Institutionalizing data interoperability and the application of common data models in a health data and research center: CIDACS' experience in Brazil

Valentina Martufi, Juliana Araújo Prata de Faria, Danilo Luis Cerqueira Dias, Elzo Pereira Pinto Junior, Roberto Carreiro, Pablo Ivan Ramos, Maurício L. Barreto

CIDACS-IGM/FIOCRUZ-BA

Background

The Center for Data and Knowledge Integration for Health (CIDACS) in Brazil, was created with the mission of contributing to the production of scientific knowledge, through the application of innovative methodologies, to expand the understanding of socio-economic and environmental determinants of health. Through the linkage of large databases, and subsequent creation of cohorts including millions of individuals, CIDACS conducts public health research that provides robust scientific evidence to support decision-makers. With the aim of consolidating actions that strengthen the globalized strategy for digital transformation in the field of health, through data interoperability and the promotion of common data models such as the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), CIDACS instituted the IDAF (Data Interoperability and Federated Analysis), as a new subgroup within its Data Platform team, responsible for the curation and production of data for CIDACS' research projects (Figure 1). This was an initiative that recognized the increasing global valorization of the application of the OMOP CDM for the generation of valuable and robust knowledge in the field of health, based on real-world data, as well as the steep learning curve for the utilization of the tools developed within the OHDSI (Observational Health Data Sciences and Informatics) community, and the importance of the contribution to their open-source development.

Methods

The IDAF is responsible for consolidating capacities, within CIDACS, to work with data interoperability and apply the OMOP CDM, for the realization of international research projects to study health phenomena that have a relevant global dimension.

The methodology employed by the IDAF includes mapping CIDACS' data to the OMOP CDM by using tools like White Rabbit and Rabbit in a Hat, implementing ETL (Extract, Transform, Load) processes, setting up OHDSI AWS (Amazon Web Services, Web API CDM by OHDSI), and configuring the Atlas tool for data analysis. The objectives of the IDAF encompass:

1- Consolidating and adopting open standards for health data (OMOP CDM Model)

2- Supporting the development of research projects utilizing common data models

3 - Establishing a solid foundation to promote data reuse (according to the FAIR principles)

4 - Consolidating interoperability processes and federated data analyses

5 - Training and disseminating information on interoperability and common data models.

Initially, the IDAF will apply the cloud-based version of the OHDSI tools, and therefore include within its competencies the creation of synthetic databases, in an effort to preserve privacy rights (Figure 2). At a later stage, it is envisioned that CIDACS will install in-house the tools necessary to implement the OMOP

CDM, and finally be able to apply them directly to the original, anonymized data.

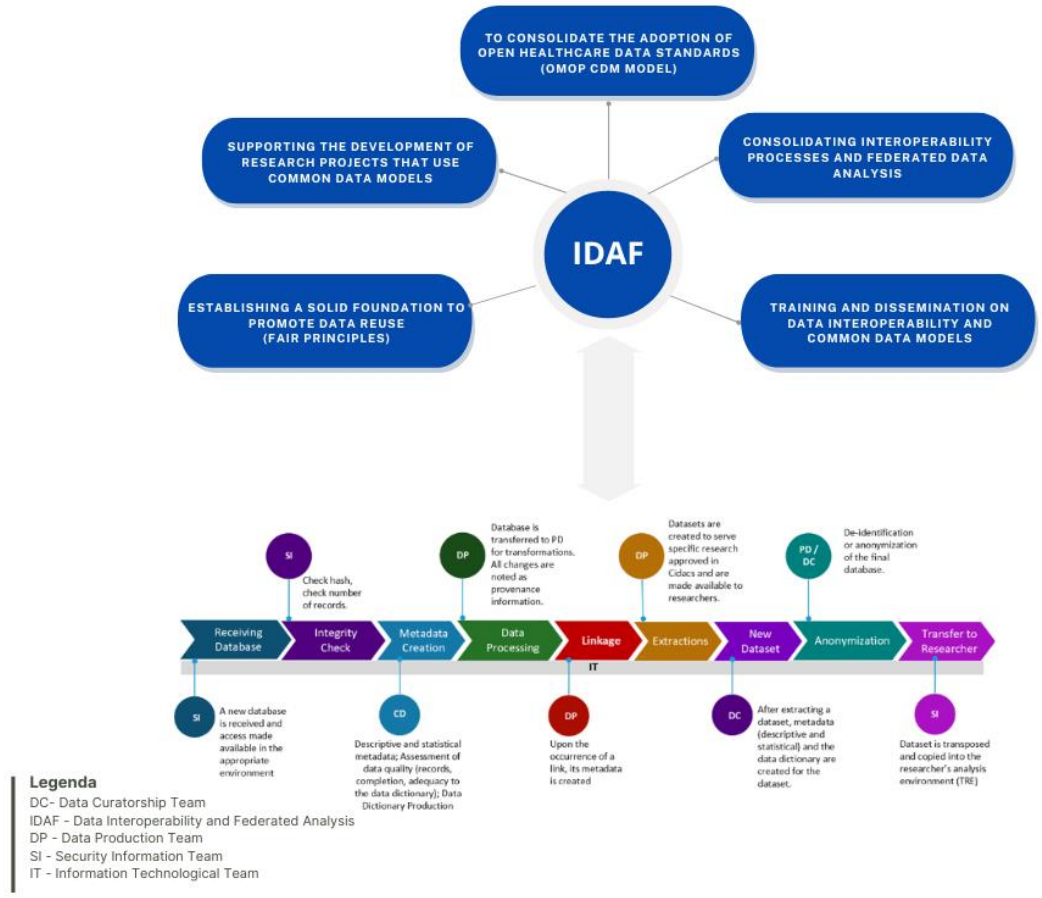

Figure 1. CIDACS' data flow and IDAF objectives

Figure 2. Steps for generating synthetic databases

Results

By using the OHDSI tools, the OMOP CDM will be applied to CIDACS' cohorts, as well as the other real-world data entrusted to CIDACS by the Brazilian government, to answer a vast variety of research

questions related to the health of the global population. Current efforts are being focused towards maternal and child health topics, such as gestational syphilis, tuberculosis in pregnancy, and pregnancy outcomes – employing data from compulsory disease notification databases, the national live birth registry, as well as datasets from social welfare programs. Nonetheless, it is envisioned that CIDACS' data may provide the opportunity to delve into a wide range of global health issues, including mental health, the impact of environmental factors and climate change on the population's health, as well as health systems characteristics and their influence on health outcomes.

Conclusion

Current nternational collaboration projects highlight IDAF's ability to work on a global scale, applying the OMOP CDM across CIDACS's large integrated databases, facilitating interoperability and federated analyses. Specifically, CIDACS' hopes to contribute to the OHDSI community the promotion of the inclusion of significant socioeconomic variables, to provide a more holistic understanding of health determinants, boosting the application of the OMOP CDM in the Global South.

References

1. Jörg Drechsler. Synthetic Datasets for Statistical Disclosure Control. Theory and Implementation. Springer, 2011.
2. Rubin, D. B. Discussion: Statistical disclosure limitation. Journal of Official Statistics. vol 9, p.p 462–468, 1993.
3. Surendra, H., Mohan H. S. A Review Of Synthetic Data Generation Methods For Privacy Preserving Data Publishing. International Journal of Scientific & Technology Reasearch. vol. 6, issue 3, 2017.
4. Elzo Pereira Pinto Junior, Priscilla Normando, Renzo Flores-Ortiz, Muhammad Usman Afzal, Muhammad Asaad Jamil, Sergio Fernandez Bertolin, Vinícius de Araújo Oliveira, Valentina Martufi, Fernanda de Sousa, Amir Bashir, Edward Burn, Maria Yury Ichihara, Maurício L Barreto, Talita Duarte Salles, Daniel Prieto-Alhambra, Haroon Hafeez, Sara Khalid, Integrating real-world data from Brazil and Pakistan into the OMOP common data model and standardized health analytics framework to characterize COVID-19 in the Global South, Journal of the American Medical Informatics Association, Volume 30, Issue 4, April 2023, Pages 643–655, https://doi.org/10.1093/jamia/ocac180