

# **POTENCIALIDADES E DESAFIOS DO CENTRO DE INTEGRAÇÃO DE DADOS E CONHECIMENTOS PARA SAÚDE - CIDACS**

Maria Yury Travassos Ichihara  
Maurício Lima Barreto  
Cidacs/IGM/FIOCRUZ

Encontros Pré-ConfOA 2017-FIOCRUZ  
20 de setembro de 2017

# DESAFIOS NA PRODUÇÃO DO CONHECIMENTO CIENTÍFICO EM CIÊNCIAS DA SAÚDE

- 1- A necessidade de produção continuada de evidências
- 2- Reduccionismo dominante
- 3- Dificuldades na Integração de diferentes bases de dados e dos diferentes níveis de Conhecimento
- 4- Dilemas entre “testar hipóteses” (hypothesis-driven research) ou “explorar dados” (exploratory research)?

# **INTEGRAÇÃO DE GRANDES BASES DE DADOS PARA PRODUZIR CONHECIMENTO**

Implica em produzir ciência baseada em dados, de forma multidisciplinar, colaborativa e focada em problemas específicos

Gera oportunidades extraordinárias para a pesquisa científica e para a produção de conhecimentos na área da saúde e das políticas públicas

Contribui para constituir um sistema de saúde em “aprendizado contínuo”, produzindo evidências, propondo soluções aos problemas de saúde e integrando a pesquisa, as políticas e a prática nos serviços de saúde

Possibilita o desenvolvimento de métodos de análises

# DESAFIOS NO USO DE GRANDES BASES DE DADOS

- Acesso a dados (proteção por razões comerciais ou culturais, ou relacionadas à privacidade pessoal)
- Fragmentação de dados (múltiplos bancos de dados, múltiplos proprietários / stakeholders)
- Heterogeneidade de dados (precisão, formato)
- Manipulação de dados (acesso a dados, gestão de dados, qualidade de dados, compartilhamento de dados);
- Privacidade e integridade dos dados (prevenção de ataques);

## Sistema Único de Saúde - SUS

SUS universal, equânime e integral

- Cria Demandas em C&T&I
- Enfatiza a importância combinada das tecnologias e dos determinantes sociais na superação de...



## Produção de evidências

Effect of a conditional cash transfer programme on childhood mortality: a nationwide analysis of Brazilian municipalities



Davide Rasella, Rosana Aquino, Carlos A T Santos, Rômulo Paes-Sousa, Mauricio L Barreto

### Summary

**Background** In the past 15 years, Brazil has undergone notable social and public health changes, including a large reduction in child mortality. The *Bolsa Família* Programme (BFP) is a widespread conditional cash transfer programme, launched in 2003, which transfers cash to poor households (maximum income US\$70 per person a month) when they comply with conditions related to health and education. Transfers range from \$18 to \$175 per month, depending on the income and composition of the family. We aimed to assess the effect of the BFP on deaths of children younger than 5 years (under-5), overall and resulting from specific causes associated with poverty: malnutrition, diarrhoea, and lower respiratory infections.

**Methods** The study had a mixed ecological design. It covered the period from 2004–09 and included 2853 (of 5565) municipalities with death and livebirth statistics of adequate quality. We used government sources to calculate all-cause under-5 mortality rates and under-5 mortality rates for selected causes. BFP coverage was classified as low (0–0.17–1%), intermediate (17–2–32–0%), high (>32–0%), or consolidated (>32–0% and target population coverage ≥100% for at least 4 years). We did multivariable regression analyses of panel data with fixed-effects negative binomial models, adjusted for relevant social and economic covariates, and for the effect of the largest primary health-care scheme in the country (Family Health Programme).

**Findings** Under-5 mortality rate, overall and resulting from poverty-related causes, decreased as BFP coverage increased. The rate ratios (RR) for the effect of the BFP on overall under-5 mortality rate were 0.94 (95% CI 0.92–0.96) for intermediate coverage, 0.88 (0.85–0.91) for high coverage, and 0.83 (0.79–0.88) for consolidated coverage. The effect of consolidated BFP coverage was highest on under-5 mortality resulting from malnutrition (RR 0.35; 95% CI 0.24–0.50) and diarrhoea (0.47; 0.37–0.61).

**Interpretation** A conditional cash transfer programme can greatly contribute to a decrease in childhood mortality overall, and in particular for deaths attributable to poverty-related causes such as malnutrition and diarrhoea, in a large middle-income country such as Brazil.

**Funding** National Institutes of Science and Technology Programme, Ministry of Science and Technology, and Council for Scientific and Technological Development Programme (CNPq), Brazil.

Published Online  
May 15, 2013  
[http://dx.doi.org/10.1016/S0140-6736\(13\)60715-1](http://dx.doi.org/10.1016/S0140-6736(13)60715-1)  
See Online/Comment  
[http://dx.doi.org/10.1016/S0140-6736\(13\)60715-1](http://dx.doi.org/10.1016/S0140-6736(13)60715-1)  
Instituto de Saúde Coletiva, Federal University of Bahia, Salvador, Bahia, Brazil (D Rasella PhD, R Aquino MD, C A T Santos PhD, Prof M L Barreto MD); Department of Exact Sciences, State University of Feira de Santana, Feira de Santana, Bahia, Brazil (C A T Santos); Institute of Development Studies, University of Sussex, Brighton, UK (R Paes-Sousa MD); and Clínica, Tecnologia e Inovação em Saúde, INCT-CITECS, Salvador, Bahia, Brazil (Prof M L Barreto).  
Correspondence to: Prof Mauricio Barreto, Instituto de Saúde Coletiva, Universidade Federal de Bahia, Caixa 41110-040, Salvador, Bahia, Brazil ([mauricio@ufba.br](mailto:mauricio@ufba.br))

OPEN ACCESS Freely available online

PLOS ONE

## Effect of the Brazilian Conditional Cash Transfer and Primary Health Care Programs on the New Case Detection Rate of Leprosy

Jolida Silva Nery<sup>1\*</sup>, Susan Martins Pereira<sup>1</sup>, Davide Rasella<sup>1</sup>, Maria Lúcia Fernandes Penna<sup>2</sup>, Rosana Aquino<sup>1</sup>, Laura Cunha Rodrigues<sup>3</sup>, Mauricio Lima Barreto<sup>3</sup>, Gerson Oliveira Penna<sup>4</sup>

<sup>1</sup> Federal University of Bahia, Institute of Collective Health, Salvador, Bahia, Brazil, <sup>2</sup> Fluminense Federal University, Health Sciences Center, Institute of Community Health, Fluminense, Rio de Janeiro, Brazil, <sup>3</sup> London School of Hygiene and Tropical Medicine, Department Infectious and Tropical Diseases, London, United Kingdom, <sup>4</sup> University of Brasília, Tropical Medicine Center, Brasília, Distrito Federal, Brazil

### Abstract

**Background:** Social determinants can affect the transmission of leprosy and its progression to disease. Not much is known about the effectiveness of welfare and primary health care policies on the reduction of leprosy occurrence. The aim of this study is to evaluate the impact of the Brazilian cash transfer (Bolsa Família Program BFP) and primary health care (Family Health Program FHP) programs on new case detection rate of leprosy.

**Methodology/Principal Findings:** We conducted the study with a mixed ecological design, a combination of an ecological multiple-group and time-trend design in the period 2004–2011 with the Brazilian municipalities as unit of analysis. The main determinants variables were the BFP and FHP coverage at the municipal level and the outcome was new case detection rate of leprosy. Leprosy new cases, BFP and FHP coverage, population and other relevant socio-demographic covariates were obtained from national databases. We used fixed effects negative binomial models for panel data adjusted for relevant socio-demographic covariates. A total of 1,358 municipalities were included in the analysis. In the studied period, while the municipal coverage of BFP and FHP increased, the new case detection rate of leprosy decreased. Leprosy new case detection rate was significantly reduced in municipalities with consolidated BFP coverage (Risk Ratio 0.79; 95% CI = 0.74–0.83) and significantly increased in municipalities with FHP coverage in the medium (72–95%) (Risk Ratio 1.05; 95% CI = 1.02–1.09) and higher coverage tertiles (>95%) (Risk Ratio 1.12; 95% CI = 1.08–1.17).

**Conclusions:** At the same time the Family Health Program had been effective in increasing the new case detection rate of leprosy in Brazil, the Bolsa Família Program was associated with a reduction of the new case detection rate of leprosy that we propose reflects a reduction in leprosy incidence.

Citation: Nery JS, Pereira SM, Rasella D, Penna MLF, Aquino R, et al. (2014) Effect of the Brazilian Conditional Cash Transfer and Primary Health Care Programs on the New Case Detection Rate of Leprosy. PLoS ONE 9(5): e96112. doi:10.1371/journal.pone.0096112



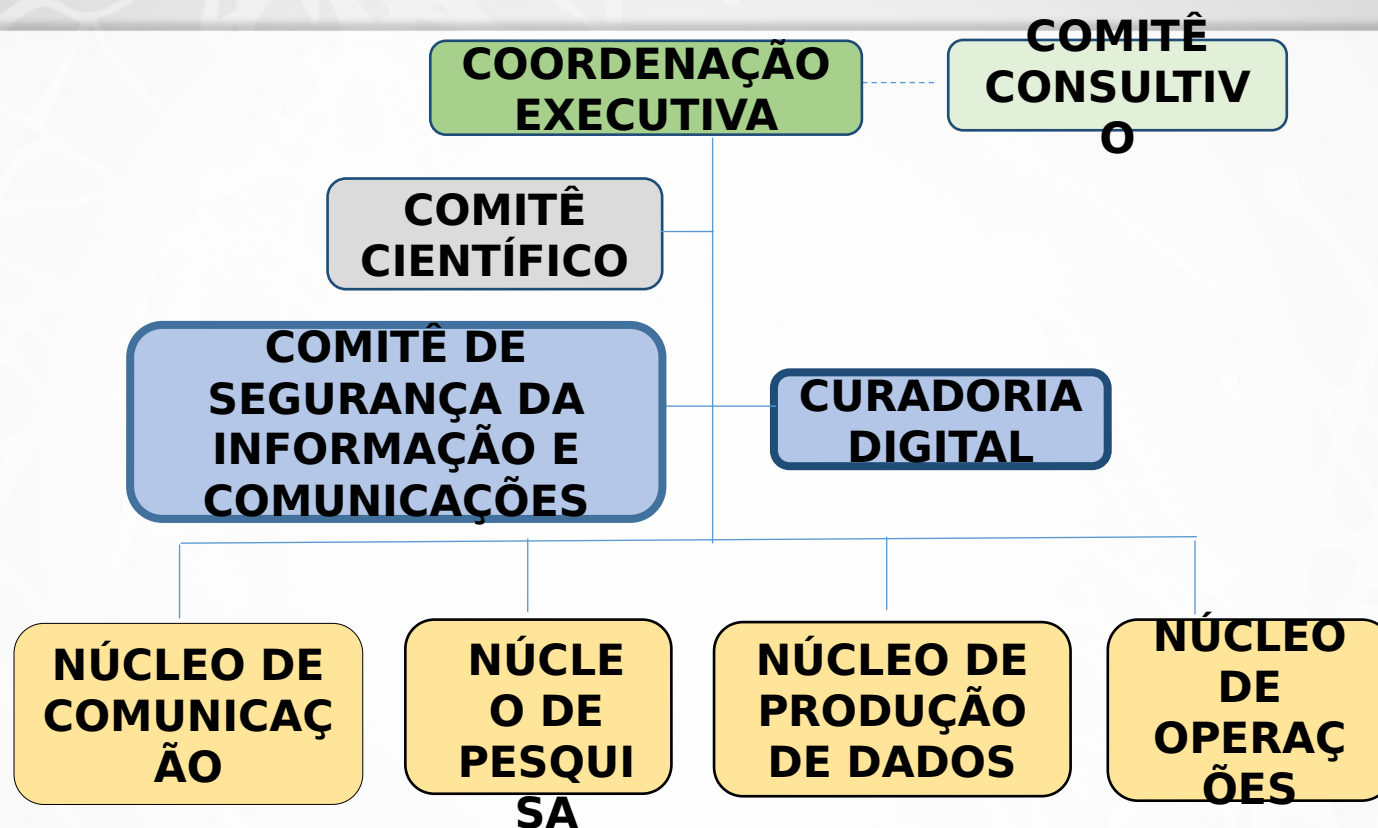
Realizar estudos e pesquisas interdisciplinares, desenvolver novas metodologias científicas e promover capacitação profissional, mediante a integração de grandes bases de dados (big data) e de conhecimentos, recorrendo a recursos computacionais de alto desempenho em ambiente seguro, com a finalidade de ampliar o campo de atuação das ciências da saúde e de apoiar tomadas de decisões em políticas públicas, em benefício da sociedade.

- Inaugurado em 07 de dezembro de 2016, vinculado ao Instituto Gonçalo Moniz/FIOCRUZ

- Atividades estratégicas de pesquisa, inovação, informação e aperfeiçoamento do SUS



- Governança compatível com o acesso, privacidade e segurança no tratamento de dados pessoais e com os conceitos de “Ciência Aberta”
- Observar o marco regulatório e as atitudes públicas relacionados ao acesso, armazenamento, preservação e tratamento de dados pessoais
- Concordância com os Padrões Éticos vigentes
- Plano Estratégico de Desenvolvimento e Sustentabilidade dialogando com os parceiros (gestores, pesquisadores, financiadores, apoiadores)



Art 196 CF/ Lei 8080  
LEI DE ACESSO A INFORMAÇÃO  
12527 (LAI)  
LEI 7845  
PORTARIA 466/CONEP  
PI 5026

POLÍTICA DE SEGURANÇA DA  
INFORMAÇÃO  
POLÍTICA DE CONTROLE DE ACESSO  
BOAS PRÁTICAS DE SEGURANÇA  
POLÍTICA DE DADOS ABERTOS



Ambiente com alto nível de segurança – física e virtual- e com grande recurso computacional

Uso de bases de dados identificadas

Capacidade para receber, armazenar, preservar, atualizar e manejar dados com alta segurança e privacidade

Desenvolver métodos para realizar vinculação de diferentes bases de dados

Produzir *datasets* desidentificados/anonimizados para análises específicas a serem usadas por

# AMBIENTE COMPUTACIONAL E SEGURANÇA

## 1) Centro de Dados - Centro de Computação de Alto Desempenho do SENAI -CIMATEC:

Abriga Cluster computacional adquirido pela FIOCRUZ- CT-INFRA-FINEP

Bases de dados identificadas

## 2) Escritório (incluindo sala segura) localizado no TECNOCENTRO :

- Uso para avaliação de qualidade de dados, vinculação, estimativa de precisão, extração de dados, desidentificação de dados e produção de datamarts

- Cluster adquirido com recursos do Bill Mellinda Gates

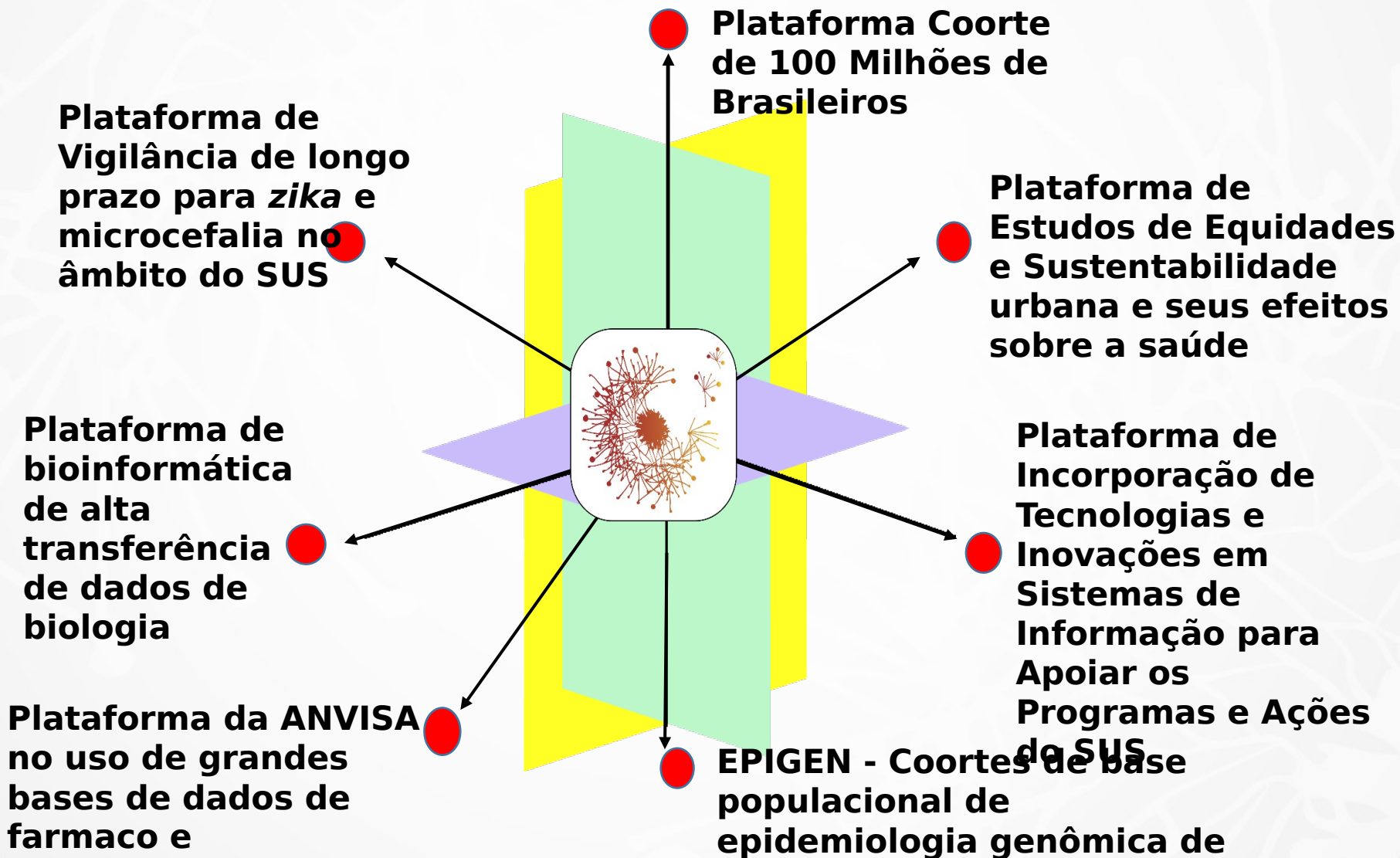


O usuário autorizado acessa dados no cluster SGI



secure private connection  
(peer-to-peer access)

# PLATAFORMAS DE PESQUISA



## Dados pessoais identificados

S

abrigados/  
produzidos



Produzido pelo  
CIDAC

**Baseline da Coorte  
100 M**

População selecionada a partir do Cadastro Único com o objetivo de realizar estudos longitudinais (2006 - 2015)

Dataset: **± 114 milhões indivíduos**

Cadastro  
Conhecer  
para incluir  
Único

**Cadastro  
Único**

Registros atualizados de indivíduos que se candidatam a programas sociais, com informações sobre condições sócio-econômica e de domicílio (2007 a 2015)

BRASIL  
PROGRAMA  
Bolsa Família

**Bolsa Família**

Registros atualizados de pagamentos do Programa condicional de transferência de renda (2004-2015: **± 64 milhões de famílias**)

beneficiárias)

Minha Casa  
Minha Vida  
GOVERNO FEDERAL

**Minha Casa e Minha  
Vida**

Registros atualizados de indivíduos beneficiados com unidades residenciais segundo a renda ( **± 2.3 milhões de unidades**)

Programa  
ÁGUA  
PARA TODOS

**Água para Todos**

Fornecer reservatórios impermeáveis para 1 milhão de famílias que vivem na região semi-árida do Brasil, garantindo o acesso a água de

**Óbitos (SIM)**

Informações atualizadas de óbitos obtidos a partir da Declaração de óbitos (2000-2015: **± 18 million óbitos**)

**Nascidos vivos (SINASC)**

Informações atualizadas de nascidos vivos obtidos a partir da Declaração de Nascidos Vivos (2000 - 2015: **± 40**)

**SINAN**

Informações atualizadas sobre 52 doenças de notificação obrigatória (1999 - 2016: **± 8 milhões de indivíduos notificados**)

**Hospitalização (SIH)**

Informações atualizadas sobre admissões hospitalares financiadas pelo SUS

**Vigilância Nutricional e**

**alimentar** atualizados dos registros para avaliação nutricional e de consumo alimentar no âmbito da



## CIDACS Arquitetura de Bigdata

### Portal de Dados

Requisição  
de Pesquisa  
em Saúde

Busca de  
datasets

Ferramentas de  
Analítica de  
dados

Segurança

### Curadoria

Metadados

Catálogo  
de  
Datasets

Vocabulário

Privacidade

### Produção de Dados

Ingestão

Tratamento e  
Processamento

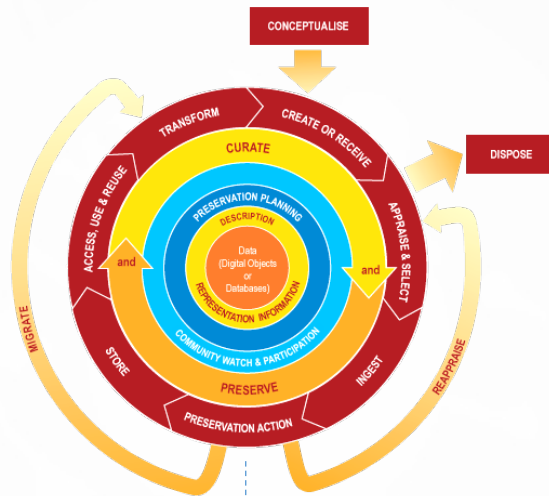
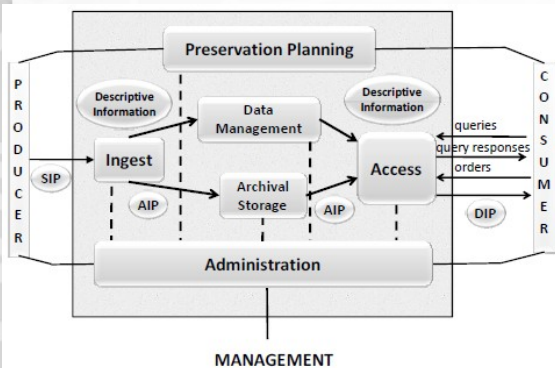
Vinculação  
(Linkage)

Extração

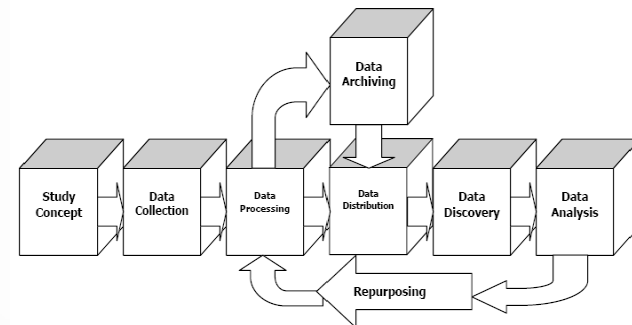
Ética



## Padrões Internacionais em uso / Propostas



**ver. 2.x Codebook: ± 350 elements**  
**ver 3.x Lifecycle: ± 1,100 elements**



revisão histórica

OAIS Model → ISO standard

Digital Curation Centre (DCC) Lifecycle Model

Data Documentation Initiative (DDI)

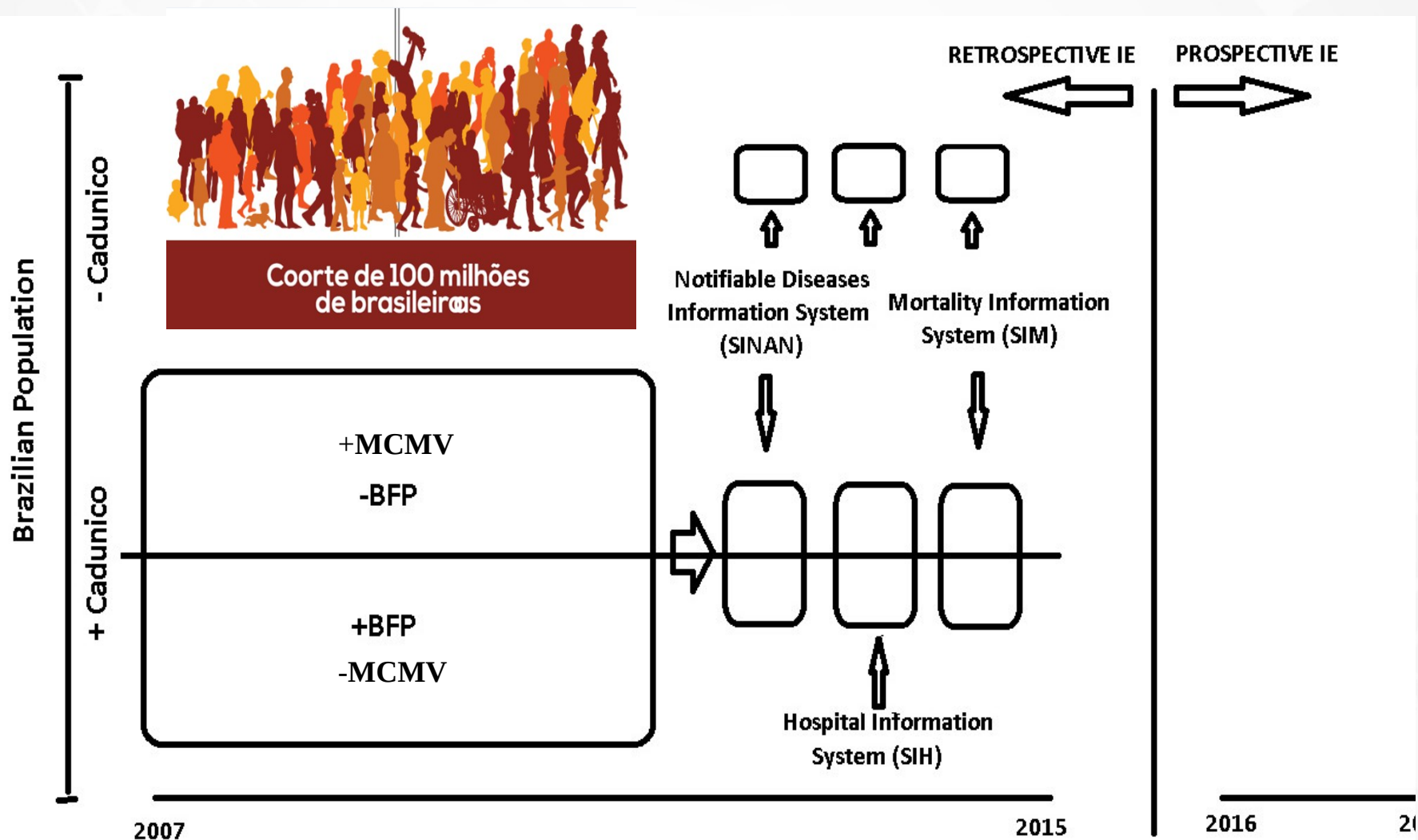


**cidacs**  
Centro de Integração de Dados  
e Conhecimentos para Saúde  
FIOCRUZ

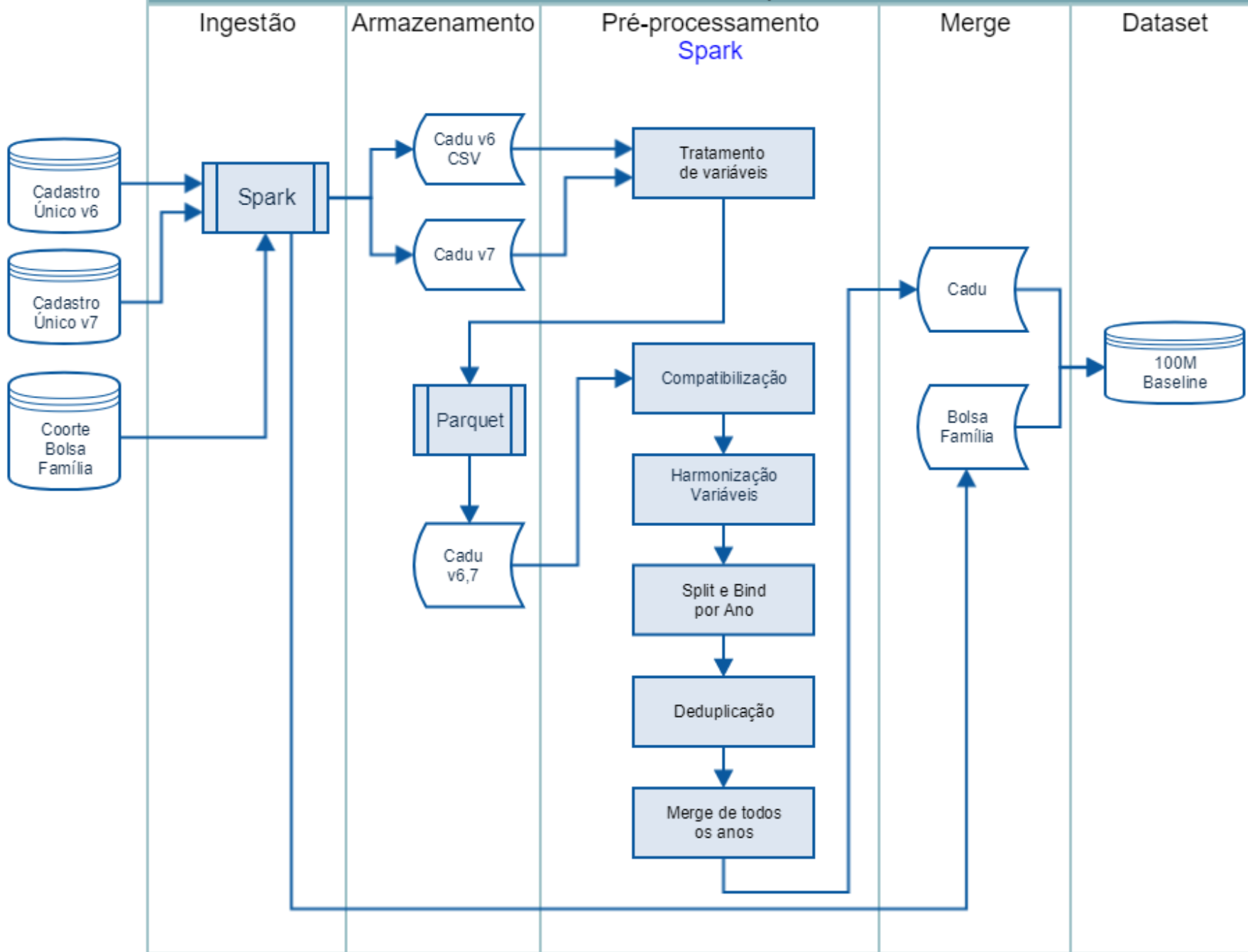


**Coorte de 100 milhões  
de brasileiras**

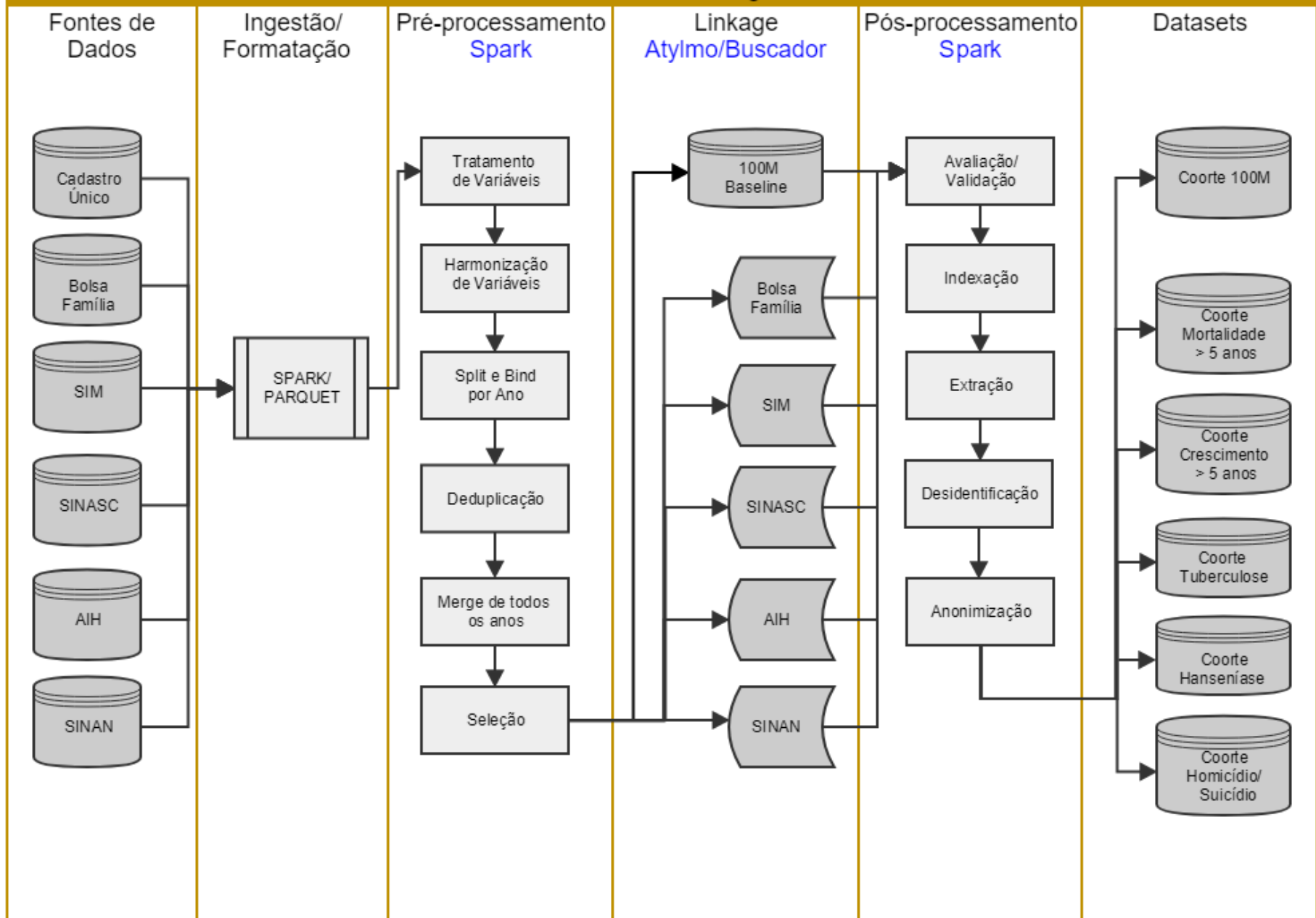
# DESENHO COORTE DE 100 MILHÕES



# 100M Baseline Pipeline



# Coorte 100M Linkage



quantos pares serão obtidos na



# AVALIAÇÃO ACURÁCIA

**Objetivo:** validar os resultados dos métodos de vinculação de dados com as bases de saúde

## Medidas epidemiológicas

Sensibilidade: capacidade do método de vincular os **pares** verdadeiramente **similares**

Especificidade: capacidade do método de não vincular os **pares** verdadeiramente **não similares**

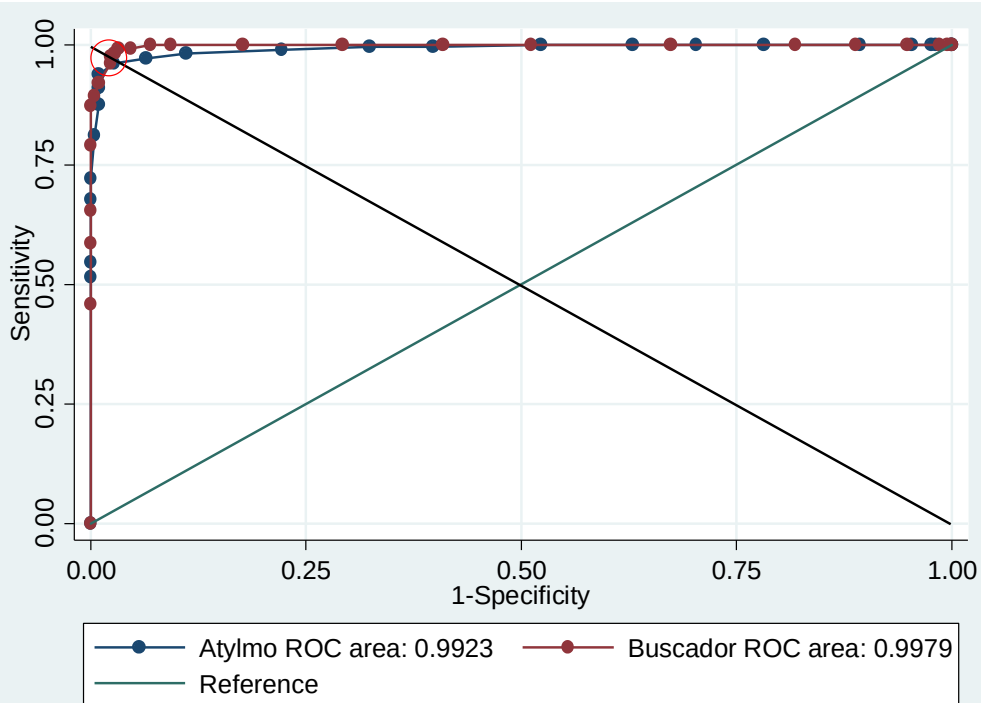
## Outras medidas:

Razão entre os números de verdadeiros vinculados e de vinculados identificados

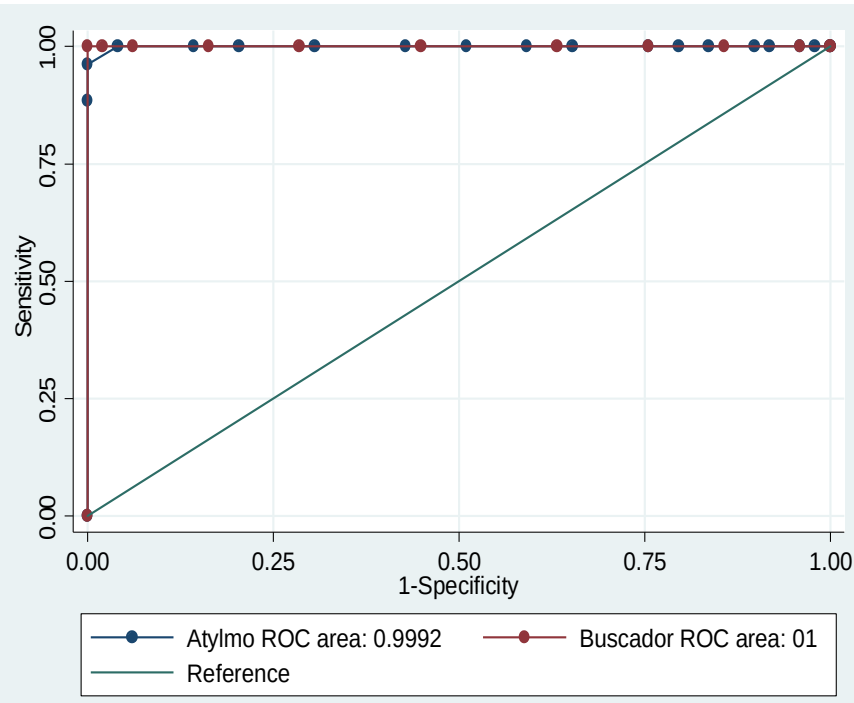
Curva ROC (*receiver operating characteristic curve*)- permite identificar o melhor ponto de corte de similaridade na vinculação das bases

	Matches	Non-matches
Linked	a (true positives)	b (false positives)
Unlinked	c (false negatives)	d (true negatives)

# SIM BA



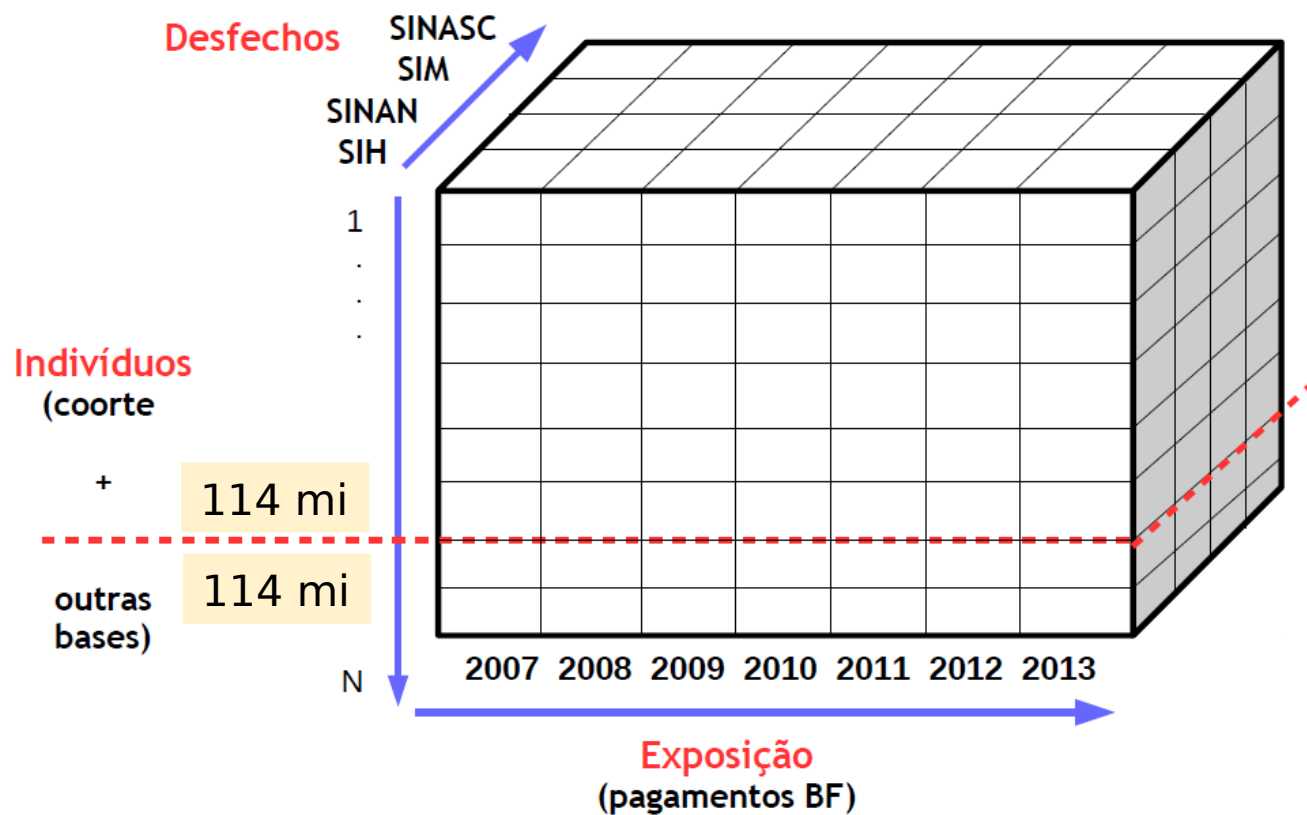
# SIM RO

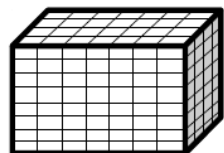


## Tamanho da base de dados

Bases	BA	SC	SE	RO
<b>CadÚnico</b>	12.036.0 10	1.988.59 9	1.447.51 2	932.909
<b>SIM (1 a 4 anos)</b>	507	132	91	75
<b>SINAN (TB)</b>	5941	2093	624	627

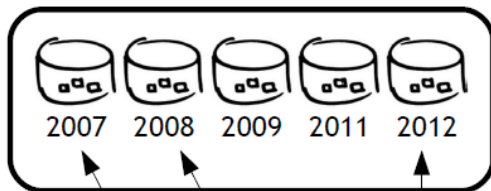
As bases estão atualizadas até o ponto de corte segundo o DCE





Individuals	Data sets						Exposition / Outcomes				
	C	B	H	N	M	A	2007	2008	2009	2011	2012
123	X						X				
	C (CadastroÚnico)	X					X	X	X	X	X
	B (Bolsa Família)							X	X		X
	H (SIH)							X	X		X
	N (SINAN)							X			X
	M (SIM)										X
	A (SINASC)						X		X		

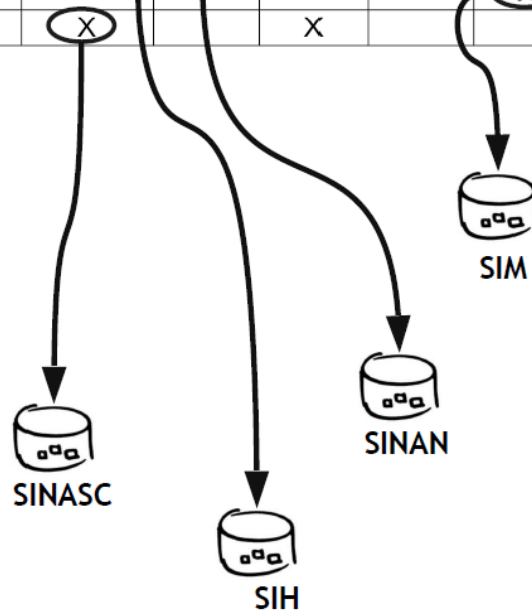
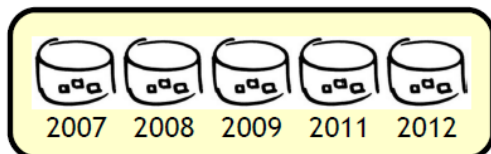
### Data sets PBF



IDPessoa NIS\_1 NIS\_2 ... NIS\_n  
<outros atributos>

Baseline (coorte)

Data sets  
Cadastro Único



x Fluxo básico de operação

# ABORDAGEM PARA DISPONIBILIZAR DADOS:

Não é tarefa trivial disponibilizar dados úteis para outros usuários a fim de realizar atividades importantes

- Tratar os dados a fim de que informações sensíveis não sejam descobertas a partir de uma publicação (técnicas de preservação de privacidade )

- Garantir a utilidade dos dados para que eventuais análises possam ser efetuadas com qualidade.

- Buscar uma solução que preserve ao máximo a utilidade das informações a serem disponibilizadas

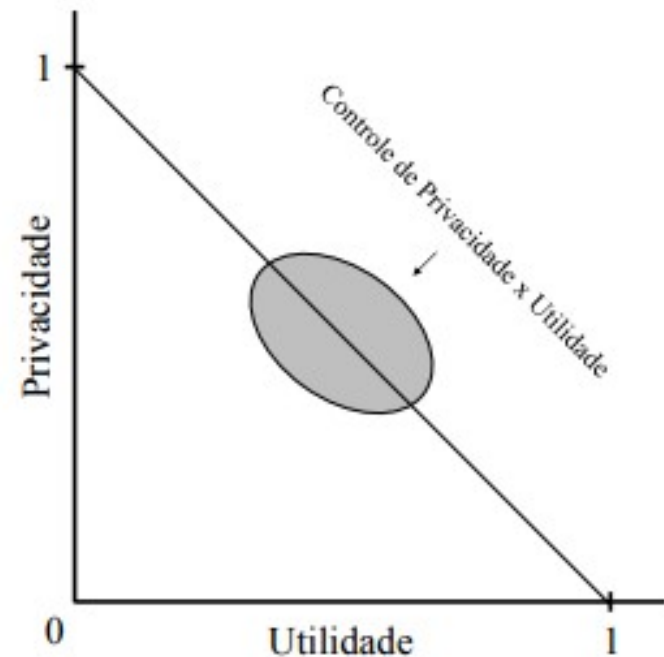


Figura 3.3. Balanceamento entre utilidade e privacidade.



# MÉTODOS PARA AVALIAR EXPERIMENTOS NATURAIS

- Regression Adjustment
- Propensity Score-Based Methods
- Difference-in-Differences
- Interrupted Time Series
- Synthetic Controls
- Instrumental Variables
- Regression Discontinuity

**Peter Craig et al. (2016). Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. *Annu. Rev. Public Health* 38:20.1-20.18**

..

**1) O Centro de integração de dados e conhecimentos é uma importante inovação para o desenvolvimento científico e para a política de dados abertos**

**2) A integração de bases de dados e sua disponibilização para gestores e pesquisadores, pode ajudar a responder a questões científicas, de avaliação de políticas e de gestão em tempo relativamente curto, a baixo custo e com grande possibilidade de superar limites** impostos por outras abordagens de investigação.

**3) É um incentivo ao desenvolvimento de novos**

- **COLABORAÇÕES CIENTÍFICAS NACIONAIS E INTERNACIONAIS**

**Nacionais:**

UFBA, FUNDAÇÃO UNIVERSIDADE BRASÍLIA  
SENAI-CIMATEC  
FUNDAÇÃO GETULIO VARGAS  
LNCC  
ICICT

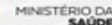
**Internacionais:**

LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE, FARR  
INSTITUTE  
UNIVERSITY OF GLASGOW

COLABORADORES CIENTÍFICOS



APOIADORES





**OBRIGADA !**

**[maria.ichihara@bahia.fiocruz.br](mailto:maria.ichihara@bahia.fiocruz.br)**

**[my.ichihara@gmail.com](mailto:my.ichihara@gmail.com)**



Ministério da Saúde

**FIOCRUZ**

**Fundação Oswaldo Cruz**

Instituto Gonçalo Moniz



# cidacs

Centro de Integração de Dados  
e Conhecimentos para Saúde

COLABORADORES CIENTÍFICOS



APOIADORES



BILL & MELINDA  
GATES foundation



SECRETARIA DE  
CIÊNCIA, TECNOLOGIA  
E INOVAÇÃO



MINISTÉRIO DA  
SAÚDE

