

# Health Data Architecture for Learning, INFOBANCO

**First published:** 01/02/2024

**Last updated:** 16/05/2025

Data source

Human

Other

Population registry

Primary care medical records

## Administrative details

### Administrative details

#### Data source ID

1111145

#### Data source acronym

INFOBANCO

#### Data holder

[Hospital Universitario 12 de Octubre](#)

#### Data source type

Other

Population registry

Primary care medical records

**Data source type, other**

Electronic health records, Complete and longitudinal EHR from Primary Care and Hospital.

---

**Main financial support**

Funding by own institution

National, regional, or municipal public funding

---

**Care setting**

Hospital inpatient care

Hospital outpatient care

Primary care – GP, community pharmacist level

Primary care – specialist level (e.g. paediatricians)

Secondary care – specialist level (ambulatory)

---

**Data source qualification**

If the data source has successfully undergone a formal qualification process (e.g., from the EMA, ISO or other certifications), this should be described.

Yes

---

**Description of the qualification**

STAGE 6 EMRAM - HIMSS

---

**Data source website**

<https://cpisanidadcm.org/infobanco/?lang=en>

## Contact details

Javier de la Cruz [javier.delacruz@salud.madrid.org](mailto:javier.delacruz@salud.madrid.org)

## Data source regions and languages

### Data source countries

Spain

---

### Data source languages

Spanish

---

### Data source regions

Madrid, Comunidad de

## Data source establishment

### Data source established

15/06/2010

---

### Data source time span

**First collection:** 13/12/2010

The date when data started to be collected or extracted.

## Publications

### Data source publications

[Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study](#)

Obtaining EHR-derived datasets for COVID-19 research within a short time: a flexible methodology based on Detailed Clinical Models

Impact of COVID-19 in patients with multiple myeloma based on a global data network

Outcomes and patterns of treatment in chronic myeloid leukemia, a global perspective based on a real-world data global network

Characteristics and outcomes of an international cohort of 600 000 hospitalized patients with COVID-19

## Data elements collected

### The data source contains the following information

#### **Disease information**

Does the data source collect information with a focus on a specific disease? This might be a patient registry or other similar initiatives.

Yes

---

#### **Rare diseases**

Are rare diseases captured? In the European Union a rare disease is one that affects no more than 5 people in 10,000.

Yes

---

#### **Pregnancy and/or neonates**

Does the data source collect information on pregnant women and/or neonatal subpopulation (under 28 days of age)?

Yes

---

## **Hospital admission and/or discharge**

Yes

---

## **ICU admission**

Is information on intensive care unit admission available?

Yes

---

## **Cause of death**

Captured

---

## **Cause of death vocabulary**

ICD-10

SNOMED CT

---

## **Prescriptions of medicines**

Captured

---

## **Prescriptions vocabulary**

ATC

other

---

## **Prescriptions vocabulary, other**

SNOMED CT

---

## **Dispensing of medicines**

Captured

---

## **Dispensing vocabulary**

ATC

other

---

## **Dispensing vocabulary, other**

SNOMED CT

---

## **Advanced therapy medicinal products (ATMP)**

Is information on advanced therapy medicinal products included? A medicinal product for human use that is either a gene therapy medicinal product, a somatic cell therapy product or a tissue engineered products as defined in Regulation (EC) No 1394/2007 [Reg (EC) No 1394/2007 Art 1(1)].

Yes

---

## **Contraception**

Is information on the use of any type of contraception (oral, injectable, devices etc.) available?

Yes

---

## **Indication for use**

Does the data source capture information on the therapeutic indication for the use of medicinal products?

Captured

---

## **Indication vocabulary**

SNOMED CT

---

## **Medical devices**

Is information on medicinal devices (e.g., pens, syringes, inhalers) available?

Yes

---

## **Administration of vaccines**

Yes

---

## **Procedures**

Does the data source capture information on procedures (e.g., diagnostic tests, therapeutic, surgical interventions)?

Captured

---

### **Procedures vocabulary**

ICD

SNOMED CT

---

### **Healthcare provider**

Is information on the person providing healthcare (e.g., physician, pharmacist, specialist) available?  
The healthcare provider refers to individual health professionals or a health facility organisation licensed to provide health care diagnosis and treatment services including medication, surgery and medical devices.

Yes

---

### **Clinical measurements**

Is information on clinical measurements (e.g., BMI, blood pressure, height) available?

Yes

---

### **Genetic data**

Are data related to genotyping, genome sequencing available?

Captured

---

### **Genetic data vocabulary**

HGNC

Other

---

### **Genetic data vocabulary, other**

HGVS

---

### **Biomarker data**

Does the data source capture biomarker information? The term “biomarker” refers to a broad subcategory of medical signs ( objective indications of medical state observed from outside the patient), which can be measured accurately and reproducibly. For example, haematological assays, infectious disease markers or metabolomic biomarkers.

Captured

---

## **Biomarker data vocabulary**

HPO

---

## **Patient-reported outcomes**

Is information on patient-reported outcomes (e.g., quality of life) available?

Yes

---

## **Patient-generated data**

Is patient-generated information (e.g., from wearable devices) available?

Yes

---

## **Units of healthcare utilisation**

Are units of healthcare utilisation (e.g., number of visits to GP per year, number of hospital days) available or can they be derived? Units of healthcare utilisation refer to the quantification of the use of services for the purpose of preventing or curing health problems.

Yes

---

## **Unique identifier for persons**

Are patients uniquely identified in the data source?

Yes

---

## **Diagnostic codes**

Captured

---

## **Diagnosis / medical event vocabulary**

Human Phenotype Ontology (HPO)



ICD

Orphacode

Other

SNOMED CT

---

**Diagnosis / medical event vocabulary, other**

OMIM; ICPC2

---

**Medicinal product information**

Captured

---

**Medicinal product information collected**

Active ingredient(s)

Dose

Route of administration

---

**Medicinal product vocabulary**

ATC

Other

SNOMED

---

**If 'other,' what vocabulary is used?**

Spanish Medication Agency Code

---

**Quality of life measurements**

Captured

---

**Quality of life measurements vocabulary**

EQ5D

other

---

## **Quality of life measurements, other**

EORTC QLQ; BREAST-Q; etc.

---

## **Lifestyle factors**

Captured

---

## **Lifestyle factors**

Alcohol use

Other

Tobacco use

---

## **Sociodemographic information**

Captured

---

## **Sociodemographic information collected**

Age

Gender

Other

# Quantitative descriptors

## Population Qualitative Data

### **Population age groups**

Paediatric Population (< 18 years)

Preterm newborn infants (0 – 27 days)

Term newborn infants (0 – 27 days)

Infants and toddlers (28 days – 23 months)

Children (2 to < 12 years)  
Adolescents (12 to < 18 years)  
Adults (18 to < 46 years)  
Adults (46 to < 65 years)  
Elderly ( $\geq$  65 years)  
Adults (65 to < 75 years)  
Adults (75 to < 85 years)  
Adults (85 years and over)

---

**Estimated percentage of the population covered by the data source in the catchment area**

100%

---

**Description of the population covered by the data source in the catchment area whose data are not collected (e.g., people who are registered only for private care)**

The proportion of the population in the catchment area registered only for private care is negligible

## Family linkage

**Family linkage available in the data source permanently or can be created on an ad hoc basis**

Permanently

---

**Family linkage available between the following persons**

Mother-child

## Population

## Population size

3100536

---

## Active population size

2903984

## Population by age group

Age group	Population size	Active population size
Paediatric Population (< 18 years)	243702	242483
Term newborn infants (0 – 27 days)	310	290
Infants and toddlers (28 days – 23 months)	13332	13358
Children (2 to < 12 years)	119991	119644
Adolescents (12 to < 18 years)	1097590	109190
Adults (18 to < 46 years)	898845	895008
Adults (46 to < 65 years)	840555	823860
Elderly (≥ 65 years)	1115263	940310
Adults (65 to < 75 years)	278428	258455
Adults (75 to < 85 years)	208666	177143
Adults (85 years and over)	628169	504712

## Median observation time

**Median time (years) between first and last available records for unique individuals captured in the data source**

7.30

---

**Median time (years) between first and last available records for unique active individuals (alive and currently registered) capt**

7.40

## Data flows and management

### Access and validation

#### **Governance details**

Documents or webpages that describe the overall governance of the data source and processes and procedures for data capture and management, data quality check and validation results (governing data access or utilisation for research purposes).

<https://cpisanidadcm.org/infobanco/?lang=en>

#### **Biospecimen access**

Are biospecimens available in the data source (e.g., tissue samples)?

No

---

#### **Access to subject details**

Can individual patients/practitioners/practices included in the data source be contacted?

No

---

## **Description of data collection**

The sources of information are the Electronic Health Records of both the Hospital and Primary Care, so that any patient contact with the health centres implies a record of data by health professionals. The data and the systems in which they are made will vary according to the type of visit required by the patient.

## Event triggering registration

### **Event triggering registration of a person in the data source**

Birth

Disease diagnosis

Start of treatment

---

### **Event triggering de-registration of a person in the data source**

Death

---

### **Event triggering creation of a record in the data source**

Any patient contact with the Hospital or Primary Care.

## Data source linkage

### **Linkage**

Is the data source described created by the linkage of other data sources (prelinked data source) and/or can the data source be linked to other data source on an ad-hoc basis?

Yes

---

### **Linkage description, pre-linked**

The combination of information is done at the patient level, by means of a unique identifier of the residents in Madrid Region.

## **Linked data sources**

### **Pre linked**

Is the data source described created by the linkage of other data sources?

Yes

---

### **Data source, other**

It is combined the information from the Hospital and Primary Care, so as to have information on all the events that occurred throughout the patient's life.

---

### **Linkage variable**

The variable that allows the unique identification of patients treated in Madrid Region is called CIPA.

---

### **Linkage completeness**

All patients

## **Data management specifications that apply for the data source**

### **Data source refresh**

Monthly

---

## **Informed consent for use of data for research**

Required for intervention studies

---

## **Possibility of data validation**

Can validity of the data in the data source be verified (e.g., access to original medical charts)?

Yes

---

## **Data source preservation**

Are records preserved in the data source indefinitely?

Yes

---

## **Approval for publication**

Is an approval needed for publishing the results of a study using the data source?

Yes

---

## **Data source last refresh**

26/04/2023

# **Common Data Model (CDM) mapping**

## **CDM mapping**

Has the data source been converted (ETL-ed) to a common data model?

Yes

## **CDM Mappings**

### **CDM name (other)**

OpenEHR

---



**Data source ETL frequency**

0,03 months

---

**Data source ETL status**

In progress

**CDM name**

i2b2

---

**CDM website**

<https://community.i2b2.org/wiki/display/BUN/i2b2+Common+Data+Model+Documentation>

---

**Data source ETL frequency**

0,03 months

---

**Data source ETL status**

Completed

**CDM name**

OMOP

---

**CDM website**

<https://www.ohdsi.org/Data-standardization/>

---

**Data source ETL CDM version**

5.3

---

## **Data source ETL frequency**

0,03 months

---

## **Data source ETL status**

In progress