

GRELL ASCENSION MEETING 2022

CONGRESS CENTRE AND AUDITORIUM OF NAVARRE “BALUARTE”



Speed up the standardisation of clinical and epidemiological data – Luxembourg's National Cancer Registry data alignment to OMOP-CDM

Michael Schnell^{1,3}, Ania Cwojdzinska⁴, David Marcic^{1,3}, Sebastiano La Ferla⁴, Nils Christian⁴, Sophie Couffignal^{1,2}, Andreas Kremer⁴, Claudine Backes^{1,2}

¹ National Cancer Registry (Registre National du Cancer (RNC)), Luxembourg Institute of Health (LIH), Strassen Luxembourg

² Department of Precision Health, Luxembourg Institute of Health (LIH), Strassen, Luxembourg

³ Data Integration and Analysis Unit, Luxembourg Institute of Health, Strassen, Luxembourg

⁴ Information Technology for Translation Medicine (ITTM S.A.), Esch-sur-Alzette, Luxembourg

OBJECTIVES

Luxembourg's National Cancer Registry (in French: *Registre National du Cancer, RNC*) includes high-resolution population-based data, including data on treatments and follow-up, of all new cancer cases diagnosed and/or treated in Luxembourg. Established to assess incidence, mortality, and survival trends of cancer cases at national level, the RNC will gain further national and international benefits by linking its data with other data sources. In this study, the RNC explores to evolve its data structure towards a more harmonised health data format, namely the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM), to further facilitate the use of clinical and epidemiological cancer-related data to improve patient outcomes.

METHODS



Figure 1: OHDSI tools, presented in order in which they are used in the ETL building process. Source: <https://www.ohdsi.org/>

The harmonisation of the National Cancer registry data and the OMOP CDM requires translating NCR categories into selected group of fitting OMOP tables. IT tools (as listed in Figure 1) and processes are set up and implemented, as those developed by the international Observational Health Data Sciences and Informatics (OHDSI) open science collaboration, to convert the RNC data into the OMOP-CDM v5.4 – a model capturing patient data in the same way across different institutions. Using Python, R and OHDSI mapping tools, a pipeline was created transforming the data from the RNC relational model (containing four main variable tables: patient, tumor, treatment and multi-disciplinary team meeting - RCP) into the OMOP-CDM.

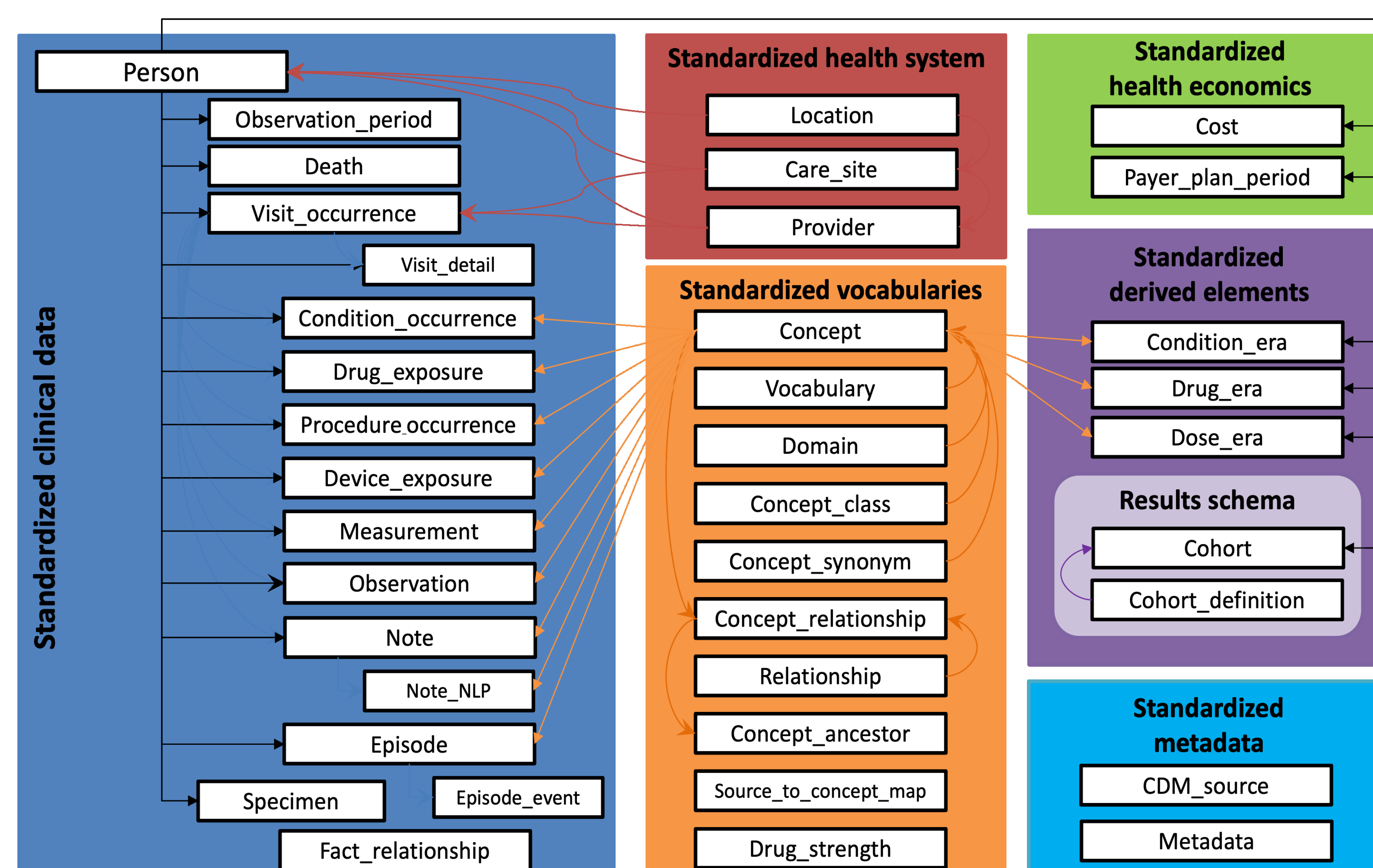


Figure 3: OMOP Common Data Model, v5.4. diagram. Source: <https://www.ohdsi.org/>

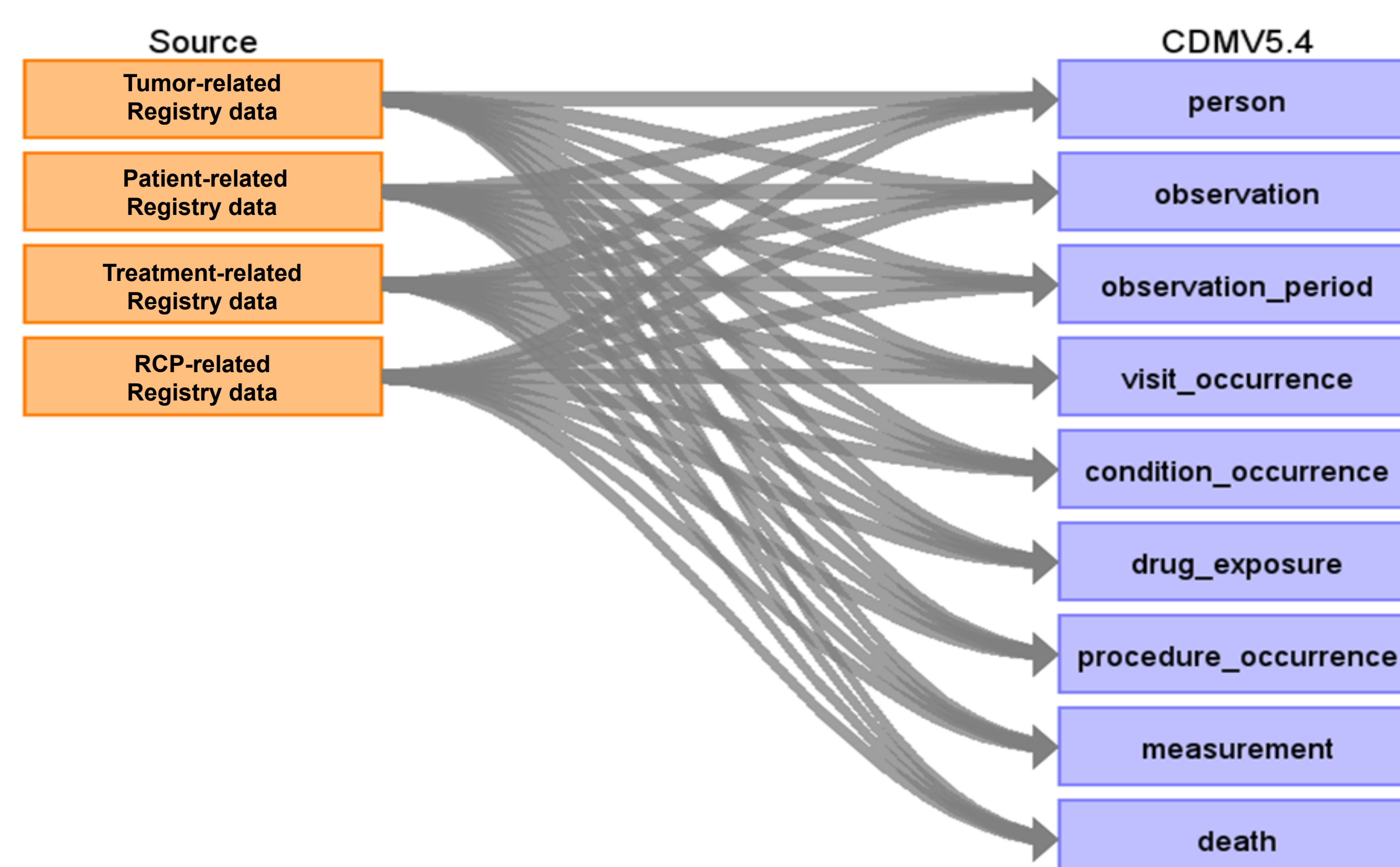


Figure 2: Presentation of the translation of the RNC data into the selected OMOP CDM tables. Source: Rabbit in a Hat output

RESULTS

During this process, about 600 variables and about 900 cancer cases from the current database are reviewed. In future, all new cancer cases collected by the RNC can be transformed into the same OMOP-CDM format by using this pipeline. The OHDSI software stack is also implemented, providing tools to the RNC describing and analysing data in the OMOP-CDM format.

CONCLUSIONS

With this new data format transformation, the RNC is broadening the use of its rich epidemiological cancer data, greatly facilitating the reuse of its data for research projects. The transformation process for the RNC is not only a database translation exercise, but a paradigm shift in the way the data is approached, as the RNC data are mainly tumor-centered, while the OMOP CDM is a person-centered model.

The framework facilitates secure and GDPR-conform collaboration and therefore enables future contributions to larger initiatives.

This data harmonisation is particularly interesting for small-sized registries or for rare diseases, as pooling of data from numerous sources increases the amount of available data.



The European Health Data & Evidence Network (EHDEN), having received funding from the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking (JU) under grant agreement No 806968, supports the data conversion process.