

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

SUS and Datasus

Unified Health System - SUS

**Federal
Government**

Health Ministry
National Council

Federative Units

State Secretary of Health
State Board

Municipality

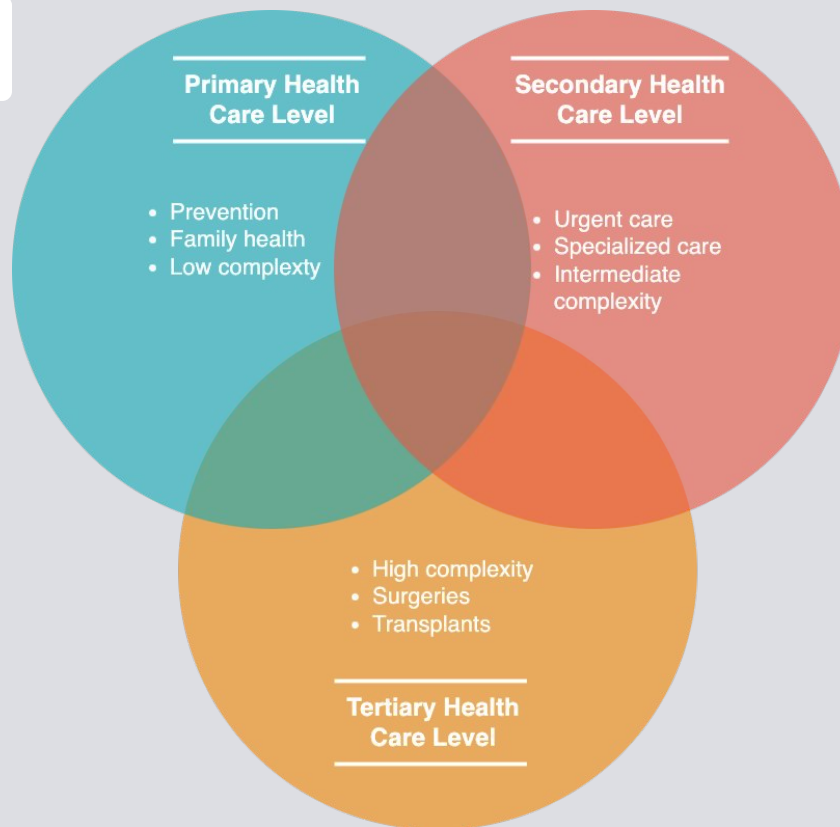
City Secretary of Health
City Council

Tripartite
Commission

Bipartite
Commission

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 ⁵

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

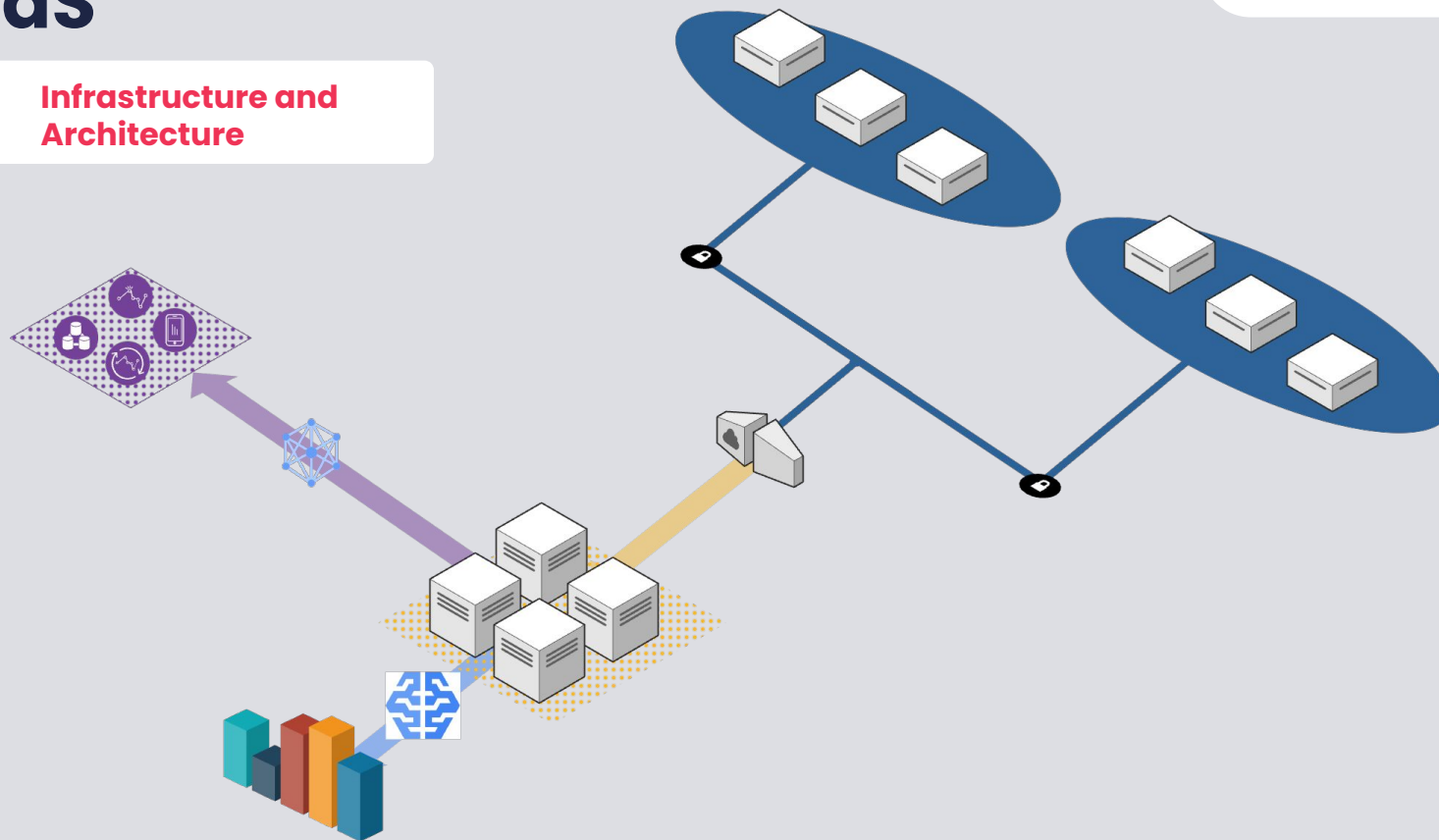
- Multiple **autonomous** information systems
- Legacy systems
- **Frequent** Schema changes
- *Large data Volume*

Objective

To provide a public health data ecosystem through

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

Infrastructure and Architecture



Infrastructure and Architecture

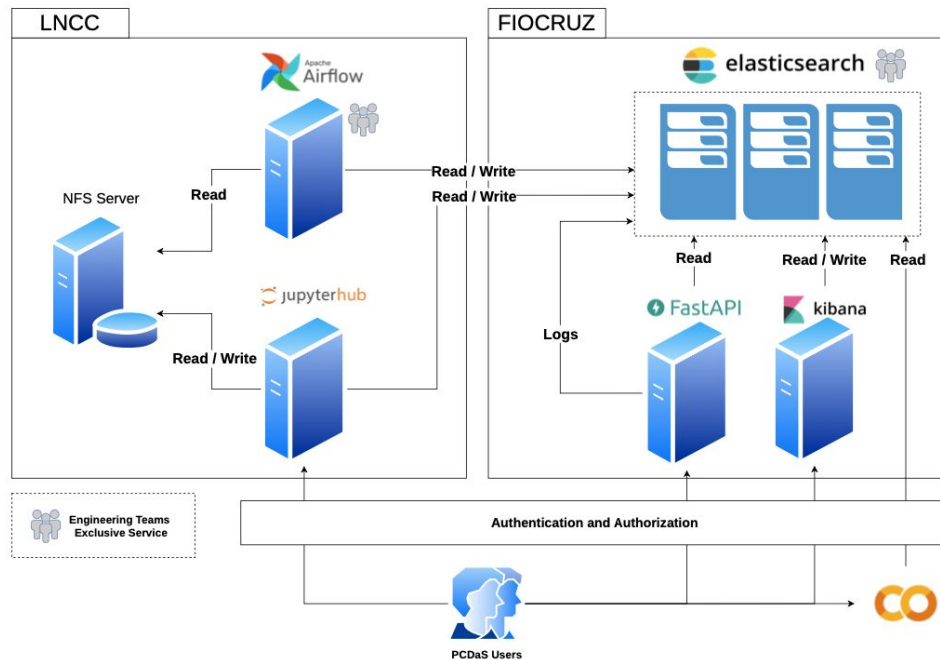


Figure 1: PCDaS Infrastructure and Architecture

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 <i>PCDaS</i> users	✓	✓	✓
Secure and dedicated remote access to <i>PCDaS</i> via JupyterHub (Python or R)		✓	✓
Promotion of academic and partner projects on the <i>PCDaS</i> website		✓	✓
Basic training in Python or R			✓
Training of the research team for the use of <i>PCDaS</i>			✓
Extraction, Transformation, and Loading (ETL) of databases relevant to the research team's interests			✓

Data Management

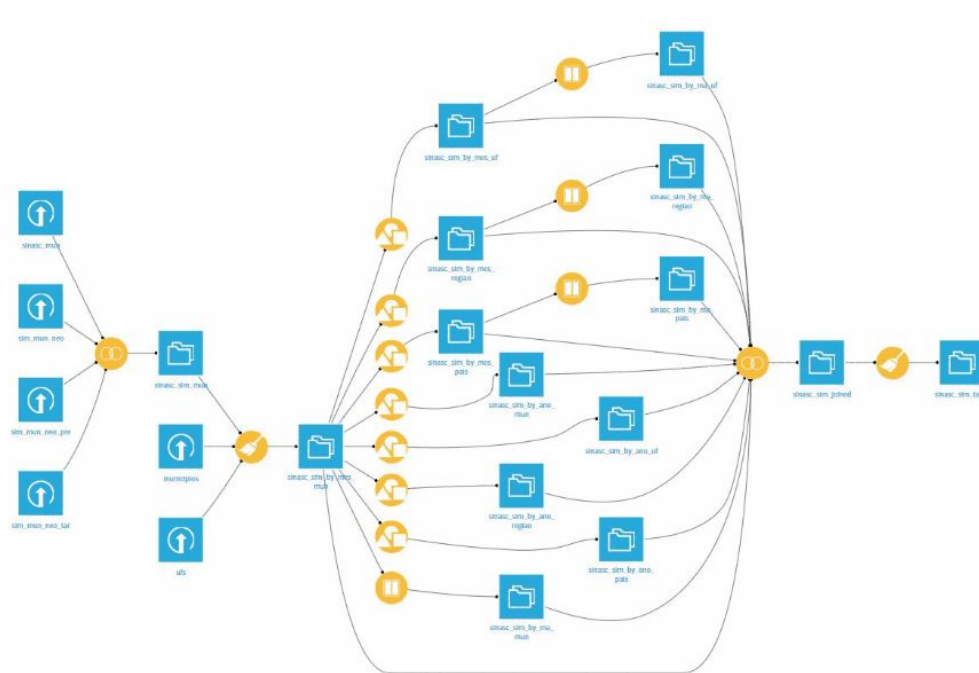
Extract



Transform



Load



Data Management

Data Transparency

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

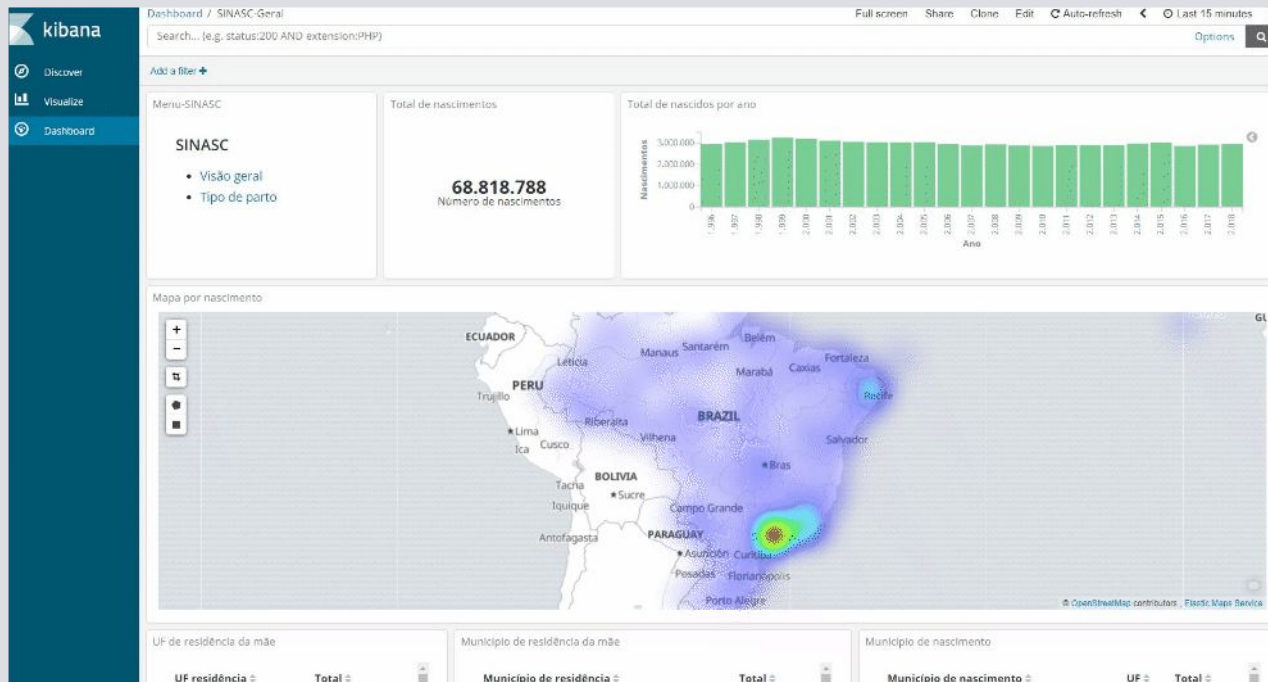
Data Protection

- 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 via open formats and **API**.

Data Analysis and Visualization



SUCCESS STORIES

Grand Challenges Brazil

WHO WE ARE CALLS PROJECTS NEWS CONTACT US



Breastfeeding in Brazil in the MATRECI model: Mapping, Trend, Clustering, and Impact

The objective of this project is to map the implementation and evolution of breastfeeding initiatives in the scope of primary health care (PHC), assessing...

[READ MORE](#)



Defining optimal gestational weight gain based on the new Brazilian charts: Impact on public health nutrition policy using linked data from SISVAN-SINASC

This project is based on a previous 2018 Grand Challenges Exploration project, in which the Brazilian gestational weight gain charts were created. To be...

[READ MORE](#)



Predictive factors of response to treatment offered by SUS in women crack cocaine users: epigenomic analyses and supervised learning algorithms

This project will identify clinical, sociodemographic, psychosocial, neurocognitive and epigenomic factors to assist in the identification of the most...

[READ MORE](#)



System for identification, monitoring, and forecasting of maternal near miss cases

The purpose of this proposal is to develop a system for



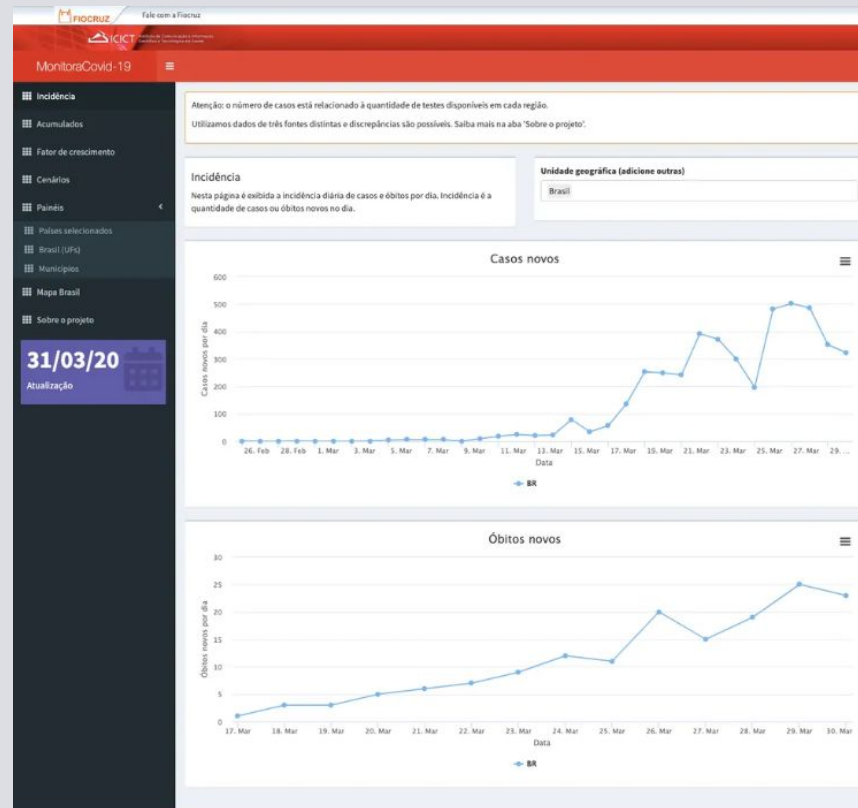
Brazilian Obstetric Observatory

The objective of this project is to create an obstetric observatory through an interactive platform for monitoring, analyzing public data and disseminating...





Make childbirth interventions more visible through information systems

This project will link the SINASC, SIM, SIGA and SIH databases to assess the role of interventions in childbirth and will also analyze the avoidability...



SUCCESS STORIES

PESQUISA NACIONAL DE SAÚDE



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!

PCD  **S** Plataforma de Ciência de Dados aplicada à Saúde

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

PCD  **S** 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