# CQF Phenotype Execution and Modeling Architecture

## Pilot Background

The identification of patient cohorts for clinical and genomic research is a costly and time-consuming process. This bottleneck adversely affects public health by delaying research findings, and in some cases by making research costs prohibitively high. To address this issue, leveraging electronic health records (EHRs) for identifying patient cohorts has become an increasingly attractive option. With the rapidly growing adoption of EHR systems due to Meaningful Use, and linkage of EHRs to research biorepositories, evaluating the suitability of EHR data for clinical and translational research is becoming ever more important, with ramifications for genomic and observational research, clinical trials, and comparative effectiveness studies. A key component for identifying patient cohorts in the EHR is to define inclusion and exclusion criteria that algorithmically select sets of patients based on stored clinical data. This process is commonly referred to as “EHR-driven phenotyping”. Phenotypes are defined over both structured data (demographics, diagnoses, medications, lab measurements) as well as unstructured clinical text (radiology reports, encounter notes, discharge summaries). Phenotyping logic can be quite complex, and typically includes both Boolean and temporal operators applied to multiple clinical events. In general, the phenotyping algorithm development process is a multi-disciplinary team effort, including clinicians, domain experts, and informaticians, and is operationalized as database queries and software, customized to the local EHR environment. The typical way to share phenotyping algorithms across institutions is through the use of informal free text descriptions of algorithm logic, possibly augmented with graphical flowcharts and simple lists of structured codes. This is due to the lack of a widely accepted and standards-based formal information model for defining phenotyping algorithms. However, implementing a phenotyping algorithm from a free-text description is itself an error-prone and time-consuming process, due to the inherent ambiguities of free text as well as the necessity for human intermediaries to map algorithmic criteria expressed as free text to database queries and code.

## Pilot Goal

The proposed project will design, build and promote an open-access community infrastructure for standards-based development and sharing of phenotyping algorithms, as well as provide tools and resources for investigators, researchers and their informatics support staff to implement and execute the algorithms on native EHR data.

Please contact Will Thompson for further details: wkt@northwestern.edu . The results of the Phase III pilot are available here:

CQF Phases II Pilot SurveyFINALwillthompson.pdf