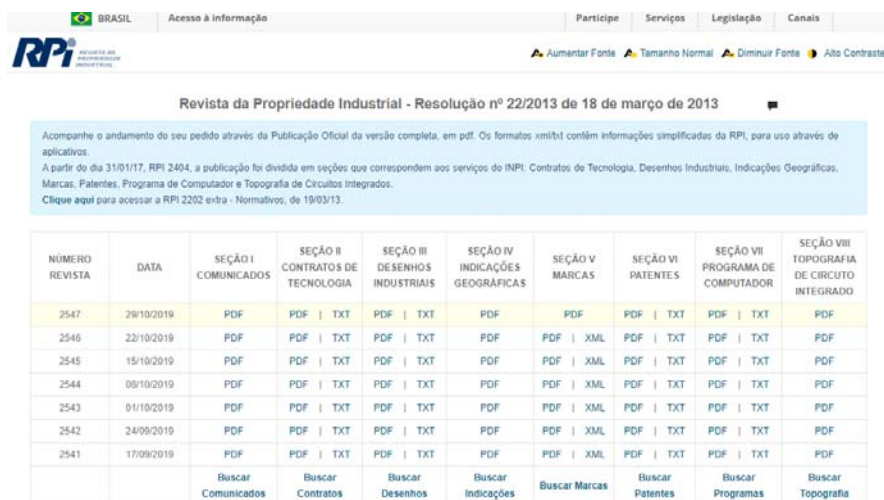
---

15 Vide: DE ARAUJO, G. G. *et al.* Bases de Dados de Patentes: uma análise a partir do portal de periódicos da CAPES. *Cadernos de Prospecção*, Salvador, v. 12, n. 5 Especial, p. 1500, 2019.

Neste momento, vislumbra-se a possibilidade de organização de um banco de dados personalizado para a pesquisa e, por isso, mais confiável, haja vista que as RPI correspondem aos veículos oficiais de publicações do INPI concernente à questão de propriedade intelectual.

Verificou-se que as RPI em formato digital (sejam as digitalizadas com camada de OCR sejam as eletrônicas desde a origem) estavam disponíveis no próprio site do INPI (<http://revistas.inpi.gov.br/rpi/>).

Figura 3 - Tela de consulta das RPI disponíveis para consulta de acordo com os serviços do INPI.



NÚMERO REVISTA	DATA	SEÇÃO I COMUNICADOS	SEÇÃO II CONTRATOS DE TECNOLOGIA	SEÇÃO III DESENHOS INDUSTRIAIS	SEÇÃO IV INDICAÇÕES GEOGRÁFICAS	SEÇÃO V MARCAS	SEÇÃO VI PATENTES	SEÇÃO VII PROGRAMA DE CÍRCULO INTEGRADO	SEÇÃO VIII TOPOGRAFIA DE CÍRCULO INTEGRADO
2547	29/10/2019	PDF	PDF   TXT	PDF   TXT	PDF	PDF	PDF   TXT	PDF   TXT	PDF
2546	22/10/2019	PDF	PDF   TXT	PDF   TXT	PDF	PDF   XML	PDF   TXT	PDF   TXT	PDF
2545	15/10/2019	PDF	PDF   TXT	PDF   TXT	PDF	PDF   XML	PDF   TXT	PDF   TXT	PDF
2544	09/10/2019	PDF	PDF   TXT	PDF   TXT	PDF	PDF   XML	PDF   TXT	PDF   TXT	PDF
2543	01/10/2019	PDF	PDF   TXT	PDF   TXT	PDF	PDF   XML	PDF   TXT	PDF   TXT	PDF
2542	24/09/2019	PDF	PDF   TXT	PDF   TXT	PDF	PDF   XML	PDF   TXT	PDF   TXT	PDF
2541	17/09/2019	PDF	PDF   TXT	PDF   TXT	PDF	PDF   XML	PDF   TXT	PDF   TXT	PDF
		Buscar Comunicados	Buscar Contratos	Buscar Desenhos	Buscar Indicações	Buscar Marcas	Buscar Patentes	Buscar Programas	Buscar Topografia

Fonte: Site do INPI. Disponível em: <http://revistas.inpi.gov.br/rpi/>.

Nele, discutem-se as diferentes bases de dados de patentes no mundo, incluindo as RPI e a consulta à base de dados do INPI. Interessante registrar que, a despeito da existência de diversas plataformas, os dados primários que alimentam estas plataformas são provenientes dos escritórios nacionais de patentes, informados por compromissos assumidos em tratados ou convênios.



Manualmente, iniciou-se um trabalho de *download* dos documentos para descoberta do grau de completude das RPI digitais em relação ao universo total de publicações. A identificação quanto ao grau de incompletude foi relativamente simples, uma vez que, desde 1972, as RPI gozam de numeração ininterrupta, de modo que a RPI1 foi publicada em 04/04/1972 e a última RPI, até a conclusão da coleta (30/10/2019), a RPI2546, foi publicada em 22/10/2019. Tal trabalho iniciou-se pela Seção VI (Patentes) porque se imaginou ser uma das mais relevantes para o trabalho, considerando o maior volume de literatura que trata o instituto em comparação com os demais.<sup>16</sup>

Neste processo, verificou-se que: (a) as RPI de 1972 e 1973 estavam digitalizadas na íntegra sem qualquer edição faltante; (b) as RPI de 1975 a 1981 (inclusive) estavam parcialmente digitalizadas, identificando-se diversas edições faltantes; e (c) a partir de 1982 não se identificou incompletude no processo de análise manual, embora algumas falhas de digitalização ou de OCR.

Constatou-se que, se se quisesse um banco de dados o mais completo possível, contendo todas as RPI disponíveis, incluindo os diversos tipos de proteções do sistema de propriedade intelectual, nomeadas pelo INPI como “serviços”, os quais correspondem a cada uma das seções, ter-se-ia que automatizar de alguma maneira o processo de *download* dos arquivos das RPI.

Para isso, iniciou-se um processo de catalogação dos arquivos de RPI disponíveis, de acordo com cada uma das seções, uma vez que, pelo *site*, não se identificou a possibilidade de, por meio de um

---

16 COULTER, M. B. *Property in ideas: The Patent question in Mid-Victorian Britain (Inventions)*. Kirksville: Truman State Univ., 1991; MACHLUP, F. *An economic review of the patent system*. Study nº 15. Washington: Committee on the Judiciary, 1958; MACHLUP, F.; PENROSE, E. The patent controversy in the nineteenth century. *The Journal of Economic History*, v. 10, n. 1, p. 1-29, 1950; MACLEOD, C. *Inventing the industrial revolution: The English patent system, 1660-1800*. Cambridge University Press, 2002; MALAVOTA, L. M. A construção do sistema de patentes no Brasil: um olhar histórico. Rio de Janeiro: Editora Lumen Juris, 2011; e MOSER, P. Patents and innovation: evidence from economic history. *Journal of Economic Perspectives*, Pittsburgh, v. 27, n. 1, p. 23-44, 2013.

filtro, consultar todas as seções de uma edição qualquer, aleatoriamente escolhida. Por isso, iniciou-se, a classificação escolhendo uma seção específica, avançando para a seguinte até a análise de todas. A partir dos dados retornados, foi elaborada uma planilha no Programa Microsoft Office Excel, licença Office 365.

Figura 4 - Consulta para identificar quais as RPI disponíveis digitalmente de 01/01/1972 a 26/10/2019, concernentes à seção de Patentes.

Filtro RPI

RPI de Patentes

Selecione o tipo Pesquisa:
☐ por número
☒ por data

Data Inicial: 01/01/1972
Data Final: 26/10/2019

Buscar

Baixar:	Número:	Nome:	Data:	Tamanho:
PDF   TXT	2546	Patentes2546.pdf	22/10/2019	4.38 MB
PDF   TXT	2545	Patentes2545.pdf	15/10/2019	4.22 MB
PDF   TXT	2544	Patentes2544.pdf	08/10/2019	4.61 MB
PDF   TXT	2543	Patentes2543.pdf	01/10/2019	6.19 MB
PDF   TXT	2542	Patentes2542.pdf	24/09/2019	5.25 MB
PDF   TXT	2541	Patentes2541.pdf	17/09/2019	4.93 MB
PDF   TXT	2540	Patentes2540.pdf	10/09/2019	5.33 MB
PDF   TXT	2539	Patentes2539.pdf	03/09/2019	2.66 MB
PDF   TXT	2538	Patentes2538.pdf	27/08/2019	6.27 MB
PDF   TXT	2537	Patentes2537.pdf	20/08/2019	4.21 MB

Fonte: Site do INPI. Disponível em: <http://revistas.inpi.gov.br/rpi/>.

Este processo foi realizado para cada uma das seções, de modo que foram identificados um total de 4702 arquivos de RPI com edições ordinárias e 1 arquivo denominado pelo próprio INPI como RPI 2202-extra com os normativos vigentes à época.

No entanto, percebeu-se que alguns arquivos, correspondentes a algumas edições, estavam em duplicidade, pois foram encontrados tanto por meio da busca dentro da seção de patentes quanto da seção de marcas. Isso pode ser explicado porque, em dado período, as seções não eram subdivididas em mais de uma publicação, quan-

do no formato físico, ou em mais de um arquivo, quando no formato eletrônico.

Dessa forma, foram excluídos os arquivos repetidos da planilha, de modo que dos 4702 arquivos, foram excluídos 630 em duplicidade, resultando em 4072 arquivos disponíveis, sendo: 1970 encontrados na seção de patentes, 910 encontrados na seção de marcas, 334 encontrados em duplicidade tanto na seção de patentes quanto de marcas, 143 encontrados na seção de contratos de transferência de tecnologia, 143 encontrados na seção de desenho industrial, 143 encontrados na seção de indicações geográficas, 143 encontrados na seção de programas de computador, 143 na seção de topografia de circuitos integrados, 143 na seção de comunicados.

A respeito dos dados acima, o que chama a atenção é que, desde o início da RPI, coexistiram as seções de patentes e a seção de marcas, de modo que, por lógica, deveria estar disponível a mesma quantidade de arquivos em cada uma destas seções.

No entanto, tal informação não se verifica. Quando se buscam os arquivos da RPI na seção de patentes, são encontrados 2304 arquivos na seção de patentes e 1244 arquivos na seção de marcas. Isso conduz à conclusão de que, por equívoco, alguns arquivos foram indexados em uma seção sem serem indexados na outra seção. Por exemplo, a RPI 551 aparece apenas na seção de Patentes, no entanto, quando se investiga o conteúdo da RPI, identifica-se uma seção destinada a tratar apenas do conteúdo marcário.

Apenas a partir da RPI 1592, de 10/09/2011, é que se constata uma separação em dois arquivos, um deles com conteúdo exclusivo relacionado a patentes e outro a marcas. Para evitar falhas, todos os arquivos disponíveis foram catalogados na planilha elaborada, ainda que em duplicidade. A duplicidade foi desconsiderada apenas quando os arquivos continham o mesmo nome, independentemente da seção indexada (marcas ou patentes).

Outro dado interessante é refletir a respeito das demais seções (contratos de transferência de tecnologia, desenho industrial, indicações geográficas, programas de computador, topografia de circuitos integrados e comunicados), as quais foram criadas concomitantemente, de modo que individualmente têm a mesma quantidade de arquivos.

Em seguida, considerando o tempo que seria gasto no *download* dos mais de 4000 arquivos, buscou-se uma alternativa mais automatizada para a questão. Assim, lançou-se mão do Anaconda, uma distribuição Python de código livre, New BSD *License*, disponível gratuitamente na internet. Foi realizada a seguinte série de comandos:

Quadro 1 – Linha de comando utilizada no Python.

```
from pathlib import Path
import requests

from openpyxl import load_workbook
arquivo_excel = load_workbook('Revistas.xlsx')
Plan1 = arquivo_excel.active

for i in range(1,4704):
    a=Plan1.cell(column=4, row=i)
    filename = Path('%s'%(a.value))
    url = 'http://revistas.inpi.gov.br/pdf/%s'%(a.value)
    response = requests.get(url)
    filename.write_bytes(response.content)
```

Fonte: Elaboração Própria.

Em síntese, a linha de comando significa que, uma vez dada a ordem, o computador baixaria do *site* do INPI todas as RPI encontradas, cujos nomes dos arquivos constaram da tabela Excel construída

anteriormente. Lembrando que foram incluídos todos os arquivos disponíveis, de todas as seções disponíveis até o dia 29/10/2019. O computador ficou 4 dias e 4 noites realizando o *download* automático dos arquivos, totalizando, ao final, um banco de dados de 126 GB, aproximadamente, composto por 4073 arquivos (4072 RPI ordinárias e 1 RPI extra). Ressalta-se que algumas revistas disponíveis apenas em formato .txt não foram baixadas automaticamente. No entanto, assim o foram manualmente, de modo a completar o *download* de todo o acervo disponível.

Finalizado este processo, foi analisada a qualidade dos arquivos baixados e verificaram-se algumas falhas na camada de OCR de algumas edições físicas que foram digitalizadas. Por exemplo, a RPI 340 não tem a camada de OCR em todas as folhas, de modo que se fizermos uma busca por palavra no documento, o resultado seria um “falso negativo”.

Além disso, foram observados alguns erros no *upload* das RPI no *site* do INPI. Por exemplo, às RPI203, RPI204, RPI205, foi atribuído o ano de 1977 como o ano da publicação. Ocorre que tal informação é inconsistente, considerando que a RPI202 é de 05/09/1974 e a RPI206 é de 03/10/1974. Assim, infere-se que as três revistas são de 1974, supondo-se que houve apenas um erro de digitação.

Acrescenta-se ainda que quando se busca especificamente pelas RPI não disponíveis no *site* do INPI, a consulta retorna com a informação de que não foram encontrados documentos.

Figura 5 – Resultado da busca pela RPI223 não disponível.

4 Códigos e Abreviações

Filtro RPI

Fechar

RPI de Patentes

Selecione o tipo Pesquisa: ☒ por número. ☐ por data.

Numero Revista: 223
 

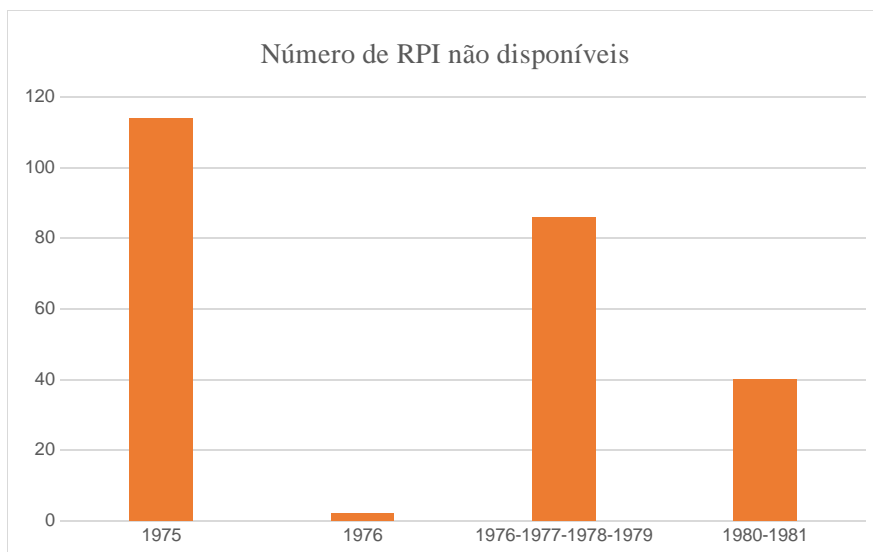
Buscar

Baixar:	Número:	Nome:	Data:	Tamanho:
Nenhum Documento encontrado com os dados informados!				

Fonte: Site do INPI. Disponível em: <http://revistas.inpi.gov.br/rpi/>.

Ocorre que, apesar das inconsistências encontradas, o banco de dados é bastante representativo, dada a robustez de seus dados.

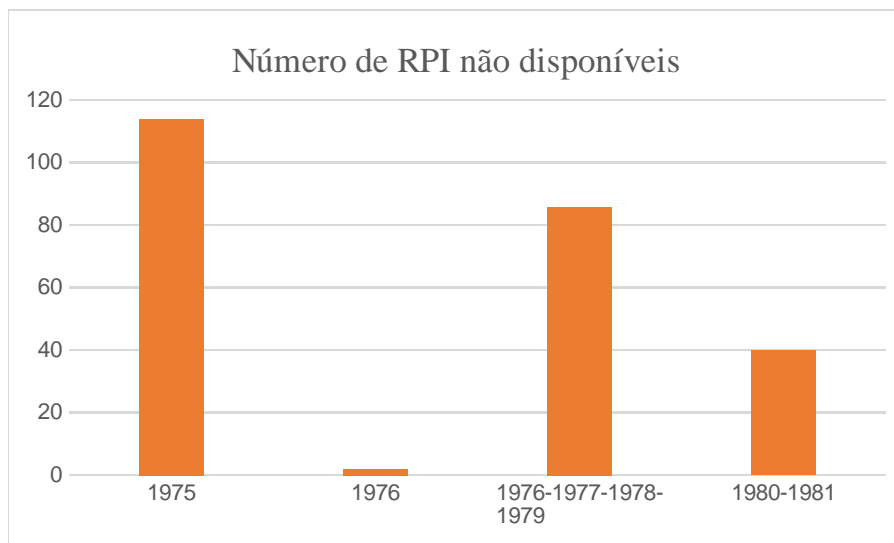
Figura 6 - Número de RPI disponíveis no site do INPI.



Fonte: Elaboração Própria.

O gráfico acima indica todas as RPI disponíveis em formato digital (digitalizadas ou eletrônicas) encontradas no *site* no INPI, considerando que, a partir de 1982, as RPI disponíveis correspondem às RPI existentes. Por sua vez, as RPI não disponíveis correspondem ao gráfico abaixo:

Figura 7 – Número de RPI não disponíveis.



Fonte: Elaboração Própria.

Pelo gráfico, observa-se que apenas no intervalo, compreendido entre 1975-1981, é que se registram publicações não disponíveis. Portanto, apenas neste período é que se incorrerá na possibilidade de “falsos negativos”, o que pode ser superado por uma complementação de outros dados disponíveis, ainda que de fontes secundárias.

### 3. Discussão de Resultados.

A partir da identificação dos limites de cada fonte de dados (RPI ou consulta à base de dados do INPI), iniciou-se a elaboração de uma planilha com a identificação de todas as referências à palavra Embraer, pontuando o instituto de propriedade intelectual envolvido, suas características, bem como o despacho relacionado, a partir dos arquivos baixados. Para solucionar a falha de OCR destes arquivos,

utilizou-se o software livre PDF-X *Change Viewer*, de propriedade da Microsoft Windows, o qual dispõe desta ferramenta. Por meio deste programa de computador, foi possível corrigir as falhas de OCR, evitando falsos negativos quando requeridas as localizações das referências à Embraer. No entanto, identificaram-se alguns arquivos digitalizados como imagem, inviabilizando a localização da palavra-chave e acrescentando novas possibilidades de falsos negativos.

A elaboração manual da planilha permitiu identificar que, diferente do que se pensava, o sistema de consulta do INPI, inicialmente rechaçado devido à limitação temporal indicada na Figura 1, poderia ser sim utilizado para facilitar a busca por dados. Apesar da limitação temporal, à medida que fossem ocorrendo novas publicações, o sistema era alimentado com informações básicas do processo original, uma vez que este número permanecia o mesmo ao longo do tempo.

Por exemplo, verificou-se que a primeira referência à Embraer encontrada nas RPI baixadas, observando a ordem cronológica das publicações, relacionada ao processo nº 000.587, de 16/02/1973, na seção “Pedidos de averbação de contratos aprovados” da RPI 78, de responsabilidade da Secretaria de Informação e Transferência de Tecnologia (à época), cuja Concedente é a Société pour le Perfectionnement des Matériels et Equipements Aérospatiaux (SOPEMA) e a Cessionária é a EMBRAER, também pode ser encontrada no sistema de consulta do INPI, utilizando os parâmetros indicados abaixo:



Figura 8 - Tela do sistema INPI correspondente ao primeiro registro da palavra Embraer encontrada nas RPI.

The screenshot displays the INPI system interface for a technology transfer registration. At the top, there is a header with the Brazilian flag, the text 'BRASIL', and 'Acesso à informação'. Below this is a blue banner with 'Instituto Nacional da Propriedade Industrial' and 'Ministério da Economia'. The main title of the page is 'Consulta à Base de Dados do INPI'. On the right, there are links for '[ Início | Ajuda? ]' and 'Anterior 607/607'. The left sidebar contains navigation links: 'Consultar por: Pesquisa Base Contratos | Finalizar Sessão'. The main content area is titled 'Transferência de Tecnologia' and lists the following details:
 

- Nº do Requerimento: 730587
- Nº do Protocolo: 1058437
- Entrada: 16/02/1973
- Categoria Contratual: SERVIÇO DE ASSISTÊNCIA TÉCNICA E CIENTÍFICA
- Nome da Cedente: SOPEMEA - SOC. POUR LE PERFECT. DES MATER. ET EQUIP. AEROSP.
- País da Cedente: FRANÇA
- Nome da Cessionária: EMBRAER - EMPRESA BRASILEIRA DE AERONÁUTICA S/A
- País da Cessionária: BRASIL(SP)
- Setor da Cessionária: CONSTRUÇÃO E MONTAGEM DE AERONAVES (35.31-9)

 Below the details, there are two tables. The first table, 'Petições', has columns: Pgo, Protocolo, Data, Img, Serviço, Cliente, Delivery, and Data. The second table, 'Publicações', has columns: RPI, Data RPI, Doc, Carta, Despacho, and Complemento do Despacho. The first row of the 'Publicações' table shows RPI 2503, Data RPI 26/12/2018, Doc -, Carta -, and Despacho 184. At the bottom, it states 'Dados atualizados até 31/03/2020 - Nº da Revista: 2569'.

Fonte: Site do INPI Disponível em: <http://revistas.inpi.gov.br/rpi/>.

Isso porque, embora o despacho original não conste no sistema, dada a limitação temporal descrita na Figura 1, os despachos subsequentes a 2009, no caso o constante da RPI 2503 indicado pela elipse, garante a inserção dos dados básicos originais no sistema, tal como o ocorrido. De semelhante modo, foram cotejadas outras referências encontradas nas RPI com o sistema de consulta do INPI e não se visualizaram inconsistências. No entanto, não havia descrições mais detalhadas do conteúdo técnico do ato averbado, o qual poderia ser estudado mais perfunctoriamente se analisada a RPI.

Além disso, observou-se que, especialmente na seção de transferência de tecnologia, cuja restrição temporal é mais severa, provavelmente por alguma orientação interna do próprio INPI, houve o arquivamento dos processos administrativos de averbação dos contratos de transferência de tecnologia que estavam vencidos.<sup>17</sup> Observa-se que a obrigação de averbar contratos de transferência de tecnologia advém tanto da Lei nº 5.772/1971<sup>18</sup> quanto da Lei nº

9.279/1996<sup>19</sup> que estabeleceram a obrigação de o INPI averbar a transferência de tecnologia, hoje principalmente para produzir efeitos em relação a terceiros.

Por conseguinte, com a publicação do despacho de arquivamento, o banco de dados do INPI foi alimentado com dados básicos originais. Tal fato ficou explícito à medida em que se coletavam as informações, pois na maioria dos casos em que houve arquivamento, eles ocorreram no mesmo período (anos 2018-2019) e o histórico dos andamentos processuais estavam incompletos. Nesse sentido, tanto um contrato de 1972 quanto um da década de 2000, foram arquivados no mesmo momento. Portanto, embora em uma primeira análise o método de consulta às RPI pudesse ser considerado como a fonte mais óbvia por ser o veículo oficial de divulgação dos atos do INPI, a ferramenta de consulta à base de dados do INPI revelou-se mais adequada para a pesquisa por diversos fatores, seja pela sua confiabilidade, completude seja pela facilidade de manuseio das informações.

Assim, a consulta à base de dados do INPI corresponde a um avanço não só tecnológico, mas institucional na medida em que facilita o acesso ao estado da técnica da propriedade intelectual, a despeito das limitações relativas aos depósitos/registros mais antigos. Desse modo, se a pesquisa de doutorado fosse relacionada ao conteúdo técnico em si, ou seja, relacionada à descrição da tecnologia, às reivindicações, desenhos técnicos etc., poder-se-ia vislumbrar uma

---

17 Os dados registrados no INPI a respeito das averbações dos contratos de transferência de tecnologia foram utilizados, por exemplo, na pesquisa de Hemais, Barros e Rosa (E.O. Contratos de transferência tecnológica: Um estudo sobre aquisição de tecnologia em Polímeros no Brasil. *Polímeros*, São Paulo, v. 14, n. 4, p. 242-250, 2004).

18 Art. 126. Ficam sujeitos à averbação no Instituto Nacional da Propriedade Industrial, para os efeitos do artigo 2º, parágrafo único, da Lei n. 5.648, de 11 de dezembro de 1970, os atos ou contratos que impliquem em transferência de tecnologia.

19 Art. 211. O INPI fará o registro dos contratos que impliquem transferência de tecnologia, contratos de franquia e similares para produzirem efeitos em relação a terceiros. Parágrafo único. A decisão relativa aos pedidos de registro de contratos de que trata este artigo será proferida no prazo de 30 (trinta) dias, contados da data do pedido de registro.

dificuldade adicional em atender a questão, visto que tal informação não foi migrada para a consulta à base de dados do INPI no período não compreendido antes da restrição indicada para cada seção.

## **Conclusão.**

Ao longo da pesquisa, verificou-se que, conquanto existam duas fontes para a coleta de dados sobre propriedade intelectual, ambas possuem limitações. Enquanto a RPI foi criada para dar publicidade aos atos oficiais do escritório nacional de patentes, a consulta à base de dados do INPI foi criada com preocupações voltadas à difusão da tecnologia, tanto que é relativamente simples a consulta à descrição técnica, reivindicações e desenho técnico disponíveis no momento da publicação da invenção ou correlato, observada a limitação temporal indicada em cada seção. Portanto, este trabalho pretendeu discutir qual das fontes seria a mais adequada para a busca de dados que subsidiem as pesquisas empíricas sobre propriedade intelectual, especialmente na área do direito e ciências sociais, menos voltadas ao aprofundamento do conhecimento tecnológico ou estado da técnica em si.

Assim, por meio dessa técnica comparativa entre as duas fontes de dados, foi possível identificar que a consulta à base de dados do INPI revelou-se como uma fonte adequada para a coleta de dados básicos da propriedade intelectual em suas diversas modalidades, já que estas não estão restritas às limitações temporais consignadas pelo próprio INPI. Portanto, de um lado, é possível considerar válidos e confiáveis os dados extraídos da plataforma. No entanto, por outro lado, caso a pesquisa esteja relacionada ao conteúdo tecnológico (descrição da tecnologia, às reivindicações, desenhos técnicos), a consulta à base de dados do INPI pode não atender satisfatoriamente, pois se restringe à limitação temporal indicada em cada uma das seções, sendo, portanto, recomendável a complementação com a consulta às RPI.

