

Anexo Metodológico – Capítulo 5

Atividade de patenteamento no Brasil e no exterior

1. Os dados fornecidos pelo Ipea e pelo INPI

A base de informações utilizada aqui é o resultado da compilação de dois bancos de dados distintos. A primeira parte foi fornecida pelo Instituto de Pesquisas Econômicas Aplicadas (Ipea), que repassou arquivo em formato “txt” fornecido pelo Instituto Nacional da Propriedade Industrial (INPI), com dados de patentes depositadas até 2003. Esse arquivo contém informações sobre os depósitos de patentes referentes à data de depósito, nome dos titulares, nome dos inventores, classificação internacional (Wipo),¹ CNPJ ou CPF do titular, CEP dos titulares, país dos titulares e número do pedido. A partir desse arquivo procedeu-se à construção de parte da base de dados utilizada neste capítulo.

Alguns problemas em relação às informações obtidas foram verificados desde o início. Em primeiro lugar, algumas patentes apresentavam o ano de depósito em 1900, o que sugere erros no preenchimento dos formulários do pedido ou erros no momento da digitalização. Os depósitos mais recentes referiam-se ao ano de 2000, configurando assim o período das informações disponíveis naquele arquivo, ao qual corresponde um total de 299749 depósitos de pedidos patentes.

O número do pedido da patente permitiu, após tratamento desta informação, a obtenção do tipo da patente requerida. Dessa forma, as patentes puderam ser identificadas como Patentes de Invenção (PI); Modelos de Utilidade (MU); Desenho Industrial (DI); ou Certificado de Adição (C). Utilizaram-se para as investigações do capítulo as patentes do tipo PI e MU, que somaram 252515 (84,2%) depósitos no período. O *site* do INPI explica as diferenças entre esses tipos: “Em razão das diferenças existentes entre as invenções, elas poderão se enquadrar nas seguintes naturezas ou modalidades: Privilégio de Invenção (PI) – a invenção deve atender aos requisitos de atividade inventiva, novidade e aplicação industrial. Modelo de Utilidade (MU) – nova forma ou disposição envolvendo ato inventivo que re-

sulte em melhoria funcional do objeto. Existe também o Certificado de Adição de Invenção, para proteger um aperfeiçoamento que se tenha elaborado em matéria para a qual já se tenha um pedido ou mesmo a Patente de Invenção. Se a criação industrializável for relacionada com a forma plástica ornamental de um objeto ou conjunto de linhas e cores, que possa ser aplicada a um produto, proporcionando resultado visual novo e original na sua configuração externa e que possa servir de tipo de fabricação, pode-se requerer um Registro de Desenho Industrial, pois nesse caso ela não poderá ser considerada uma patente.

A partir dessas definições, a escolha de trabalhar com PIs e MUs reflete a necessidade de lidar apenas com os registros que resultam de atividade inovativa mais significativa. Dessa forma, excluem-se os Desenhos Industriais, relacionados exclusivamente à forma e não considerados como patentes pelo INPI, e os Certificados de Adição, cuja utilização levaria este capítulo a tomar qualquer alteração em uma patente já existente como atividade inovativa significativa, independente do conteúdo desta alteração.

A informação referente ao país dos depositantes permitiu classificar os pedidos de patentes entre pertencentes a titulares residentes e não residentes. Assim, dos 252515 depósitos dos tipos PI e MU no INPI, 91372 (36,2%) tinham como primeiro titular uma empresa ou indivíduo residente no Brasil, enquanto 161143 (63,8%) apresentavam o primeiro titular não residente.

Entre os residentes, 77875 (85,2%) pedidos de patentes apresentavam alguma informação no campo do CNPJ ou CPF do titular. Destes, 55032 (70,7%) tinham estrutura compatível com o CPF (nove dígitos mais dois dígitos verificadores), caracterizando o primeiro titular como “pessoa física”, enquanto 22843 (29,3%) apresentavam estrutura de CNPJ (doze dígitos mais dois dígitos verificadores), caracterizando o titular como “pessoa jurídica”. Restaram 13497 (14,8%)

1. World Intellectual Property Organization (Wipo) é uma agência da Organização das Nações Unidas (ONU) criada em 1967 para estimular a proteção da propriedade intelectual em âmbito mundial, com base na cooperação entre países e a colaboração de organizações internacionais. Conta hoje com 184 países-membros (mais informações em <www.wipo.org>).

das patentes de residentes que não continham informações que pudessem auxiliar a classificação do titular quanto à natureza jurídica, seja por ausência total ou parcial da informação.

O problema da falta de informação ou de inexatidão de informações foi ainda mais grave no campo de CEP (Código de Endereçamento Postal). Esse campo é importante para ajudar a localizar territorialmente os pedidos de patentes. Contudo, entre os 91372 depósitos de residentes, 60662 (66,4%) não informavam o CEP, sendo que os demais 30710 necessitavam de algum tratamento (busca nos dados da Rais, busca na internet etc.) para se obter a localização do titular a partir desse campo.

A primeira restrição colocada às informações recebidas do INPI por via do Ipea foi relativa ao período. Para fins de análise neste capítulo, utilizaram-se os pedidos de patentes referentes ao intervalo de anos entre 1980 e 1999, inclusive. Embora este conjunto de dados apresentasse informações até o ano 2003, as informações dos quatro anos finais (2000 a 2003) foram descartadas. Tal procedimento foi adotado em virtude da melhor qualidade dos dados recebidos posteriormente do próprio INPI para o período 2000 a 2007.²

Assim, foram aproveitados 179692 depósitos de pedidos de patentes dos tipos PI e MU (71,16% do total de registros), sendo que 66609 (37,1%) eram de residentes e 113083 (62,9%) de não residentes. Entre os pedidos de residentes, 38686 (58,1%) tinham no primeiro titular uma “pessoa física”, enquanto as “pessoas jurídicas” apareciam como primeiro titular em 16371 (26,6%). Ainda 11552 (17,34%) dos pedidos de patentes não apresentavam informação nos campos CNPJ/CPF.

O segundo conjunto de dados utilizados foi resultado de uma solicitação da FAPESP junto ao INPI, exclusivamente para subsidiar o presente capítulo. Foram fornecidos dois arquivos no formato “txt”, totalizando 1 321 320 registros, que correspondem a 1 585 94 documentos, entre Patentes de Invenção, Modelos de Utilidade, Desenho Industrial e Certificado de Adição (depositados de 2000 a 2007).

Os depósitos dos tipos PI e MU somaram 126345 (79,67%) pedidos de patentes. Deste total, 44689 (35,4%) tinham residentes como primeiro titular, enquanto os outros 81656 (64,6%) foram depositados por não residentes.

Este conjunto de dados de 2000 a 2007 também trouxe informações sobre o CNPJ/CPF dos depositantes, o que permitiu posteriormente classificá-los quanto à sua natureza. Assim, 39411 (88%) depósitos de residentes continham alguma informação neste campo, dos quais 27593 (61,2%) apresentavam estrutura de CPF e 11818 (26,4%) tinham estrutura de CNPJ. Os demais 5278 (11,8%) dos pedidos não continham informação sobre CNPJ/CPF.

Em virtude do período de sigilo de 18 meses inerente ao processo de concessão de patentes, e tendo em conta que a busca nos bancos de dados do INPI foi realizada em meados de 2007, as informações para 2006 e 2007 não correspondem ao total de pedidos, de fato, depositados para estes anos. Dessa forma, foram utilizados apenas os depósitos de patentes referentes aos anos 2000 a 2005, cujo período de sigilo já não restringia as informações disponíveis.

O período 2000 a 2005 contabiliza um total de 118648 depósitos dos tipos PI ou MU. Deste total, 40173 (33,9%) são de residentes e 78475 (66,1%) são de titulares não residentes. Entre os pedidos de residentes, 25046 (62,3%) são de titularidade de “pessoa física”, enquanto 10615 (26,5%) têm uma “pessoa jurídica” como primeiro titular. Os 4513 (11,2%) pedidos de patentes restantes não continham informação sobre CNPJ/CPF, o que não permitiu a classificação da natureza jurídica.

Diferentemente do conjunto inicial de informações referentes ao período 1980 a 1999, para os anos de 2000 a 2005 não foram fornecidas informações sobre o CEP do depositante/inventor, sendo, contudo, informados seus estados e respectivos municípios. Entre os pedidos de patentes de residentes, porém, 606 não informavam o estado e 4653 não informavam o município.

2. Tratamento dos dados fornecidos pelo INPI

Para gerar o banco de dados utilizado neste capítulo, as informações obtidas junto ao INPI foram tratadas visando cumprir três objetivos. O primeiro foi possibilitar o uso da maior quantidade de dados possível, buscando completar as informações não disponíveis nos arquivos recebidos. O segundo referia-se à melhora da qualidade das informações recebidas. Por fim, buscou-

2. Na verdade, devido ao período de sigilo de 18 meses a que está sujeito o pedido de patente e ao fato de este capítulo trabalhar com os pedidos depositados junto ao INPI – e não com as patentes concedidas –, ao menos os dois últimos anos desta base de dados não correspondem ao total de pedidos depositados nos seus respectivos anos. De fato, um pedido depositado em 1º de janeiro de 2001 só poderia aparecer nos dados a partir de julho de 2002. Assim, como os dados foram extraídos do sistema do INPI em algum momento de 2003, a partir do ano de 2001 não havia garantias de que estivessem registrados todos os pedidos depositados no período. Além disso, como a nova base de dados fornecida (2000 a 2007) apresentava maior consistência nas informações, preferiu-se trabalhar com este segundo conjunto de dados no limite do que ele permitia, ou seja, até o ano 2000, em detrimento do mesmo ano da base anterior.

se agregar informações externas àquelas fornecidas pelo INPI (Rais, internet) que possibilitassem uma análise mais rica em termos das atividades tecnológicas realizadas no Brasil entre 1980 e 2005.

Para cumprir o primeiro objetivo, foi realizado um trabalho de harmonização dos dados, que permitiu inserir informações em campos originalmente vazios em um registro qualquer a partir do conteúdo presente no mesmo campo de um registro com informações mais completas. Assim, por exemplo, se dois pedidos de patente possuíam o mesmo CNPJ e um deles não informava a localização do titular, então estas informações foram inseridas com base no registro mais completo, sob a hipótese de ser a mesma localização.

Em relação ao segundo objetivo, buscou-se refinar o banco de dados para que este possibilitasse a construção de estatísticas mais confiáveis, uma vez que as informações originais apresentavam muitos erros de preenchimento e grafia. Dessa forma, depósitos com CNPJ iguais, porém com divergências quanto ao nome do titular, tiveram tais nomes harmonizados com base no mais completo e, quando a diferença foi muito grande, recorreu-se ao serviço de consulta à situação cadastral de CNPJs da Receita Federal do Brasil.

Outro procedimento importante foi a obtenção dos estados e municípios dos titulares a partir dos CEPs dos pedidos de patentes depositados entre 1980 e 1999. Em primeiro lugar, utilizando-se o método descrito no parágrafo anterior, foram corrigidas as informações fornecidas de forma a se obter, em tantos depósitos quanto possíveis, um número compatível com a estrutura de um CEP. Dessa forma, foram tratados os 91732 depósitos, sendo que 60662 não continham a informação do CEP, enquanto os demais 30710 precisaram de alguma conferência. Quando esse tratamento não foi possível, a informação foi descartada e os campos de localização entre estados e municípios receberam valor “NA”.

Em segundo lugar, fez-se uso do “Endereçador” da Empresa Brasileira de Correios e Telégrafos para transformar os CEPs em estados e municípios. Trata-se de um CD adquirido junto aos Correios, capaz de fazer essa correspondência. Em seguida, foi realizado o mesmo procedimento utilizado anteriormente para completar informações ausentes em CNPJs iguais.

Esta parte da montagem do banco de dados foi particularmente trabalhosa. As informações fornecidas pelo INPI (segundo conjunto de dados) apresentavam muita inconsistência nos campos de estado e município. Em muitos casos, faltava o estado, o que exigia uma busca pelo município para completar a informação do campo. Em outros casos faltava o município, o que dificultava a complementação das informações, em particular para titulares “pessoa física”, em virtude da impossibilidade de informações sobre o CPF. Além disso, os dados apresentavam muitas inconsistências,

no sentido de que os campos foram preenchidos com informações diferentes de seus atributos, isto é, em vez de informar o nome do município, o titular fornecia o bairro/localidade, ou mesmo o município seguido do bairro, no mesmo campo. Dessa forma, foi preciso examinar registro por registro para a “limpeza” do banco de dados. Houve casos em que a informação contida no campo município não era compatível com a informação no campo estado, ou seja, não havia naquele estado um município com o nome informado. Em alguns casos, foi possível corrigir o dado de estado, quando se tratava de municípios de nome único no país. Nos demais casos, a informação do município foi perdida.

É importante esclarecer a diferença entre os dois conjuntos de dados utilizados, no que se refere às informações sobre a localização dos titulares dos pedidos de patentes. O primeiro conjunto de dados, cobrindo o período de 1980 a 1999, não informava diretamente a localização dos titulares. Tal localização só foi possível, inicialmente, por meio da conversão dos códigos de CEP encontrados (incluindo a complementação e correção descritas anteriormente) em municípios e estados, graças ao “Endereçador”. O segundo conjunto de dados, cobrindo o período de 2000 a 2005, forneceu diretamente o nome dos estados e municípios, embora houvesse também a necessidade de complementação e correção das informações. Assim, os 4653 depósitos que, originalmente, não informavam o município do titular e os 606 que não informavam o estado foram comparados aos demais registros com CNPJ/CPF iguais para que os dados faltantes fossem completados.

Após serem tratados, esses dois conjuntos de dados foram combinados e harmonizados para que as informações recuperadas fossem consistentes ao longo de todo período entre 1980 e 2005.

Para alcançar o terceiro objetivo, recorreu-se a duas outras fontes. Em primeiro lugar, utilizou-se a Relação Anual de Informações Sociais, com a identificação dos CNPJs de cada estabelecimento (Rais-ID), fornecida pelo Ministério do Trabalho e Emprego (MTE). Essa base de dados permitiu obter a Classificação Nacional de Atividades Econômicas (CNAE), por meio do CNPJ informado no pedido de patente.

Contudo, alguns cuidados e algumas correções foram necessários para que a Classe CNAE pudesse ser usada. Foram encontrados, após a correspondência com os dados da Rais, alguns CNPJs com a Classe CNAE “Sedes de empresas e unidades administrativas locais”, o que de fato não representa a real atividade da empresa depositante. Dessa forma, mais uma vez foi utilizado o serviço de consulta à situação cadastral de CNPJs da Receita Federal do Brasil, com o objetivo de conferir a correta classificação de todos os CNPJs nessa condição, bem como obter a classificação quando esta não constava na Rais. Foi atribuída ao titular, então, a

Classe CNAE informada pela Receita Federal, convertida para a Classe CNAE 1.0, com base na tabela de correspondência disponibilizada pelo IBGE.³

Quando não foi possível obter uma informação segura a partir dos dados da Receita Federal, optou-se por atribuir às empresas titulares a classificação da principal atividade econômica do grupo, obtida pelo radical do CNPJ informado e comparado com as demais informações do banco de dados, como localização e ano de depósito.

Além disso, a Rais-ID também foi utilizada para obter os estados e municípios dos titulares, o que serviu ao mesmo tempo para completar campos vazios e corrigir informações imprecisas que constavam das bases de dados originais.

A segunda fonte foi o *Science & Technologie: indica-*

teurs 2006 (OST, 2006), que informou o algoritmo para agregação das classes tecnológicas dos pedidos de patentes em “domínios tecnológicos”. Dessa forma, foi possível identificar, por exemplo, um pedido classificado originalmente, com base na Classificação Wipo, em “peptídeos” (classes C07K) como sendo um pedido de patente em biotecnologia. Além disso, foi possível a agregação das centenas de classes em apenas sete “domínios tecnológicos” e 30 “subdomínios”.

Após o tratamento dado aos dois arquivos recebidos do INPI, as informações foram compiladas em dois bancos de dados com pedidos de patentes depositados entre 1980 e 2005, sendo o primeiro, e mais completo, referente aos pedidos de residentes no Brasil, e o segundo com informações dos pedidos depositados por titulares não residentes, no mesmo período.

3. As informações contidas no site da Receita Federal utilizam a classificação da CNAE 2.0. A Classificação Nacional de Atividades Econômicas (CNAE 2.0), elaborada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e utilizada neste capítulo, corresponde à International Standard Industrial Classification (ISIC), em sua quarta revisão, a qual é adotada pela Organização das Nações Unidas (ONU).