

OMOP's evolution in the data reuse strategy of Hospital Universitario 12 de Octubre

Noelia García, Miguel Pedrera, Blanca Baselga, Paula Rubio, Alberto Tato, Cristina Díaz, Tomás González, Bruno Diez, Víctor Quirós, M^a Teresa García, Juan Luis Cruz, José Luis Bernal, Pablo Serrano

Background

In recent years, studies based on real-world data (1) have played a very important role. However, the COVID-19 pandemic has magnified its value and it is now a key tool in decision-making processes, treatment and vaccine monitoring and disease characterisation, among others.

The Hospital Universitario 12 de Octubre (H12O) (2) has been working on the standardisation and reuse of clinical data since 2017 with the implementation of a first i2b2 repository (3) for the centralisation of information from its systems. At the same time, the Hospital has developed a methodology for the reuse of the Electronic Health Record (EHR) (4), which has allowed it to participate in numerous projects, even during the pandemic. Among them, the 4CE initiative (5), the automatic data upload to ISARIC Case Report Form (6), and two EHDEN calls (7). Thanks to the latter, the Hospital has come into contact with the OMOP model (8) and has created several instances, using different data sources and focusing on different clinical conditions. Based on this trajectory, it has been possible to identify limitations and points for improvement, and within the INFOBANCO project (9), developed by the Hospital for implementing an infrastructure based on the agnostic use of standards for the reuse of clinical information in research, a new OMOP repository will be built to overcome the limitations encountered in previous instances.

Methods

The first OMOP repository at H12O emerged within the European network HONEUR (10) and was based on data of patients with haematological diseases contained in the Hospital's i2b2 repository. This i2b2 repository contains standardized information from both Hospital and Madrid Primary Care (PC) (11) systems and it is updated daily. This first proposal arose as a methodological challenge, since the aim was to understand the capacities and compatibilities of each type of repository.

Following this same line, an OMOP was also created within the European EHDEN network, in which the whole i2b2 repository was transformed to the OMOP model.

The next OMOP instance created in the H12O was also within the EHDEN framework, in collaboration with Madrid PC, in the COVID-19 Rapid Collaboration Call launched during the pandemic. In this case, the EHR was used directly as source of information, because of the speed with which it was adapted to meet the information needs that emerged in this context, thanks to the work methodology that had been developed previously in terms of information standardisation (12).

OMOP's latest proposal, which aims to be the most complete and highest quality instance of H12O, is part of the INFOBANCO project, being developed by the Hospital. This project is based on the creation of a regional data network architecture for health system learning, conceived as a standardised health data repository. This architecture will function as a platform that provides services to clinicians, managers and researchers, enabling the combination of data from multiple sources, and which will be equipped with tools for data governance, collection, transformation, interrogation, visualisation and analysis for knowledge and decision support. Two of the main components of this infrastructure will be an openEHR

repository (13) and an OMOP repository. The openEHR repository will allow the homogeneous storage of patient’s EHR information in a standardised and semantically enriched format through the use of archetypes, together with the use of clinical terminologies (14). This openEHR repository will be the source of data of the OMOP, as it offers a complete answer to the needs of registration, persistence and exploitation. Figure 1 shows the main infrastructure of INFOBANCO project.

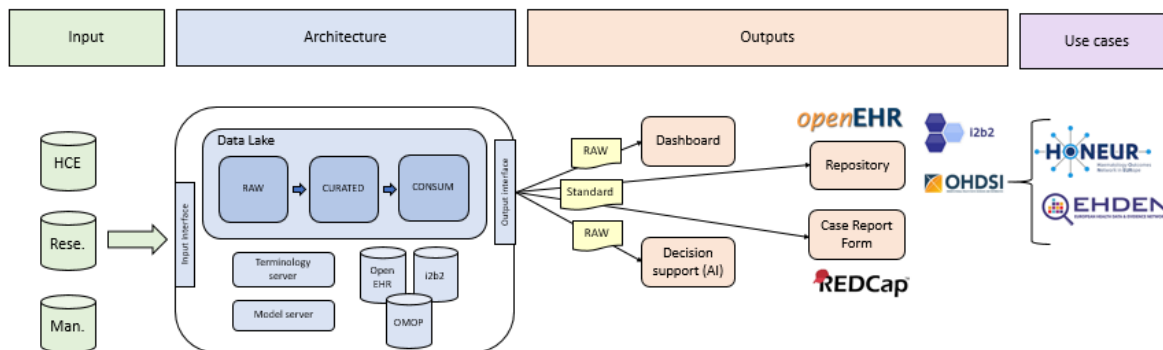


Figure 1. INFOBANCO infrastructure

Results

The OMOP instances populated by the i2b2 repository were useful from a methodological perspective, providing knowledge about limitations of each model. On the one hand, i2b2 is a more flexible repository as it is not conditioned to specific information domains or terminologies. However, this means that each i2b2 repository is particular to each organisation, making it more difficult data combination. In our case, due to the similarities between our i2b2 repository and the OMOP model, the transformations were not tedious. Nevertheless, it was observed that in this double transformation granularity and data quality were lost.

In the case of the OMOP created for COVID-19, based on EHR data, the data quality issue was overcome. Nevertheless, it was only useful for carrying out studies around COVID-19 due to the limitation applied in concepts and patients of this condition. This same issue was found in the HONEUR’s project OMOP.

The following table shows the amount of data loaded in each OMOP as well as the limitations found in each.

		HONEUR	i2b2	COVID
Data loaded	Person	1054	2,891,589	260,152
	Death	536	166,268	4571
	Condition_occurrence	53,283	14,187,889	1,940,499
	Drug_exposure	158,309	45,656,569	2,556,679
	Measurement	97,229	923,518,665	775,259
	Observation	3853	135,031	413,004

	Visit_occurrence	223,915	2,518,977	586,118
	Procedure_occurrence	3554	2,518,977	673,013
Data quality issue		Yes	Yes	No
Cohort restriction		Yes	No	Yes

Table 1. Overview of OMOP projects developed at H12O

One of the objectives of INFOBANCO project is to create an OMOP that overcomes the barriers already identified. To this end, the source of information will be the openEHR repository, also developed within this project, and which will be fed by all the information contained in the EHR of the Hospital and Madrid PC, as well as other information systems, without restrictions on specific cohorts. In this way, the OMOP data source will be based on archetypes and clinical terminologies and will therefore be of high quality. Data communication standards, such as CDISC ODM, HL7 FHIR, or openEHR itself, will be used to load data from openEHR to the OMOP repository.

Conclusion

The trajectory of projects and the learning gained has resulted in H12O INFOBANCO project. Its vision is to create an infrastructure for the persistence, exploitation and analysis of clinical data based on a standards-agnostic use, being the OMOP CDM one of the pillars of this platform. Thus, from an openEHR repository, which will provide a standardised EHR, the H12O OMOP instance will be fed without concept or cohort limitations. From this system and using OHDSI software, our researchers will be able to participate in studies with the guarantee that the data they are using meets quality criteria and that the information is accurate and unambiguous.

References/Citations

1. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med* [Internet]. 2016 Dec 8 [cited 2022 Jun 23];375(23):2293–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/27959688/>
2. Hospital 12 de octubre | Hospital Universitario 12 de Octubre [Internet]. [cited 2022 Jun 23]. Available from: <https://www.comunidad.madrid/hospital/12octubre/>
3. i2b2: Informatics for Integrating Biology & the Bedside [Internet]. [cited 2022 Jun 23]. Available from: <https://www.i2b2.org/>
4. Pedrera-Jiménez M, García-Barrio N, Cruz-Rojo J, Terriza-Torres AI, López-Jiménez EA, Calvo-Boyero F, et al. Obtaining EHR-derived datasets for COVID-19 research within a short time: a flexible methodology based on Detailed Clinical Models. *J Biomed Inform* [Internet]. 2021 Mar 1 [cited 2022 Jun 23];115. Available from: <https://pubmed.ncbi.nlm.nih.gov/33548541/>
5. Brat GA, Weber GM, Gehlenborg N, Avillach P, Palmer NP, Chiovato L, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. *npj Digit Med* 2020 31 [Internet]. 2020 Aug 19 [cited 2022 Jun 23];3(1):1–9. Available from: <https://www.nature.com/articles/s41746-020-00308-0>
6. Abbas A, Abdukahil SA, Abdulkadir NN, Abe R, Abel L, Absil L, et al. The value of open-source clinical science in pandemic response: lessons from ISARIC. *Lancet Infect Dis* [Internet]. 2021 Dec 1 [cited

2022 Jun 23];21(12):1623–4. Available from: <http://www.thelancet.com/article/S147330992100565X/fulltext>

7. European Health Data Evidence Network – ehden.eu [Internet]. [cited 2022 Jun 23]. Available from: <https://www.ehden.eu/>

8. OMOP Common Data Model – OHDSI [Internet]. [cited 2022 Jun 23]. Available from: <https://www.ohdsi.org/data-standardization/the-common-data-model/>

9. Infobanco - Compra Pública de Innovación [Internet]. [cited 2022 Jun 23]. Available from: <https://cpisanidadcm.org/infobanco/>

10. Home - HONEUR [Internet]. [cited 2022 Jun 24]. Available from: <https://portal.honeur.org/>

11. Gerencia de Atención Primaria | Comunidad de Madrid [Internet]. [cited 2022 Jun 23]. Available from: <https://www.comunidad.madrid/servicios/salud/gerencia-atencion-primaria>

12. Pedrera MP, Garcia N, Blanco A, Terriza AI, Cruz J, Lopez EA, et al. Use of EHRs in a Tertiary Hospital During COVID-19 Pandemic: A Multi-Purpose Approach Based on Standards. *Stud Health Technol Inform* [Internet]. 2021 Jul 1 [cited 2022 Jun 23];281:28–32. Available from: <https://pubmed.ncbi.nlm.nih.gov/34042699/>

13. openEHR Home [Internet]. [cited 2022 Jun 23]. Available from: <https://www.openehr.org/>

14. Pedrera-Jiménez M, Spanish Expert Group on EHR standards, Kalra D, Beale T, Muñoz-Carrero A, Serrano-Balazote P. Can OpenEHR, ISO 13606 and HL7 FHIR work together? An agnostic perspective for the selection and application of EHR standards from Spain. 2022 May 25 [cited 2022 Jun 23]; Available from: /articles/preprint/Can_OpenEHR_ISO_13606_and_HL7_FHIR_work_together_An_agnostic_perspective_for_the_selection_and_application_of_EHR_standards_from_Spain/19746484/1