

Metodologia para coleta de dados relacionados aos Grupos de Pesquisa da Fiocruz

Para esta análise foram utilizados dados públicos disponíveis no Diretório dos Grupos de Pesquisa, do CNPq (<https://lattes.cnpq.br/web/dgp>). A coleta/extração foi feita no dia 05 de novembro de 2022 utilizando um *webcrawler* no sistema do CNPq. O resultado foi salvo em páginas para consulta e convertidas para o formato xml (Excel). Todos os dados do diretório foram extraídos. Para seleção dos grupos da Instituição, um filtro no campo “Instituição do grupo” buscando pelo termo Fundação Oswaldo Cruz – Fiocruz, foi realizado.

Nesta análise, foram considerados os grupos de pesquisa ativos e inativos desde 1975 até 2022. 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, Unidade Padronizada¹, UF, Linhas de pesquisa, Colaboradores estrangeiros, Área de Pesquisa, Subárea de Pesquisa, Nome dos Pesquisadores, Titulação dos Pesquisadores.

Estes dados foram indexados no *Elastic Search* e transformados em visualizações no software kibana.

¹ Para harmonização dos dados referentes as unidades da Fiocruz a que pertencem cada grupo, um dicionário para padronização foi desenvolvido. Os pesquisadores que colocaram na descrição o nome do laboratório ou o departamento que trabalham tiveram estes substituídos pela sua unidade. A busca pela unidade pertencente do líder do grupo de pesquisa foi feita baseada no banco de dados da coordenação-Geral de Gestão de Pessoas (Cogepe).