

APLICAÇÃO DA FERRAMENTA GOOGLE LOOKER STUDIO PARA ANÁLISE E VISUALIZAÇÃO DE DADOS DE UM SISTEMA DE GESTÃO DE PESQUISA AMBIENTAL

Rodrigo Dias DUTRA

RESUMO

Este trabalho é uma proposta para aplicação de uma ferramenta para análise e visualização de dados de um sistema de gestão de pesquisa ambiental. Adotou-se a ferramenta de *Business Intelligence* (BI), intitulada Google Looker Studio, com base em conceitos de sistemas de informação na gestão pública, *Business Intelligence*, ETL (*Extrat, Transform and Load*) e visualização de dados. Utilizou-se na aplicação a base de dados do CadGP (Cadastro e Gestão de Pesquisas), que é uma plataforma de gestão de projetos de pesquisa desenvolvidos em Unidades de Conservação estaduais do SIEFLOR (Sistema Estadual de Florestas do Estado de São Paulo), bem como no âmbito do Instituto de Pesquisa Ambientais da Secretaria de Meio Ambiente, Infraestrutura e Logística do Estado de São Paulo. A base de dados abrange projetos nas áreas de Biodiversidade, Botânica, Ciências Florestais e Geociências, sendo esta última enfatizada no presente trabalho. Apresentou-se os resultados e dificuldades no processo ETL e na análise e visualização de dados no Google Looker Studio. As ferramentas utilizadas, todas disponíveis de forma gratuita, mostraram-se eficientes na representação das informações dos projetos e permitem subsidiar o processo de gestão de pesquisa.

Palavras-chaves: Painel de Dados; Business Intelligence (BI); Visualização de Dados; Google Looker Studio; Geociências.

ABSTRACT

APPLICATION OF THE GOOGLE LOOKER STUDIO TOOL FOR ANALYZING AND VISUALIZING DATA FROM AN ENVIRONMENTAL RESEARCH MANAGEMENT SYSTEM. This study presents a proposal for the implementation of a tool for the analysis and visualization of data derived from an environmental research management system. A business intelligence (BI) tool, Google Looker Studio, was selected based on its alignment with the conceptual framework of information systems in public management, business intelligence, ETL (Extract, Transform and Load), and data visualization. The CadGP (Research Registry and Management) database, which is used for the management of research projects developed in state conservation units of the SIEFLOR (São Paulo State Forest System) areas, as well as within the scope of the Environmental Research Institute of the Secretariat for the Environment, Infrastructure, and Logistics of the State of São Paulo, was employed in this study. The database encompasses projects pertaining to Biodiversity, Botany, Forest Sciences, and Geosciences, with the latter being focused in this work. The outcomes and challenges encountered during the ETL process, along with the data analysis and visualization in Google Looker Studio, were elucidated. The tools utilized, all freely available, proved to be efficient in portraying project data and facilitating the research management process.

Keywords: Dashboards; Business Intelligence (BI); Data Visualization; Google Looker Studio; Geosciences.

RESUMEN

APLICACIÓN DE LA HERRAMIENTA GOOGLE LOOKER STUDIO PARA EL ANÁLISIS Y VISUALIZACIÓN DE DATOS DE UN SISTEMA DE GESTIÓN DE LA INVESTIGACIÓN MEDIOAMBIENTAL. Este estudio presenta una propuesta para la implementación de una herramienta para el análisis y visualización de datos derivados de un sistema de gestión de la investigación ambiental. Se seleccionó una herramienta de inteligencia de negocios (BI), Google Looker Studio, en función de su alineación con el marco conceptual de los sistemas de información en la gestión pública, inteligencia de negocios, ETL (Extract, Transform and Load) y visualización de datos. En este estudio se utilizó la base de datos CadGP (Registro y Gestión de la Investigación), que se utiliza para la gestión de los proyectos de investigación desarrollados en las unidades estatales de conservación del SIEFLOR (Sistema Forestal del Estado de São Paulo), así como en el ámbito del Instituto de Investigación Ambiental de la Secretaría de Medio Ambiente, Infraestructura y Logística del Estado de São Paulo. La base de datos engloba proyectos relativos a Biodiversidad, Botánica, Ciencias Forestales y Geociencias, siendo esta última la enfocada en este trabajo. Los resultados y desafíos encontrados durante el proceso ETL, junto con el análisis de datos y visualización en Google Looker Studio, fueron elucidados. Las herramientas utilizadas, todas ellas de libre acceso, demostraron su eficacia a la hora de representar los datos del proyecto y facilitar el proceso de gestión de la investigación.

Palabras clave: Cuadros de mando; Business Intelligence (BI); Visualización de datos; Google Looker Studio; Geociencias.

1 INTRODUÇÃO

Os órgãos públicos geram uma grande quantidade de dados por meio dos seus sistemas de informação, banco de dados, relatórios, planilhas entre outros. Promover a organização dos dados e torná-los em informações estratégicas é um grande desafio e, para isso, o uso de tecnologias são imprescindíveis para esse processo, pois oferece às organizações públicas mais dinamismo, disponibilidade e intercâmbio dessas informações (CHIAVENATO 2010). Tais dados necessitam de tratamento adequado para a divulgação ao público em geral, de forma transparente, visando tomadas de decisões assertivas e maior eficiência dos recursos públicos.

A implantação de sistemas de informação é fundamental para que o setor público possa melhor gerenciar seus dados internos ou externos, proporcionando um fluxo de dados consistente para que seus gestores, servidores e público em geral possam alimentar, acessar e usufruir dos dados armazenados, transformando-os em informação estratégica para tomadas de decisões.

Transformar dados brutos em informações para tomadas de decisões com o uso de ferramen-

tas tecnológicas é o conceito mais utilizado para definir *Business Intelligence* (BI). A aplicação dessas ferramentas auxilia as empresas a compreenderem melhor seus dados e usá-los para melhorar sua performance (FERREIRA 2023). Na mesma linha, ALBUQUERQUE (2021) expõe que o sistema de BI corresponde a técnicas, processos, ferramentas e métodos que oferecem ao usuário instrumentos para análises de cenários de mercado. O *Business Intelligence* (BI) possui regras para formatação adequada de dados gerados pelas empresas, transformando-os em depósitos estruturados de informações, além de apoiar os gestores empresariais no processo de tomada de decisão estratégica (BEZERRA 2015). Antes de qualquer análise é preciso identificar de onde provêm os dados, como serão captados, extraídos e reorganizados. A escolha correta de uma ferramenta de BI é de suma importância para que a organização não tenha prejuízos em recursos pessoais, financeiros, tempo e, principalmente, a perda de dados valiosos. É uma decisão técnica que pode definir o sucesso ou o fracasso de um projeto de análise de dados (LINHARES 2023).

O CadGP (Cadastro e Gestão de Pesquisas) pode ser entendido como um sistema de informa-

ção, pois é uma plataforma de gestão de projetos de pesquisa, onde pesquisadores vinculados a instituições do Sistema Ambiental Paulista e externos, submetem seus projetos, seguindo um fluxo de tarefas de cadastramento até seu encerramento. Todos os projetos desenvolvidos em Unidades de Conservação estaduais do SIEFLOR (Sistema Estadual de Florestas do Estado de São Paulo, conforme Decreto nº 51.453, 29/12/2006) estão cadastrados nesta plataforma. Além destes, também há projetos desenvolvidos em outras áreas de interesse no Estado, coordenados por pesquisadores do Instituto de Pesquisas Ambientais (IPA). Esse portal foi desenvolvido pelo IPA dentro do Sistema Integrado de Gestão Ambiental (SIGAM), plataforma de serviços da Secretaria de Meio Ambiente, Infraestrutura e Logística (SEMIL) do Governo do Estado de São Paulo. O CadGP abrange projetos nas áreas de Biodiversidade, Botânica, Ciências Florestais e Geociências, sendo esta última enfatizada no presente trabalho.

O objetivo deste estudo foi aplicar o conceito de *business intelligence* (BI) no sistema CaGP de modo a aprimorar a organização e apresentação de dados. Testou-se o uso da ferramenta Google Looker Studio para análise e visualização de dados em Geociências oriundos da plataforma CadGP, com o intuito de gerar relatórios mais dinâmicos e interativos visando auxiliar seus gestores a obterem novos *insights*, criar estratégias e um melhor aproveitamento na análise desses dados.

2 O PROCESSO DE ANÁLISE E VISUALIZAÇÃO DE DADOS NA PLATAFORMA CADGP

2.1 Plataforma CadGP

A plataforma CadGP possui dados de diversas áreas relacionadas ao Meio Ambiente, oriundos do cadastro e submissão de projetos de pesquisas, sendo muito acessada por pesquisadores, estudantes e profissionais em geral de diversas instituições em Ciências. A estrutura do CadGP permite ao usuário acessar um cadastro por meio de campos de preenchimento que estão divididos por abas de Projeto (Título do Projeto de Pesquisa, Data de Execução, Áreas de conhecimento científico do CNPq), Responsável (Nome do Autor, Modalidade da Pesquisa e Bolsa), Local (Área SIEFLOR e Município), entre outras.

Para a análise dos dados, os gestores da plataforma CadGP têm acesso ao Relatório Geral em planilha Excel, contendo todos os dados dos projetos de

pesquisa cadastrados ou em fase de cadastramento. Atualmente o número total de projetos da base de dados é de 1132.

Para fins deste trabalho foram selecionados projetos de pesquisas na área de Geociências, de acordo com as áreas do conhecimento CNPq, cadastrados pelos pesquisadores da instituição (IPA/SEMIL) e público externo. Considerou-se apenas os projetos em andamento, totalizando 35.

2.2 Processo ETL com Excel e Power Query

O processo ETL (extração, transformação e carregamento/*extract, transform, and load*) deve integrar a ferramenta de BI adotada por uma organização, pois possibilita filtrar, classificar e processar dados de fontes internas ou externas, facilitando, posteriormente, sua disponibilização. Caso uma empresa já possua sistemas transacionais que necessitam de migração de dados, o processo ETL já poderá ser implantado (GAMA & ABREU 2008). A extração de dados pode ser realizada de diversas fontes tais como planilhas eletrônicas, documentos de texto, banco de dados, páginas de internet entre outras (JUNIOR *et al.* 2014). Para que um processo de ETL tenha êxito, seu funcionamento irá depender do que cada organização definir como necessidade para transformar os dados confiáveis, definindo as regras para cada passo de execução, de modo que a ferramenta execute a tarefa no tempo programado (JUNIOR *et al.* 2014).

Para este estudo utilizou-se o *software* Excel e a ferramenta Power Query da Microsoft para migração e tratamento dos dados no processo ETL (Figura 1), o que proporcionou um tratamento assertivo, estruturado e de confiabilidade para posterior visualização de dados na ferramenta BI Google Looker Studio.

O relatório extraído do CadGP contém os campos: ID do projeto, Título do Projeto, Datas (início, término), Grande Área, Área, Subáreas e Especialidades das pesquisas, Nome do Pesquisador Responsável, Equipe do Projeto, Nome da Instituição, Unidade e Núcleo de Pesquisa, Local da Pesquisa e Município e Status do Projeto.

No Excel e no Power Query foram realizadas as tarefas de limpeza de dados (remover ID de títulos de projetos repetidos e remoção de células vazias), padronização (formatar palavras, excluir erros de grafia e acentuação, uso de delimitadores para dividir colunas, filtro do status do projeto) e uniformização (normalizar as datas para um único padrão). A planilha foi exportada para o Power Query, onde deu-se início ao tratamento de dados conforme o fluxo exposto na figura 2.

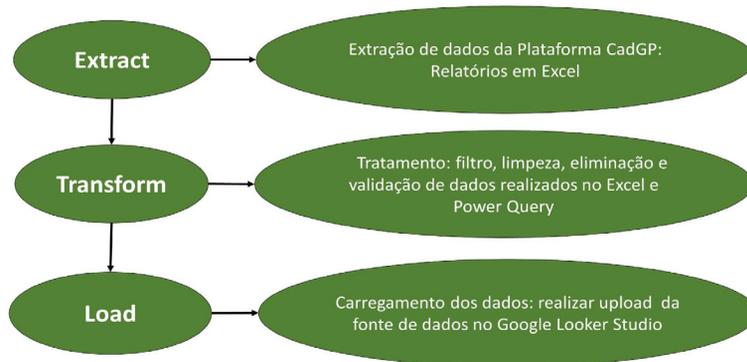


FIGURA 1 – Processo ETL no tratamento de dados da Plataforma CadGP.

FIGURE 1 – ETL Process in Data Processing of the CadGP Platform.



FIGURA 2 – Tratamento de dados no Power Query.

FIGURE 2 – Data Processing in Power Query.

2.3 Análise e visualização dos dados no Google Looker Studio

Tornar a interpretação de dados de fácil compreensão é um dos objetivos da visualização de dados e para isso, os gráficos, mapas, tabelas e *templates* são primordiais para alcance desse objetivo. A visualização de dados permite traduzir jargões técnicos em algo compreensível para o público leigo, sendo possível interpretar instantaneamente a partir de um gráfico ou tabela um conjunto de dados em tempo real (STORM 2022). A escolha de componentes para uma visualização de dados correta depende não somente dos dados tratados e estruturados, mas também de um visual acessível e claro para facilitar seu entendimento.

O Google Looker Studio é uma ferramenta para a visualização de dados *online* de acesso fácil e intuitivo que permite a confecção de painéis de dados, relatórios, formatação dos dados, dimensões e métricas (GOOGLE 2024). Adotou-se um

fluxo de tarefas para análise e visualização dos dados, conforme ilustrado na figura 3.

Para transformar os dados planilhados em representações gráficas seguiu-se as etapas de escolha de Fontes de Dados (planilha matriz), Dimensão (categoria dos dados), Métrica (conjunto de valores da planilha) e gráfico (opções de representações dos dados), conforme ilustrado na figura 4a. As representações gráficas disponíveis no Google Looker Studio permitem ao usuário a escolha que melhor represente seus dados, possibilitando abordagens de dados operacional, performance, desempenho, tático e estratégico (Figura 4b). O acesso aos dados é dinâmico, permitindo elaborar filtros, listas e controles para a sua consulta.

Para análise de dados em Geociências escolheu-se diversas representações de dados, de acordo com suas dimensões e métricas:

- a) Gráfico de Visão Geral (Número de projetos em andamento)

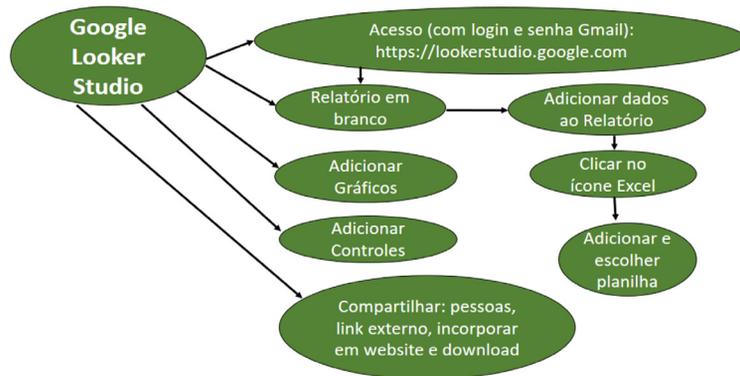


FIGURA 3 – Fluxo de tarefas no Google Looker Studio.
 FIGURE 3 – Task Flow in Google Looker Studio.

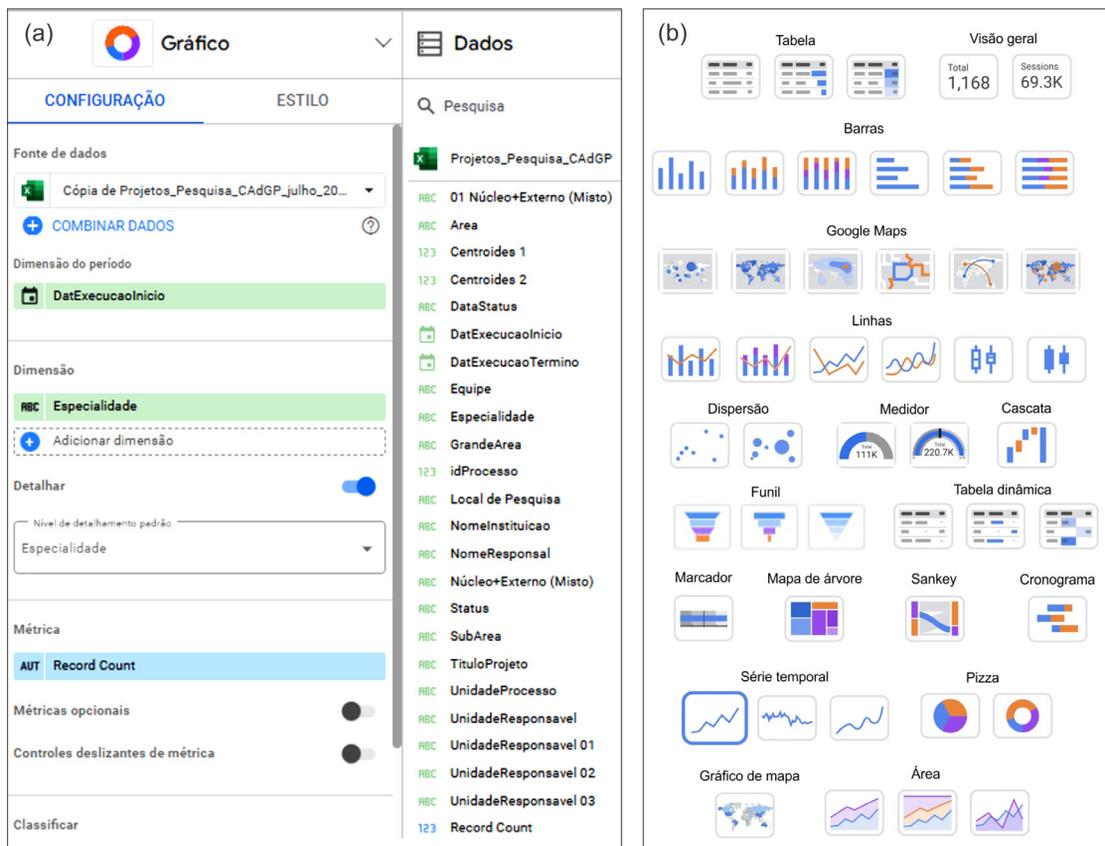


FIGURA 4 – Funcionalidades e padrões do Google Looker Studio: a) Menu de dados aplicado à fonte de Projetos de Pesquisa; b) Estilos de representações gráficas utilizadas. Fonte: Google Looker Studio.

FIGURE 4 – Functionalities and standards of Google Looker Studio: a) Data menu applied to the Research Projects source; b) Styles of graphical representations used. Source: Google Looker Studio.

b) Gráficos de Rosca (Número de área, sub-área e especialidades do conhecimento CNPq/ Projetos de instituições externas / fontes de recursos dos projetos de pesquisa).

c) Gráfico de Pizza (Tipos de projetos de pesquisa).

d) Tabela, Controles de Período e Controle de Filtros Avançados (Projetos de pesquisa internos, mistos e externos).

e) Mapa de Balão - Google Maps (Locais de Pesquisa em áreas SIEFLOR e municípios onde

foram realizados os projetos de pesquisa). Não se adotou neste trabalho para fins de localização geográfica as coordenadas (latitude e longitude) das áreas de pesquisa; ao invés disso, utilizou-se o recurso de “Informações Geográficas”, a partir da opção “Área de Designação de Mercado”, que corresponde a uma área mais regional dentro de um Estado ou Cidade (Figura 5). Esta ferramenta fornece uma localização automática, desde que o nome do local esteja gramaticalmente correto. Em geral, a ferramenta atribui a sede do município ou da unidade de conservação associada ao projeto de pesquisa. A representação espacial de projetos de abrangência regional, que abrangem vários municípios ou unidades de conservação, fica prejudicada neste formato. O ideal seria a representação por feições de polígonos, mas esta função não está automatizada.

3 RESULTADOS

As ferramentas de BI utilizadas neste estudo permitiram a elaboração de um painel de dados dinâmico e intuitivo, contendo informações das áreas do conhecimento de maior abrangência em Geociências, bem como a distribuição espacial dos projetos de pesquisa em andamento nas áreas SIEFLOR e municípios, ilustradas a seguir (Figura 6).

De acordo com o gráfico da Grande Área CNPq, 236 dos projetos de pesquisas em andamento cadastrados na plataforma CadGP são de Ciências Biológicas, seguidos por Ciências Exatas e da Terra com 48 projetos e Ciências Agrárias com 41. Em Ciências Exatas e da Terra inserem-se os projetos em Geociências e áreas afins que representam um total de 35 projetos em andamento. Considerando a Subárea CNPq, observa-se a seguinte distribuição: 17 em Geologia, 16 em Geografia Física e 2 em Meteorologia. As três “Especialidades CNPq” mais abordadas são a Hidrogeologia, Geomorfologia e Hidrogeografia (Figura 6). Ressalta-se que a maior parte dos projetos em Hidrogeologia são desenvolvidos fora de unidades de conservação, em áreas onde a captação de águas subterrâneas é crucial para o abastecimento público.

Os projetos podem ser categorizados em função do tipo de coordenação: internos (somente pesquisadores do IPA/SEMIL), externos (entidades externas ao IPA) e mistos (pesquisadores de outras instituições + representante IPA) (Figura 7). Dos 35 projetos em andamento, 11 foram submetidos por pesquisadores externos, 14 por pesquisadores internos e 11 por representante IPA + externo (misto). Em relação às instituições externas responsáveis por projetos de pesquisa (total de 17 entidades), verifica-se a predominância de projetos desenvolvidos pelo Instituto de Geociências e a Faculdade de

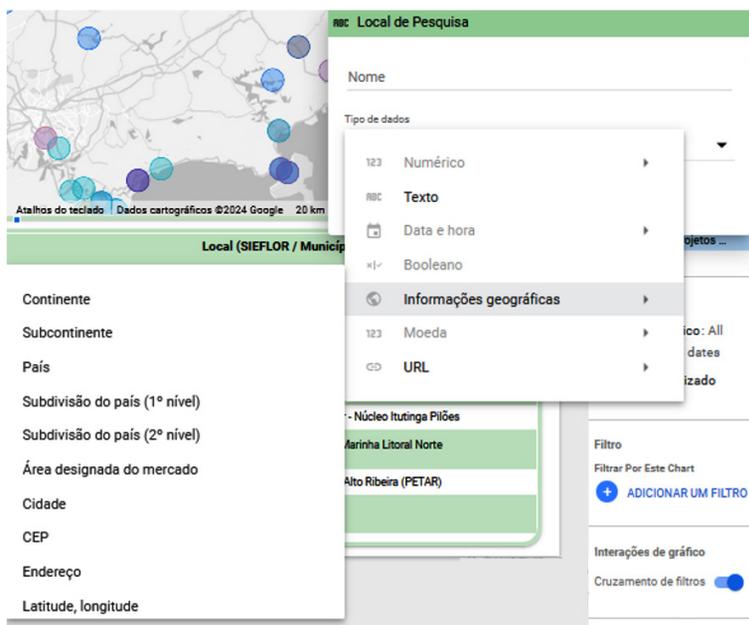


FIGURA 5 – Definição das informações geográficas do mapa. Fonte: Google Looker Studio

FIGURE 5 – Definition of the map’s geographic information. Source: Google Looker Studio.

As ferramentas de visualização permitem a apresentação da distribuição espacial das pesquisas (Figura 8). Esta visualização espacial indica pontualmente a localização da área estudada no projeto de pesquisa, bem como permite distinguir as regiões com maior concentração de estudos. Verifica-se uma concentração expressiva de projetos na porção centro-leste do estado de São Paulo.

O mapa de distribuição geográfica permite uma navegabilidade interativa, onde é possível acessar e filtrar as informações de cada projeto (título, local da pesquisa, responsável, quantidade de projetos em desenvolvimento, entre outras).

Com o término do desenvolvimento do Painel de Dados é possível compartilhá-lo gratuitamente por *e-mail* (somente contas Gmail), link externo, incorporar em site por código/ URL ou *download* em PDF:

4 DISCUSSÃO E CONSIDERAÇÕES FINAIS

A aplicação do Google Looker Studio, conjugada com as ferramentas de ETL, mostrou-se efi-

ciente para a análise e visualização dos dados da Plataforma CadGP. Ressalta-se que, além de serem ferramentas gratuitas, seu uso é acessível e intuitivo, mesmo para um usuário iniciante, podendo ser aplicado em outras bases de dados de projetos de pesquisa.

O processo ETL, que antecede o desenvolvimento do painel de dados, demandou um tempo maior (cerca de 70% do tempo total) para que os dados fossem tratados e carregados, até torná-los confiáveis. As ferramentas de ETL do Excel e Power Query mostraram-se eficientes para o tratamento dos dados.

A maior dificuldade encontrada na etapa de visualização de dados foi atribuída à escolha de mapas no Google Looker Studio (Google Maps). Dentre as opções, escolheu-se o Mapa de Balão pela forma automática que a ferramenta efetua a geolocalização das áreas SIEFLOR e municípios onde os projetos de pesquisa em Geociências estão em andamento. As demais ferramentas de mapas disponíveis no Google Looker Studio (Preenchimento, Calor, Linhas, Conexão e Combinação) não foram ade-

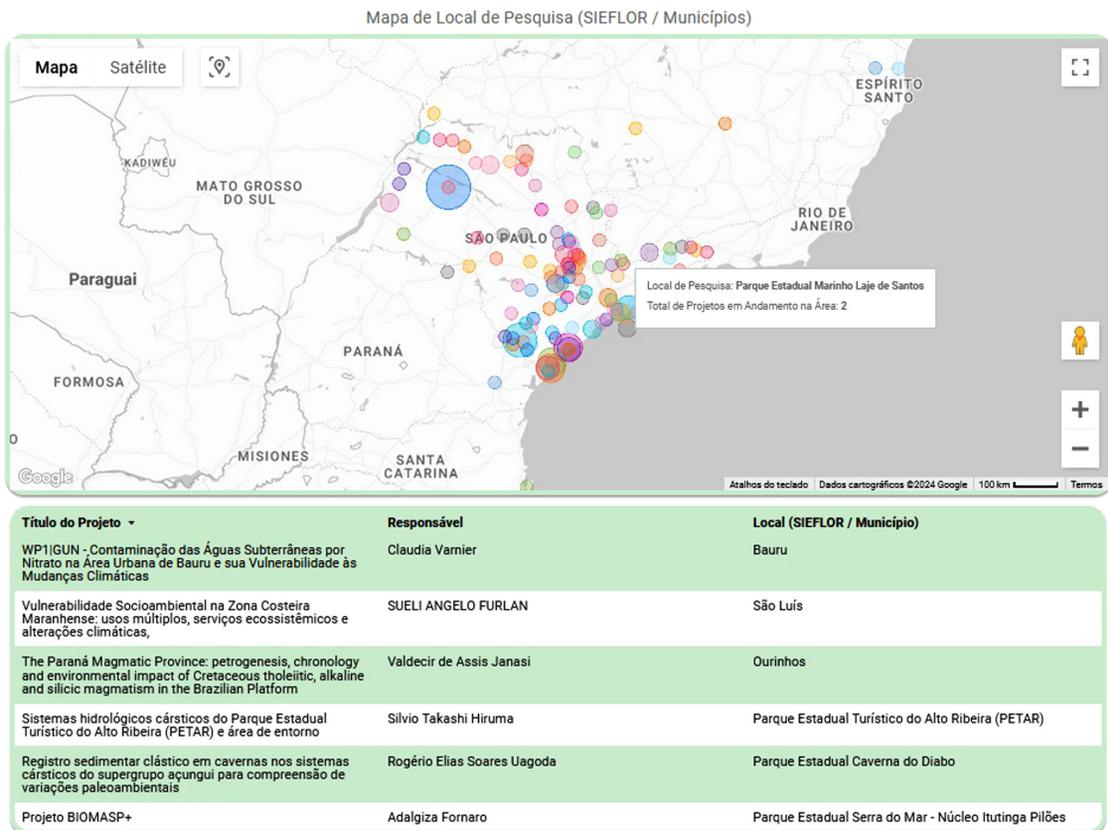


FIGURA 8 – Mapa de distribuição geográfica e lista de consultas de projetos de pesquisa.

FIGURE 8 – Map of geographic distribution and list of research project queries.

quadas para esse fim. Observa-se, no entanto, que o Mapa de Balão não permite a visualização de limites e polígonos complexos, havendo a necessidade de se integrar o Google Looker Studio a ferramentas de geolocalização e geoprocessamento, presentes no ArcGis e QGis.

Por fim, demonstrou-se que a adoção das ferramentas de Business Intelligence (BI) por uma organização do setor público é viável e necessária, pois grandes volumes de dados são gerados diariamente e sua visualização e análise é essencial para a melhor gestão de pesquisa, beneficiando as tomadas de decisões de gestores e para divulgação ao público externo.

5 AGRADECIMENTOS

Aos pareceristas e editores de Derbyana pelos comentários que enriqueceram o trabalho. À diretora do Centro de Gestão de Pesquisas (CGP/CadGP/IPA), Silvana Cristina Muniz de Souza, pela sua colaboração no desenvolvimento desta pesquisa.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ALBUQUERQUE, J.C.M. 2015. *Sistemas de informação e comunicação no setor público – Volume 1*. Florianópolis, Universidade Federal de Santa Catarina - UFSC, Universidade Aberta do Brasil - UAB. Programa Nacional de Formação em Administração Pública-PNAP. Bacharelado em Administração Pública, 150 p. Disponível em <https://canal.cecierj.edu.br/recurso/13826>
- BEZERRA, A.A.; SIEBRA, S.A. 2015. Implantação e Uso de Business Intelligence: Um Relato de Experiência no Grupo Provider. *GESTÃO.Org - Revista Eletrônica de Gestão Organizacional*, Rio de Janeiro, 13(n. esp.): 233-243.
- CHIAVENATO, I. 2010. *Administração de recursos humanos*. Campus, 12ª. ed., Rio de Janeiro.
- CHIAVENATO, I. 2003. *Introdução à teoria geral da administração*. Elsevier Brasil, 7. ed. Rio de Janeiro, 634 p.
- FERREIRA, E. 2020. *Business Intelligence e Data Warehouse: Guia Para Iniciantes*. Kindle Direct Publishing, 187 p.
- GAMA, F.S.G.; ABREU, F. 2008. Desmistificando o conceito de ETL. *Revista de Sistemas de Informação*, 2(2): 1-10.
- GOOGLE CLOUD. *O Google Looker Studio*. Disponível em <https://cloud.google.com/looker-studio/>. Acessado em 21 set. 2024.
- GOVERNO DO ESTADO DE SÃO PAULO. *Decreto nº 51.453, de 29 de dezembro de 2006. Cria o Sistema Estadual de Florestas – SIEFLOR e dá providências correlatas*. Assembleia Legislativa do Estado de São Paulo (ALESP), São Paulo, 29 dez. 2006.
- JUNIOR, F.C.; PARRÃO, J.A.O.; LANGHI, P.J.P. 2014. *Business intelligence – Desmistificando a camada ETL (extração, transformação e carga)*. Cultura Editores, São Paulo, 240 p.
- LINHARES, T. 2023. *Guia para se Tornar um Analista de Dados*. Linhares Corp. 1ª.ed., 54 p.
- STORM, J. 2022. *Design Mind for Data Visualization: Learn to Use Information and Graphic Design Principles to Produce Engaging Data Stories Your Audience Can't Ignore*. Publishing House LLC, 216 p.

Endereço do autor:

Rodrigo Dias Dutra* ([id 0009-0009-8217-2335](https://orcid.org/0009-0009-8217-2335)) – Núcleo de Acompanhamento de Pesquisas Institucionais, Instituto de Pesquisas Ambientais/SEMIL, Rua Joaquim Távora, 822, Vila Mariana, CEP 04015-011, São Paulo, SP, Brasil. *E-mail*: rodrigo.dutra@sp.gov.br

*Autor correspondente

Artigo submetido em 4 de julho de 2024, aceito em 10 de dezembro de 2024.



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License.