

# Clinical Research Informatics at InCor

Daniel M. Lima

Prof. Dr. José F. Rodrigues Jr.

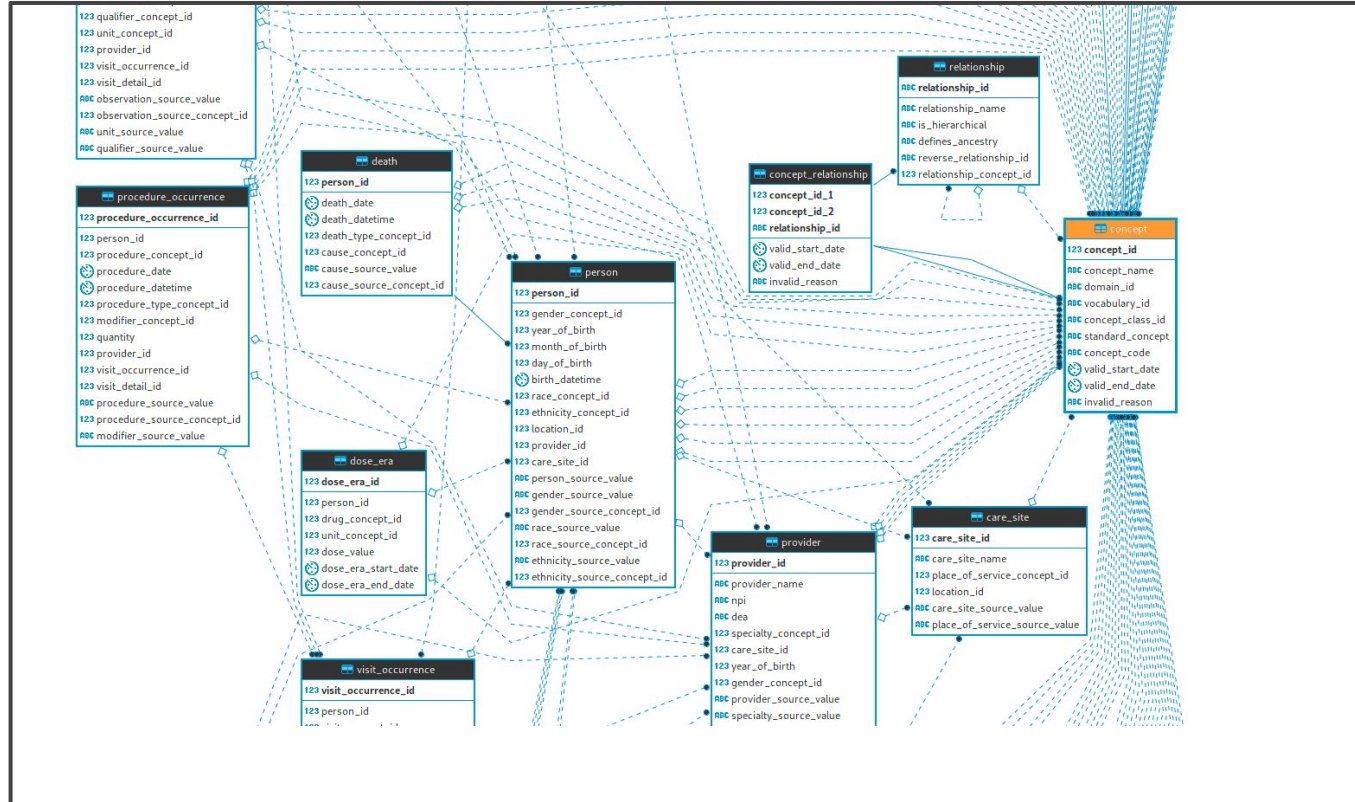
Profa. Dra. Agma J. M. Traina

Prof. Dr. Marco A. Gutierrez

Prof. Dr. Fábio A. Pires

[danielm@usp.br](mailto:danielm@usp.br), { [junio](#), [agma](#) }@icmc.usp.br, { [marco.gutierrez](#), [fabio.antero](#)

# Standardization!



- > AG\_TRANSFUSIONAL
- > AMBULATORIO
- > APEX\_030200
- > APEX\_040100
- > APEX\_040200
- > CAIXA
- > CALLCENTER
- > CIR\_CIRURGIA
- > CMP
- > COCI
- > COM
- > COP\_SEG
- > CTXSYS
- > DBA\_FABIO
- > DBA\_VALDEMIR
- > EASY
- > ENFERMAGEM
- > FABIANO
- > FONETICA
- > GENEHY
- > GREA
- > GSMADMIN\_INTERNAL
- > GUICHELATTES
- > HELPDESK
- > HISTORICO
- > INFOSAÚDE
- > INFOSAÚDE\_S13
- > INF\_GER
- > INTEGRACAO
- > JAVA\_ADM
- > JOB\_OPER
- > KIT\_INCOR
- > LAUDO
- > LAUDO\_S13
- > MAGIC
- > MAILING
- > MANUT\_HIST
- > MASS
- > MAT
- > MOBILEADMIN
- > MYSQL\_DRIVER
- > NETI HC S13

```
INSERT INTO person
--CREATE OR REPLACE VIEW v_person AS
SELECT
    person.id,
    gender.id,
    EXTRACT(YEAR FROM paci_dt_nasc) AS year_of_birth,
    NULL AS month_of_birth,
    NULL AS day_of_birth,
    NULL AS birth_datetime,
    coalesce(color.id, 0) AS race_concept_id,
    0 AS ethnicity_concept_id,
    NULL AS location_id,
    NULL AS provider_id,
    NULL AS care_site_id,
    NULL AS person_source_value,
    NULL AS paciente.paci_tp_sexo AS gender_source_value,
    NULL AS paciente.paci_sg_cor AS gender_source_concept_id,
    NULL AS paciente.paci_sg_cor AS race_source_value,
    NULL AS paciente.paci_sg_cor AS race_source_concept_id,
    NULL AS paciente.paci_sg_cor AS ethnicity_source_value,
    NULL AS paciente.paci_sg_cor AS ethnicity_source_concept_id
FROM PACIENTE.PAC_PACIENTE paciente
JOIN keys person
    ON person.rel = 1302
    AND person.src = paci_id
LEFT JOIN gender
    ON gender.src = paciente.paci_tp_sexo
LEFT JOIN color
    ON paci_sg_cor = color.src
```

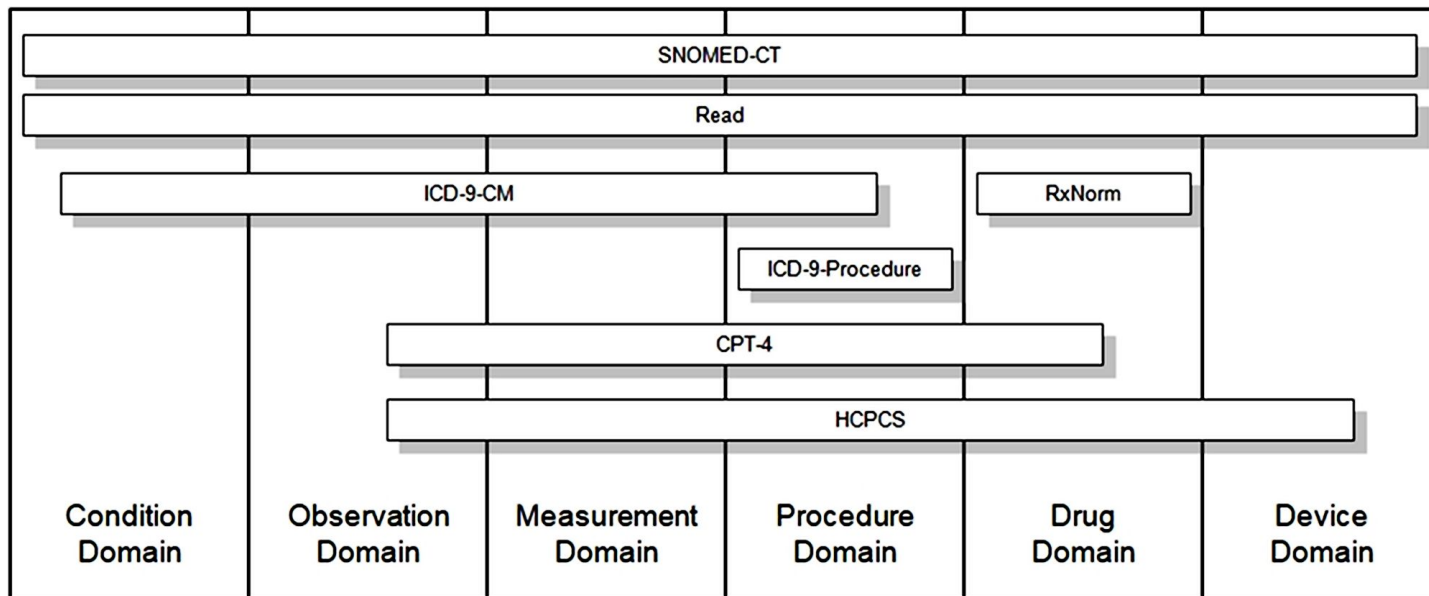
person

keys

```
INSERT INTO visit_occurrence
-- CREATE OR REPLACE VIEW v_visit_occurrence AS
SELECT
    item.id,
    visit_occurrence.id,
    person.id,
    visit_id,
    visit_concept_id,
    trunc(adm_dt_ingresso, '00') AS visit_start_date,
    NULL AS visit_start_datetime,
    trunc(coalesce(said_dt_hr, adm_dt_fin_an_hosp_dia, adm_dt_ingresso), '00') AS visit_end_date,
    NULL AS visit_end_datetime,
    32035 AS visit_type_concept_id,
    NULL AS provider_id,
    NULL AS care_site_id,
    NULL AS visit_source_value,
    NULL AS visit_source_concept_id,
    NULL AS admitting_source_value,
    NULL AS admitting_source_concept_id,
    NULL AS discharge_to_concept_id,
    NULL AS discharge_to_source_value,
    NULL AS preceding_visit_occurrence_id
FROM KEYS item
JOIN PACIENTE.ADM_ADMISSAO
    ON item.rel = 2478 -- (SELECT id FROM tables WHERE schem = 'PACIENTE' AND name = 'ADM_ADMISSAO')
    AND item.src = adm_nr || ';' || adm_an || ';' || adm_inst_cd
JOIN KEYS person
    ON person.rel = 1302 -- (SELECT id FROM tables WHERE schem = 'PACIENTE' AND name = 'PAC_PACIENTE')
    AND person.src = to_char(adm_paci_id)
JOIN visit
    ON visit.src = adm_tp
LEFT JOIN PACIENTE.SAI_SAIDA
    ON said_adm_nr = adm_nr
    AND said_adm_ano = adm_ano
    AND said_inst_cd = adm_inst_cd
    AND said_dt_hr_canc IS NULL
```

visit

# OMOP Domains and Vocabularies

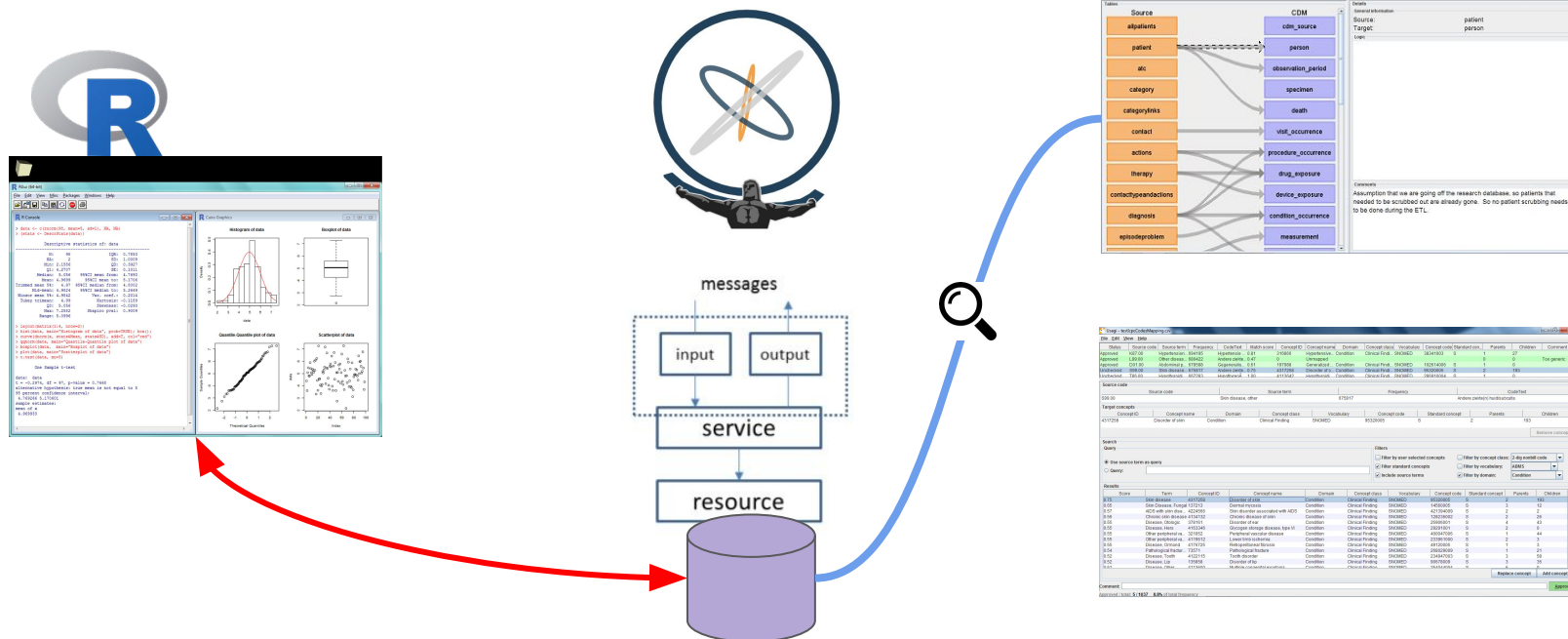


# Current Tools using the OMOP Common Data Model

# ACHILLES

## ATLAS + WebAPI

## Rabbit + Usagi



## 2-week cronogram

- 07.1: Study related OMOP work
- 07.2: Setup environment
- 08.1: Plan ETL
- 08.2: Load 1% Sample: Person, Visit, Condition, Procedure
- 09.1: Load 10% Sample: +Drug\_Exposure, Measurement
- 09.2: Load 100% Sample: impute Observation/Condition/Drug Eras
- 10.1: Replicate reference cohort for evaluation
- 10.2: Online updates and prepare for SI3\_Production
- 11.1: Draft paper / technical report

# InCor ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Incidence Rates

Profiles

Estimation

Prediction

Jobs

Configuration

Feedback

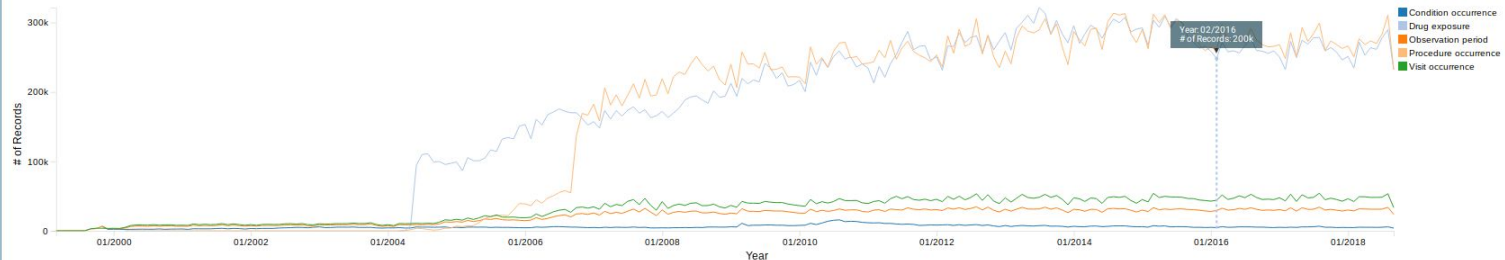
## Data Sources reports

InCor OMOP CDMv5

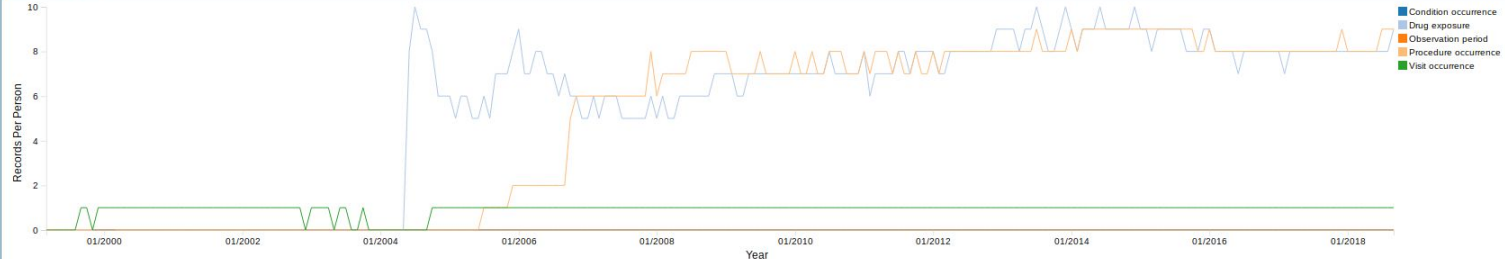
Data Density

### Data Density report (INCOR-CDM5)

#### Total Rows

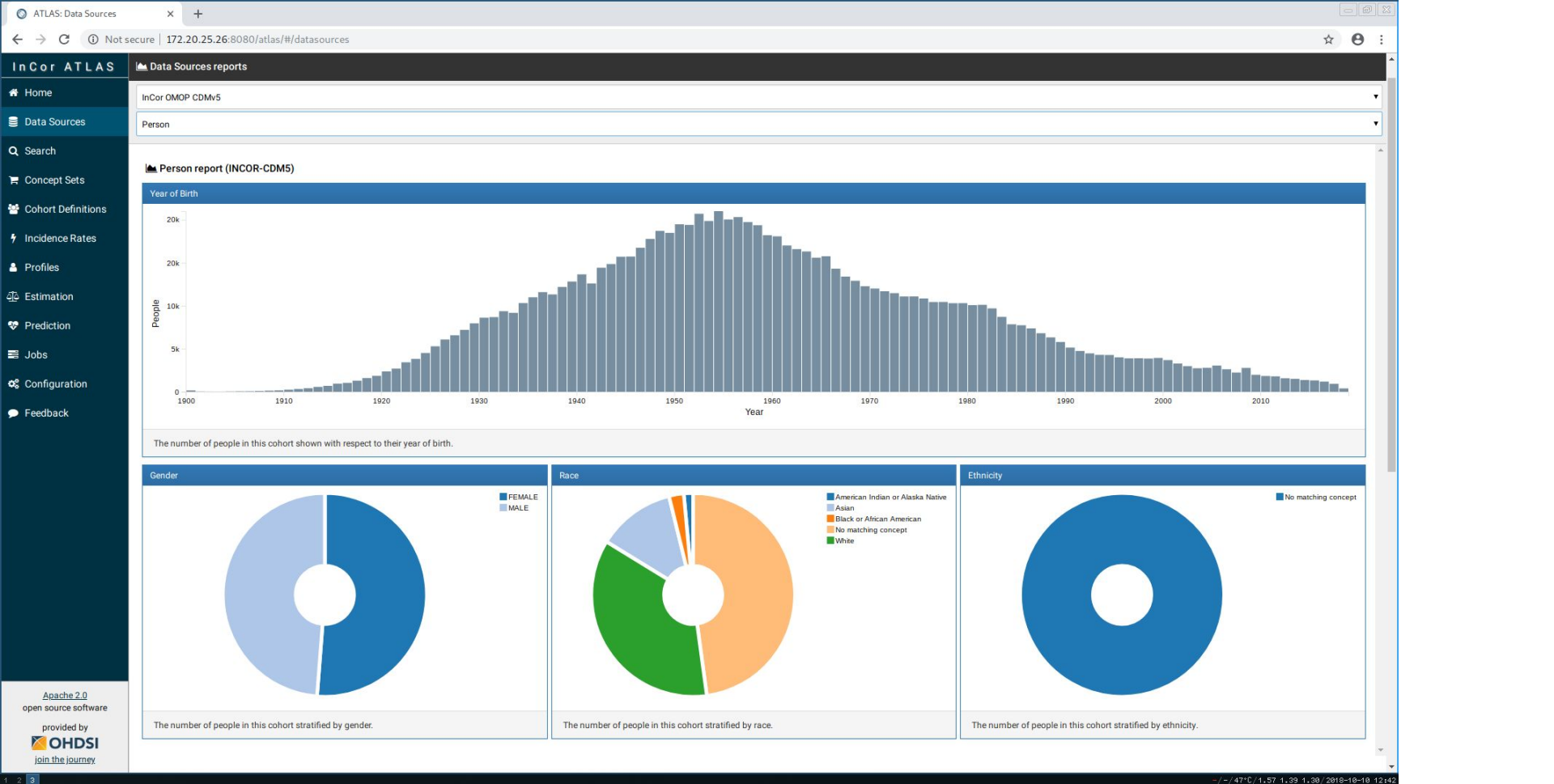


#### Records Per Person



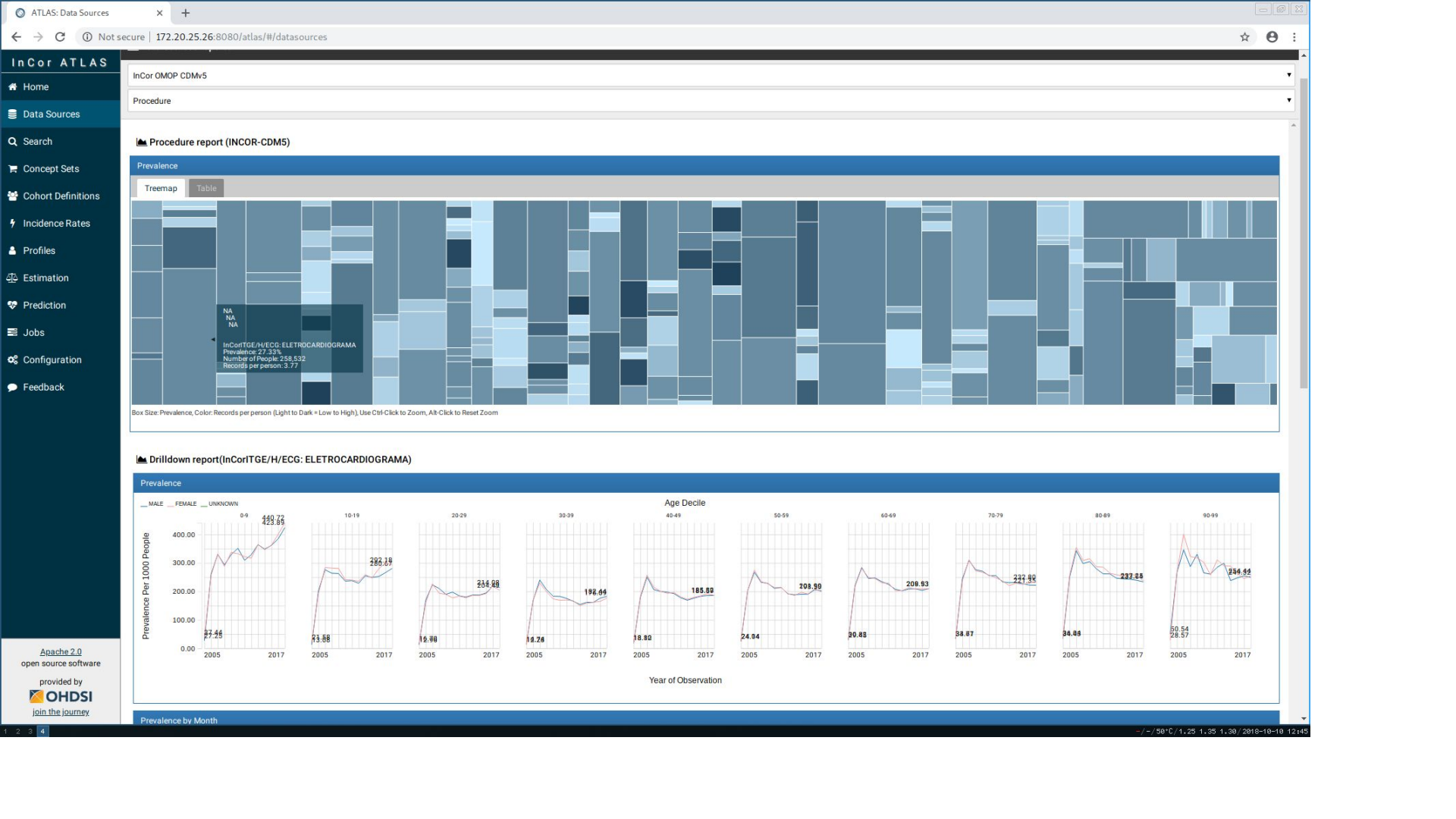
#### Concepts Per Person











ATLAS: Cohort Definitions

Not secure | 172.20.25.26:8080/atlas/#/cohortdefinition/6

Atenção: projeto em desenvolvimento – Pesquisa / Informática / InCor / HCFMUSP

InCor ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Incidence Rates

Profiles

Estimation

Prediction

Jobs

Configuration

Feedback

Cohort #6

DCV Abrahao 2016

Definition

Concept Sets

Generation

Reporting

Export

Warnings 13

enter a cohort definition description here

Initial Event Cohort

?

People having any of the following:

+ Add Initial Event

a visit occurrence of Qualquer Admissão

+ Add criteria attribute

Delete Criteria

with continuous observation of at least 0 days before and 0 days after event index date

Limit initial events to: all events per person.

Initial event inclusion criteria: From among the initial events, include:

having all of the following criteria:

+ Add criteria to group

Limit cohort of initial events to: all events per person.

Remove initial event inclusion criteria

Additional Qualifying Inclusion Criteria

?

New qualifying inclusion criteria

18+

Copy

Delete

1. 18+

enter an inclusion rule description

having all of the following criteria:

+ Add criteria to group

with the following event criteria:

+ Add criteria attribute

Delete Criteria

with age Between 18 and 80

Limit qualifying cohort to: all events per person.

Cohort Exit Criteria

?

Add a cohort exit criteria:

Based on a fixed time period relative to initial event start or end date

Based on the end of an era of persistent exposure to any drug within a defined concept set

The minimum date from amongst the selected cohort exit criteria occurring after the cohort entry date will be selected as the end date for the person's episode.

Apache 2.0

open source software

provided by

OHDSI

join the journey

1 2 3 4

~/ / 47°C / 1.21 1.35 1.31 / 2018-10-10 12:46

ATLAS: Cohort Definitions

Not secure | 172.20.25.26:8080/atlas/#/cohortdefinition/6

InCor ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Incidence Rates

Profiles

Estimation

Prediction

Jobs

Configuration

Feedback

Apache 2.0

open source software

provided by

OHDSI

join the journey

Available CDM Sources

Source Name	Generation Status	People	Records	Generated	Generation Duration
InCor OMOP CDMv5	COMPLETE	23,339	23,339	10/26/2018 2:48:20 PM	470.431s

Generate

View Reports

Inclusion Report

Cohort Features

By Events

By Person

Inclusion Report for InCor OMOP CDMv5

	Match Rate	Matches	Total
Summary Statistics:	2.76%	26,065	946,031

Inclusion Rule	N	% Satisfied	% To-Gain
1. Dx	497,263	52.56%	0.00%
2. 18+	846,249	89.45%	0.19%
3. M/F	886,264	93.68%	0.18%
4. DCV	108,188	11.44%	1.65%
5. 2a admissao	436,036	46.09%	0.01%
6. evento subsequente	48,898	5.17%	4.93%

Population Visualization

Switch to attrition view

1

2

3

gldi100% / ▲ 65.56% 11:43:36 / 46°C / 0.84 1.10 1.05 / 2018-11-09 09:12

ATLAS: Cohort Definitions

Not secure | 172.20.25.26:8080/atlas/#/cohortdefinition/6

☆ ⓘ ⚙

InCor ATLAS

Atenção: projeto em desenvolvimento – Pesquisa / Informática / InCor / HCFMUSP

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Incidence Rates

Profiles

Estimation

Prediction

Jobs

Configuration

Feedback

Cohort #6

DCV Abrahao 2016

Definition ⓘ Concept Sets Generation Reporting Export Warnings 13

Available CDM Sources

Source Name	Generation Status	People	Records	Generated	Generation Duration
InCor OMOP CDMv5	COMPLETE	23,339	23,339	10/26/2018 2:48:20 PM	470.431s

Inclusion Report Cohort Features

By Events By Person

Inclusion Report for InCor OMOP CDMv5

Inclusion Rule	Summary Statistics:	Match Rate	Matches	Total	% Diff
			N	% Remain	
1. Dx			497,263	52.56%	47.44%
2. 18+			444,307	46.97%	5.60%
3. M/F			417,555	44.14%	2.83%
4. DCV			95,115	10.05%	34.08%
5. 2a admissao			72,718	7.69%	2.37%
6. evento subsequente			26,065	2.76%	4.93%

Attrition Visualization

Switch to intersect view

Apache 2.0  
open source software

provided by  
OHDSI  
join the journey

1 2 3

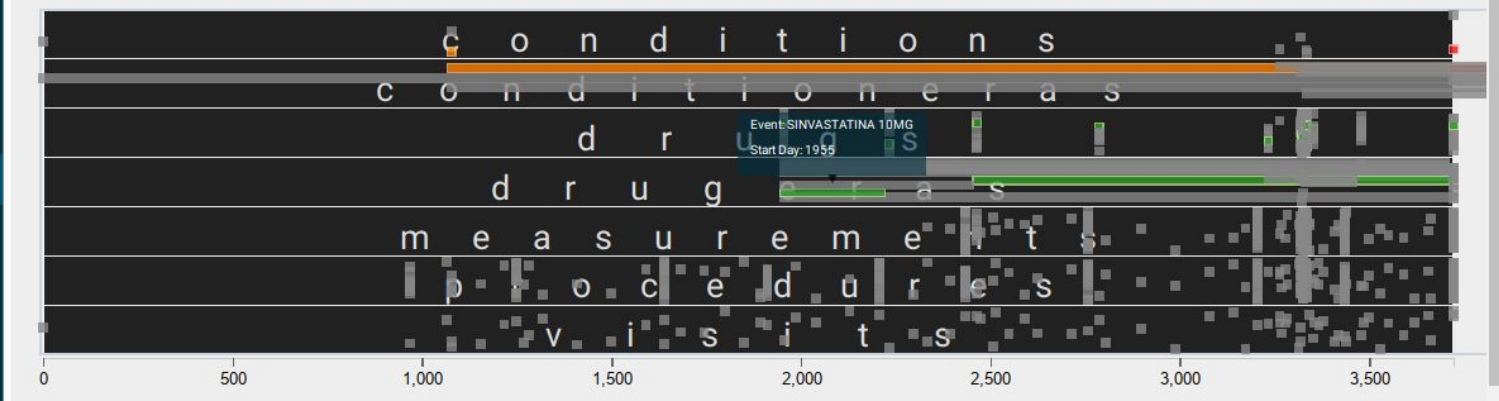
gldi100% / ▲ 65.64% 11:42:12 / 46°C / 0.67 1.04 1.03 / 2018-11-09 09:13

- InCor ATLAS
- Home
- Data Sources
- Search
- Concept Sets
- Cohort Definitions
- Incidence Rates
- Profiles
- Estimation
- Prediction
- Jobs
- Configuration
- Feedback

Profiles

INCOR-CDM5 4274163155117

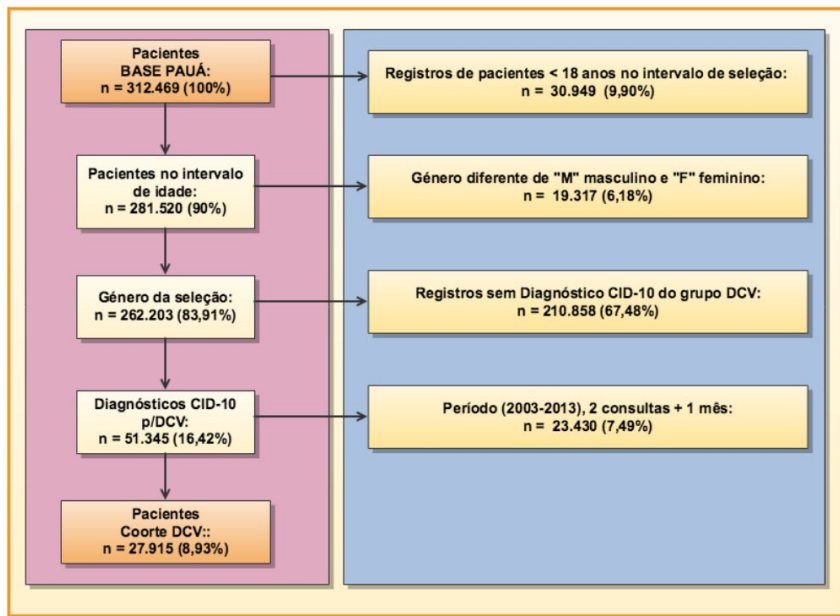
MALE | 944 events | Age 69 at index | Cohort: DCV Abrahao 2016



Column visibility Copy CSV Show 15 entries Filter:

Showing 1 to 15 of 944 entries

Domain	Concept Id	Concept Name	Domain	Start Day	End Day
procedure (316)	2000014443	Defeito de coagulacao nao especificado	condition	0	0
measurement (285)	2000014443	Defeito de coagulacao nao especificado	conditionera	0	34610
drug (217)	9203	Emergency Room Visit	visit	0	0
visit (79)	9202	Outpatient Visit	visit	968	968
druga (29)					
conditionera (9)					
conditionera (9)					



Abrahão et al. (2016)

Table 5 – Varying condition start periods.

(years)	2003-2013	2000-2013	2000-2016
Criteria			
Initial		778,015	
Dx, 18+, M/F		303,847	
CVD	45,710	49,942	63,656
People	39,498	43,293	54,126

Table 6 – Varying 2<sup>nd</sup> visit event start after index.

(days)	All	365	180	90
Criteria				
Initial		778,015		
Dx, 18+, M/F		303,847		
CVD	45,710	44,228	43,950	43,667
People	39,498	35,457	32,767	29,414

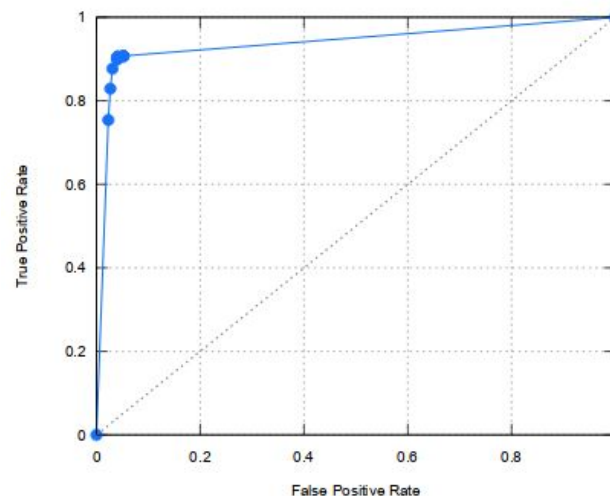


Figure 1 – Empirical ROC curve for Table 8 (AUC = 0.938).

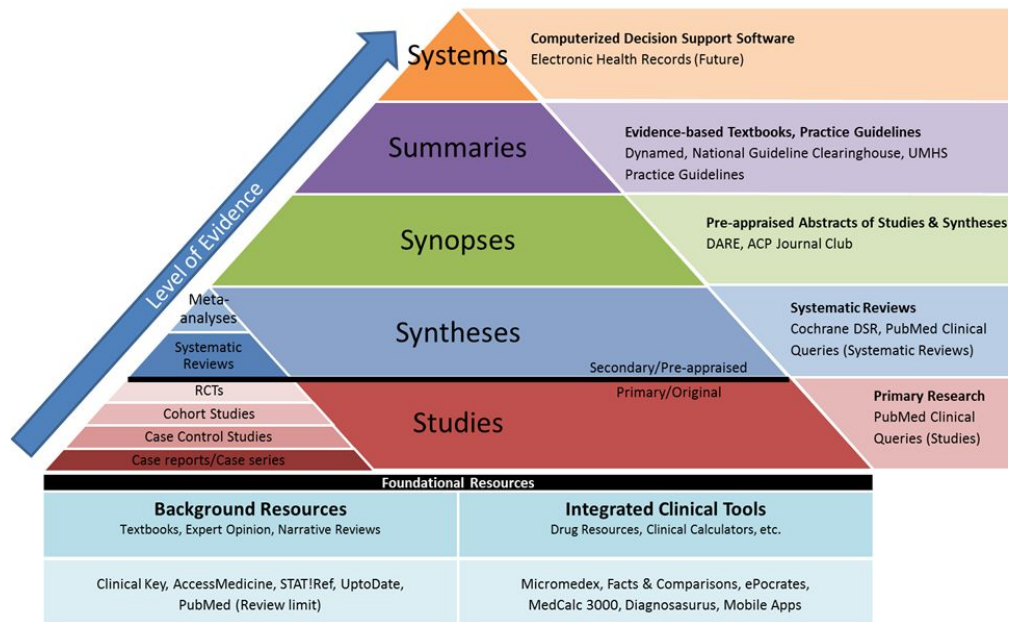


# Levels of Evidence

Level of Evidence	Type of Study
1a	Systematic reviews of randomized clinical trials (RCTs)
1b	Individual RCTs
2a	Systematic reviews of cohort studies
2b	Individual cohort studies and low-quality RCTs
3a	Systematic reviews of case-controlled studies
3b	Individual case-controlled studies
4	Case series and poor-quality cohort and case-control studies
5	Expert opinion based on clinical experience

Adapted from: Sackett DL et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd ed. Churchill





### Univ. Michigan. Sources for Evidence Based Research

Modified from: Haynes RB. Of studies, syntheses, synopses, summaries and systems: the "5S" evolution of services for evidence-based health care decisions. ACP J Club. 2006 Nov-Dec;145(3):A8-9.



Logged in as [username] | Log out

My Projects or Control Center

Project Home

Project Setup

Project status: **Development**

---

**Data Collection** Edit instruments

**Record Status Dashboard**  
- View data collection status of all records

**Add / Edit Records**  
- Create new records or edit/view existing ones

---

Data Collection Instruments:

Basic Demography Form

---

**Applications**

Calendar

## Personnel Registration Forms



Project Home



Project Setup



Other Functionality



Project Reports

Project status: Development



Complete!

Not complete?

### Main project settings

Enable

Use longitudinal data collection with repeating forms?

Enable

Use surveys in this project?

[VIDEO: How to create](#)



Complete!

Not complete?

### Design your data collection instruments

Add or edit fields on your data collection instruments. This may be done by [online method](#) or by uploading a Data Dictionary (offline method), in which both. Quick links: [Download PDF of all data collection instruments](#) OR [Download Dictionary](#)

➤ BRFSS 2009 Section 7: Hypertension Awareness ★	32
➤ BRFSS 2009 Section 8: Cholesterol Awareness ★	408
➤ BRFSS 2009 Section 9: Cardiovascular Disease Prevalence ★	14
➤ BRFSS 2013 Seccion 1 Estado De Salud ★	5
➤ BRFSS 2013 Seccion 2 Das En Los Que Se Siente Bien ★	5
▼ BRFSS 2013 Seccion 3 Acceso A Cuidados Medicos ★	4

**Details:**

**Institution:** REDLOC

**Contact:** Brenda Minor

**Contact email:** [brenda.l.minor@vanderbilt.edu](mailto:brenda.l.minor@vanderbilt.edu)

**Submitted by:** Brenda Minor



**Description:** La encuesta telefonica de Vigilancia de Comportamientos Riesgosos es el Sistema mas grande de salud en el mundo. El cuestionario captura condiciones de salud y comportamientos riesgosos en los Estados Unidos desde 1984. <http://www.cdc.gov/brfss/>

**Acknowledgement:** Centros para el Control y la Prevención de Enfermedades (CDC). Sistema de Vigilancia de Factores de Riesgo del Comportamiento Cuestionario . Atlanta, Georgia: Estados Unidos. Departamento de Salud y Servicios Humanos, Centros para el Control y la Prevención de Enfermedades. 2012.

**Terms of use:** Los datos y materiales fueron creados por agencias federales, por lo tanto son publicas y pueden ser reproducidas sin autorizacion necesaria.

**Last updated:** November 3, 2014

Visualizar o questionário

 [View as web page](#)  
 [View as PDF](#)

Import into my REDCap project

Importar o questionário para o projeto

## Viewing instrument as web page

The table below displays the data collection instrument as it would appear when viewed in REDCap.

### BRFSS 2013 Seccion 3 Acceso A Cuidados Medicos

Editing existing record

3.1 ¿Tiene algún tipo de cobertura de seguro médico, como seguro de salud, planes prepagos como los que brindan las HMO (organizaciones de atención médica administrada) u otros planes gubernamentales como Medicare o Servicios de Salud a Poblaciones Indígenas?

- ☐ Sí
- ☐ No
- ☐ No sabe/No está seguro
- ☐ Se niega a contestar

reset value

<div><span style="color:red;">Si es un estado bajo PPHF, pase a la pregunta 1 del Módulo 4, de lo contrario continúe.</div>

3.2 ¿Hay alguna persona a la que usted considere su médico o proveedor de atención médica de confianza?

- ☐ Sí
- ☐ No

reset value

3.2.1 ¿Hay más de una o no hay ninguna persona a la que considere su médico o proveedor de atención médica personal?"

- ☐ Sí
- ☐ Más de una
- ☐ No
- ☐ No sabe/No está seguro
- ☐ Se niega a contestar

reset value

3.3 En los últimos 12 meses, ¿hubo algún momento en que necesitó consultar a un médico pero no pudo hacerlo por razones económicas?

- ☐ Sí
- ☐ No
- ☐ No sabe/No está seguro
- ☐ Se niega a contestar

reset value

<div><span style="color:red;">Si es un estado bajo PPHF, pase a la pregunta 3 del Módulo 4 de lo contrario continúe.</div>

3.4 ¿Cuánto hace que fue al médico para hacerse un chequeo de rutina? Un chequeo de rutina es un examen físico general, que no se realiza a consecuencia de una

- ☐ En el último año (hace menos de 12 meses)
- ☐ En los últimos 2 años (hace 1 año pero menos de 2)
- ☐ En los últimos 5 años (hace 2 años pero menos de 5)
- ☐ Hace 5 años o más <p>
- ☐ No sabe/No está seguro

Close

### Add New Field

You may add a new project field to this data collection instrument by completing the fields below and clicking the Save button at the bottom. When you add a new field, it will be added to the form on this page. For an overview of the different field types available, you may view the [Field Types video \(4 min\)](#).

Field Type: Text Box (Short Text, Number, Date/Time, ...)

Field Label [How to use Piping](#)

Nome

Texto da questão, a pergunta em si

Field Annotation (optional) [Learn about Action Tags](#)

Explanatory notes - not displayed on any page

Variable Name (utilized during data export)

nome

ONLY letters, numbers, and underscores

☐ Enable auto naming of variable based upon its Field Label?

Validation? (optional) --- None ---

— Or —

Enable searching within a biomedical ontology [?](#)

— choose ontology to search —

Required?\* ☐ No ☒ Yes **Escolha "Yes" se a pergunta for obrigatória**

\* Prompt if field is blank

Identifier? ☐ No ☒ Yes

Does the field contain identifying information (e.g., name, SSN, address)?

Custom Alignment Right / Vertical (RV)

Align the position of the field on the page

Field Note (optional)

Small reminder text displayed underneath field

**Nome da questão que será guardada no banco de dados. Forneça um nome simples, significativo e sem acentuação.**

**"Yes", quando o campo coleta dados de identificação, Nome, endereço, e-mail, etc.**

**Alinhamento**

**Dicas sobre o assunto da questão, essa informação não será mostrada para o entrevistado.**

**Save** **Cancel**

## Exporting "Selected instruments and/or events (all records)"

Select your export settings, which includes the export format (Excel/CSV, SAS, SPSS, R, Stata) and if you wish to perform de-identification on the data set.

### Choose export format

☐  CSV / Microsoft Excel (raw data)

☒  CSV / Microsoft Excel (labels)

☐  SPSS Statistical Software

☐  SAS Statistical Software

☐  R Statistical Software

☐  Stata Statistical Software

### De-identification options (optional)

The options below allow you to limit the amount of sensitive information that you are exporting out of the project. Check all that apply.

#### Known Identifiers:

- ☐ Remove all tagged Identifier fields (tagged in Data Dictionary)
- ☐ Hash the Record ID field (converts record name to an unrecognizable value)

#### Free-form text:

- ☐ Remove unvalidated Text fields (i.e. Text fields other than dates, numbers, etc.)
- ☐ Remove Notes/Essay box fields

#### Date and datetime fields:

- ☐ Remove all date and datetime fields
- OR —
- ☐ Shift all dates by value between 0 and 364 days  
(shifted amount determined by algorithm for each record) [What is date shifting?](#)
- ☐ Also shift all survey completion timestamps by value between 0 and 364 days  
(shifted amount determined by algorithm for each record)

[Deselect all options](#)

### Additional export options

- ☒ Export Data Access Group name for each record (if record is in a group)?
- ☒ Export survey Identifier field and survey timestamp field(s)?

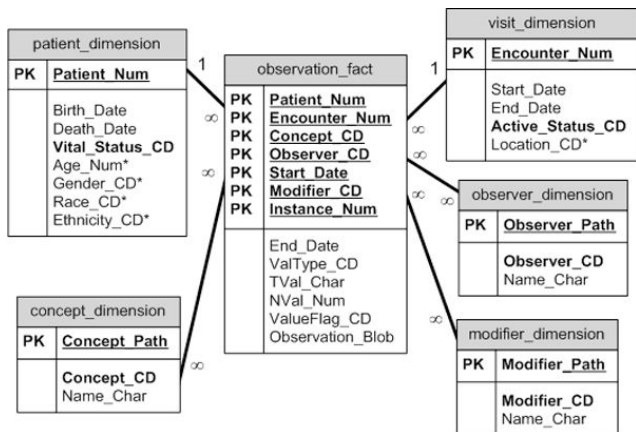
Export Data

Cancel



# i2b2

Informatics for Integrating Biology and the Bedside



The screenshot shows the **Mapping Editor** interface. The top bar indicates the project is **Staging i2b2: i2b2idrttargetproje** and the instance is **new target instance Version: 2**. The target project is **i2b2idrttargetproject2**.

The left pane shows the project structure:

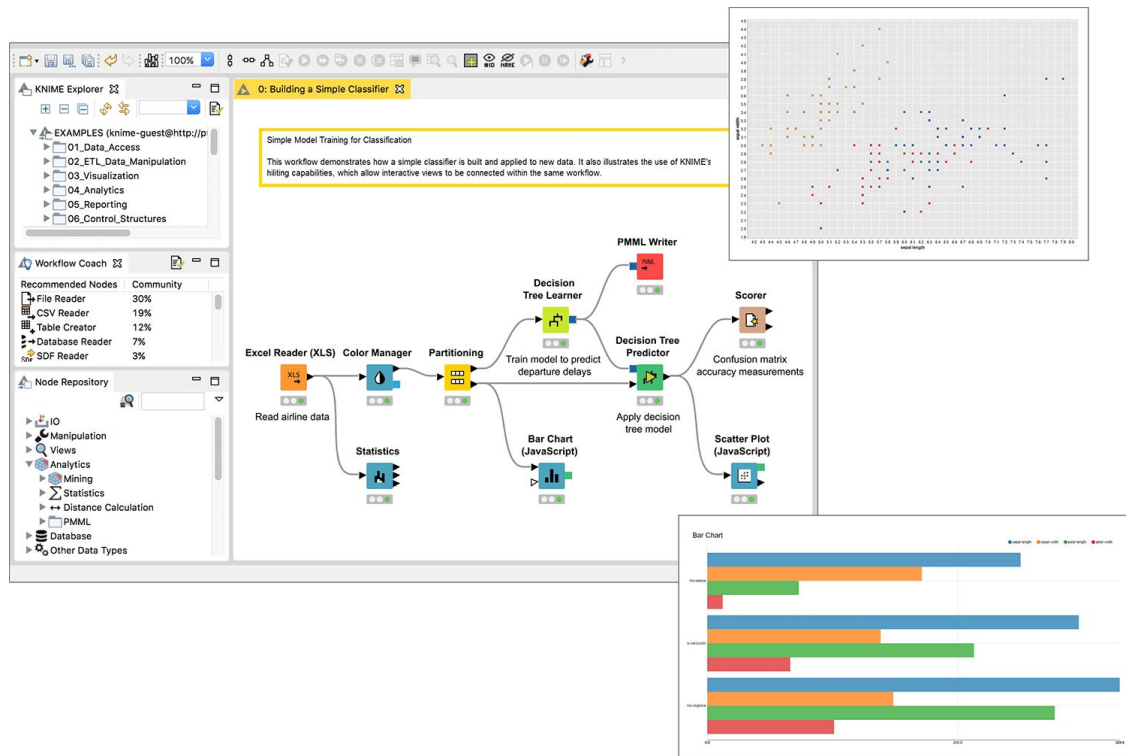
- i2b2 - [500]
  - biospecimen data
    - LIMS Export
      - Identifiers
      - Location
      - Metadata
      - Specimen type
      - SPREC
        - Fluid Specimen
        - Long-term storage
      - Solid Specimen
        - Cold ischemia time
        - Fixation Time
        - Fixation Type
        - Type of collection
        - Type of sample
        - Warm ischemia time
  - patient data

The right pane shows the **Target-Ontology** structure:

- Target-Ontology
  - Sespi Study
    - Biospecimen Data
      - CRFs
        - 001
        - 002
        - 003
        - 004
        - 005
        - 006 SAPII-Score
        - 009 Discharge
        - 002 Day 7
        - 003 Discharge
        - 004 Relocation

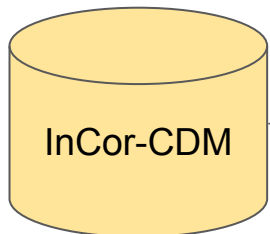
A context menu is open over the **CRFs** folder, showing options: **Sort Children**, **Add Folder**, **Add Item**, **Expand All Children**, **Collapse All Children**, **Hide/Activate**, **Delete**, and **Rename**.







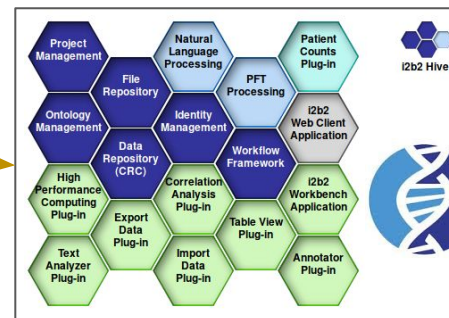
Deidentified records



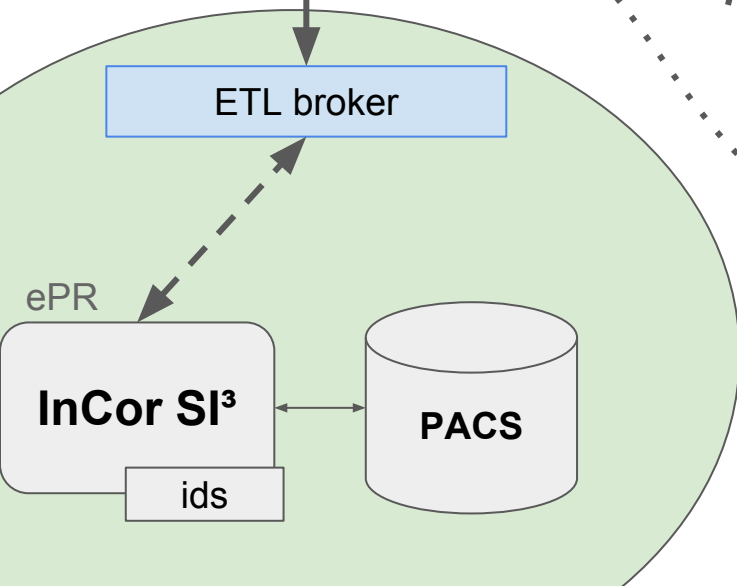
Visual Analytics  
Cohort Selection  
Model Execution



Study Specification  
Survey Forms



Bio Lab Data  
Genomics



Analytics  
workbench