Data Science Platform applied to Health in contribution to the Brazilian Unified Health System

PCDaS/ICICT/FIOCRUZ/DEXL/LNCC

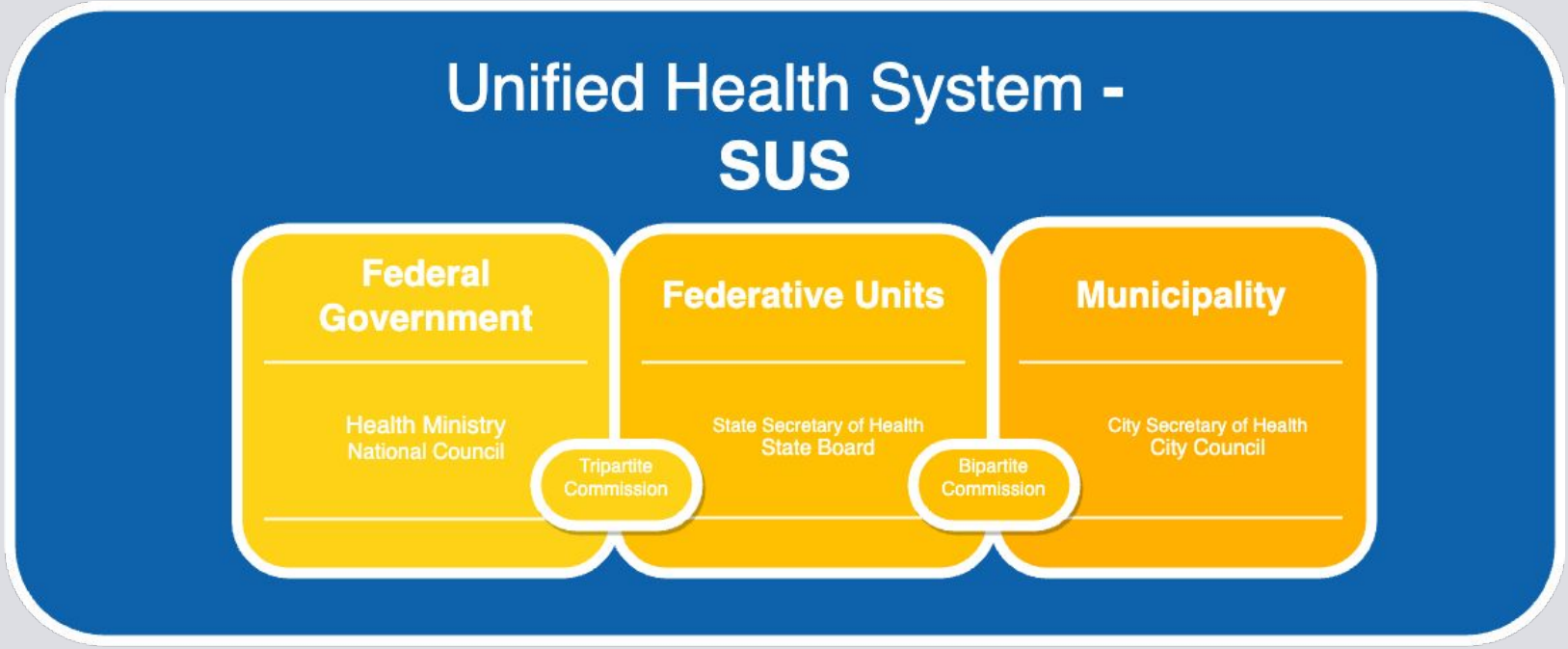
Victor Ribeiro, M.Sc

Agenda

- Brazilian public health
- Health Information Systems **(HIS)**
- Motivation
- Infrastructure and Architecture
- Services
- Data Management
- Success Cases
- Final remarks

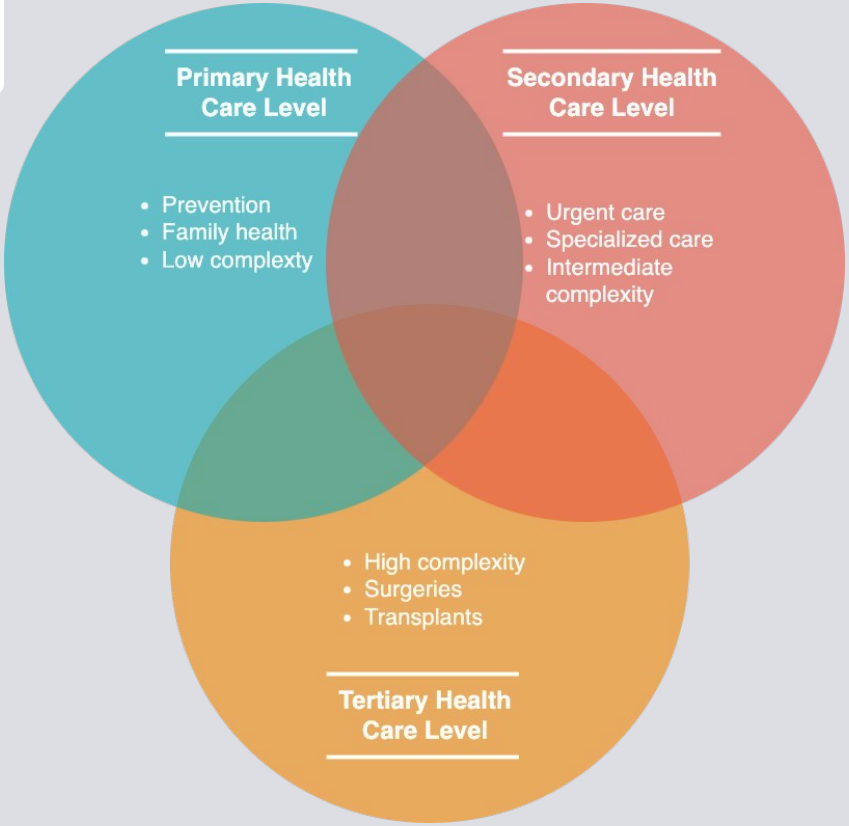
INTRODUCTION

SUS and Datasus



INTRODUCTION

SUS and Datasus



INTRODUCTION

DataSUS and Health Information Systems

SINASC

Sistema de Informação de Nascidos Vivos (1994)

- Based on birth declaration
- Information about **mothers health**, **prenatal care** and **newborn health**
- **Yearly** updated

68,818,788 **entrances**

SIM

Sistema de Informação de Mortalidade (1996)

- Based on death registres
- Information about **where** and **how** *death occurred*.
- **Yearly** updated

24,996,243 **registres**

INTRODUCTION

Brazilian Information Systems

CNES

Cadastro Nacional de Estabelecimentos de Saúde (2005)

- Every **health facilities** within Brazil

- **1070 attributes for each health facility:**

  Infrastructure, *experts*, *professional teams*, *equipments* information, *beds* and *services*.

- Monthly **updated**

467,118 *facilities until* **2020.**

Motivation

- Multiple ***autonomous*** information systems
- Legacy systems
- ***Frequent*** Schema changes
- *Large data Volume*

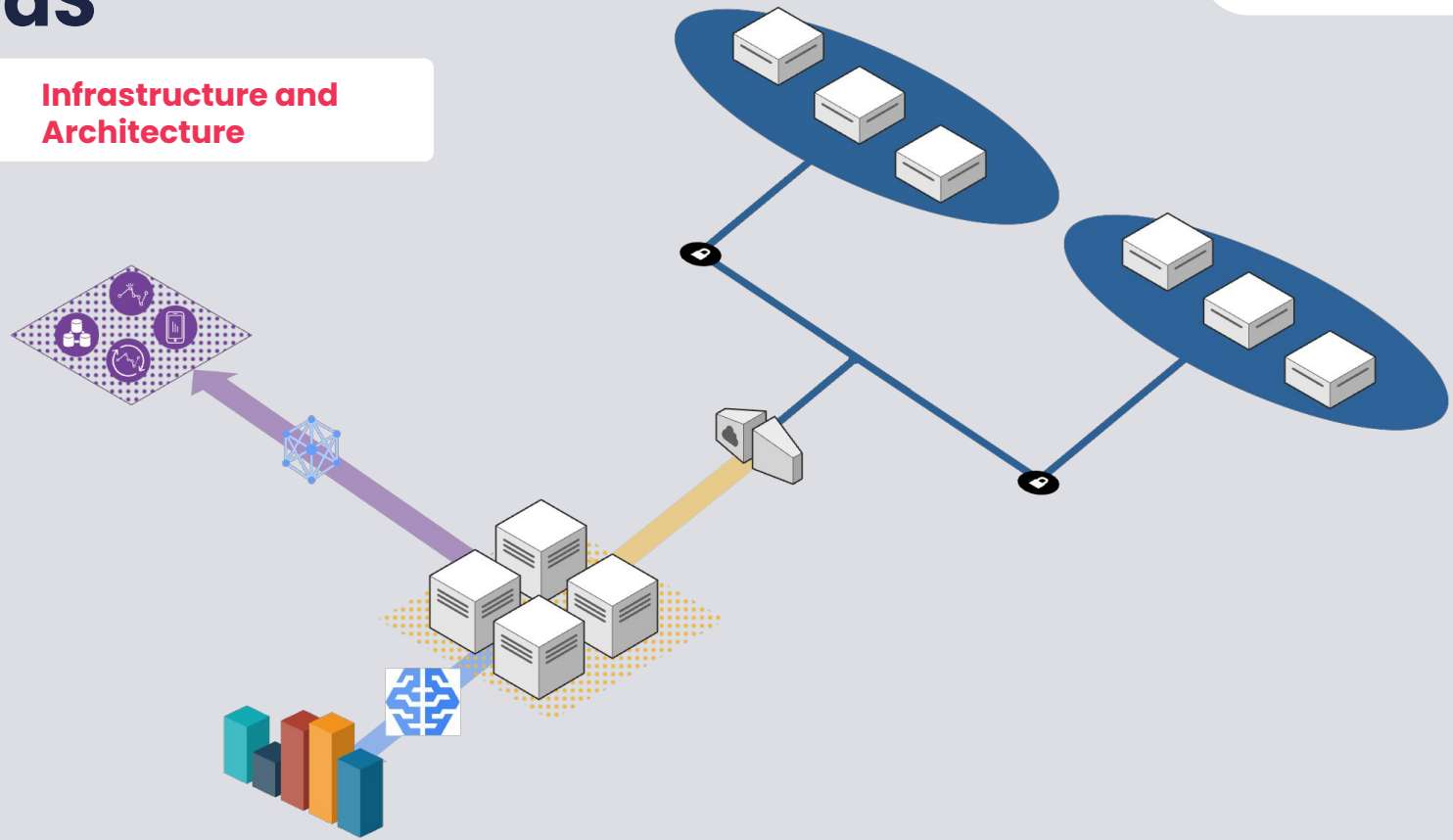
Objective

To provide a public health data ecosystem through

- Public Health data enrichment
- Research applications' support
- Structured datasets
- Notebook based infrastructure for researchers

PCDaS

Infrastructure and Architecture



PCDaS

Infrastructure and Architecture

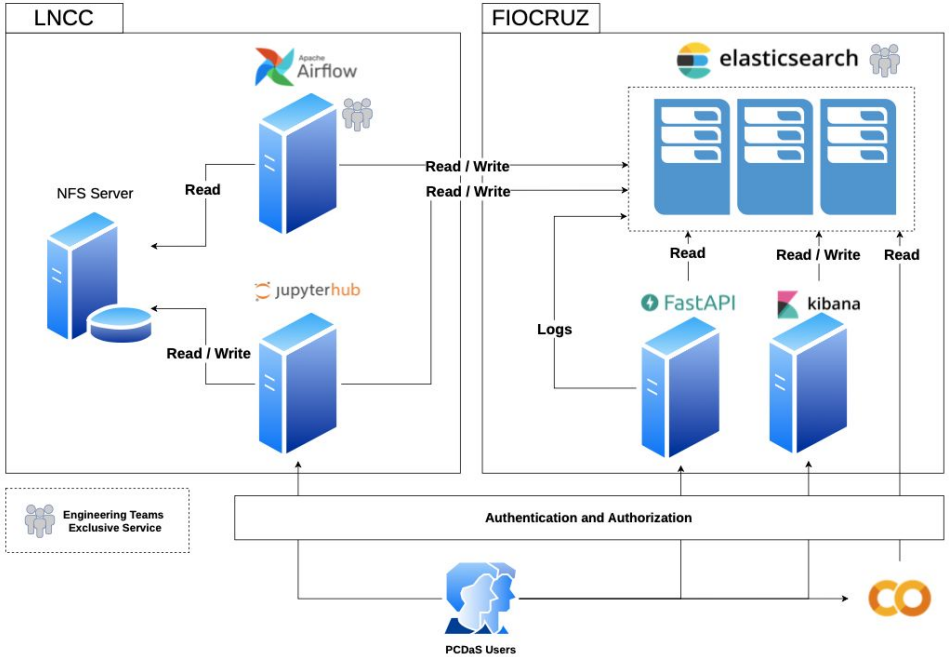


Figure 1: *PCDaS* Infrastructure and Architecture

FIOCRUZ Fundação Oswaldo Cruz DEXL DATA EXTREME LAB LNCC

PCDaS

Services

Table 2: Services by user category

Service	Basic Users	Academic Users	Partner Users
Support for data mining and predictive analysis through tutorials on Google Colab	✓	✓	✓
Access to the community of researchers, data scientists, and *PCDaS* users	✓	✓	✓
Secure and dedicated remote access to *PCDaS* via JupyterHub (Python or R)		✓	✓
Promotion of academic and partner projects on the *PCDaS* website		✓	✓
Basic training in Python or R			✓
Training of the research team for the use of *PCDaS*			✓
Extraction, Transformation, and Loading (ETL) of databases relevant to the research team's interests			✓

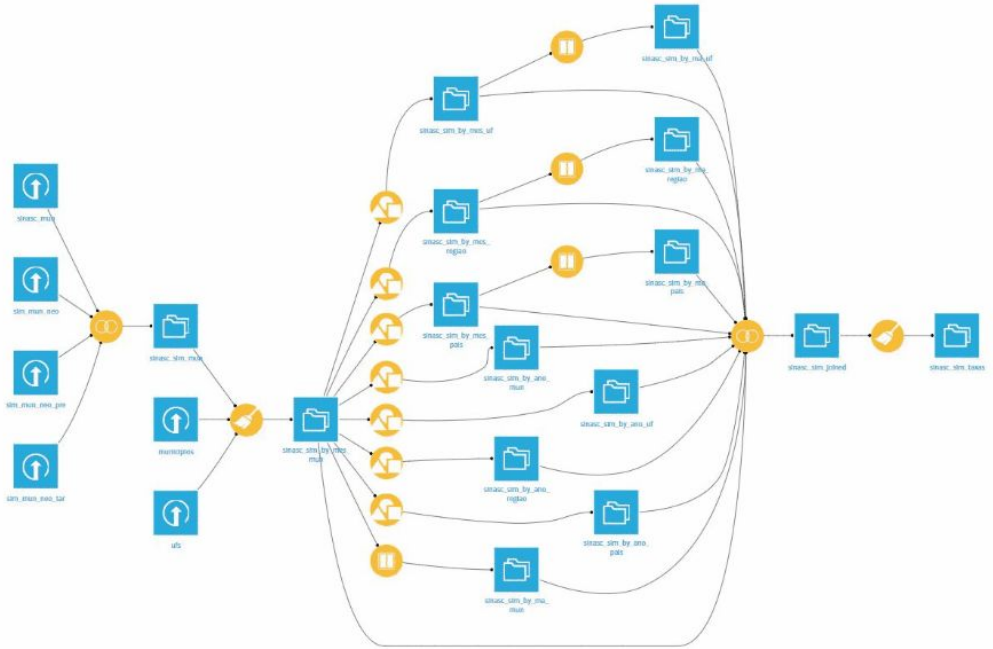
PCDaS

Data Management

Extract

Transform

Load



Data Management

PCDaS Plataforma de Ciência de Dados aplicada à Saúde

Data Transparency

- User centric process
- Documentation about data lifecycle: ***ETL***; enrichment; variables **dictionary**

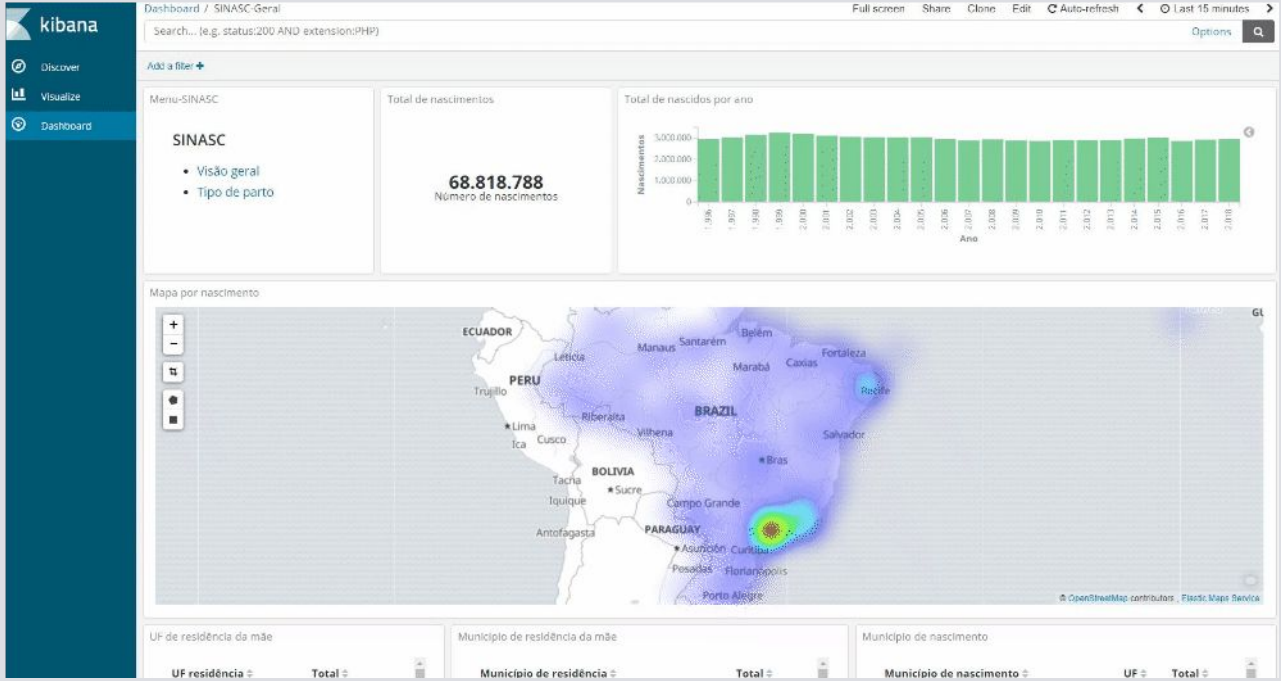
Data Protection

- Following Brazilian guidelines on research using human data, we adopt a Data Use Commitment Term (**TECUD**)

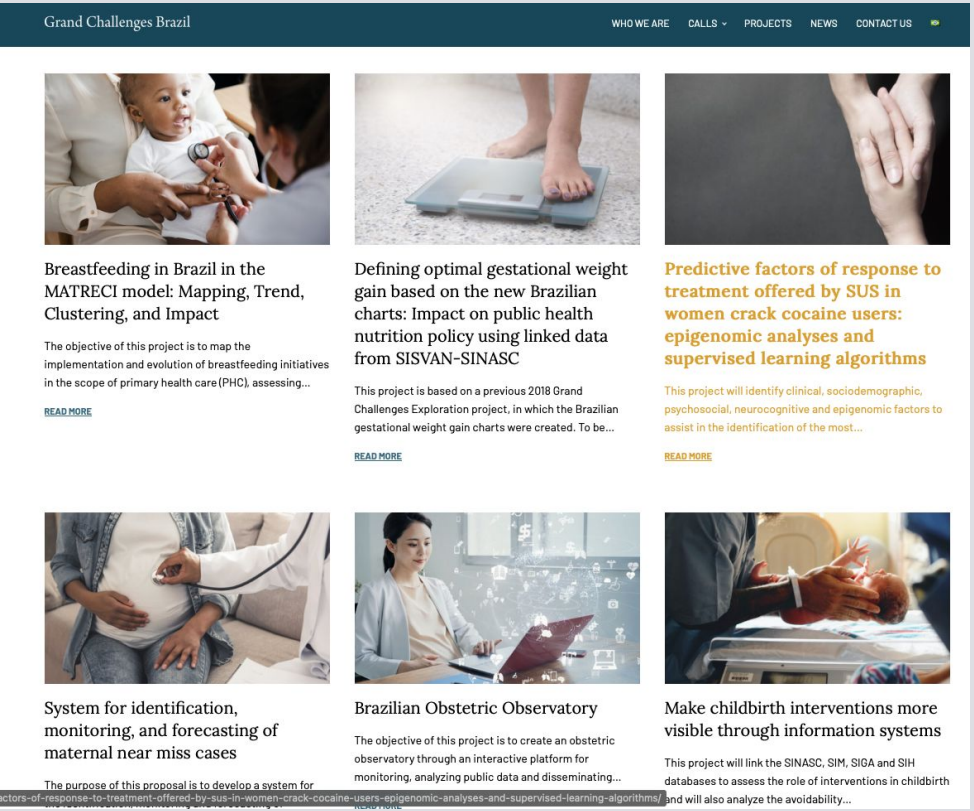
Data Fairness

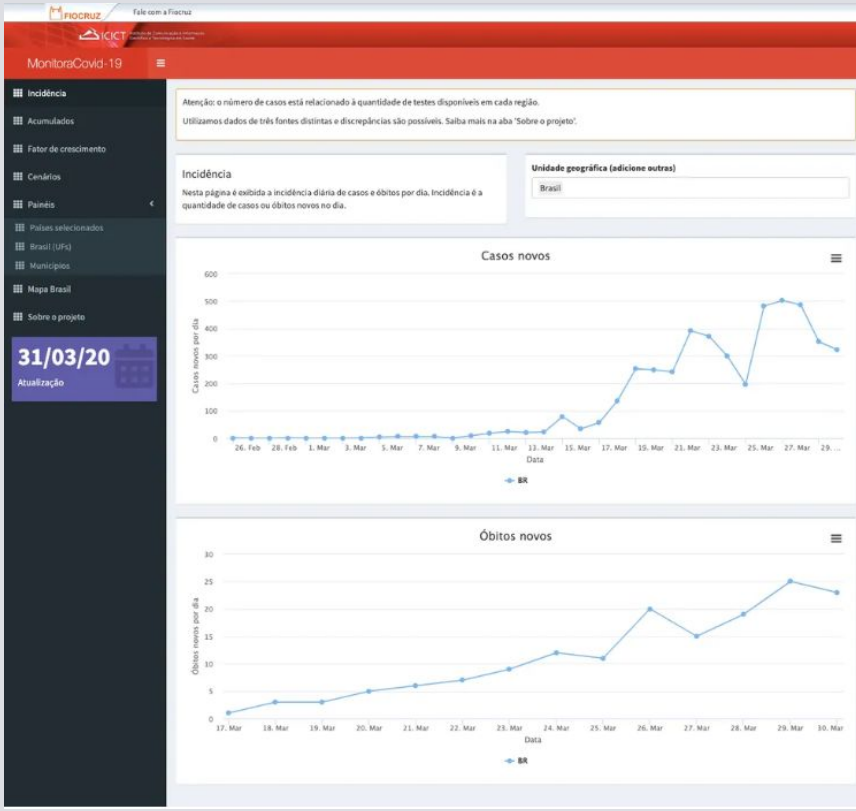
- **F**indable, **A**ccessible, **I**nteroperable and **R**eusable
- Shared **data provenance** and enrichment process.
- Data available vai open formats and ***API***.

Data Analysis and Visualization



SUCCESS STORIES





SUCCESS STORIES

PROJETO DE PESQUISA E DESENVOLVIMENTO TECNOLÓGICO

Pesquisa Nacional de Saúde (PNS)

A Pesquisa Nacional de Saúde - PNS é uma pesquisa de base domiciliar realizada em 2013 e 2019 com o objetivo de produzir dados em âmbito nacional sobre a situação de saúde e...

ibge | inquérito | pns

MAIS SOBRE O PROJETO →

PROJETO DE PESQUISA E DESENVOLVIMENTO TECNOLÓGICO

Ciência de Dados aplicada à Ouvidoria-Geral do SUS – OuvSUS

A Ouvidoria-Geral do SUS (OuvSUS) tem como objetivo principal garantir e ampliar o acesso do cidadão na busca efetiva de seus direitos, atuando como ferramenta de gestão e...

fiocruz | instituto aggeu magalhães | ministério da saúde | ouvidoria

MAIS SOBRE O PROJETO →

PROJETO DE PESQUISA E DESENVOLVIMENTO TECNOLÓGICO

Acessibilidade Geográfica a Serviços de Saúde no Brasil

O projeto visa criar cenários a respeito do deslocamento populacional em busca de atendimento avaliando a relação entre a distância e o tempo de viagem até a unidade...

acessibilidade

MAIS SOBRE O PROJETO →

Final remarks

- PCDaS provides high availability, clean and organized datasets **for public health researchers.**

- PCDaS team **trained** hundreds of people public health researchers, **strengthening** their *data-science* literacy.

- As future works we intend to: (i) expand partnerships; (ii) provide data-science training; (iii) add more datasets; and (iv) continuously improve our software ecosystem

Obrigado!

PCDaS Plataforma de Ciência de Dados aplicada à Saúde

Plataforma: **bigdata.icict.fiocruz.br**
E-mail: bigdata.saude@icict.fiocruz.br

PCDaS Plataforma de Ciência de Dados aplicada à Saúde

ICICT Instituto de Comunicação e Informação Científica e Tecnológica em Saúde

FIOCRUZ Fundação Oswaldo Cruz

DEXL DATA EXTREME LAB

CAPES

LNCC

BILL & MELINDA GATES foundation

FioTec

CEFET/RJ

SUS MINISTÉRIO DA SAÚDE

MINISTÉRIO DA EDUCAÇÃO

GOVERNO FEDERAL BRASIL UNIÃO E RECONSTRUÇÃO