Connecting Computable Phenotypes with Multiple Health IT Standards Using the Phenoflow Library

Martin Chapman[1], Luke V. Rasmussen[2], Jennifer A. Pacheco[2], Vasa Curcin[1]
[1]King's College London, London, United Kingdom; [2]Northwestern University, Chicago, Illinois, USA

Background and Problem Statement

Electronic health record-based phenotype definitions are abstract outlines of the functionality required to autonomously extract individuals from a population who exhibit the same disease or condition. The *Phenoflow* library has been shown to improve the portability of phenotype definitions by autonomously converting standardised definitions to a computable form for local execution[1]. However, in order to be executed locally, computable phenotypes also need to interoperate with the various health standards that might be adopted across different sites, which is often not the case. For example, the Observational Health Data Sciences and Informatics (OHDSI) tools also allow researchers to export computable forms of cohort definitions in the form of SQL statements[2], however they are all tied to the Observational Medical Outcomes Partnership (OMOP) common data model.

Phenoflow: System Design and Innovative Features

One of the main reasons many computable phenotypes are directly tied to a specific data source is the fact that the extraction of data and the application of definition logic are an atomic process. This is true of SQL implementations that simultaneously extract patient data and aim to identify whether those patients have the condition of interest. Instead, we leverage the modular standard adopted by Phenoflow to make these two activities distinct: the first step in a phenotype is designated as a *connector*, designed solely to extract data from a particular source. Owing to a standard interface between steps and automated mapping between coding schemes, connectors can be swapped without impacting the remaining logic of the definition (Figure 1). Thus, the same logic can be used with different data sources, enabling definitions to be reused across sites, irrespective of the standard used.

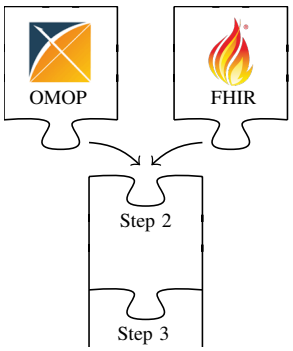

Figure 1: Phenoflow phenotype structure

In practice, when a new definition is added to the Phenoflow library, four connectors are automatically generated: Fast Healthcare Interoperability Resources (FHIR), OMOP, i2b2 and CSV (the generic default). Each connector is customised so that it can interface with the other steps of the definition. As such, prior to downloading a definition through the library, a user selects the connector that matches the standard adopted by their target dataset. Once downloaded as an executable workflow, the connector can be modified locally, in order to add details such as login credentials. In addition, using Phenoflow's programming interface, users can submit custom connectors to the library, allowing them to execute definitions against standards not currently supported by the library.

Deployment

The Phenoflow library is currently live (https://kclhi.org/phenoflow) and hosts 334 phenotype definitions at the time of writing. These definitions have been contributed from sources across the UK and the US, and each can be customised for different standards. In addition, the source of the library is available (https://github.com/kclhi/phenoflow), allowing users to run local copies of the library (within secure environments). Connectors containing proprietary information can thus be kept private, while still attaching to publicly available definitions. Bespoke connectors are currently being developed by a number of different sites.

References

[1] Chapman M, Rasmussen L, Pacheco J, Curcin V. Phenoflow: A Microservice Architecture for Portable Workflow-based Phenotype Definitions. In: Proceedings of AMIA Joint Summits on Translational Science; 2021. p. 142-51.

[2] Observational Health Data Sciences and Informatics. The Book of OHDSI. Independent; 2019.