The Phenotype Execution and Modeling Architecture (PhEMA) – A Standards-Based Composition of Software for Phenotype Algorithm Development

Luke V. Rasmussen[1]; Richard C. Kiefer[2], Huan Mo, MD, MS[3], William K. Thompson, PhD[4], Guoqian Jiang, MD, PhD[2], Jennifer A. Pacheco[1], Peter Speltz[3], Jie Xu, MS[1], Joshua C. Denny, MD, MS[3], Enid Montague, PhD[1], Jyotishman Pathak, PhD[2]

[1]Northwestern University, Chicago, IL; [2]Mayo Clinic, Rochester, MN; [3]Vanderbilt University, Nashville, TN; [4]NorthShore University HealthSystem, Evanston, IL;

Abstract

The use of a standards-based, modular architecture for the development of phenotype algorithms enhances the interoperability of electronic health record (EHR) systems to allow the dissemination of algorithms across institutions. Here we describe the implementation of previously proposed modules of a comprehensive solution for the development, validation, execution and dissemination of EHR-driven phenotype algorithms.

Introduction: The use of electronic health records (EHRs) for research has been a focus of the biomedical informatics research community, with many researchers and consortia describing methodologies for effective use of EHR data, as well as challenges discovered along the way[1, 2]. To aid in the development of phenotype algorithms using clinical data, several software solutions have been provided to the informatics community, including the Informatics for Integrating Biology and the Bedside (i2b2)[3] and Observational Health Data Sciences and Informatics (OHDSI)[4].

We previously described the Phenotype Execution and Modeling Architecture (PhEMA) – a modular software architecture that relies on components that interoperate using standard formats and interfaces[5]. These components are logically separated to complete a specific task, such as executing an algorithm and collecting the results. Having referenced existing systems in development of our proposed architecture, we noted limitations and gaps that we sought to address –specifically around increasing the use of standards and providing flexibility in configuring components to meet each institution's needs.

Methods: The PhEMA development team has identified available software systems for many of the proposed seven architecture components (Library for Artifacts, Authoring, Clinical Data Repository, Execution, Validation, Data Model Services and Terminology Services) for EHR phenotyping, and developed new software in the absence of existing solutions. During development, we designed the systems around concrete interfaces and specifications, but were agnostic to the choice of a particular programming language or development environment.

Results: The PhEMA solution includes one or more implemented components, as shown in Figure 1.
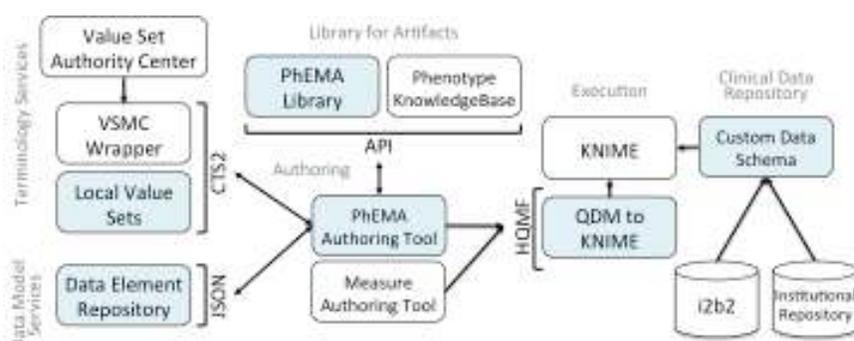

Figure 1. Implemented components of the Phenotype Execution and Modeling Architecture (PhEMA). Blue boxes indicate newly developed software, while white are existing solutions.

A demonstration system and source code are available from the project website (http://projectphema.org). Briefly, each of the implemented solutions is as follows:

Terminology Services – Our use of the Quality Data Model (QDM) relies on value sets (collections of terms to represent concepts, derived from standard vocabularies). We not only provide users with read access to the NLM-hosted Value Set Authority Center (VSAC) for existing value sets, but also provide a separate read/write instance of a repository for custom value sets. Both repositories leverage the Common Terminology Services 2 (CTS2) standard[6] (the VSAC CTS2 service utilizes a CTS2 wrapper [VSMC]). The authoring tool may be configured to use one or both repositories during installation.

Data Model Services – The Data Element Repository (DER) has been developed to provide RDF-based model representations via a SPARQL endpoint[7]. A set of REST services were implemented to return the requested terms so applications would not have to write their own queries. Currently, the repository provides computable representations of data elements extracted from QDM and HL7 Fast Healthcare Interoperability Resources (FHIR), and has been extended to also include value set recommendations taken from existing eCQMs.

Library for Artifacts – Two options are currently available. The first is a lightweight library component, which is a simple MongoDB-backed store. The second integrates with the Phenotype KnowledgeBase (PheKB, http://www.phekb.org/). The former allows adopters of the architecture to have a standalone library to use for internal algorithm development. The latter facilitates broader sharing and dissemination of algorithm definitions by using an established, publicly accessible website. Both use an API that was defined for this project.

Authoring – We developed a web application that allows the graphical composition of algorithm logic. It utilizes the Terminology Services and Data Model Services to provide the definitions of the models used in authoring, including properties, value sets and terms. The use of external standard services has allowed the authoring tool to quickly include new models, such as FHIR. Given that the output of the Authoring component is an XML document following the HL7 HQMF standard (with HQMF being an instantiation of the QDM), we are also able to support the Measure Authoring Tool (MAT) as another Authoring component in this architecture.

Execution –We position KNIME as a customizable workflow management system for complex phenotype algorithms to orchestrate multiple executions, although it also offers execution environments for extracted EHR data in the absence of alternatives. The eM2Kn (eMeasures to KNIME) translator[8] solution provides executable workflows for the algorithms, and computes the extracted EHR data in the KNIME environment.

Clinical Data Repository – The KNIME workflows from eM2Kn offer customizable interfaces for non-standardized local data repositories. This solution still requires a SQL-level mapping in the KNIME workflow to the local schema. Ongoing work will provide a more automated execution service against the i2b2 platform.

Discussion: The use of multiple solutions for many of the architecture components provides potential users with options in how they choose to use and configure the system at their institution. It provides a balance of using publicly accessible components (VSAC, DER, PheKB), obviating the need to install all of these locally. Likewise, the flexibility within the Execution component gives users the option of using an existing data warehouse schema (at the cost of requiring mapping to that schema), or leveraging an existing i2b2 implementation (as many institutions have already or are in the process of adopting this platform). Furthermore, the reliance on standards over technology allowed teams to work more effectively on solutions while ensuring interoperability.

Conclusion: The use of a modular architecture based on standards provides not only a software solution that may be adopted by institutions to more readily facilitate phenotype algorithm development, but does so in a way that collaborates with existing initiatives and solutions. This allows an institution to optimize their phenotyping infrastructure based on their needs and use cases.

References
1. Denny JC. Chapter 13: Mining electronic health records in the genomics era. PLoS Comput Biol. 2012;**8**(12):e1002823.
2. Rasmussen LV. The electronic health record for translational research. J Cardiovasc Transl Res. 2014 Aug;**7**(6):607-14.
3. Murphy SN, Mendis M, Hackett K, et al. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. AMIA Annu Symp Proc. 2007:548-52.
4. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;**216**:574-8.
5. Rasmussen LV, Kiefer RC, Mo H, et al. A Modular Architecture for Electronic Health Record-Driven Phenotyping. AMIA Jt Summits Transl Sci Proc. 2015;**2015**:147-51.
6. Common Terminology Services 2. [cited September 23, 2015]; Available from: http://www.omg.org/spec/CTS2/1.0/
7. Jiang G, Solbrig HR, Kiefer R, et al. A Standards-based Semantic Metadata Repository to Support EHR-driven Phenotype Authoring and Execution. Stud Health Technol Inform. 2015;**216**:1098.
8. Mo H, Pacheco JA, Rasmussen LV, et al. A Prototype for Executable and Portable Electronic Clinical Quality Measures Using the KNIME Analytics Platform. AMIA Jt Summits Transl Sci Proc. 2015;**2015**:127-31.