



Transforming the Hospital i2b2 data repository into OMOP Common Data Model

M^a Teresa García Morales¹, Miguel Pedrera Jimenez², Noelia García Barrio², Diego Boscá Tomás³, Pablo Serrano Balazote², David Lora Pablos², Agustín Gómez de la Cámara²

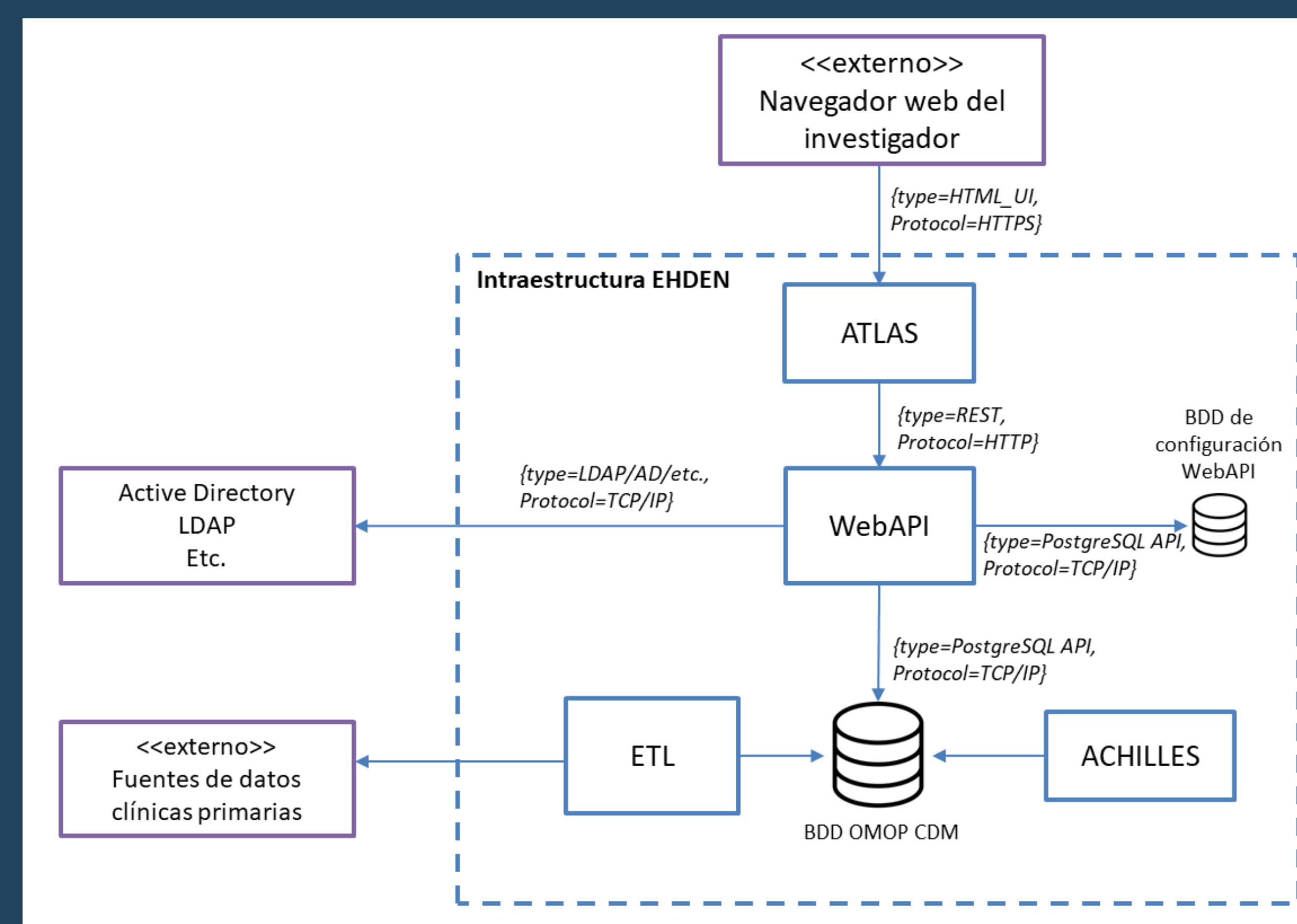
¹Instituto de Investigación Biomédica del Hospital Universitario 12 de Octubre; ²Hospital Universitario 12 de Octubre, Madrid, Spain; ³VeraTech for Health, Valencia, Spain

RESULTS

The main result of the work developed is an OMOP database from an i2b2 instance. In addition, both WebAPI and RStudio have been installed. RStudio is connected to the OMOP database, so studies can be executed in an easy way.

TABLERNAME	COUNT
measurement	923,518,665
visit_occurrence	57,146,821
drug_exposure	45,656,569
condition_occurrence	14,187,889
condition_era	6,798,843
person	2,891,589
drug_era	2,742,047
procedure_occurrence	2,518,977
observation_period	933,722
provider	213,441
death	166,268
observation	135,031
care_site	438
location	426

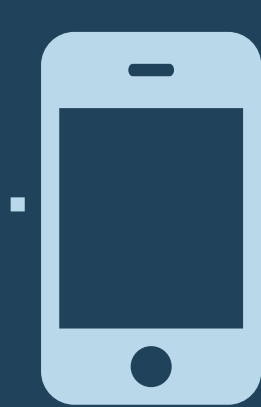
Number of records in clinical data tables



OMOP CMD H12O Infrastructure

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	1944	50	1994	97%	284	3	287	99%	2228	53	2281	98%
Conformance	656	25	681	96%	104	0	104	100%	760	25	785	97%
Completeness	376	10	386	97%	11	4	15	73%	387	14	401	97%
Total	2976	85	3061	97%	399	7	406	98%	3375	92	3467	97%

Data Quality Dashboard summary, generated at 2021-12-02



Scan QR to download the poster



Scan QR to access DataDoce GitHub

DISCUSION AND CONCLUSIONS

During the development of the OMOP instance, some delays occurred, mainly due to the amount of data contained in the i2b2 repository, which caused some processes to require more execution time. For example, it took 12 days to run the DQD, as the Measurement table contains about 1 billion results.

For the time being, manual updating of the data is contemplated. However, with the H12O INFOBANCO Project. INFOBANCO project aims to develop a regional data network architecture for combining information generated by different sources in order to provide tools for governance, collection, transformation, interrogation, visualization and analysis of data for the obtaining knowledge and support for decision making.

<https://cpisanidadcm.org/infobanco/>



BACKGROUND

The Hospital Universitario 12 de Octubre and Madrid Primary Care i2b2 repository contains information of 2.800.000 patients and integrates diagnoses, procedures, clinical findings, laboratory tests, medications and histopathology data. In addition, all data contained is encoded in different terminology and classification standards depending on the type of clinical concept.

METHODS

For the EHDEN First Call, the i2b2 repository of the H12O has been transformed to the OMOP Common Data Model (CDM) version 5.3.1. This project has been developed in collaboration with the SME VeraTech for Health 12.

This process was made with the support of OHDSI tools: White Rabbit and Rabbit-In-A-Hat for the ETL design, ATHENA and Usagi for the vocabulary correspondence and Achilles and DQD for quality assurance of data and ETL.

<https://www.ohdsi.org/software-tools/>