

Metodologia para coleta de dados dos Grupos de Pesquisa CNPq

Os dados utilizados para criação deste indicador são dados públicos e disponíveis no Diretório dos Grupos de Pesquisa, do CNPq (<https://lattes.cnpq.br/web/dgp>). A coleta/extração foi feita no dia 17 de junho de 2025 utilizando um programa *Web crawler* especialmente desenvolvido para indexar o conteúdo do site do diretório de grupos do CNPq.

O uso do *Web crawler* foi escolhido por não ser conhecido nenhum serviço Web para extração direta do Diretório do CNPq. Desta forma, o programa *Web crawler* captura todos dados do diretório independentemente do vínculo institucional. Após a obtenção dos dados em formato HTML, estes foram convertidos para o formato Excel.

Um novo programa lê a planilha excel e realiza a seleção dos grupos de pesquisa coletados que pertençam à Fiocruz através de um filtro aplicado ao campo “Instituição do grupo” buscando pelo termo Fundação Oswaldo Cruz – Fiocruz. Não houve filtros de datas, tendo sido considerados todos os grupos de pesquisa ativos e inativos desde 1975 até a data da coleta.

Os campos analisados foram: Id do grupo de pesquisa, Grupo de pesquisa, Situação do grupo, Ano de formação, Líder(es) do grupo, Instituição do grupo, UF, Linhas de pesquisa, Colaboradores estrangeiros, Área de Pesquisa, Subárea de Pesquisa, Nome dos Pesquisadores, Titulação dos Pesquisadores.

Os nomes das unidades da Fiocruz a que pertenciam os grupos filtrados, foram padronizados. Ainda algumas unidades não foram indicadas pelos líderes dos grupos no momento de seu cadastro no CNPq, casos em que permaneceu o nome mais geral: Fiocruz. Os status dos grupos foram agrupados da seguinte forma:

"Aguardando certificação": "Ativo",

"Certificado": "Ativo",

"Certificado - Não-atualizado há mais de 12 meses": "Ativo",

"Em preenchimento": "Em preenchimento",

"Excluído": "Inativo"

Após a aplicação da regra, os resultados foram consolidados. Em seguida, a planilha foi usada como base para construir uma visualização interativa publicada junto com esta metodologia no site: observatorio.fiocruz.br