Procedimento metodológico para coleta de dados e elaboração de painel de Patentes

Data de Início do projeto: 2008 a outubro de 2026

1. Tema e Contexto

O Observatório Fiocruz em Ciência, Tecnologia e Inovação em Saúde é uma iniciativa da Coordenação de Informação e Comunicação da Vice-Presidência de Educação, Informação e Comunicação da Fundação Oswaldo Cruz (Fiocruz). Sua proposta é transformar dados provenientes de diversas fontes em informações com alto valor agregado para a Instituição e para a sociedade através da construção de indicadores que tenham como premissa favorecer a tomada de decisão e proporcionar maior transparência à sociedade, fortalecendo assim direta e indiretamente as ações que compõe o Sistema Único de Saúde. Este documento descreve os procedimentos adotados na coleta, transformação e disponibilização do indicador de Patentes. Dashboards ou painéis são formatos de visualizações que permitem explorar os dados e analisar as relações entre eles a partir de simples cliques e/ou filtros interativos. O objetivo desta descrição metodológica, além de dar transparência ao indicador é promover o acesso à informação e a reprodutibilidade da coleta e do processamento do conjunto de publicações que compõe Patentes no período de cobertura de dados informado.

2. Fontes de Dados

Esta seção descreve as fontes institucionais e tecnológicas utilizadas, os conteúdos agregados e a referência temporal da base disponibilizada no painel.

2.1. Plataforma/Sistema

A principal fonte de dados é a Coordenação de Gestão Tecnológica da Fiocruz (Gestec), Núcleo de Inovação Tecnológica responsável pela gestão e manutenção do patrimônio intelectual da Fundação e utiliza, quando necessário, a base institucional da plataforma Questel Orbit como um complemento de informações. Ela pode ser utilizada para consulta e acompanhamento de documentos de patentes, famílias, classificações, status e dados bibliográficos.

2.2. Conteúdo Agregado

As fontes agregam informações administrativas e técnicas relacionadas às patentes da Fiocruz. Entre os campos descritos estão "Referência Fiocruz", "Nº do Pedido Prioritário", "Título", "Inventores" e "Documentos", além de informações sobre famílias de patentes, datas de depósito, países, instituições externas, unidades da Fiocruz, classificações e status. Para fins de apresentação no painel, os status foram interpretados e agrupados conforme as regras descritas a seguir:

- **Requerido ou depositado/pendente:** indica que o pedido de patente foi formalmente depositado no Instituto Nacional da Propriedade Industrial (INPI) ou em repartição oficial de propriedade intelectual no Brasil ou no exterior e permanece aguardando análise técnica.
- **Concedido:** indica que a análise técnica concluiu pelo atendimento aos requisitos aplicáveis e que o título de patente foi concedido, conferindo ao titular os direitos associados à proteção durante o período de vigência.
- **Não vigente:** indica que a proteção deixou de vigorar, por motivos como término do prazo legal, ausência de pagamento de anuidades, renúncia ou outra causa registrada pela repartição competente.
- **Não requerido:** indica que não houve depósito formal para a invenção ou tecnologia. Essa categoria foi excluída do conjunto de dados utilizado no painel, conforme a regra metodológica do documento de origem.

De modo geral, os registros classificados como “Concedido” e “Requerido” foram apresentados como documentos de patente vigentes, enquanto “Não vigente” corresponde aos documentos de patente não vigentes.

3. Procedimento Metodológico para Coleta e Tratamento

Esta seção detalha as etapas de obtenção, limpeza, padronização, agrupamento, deduplicação e organização relacional dos registros de patentes.

3.1. Coleta dos Dados

Recebemos da Gestec uma planilha com os dados atualizados de Patentes. A Curadoria foi realizada diretamente pela área, ficando sob responsabilidade do Observatório a construção do painel em BI. O recorte temporal do indicador abrange registros de 2008 até setembro de 2026.

3.2. Tratamento dos Dados

Realizamos diversas rodadas de análises com a área responsável (Gestec) com o objetivo de entendermos a base fornecida, bem como o significado de cada metadado, garantindo a representação coerente das invenções, dos pedidos e dos documentos de patente, evitando duplicidades e inconsistências de escrita.

3.2.1. Família

É o conjunto de pedidos de patente e patentes concedidas em mais de um país, referentes a mesma invenção e requeridos pelo(s) mesmo(s) depositante(s). Cada Referência Fiocruz corresponde a uma família de patentes.

3.2.2. Número Pedido Prioritário

Contém apenas o primeiro pedido da família.

3.2.3. Número Pedido/ Patente

É o número total de pedidos de patente e patentes concedidas de titularidade da Fiocruz. É o documento de patente.

3.3. Filtragem Inicial

Na filtragem inicial, foram excluídos os registros classificados como “Não requerido”, pois não representam pedidos formalmente depositados. Também foram preservadas as classificações necessárias à distinção entre documentos vigentes e não vigentes.

3.4. Padronização de Campos Textuais

Os campos textuais foram padronizados para uniformizar a escrita e facilitar a integração das fontes, incluindo uma rotina em cada atualização, destinada a detectar situações em que múltiplos nomes completos aparecem sem separadores explícitos.

3.5. Agrupamento e Deduplicação

Diferentes documentos pertencentes à mesma família de patente foram agrupados, de modo que a análise pudesse considerar a invenção e, quando necessário, suas múltiplas proteções em diferentes países ou jurisdições. Linhas duplicadas foram removidas para preservar a integridade da base e assegurar a contagem correta dos registros.

3.6. Chave Composta e Atomicidade

Foi criada a chave composta "Referencia_NumeroPedido", formada pela união dos campos "Referência Fiocruz" e "Nº do Pedido". Essa chave foi utilizada para identificar cada combinação de referência institucional e pedido, garantir a atomicidade dos registros e viabilizar a contagem distinta dos documentos de patente.

3.7. Armazenamento e Estruturação

O tratamento resultou em um modelo dimensional composto por sete tabelas. A coluna "Referencia_NumeroPedido" atua como chave de relacionamento na maior parte da estrutura. A tabela de inventores constitui uma exceção e utiliza a "Referência Fiocruz" como campo de ligação, conforme a granularidade disponível para esse conjunto de dados.

4. Elaboração do Painel/Relatório

Com o objetivo de ampliar a compreensão do portfólio de patentes e facilitar a análise exploratória, foi desenvolvido um painel interativo no Microsoft Power BI.

4.1. Princípios e Fontes

O painel utiliza como fontes as bases da Gestec e da Questel Orbit após as etapas de integração, padronização, agrupamento, deduplicação e estruturação. A organização do indicador permite analisar o portfólio a partir de elementos administrativos, técnicos e relacionais, preservando as restrições de confidencialidade aplicáveis.

4.2. Modelagem e Transformação

A modelagem adotada é dimensional e está organizada em sete tabelas relacionadas. A tabela central "NPedido" concentra informações dos pedidos e se relaciona com tabelas destinadas à contagem de documentos, países, instituições externas, unidades da Fiocruz, referências e pedidos prioritários e inventores. Predominam relacionamentos de um-para-muitos (1:N), utilizando "Referencia_NumeroPedido" como chave principal, com a exceção já descrita para a tabela de inventores.

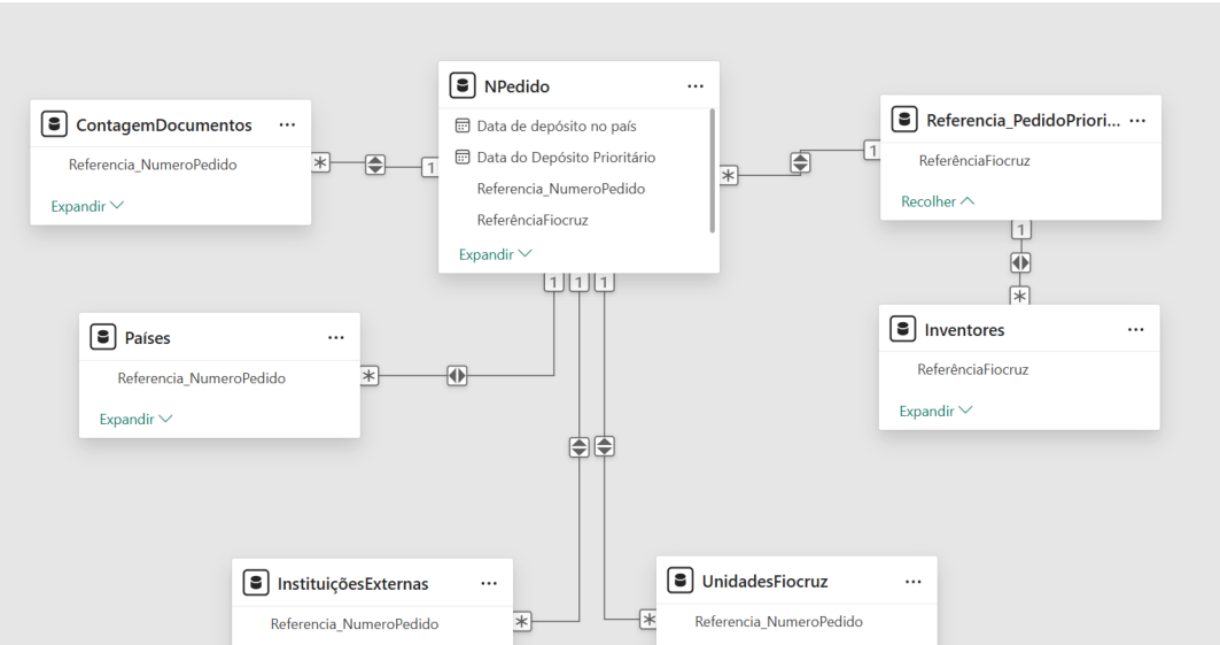


Figura 1 – Modelo de dados do painel de Patentes. Fonte: documento metodológico original.

4.3. Identidade Visual e Funcionalidade

O painel foi estruturado para oferecer exploração interativa das informações de patentes. As visualizações permitem consultar dados por status, pedidos, documentos, países,

unidades da Fiocruz, instituições externas, inventores e demais dimensões presentes no modelo.

4.4. Sigilo Estratégico

Para assegurar o sigilo estratégico, foi implementada uma camada de filtragem que omite informações detalhadas das patentes classificadas como confidenciais. Campos sensíveis, como "Nº do Pedido Prioritário", "Título", "Inventores" e "Documentos", não são exibidos nas visualizações quando associados a registros confidenciais.

5. Análise Avançada: Famílias e Relações do Portfólio

O agrupamento por famílias de patentes permite analisar uma mesma invenção sem confundir suas diferentes proteções documentais. Essa abordagem favorece a leitura do portfólio em dois níveis complementares: o nível da invenção ou família e o nível dos pedidos e documentos associados.

A estrutura relacional também possibilita examinar a distribuição dos pedidos por países, unidades da Fiocruz, instituições externas e inventores, bem como comparar documentos vigentes e não vigentes. Essas dimensões ajudam a compreender a abrangência territorial, institucional e colaborativa do patrimônio intelectual da Fiocruz.

6. Considerações Técnicas

Esta seção reúne os principais cuidados, decisões de modelagem e limitações documentais identificados no procedimento.

- **Integração de fontes:** a base combina informações institucionais da Gestec com dados técnicos e bibliográficos da Questel Orbit. A consistência da integração depende da correspondência correta entre referências, números de pedidos e demais identificadores.
- **Qualidade dos nomes:** foi implementada uma rotina para detectar múltiplos nomes completos sem separadores, reduzindo inconsistências em campos de inventores e outros registros nominais.
- **Famílias de patentes:** o agrupamento deve preservar a distinção entre a invenção e seus diferentes pedidos ou documentos, evitando tanto a supercontagem quanto a perda de detalhes necessários à análise territorial e documental.
- **Chave de relacionamento:** a chave "Referencia_NumeroPedido" sustenta a atomicidade e a maior parte dos relacionamentos do modelo. Alterações nos campos

que a compõem devem ser avaliadas para evitar quebra de vínculos ou duplicação de registros.

- **Granularidade dos inventores:** a tabela de inventores utiliza "Referência Fiocruz" em vez da chave composta, devendo ser tratada conforme sua granularidade específica para evitar relações ambíguas ou multiplicação indevida de linhas.
- **Confidencialidade:** os campos sensíveis associados a registros confidenciais devem permanecer ocultos no painel público.