

Procedimento metodológico para coleta, tratamento e visualização de dados para elaboração de painel de Instrumentos de Cooperação Internacional

1. Tema e Contexto

1.1. Tema

Este painel apresenta um conjunto de informações sobre a gestão de parcerias e acordos internacionais da Fiocruz, monitorando os Instrumentos de Cooperação Internacional (ICI), que abrangem Memorandos de Entendimento (MdE), Convênios e outros instrumentos. A ferramenta reúne dados estruturados formando uma série histórica com informações de instrumentos vigentes a partir do ano de 2015, oferecendo um panorama geográfico e institucional consolidado.

São apresentadas informações sobre as principais características desses projetos: ano, número do processo, título, *status* de validação, períodos de vigência (datas de início e término originais e atuais), unidades gestoras, coordenadores e representantes legais da Fiocruz. Além disso, o painel mapeia os países envolvidos e detalha as instituições parceiras, agrupando-as por tipologia (como Associação, Empresa, Instituto, Organismo Multilateral, Órgão de Governo ou Universidade).

Essa análise detalhada, confiável e atualizada oferece uma base quantitativa e qualitativa fundamental para subsidiar ações da Coordenação Executiva de Convênios e Mobilidade Internacional, além de apoiar o acompanhamento gerencial e institucional.

1.2. Contexto

Nesse contexto, o Sistema de Apoio à Gestão Estratégica (SAGE) constitui-se como a ferramenta oficial da instituição e repositório de dados internos, responsável por manter o registro de projetos, convênios e informações sobre parcerias internacionais da Fiocruz.

1.3. Objetivo do Trabalho e Escopo

Com o objetivo de automatizar, estruturar e visualizar os dados disponíveis no SAGE, foi desenvolvido um painel interativo (Dashboard). O procedimento metodológico detalhado a seguir abrange a coleta direta no banco de dados, limpeza, classificação e integração com planilhas de suporte, garantindo a consistência das informações.

2. Fontes de Dados

2.1. Plataforma/Sistema

A fonte principal de dados provém do **Sistema SAGE**, através de conexão direta ao banco de dados MySQL. Como fonte secundária, utiliza-se o ambiente de armazenamento institucional (SharePoint), que aloca as planilhas de apoio e dicionários para classificação de parcerias mantidas pela equipe do CRIS. A base agrega campos essenciais como

identificação (Ano, Processo, Título, Status), vigência e representação (Coordenadores, Instituições Parceiras, Unidades e Países).

2.2. Acesso

O acesso aos dados principais é feito de forma automatizada no *software* Microsoft Power BI, utilizando *scripts* que se conectam via IP ao banco MySQL do SAGE. A inserção de dados complementares de classificação de instituições é feita de forma manual na planilha armazenada em nuvem antes da atualização do painel.

3. Procedimento Metodológico para Coleta e Tratamento

3.1. Coleta dos Dados e Filtragem Inicial

A coleta de dados foi centralizada e programada no Power BI (*Power Query*) em módulo principal através de linguagem estruturada SQL. A extração consulta simultaneamente as tabelas de Projetos (ProjetoCooperacao), Tipos de Projeto, Instituições, Países, Unidades, Status e Colaboradores do SAGE.

A filtragem inicial é rigorosamente processada de forma automatizada na instrução da consulta SQL, onde são extraídos estritamente os projetos que atendem de forma simultânea aos seguintes critérios:

- Pertencem ao **Grupo de Projetos de Cooperação Internacional** (identificador = 2).
- Encontram-se com **status validado no sistema** (códigos 2, 3 e 6).
- São os **instrumentos principais ou originais**, caracterizados pelo campo Aditivo contendo '0' ou '00'.
- Possuem **vigência válida** no período de análise (Data de término original ou atual maior ou igual ao ano de 2015).

3.2. Tratamento dos Dados, Consolidação e Harmonização de Registros

Para garantir a representação única de cada projeto e contornar a multiplicidade de linhas gerada quando um instrumento possui mais de uma unidade gestora, múltiplos coordenadores ou mais de uma instituição parceira, utilizou-se a função de agregação (GROUP_CONCAT) nativa na *query* SQL. As múltiplas ocorrências são concatenadas textualmente e separadas por ponto e vírgula (;), selecionando apenas ocorrências distintas (DISTINCT) e ordenando-as alfabeticamente. Desta forma, previne-se a duplicação de dados e o projeto consolida-se em uma única linha matricial no painel.

Os registros extraídos são, na sequência, validados em face do arquivo "classificacao-da-instituicao-parceira". Identificam-se novos registros importados na base principal, garantindo que recebam manualmente a sua tipologia correspondente (Ex: Associação, Empresa, Instituto, Organismo Multilateral) e abrangência (Nacional ou Internacional) no arquivo auxiliar.

3.3. Armazenamento Os dados principais utilizados na construção do painel são consultados de maneira dinâmica e estruturada diretamente do banco de dados SAGE, abolindo o armazenamento de extrações estáticas. O modelo consolidado (banco SAGE + dicionário SharePoint) fica armazenado e gerenciado no próprio ambiente do Power BI.

4. Elaboração do Painel/Relatório de visualização dos dados

4.1. Princípios e Fontes

A elaboração visual seguiu os princípios de padronização estipulados pelo Observatório da Fiocruz. O painel consolida a conexão SQL automatizada (ao servidor MySQL no IP "157.86.96.96", banco "sage") e o carregamento do arquivo de classificação auxiliar.

4.2. Modelagem e Transformação

Esta etapa utiliza o Power Query Editor para preparar os dados de cooperação internacional. A modelagem central é realizada logo na extração (via *query*), estabelecendo o relacionamento (através de comandos INNER JOIN e LEFT JOIN) entre a tabela principal ProjetoCooperacao e suas diversas dimensões de apoio (como TipoProjetoCooperacao, InstituicaoProjetoCooperacao, EquipeProjetoCooperacao, Pais, etc.).

O tratamento textual padroniza como as razões sociais parceiras, siglas institucionais e nomes de colaboradores ou representantes legais são extraídos e concatenados, criando colunas prontas para consumo visual (ex: "unidades_gestoras", "inst_parceira", "pais", "rep_legal_fiocruz" e "coordenador_fiocruz").

4.3. Identidade Visual e Funcionalidade A interface e a identidade visual reforçam a padronização institucional. A ferramenta engloba filtros interativos voltados à exploração de informações vigentes de 2015 até o ano atual. O painel interativo é submetido a uma rigorosa etapa de validação cruzada entre o Observatório e a equipe técnica do CRIS. Ao confirmar a consistência de totais, contagens e inexistência de inconsistências, o relatório é atualizado no respectivo *workspace* e publicado no site oficial do Observatório.

5. Considerações Técnicas

Para assegurar a confiabilidade contínua da atualização do painel, estabeleceu-se as seguintes prerrogativas:

- **Integridade das Fontes Auxiliares:** Qualquer alteração na planilha de classificação deve manter estritamente os nomes das colunas, acentuação e tipo de dados originais. Alterações indevidas na estrutura corrompem a leitura do Power BI.
- **Prevenção de Erros de Amostragem:** Instituições parceiras sem correspondência na planilha de classificação geram anomalias na visualização dos indicadores, sendo obrigatório atualizar o arquivo auxiliar antes de cada atualização do painel.

- **Validação de Preenchimento:** Registros com falha na inserção das datas no SAGE ou falta de identificação do país/parceira exigem verificação e correção direta pela equipe na origem do banco de dados (SAGE), garantindo a integridade do relatório.