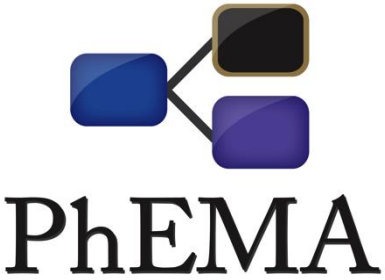


MAYO CLINIC

VANDERBILT UNIVERSITY MEDICAL CENTER

NORTHWESTERN UNIVERSITY

Weill Cornell Medicine

The Phenotype Execution and Modeling Architecture (PhEMA)

http://www.projectphema.org
http://github.com/phema

Technical System/Platform Demonstration

at MCBK meeting

July 19, 2019

Jennifer A. Pacheco

japacheco@northwestern.edu

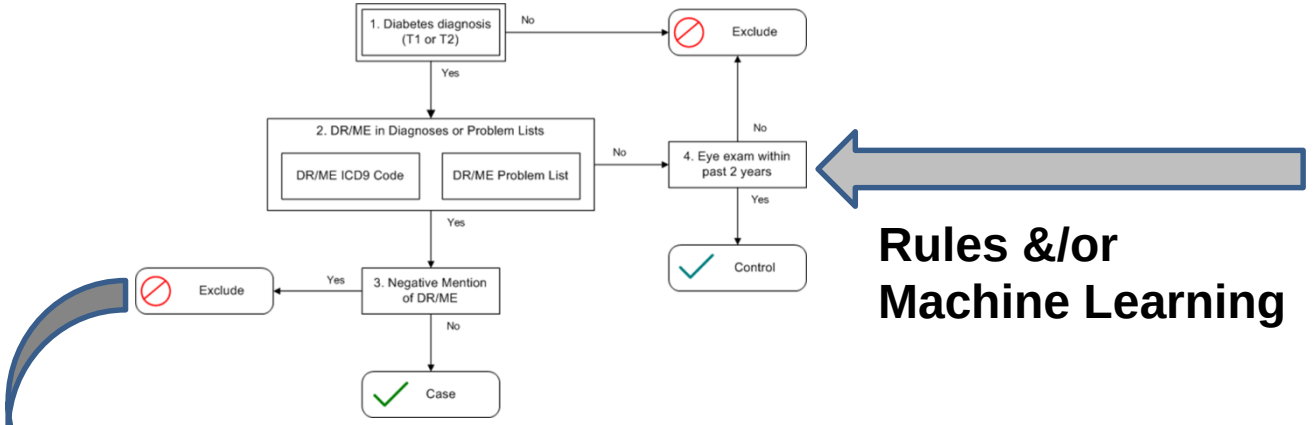
Outline

- What is PhEMA:
  - Background/Overview inc. some of our publications explaining in more detail (w/ PubMED IDs)
  - How PhEMA Authors & Executes phenotype algorithms
  - Work in progress
- Intro. & demo. of :
  - Authoring &
  - Executing w/ :
    - command line
    - GUI in progress (prototype)

  … a phenotype alg.

PhEMA

EHR-driven Phenotyping Algorithms



Rules &/or Machine Learning

Phenotype Algorithm

Evaluation

Visualization

Transform

Transform

Data

NLP &/or SQL

Mappings

PhEMA

Key lessons learned from eMERGE

Electronic Medical Records for Genetic Research: Results of the eMERGE Consortium

Abel N. Kho,[1,2*] Jennifer A. Pacheco,[1] Peggy L. Peissig,[3] Luke Rasmussen,[3] Katherine M. Newton,[4,5] Noah Weston,[4] Paul K. Crane,[6] Jyotishman Pathak,[7] Christopher G. Chute,[7] Suzette J. Bielinski,[7] Iftikhar J. Kullo,[8] Rongling Li,[9] Teri A. Manolio,[9] Rex L. Chisholm,[1] Joshua C. Denny[10]

- Algorithm design is a non-trivial an iterative process that requires experts, & manual chart review to confirm
- Standardized data representation key for transportability, …but some structured data (ICD Diagnoses) not as reliable as NLP (Natural Language Processing) of clinical records
- Initial plan of sharing SQL, XML, near executable pseudo-code didn't work: sharing flowcharts etc. helped… but still reliant on human interpretation & definitions often ambiguous
- … **need a way to represent algorithms as both executable & human readable**

[Kho et al. Sc. Trans. Med 2011; 3(79): 1-7].
https://www.ncbi.nlm.nih.gov/pubmed/**21508311**

PhEMA

Review and evaluation of electronic health records-driven phenotype algorithm authoring tools for clinical and translational research

RECEIVED 5 February 2015
REVISED 11 May 2015
ACCEPTED 15 May 2015
PUBLISHED ONLINE FIRST 29 July 2015

Jie Xu[1], Luke V Rasmussen[2], Pamela L Shaw[3], Guoqian Jiang[4], Richard C Kiefer[4], Huan Mo[5], Jennifer A Pacheco[6], Peter Speltz[5], Qian Zhu[7], Joshua C Denny[5], Jyotishman Pathak[4], William K Thompson[8], Enid Montague[1]

ABSTRACT

Objective To review and evaluate available software tools for electronic health record–driven phenotype authoring in order to identify gaps and needs for future development.

Materials and Methods Candidate phenotype authoring tools were identified through (1) literature search in four publication databases (PubMed, Embase, Web of Science, and Scopus) and (2) a web search. A collection of tools was compiled and reviewed after the searches. A survey was designed and distributed to the developers of the reviewed tools to discover their functionalities and features.

Results Twenty-four different phenotype authoring tools were identified and reviewed. Developers of 16 of these identified tools completed the evaluation survey (67% response rate). The surveyed tools showed commonalities but also varied in their capabilities in algorithm representation, logic functions, data support and software extensibility, search functions, user interface, and data outputs.

Discussion Positive trends identified in the evaluation included: algorithms can be represented in both computable and human readable formats; and most tools offer a web interface for easy access. However, issues were also identified: many tools were lacking advanced logic functions for authoring complex algorithms; the ability to construct queries that leveraged un-structured data was not widely implemented; and many tools had limited support for plug-ins or external analytic software.

Conclusions Existing phenotype authoring tools could enable clinical researchers to work with electronic health record data more efficiently, but gaps still exist in terms of the functionalities of such tools. The present work can serve as a reference point for the future development of similar tools.

https://www.ncbi.nlm.nih.gov/pubmed/26224336

PhEMA

Desiderata for computable representations of electronic health records-driven phenotype algorithms

RECEIVED 9 February 2015
REVISED 20 June 2015
ACCEPTED 24 June 2015
PUBLISHED ONLINE FIRST 5 September 2015

Huan Mo[1], William K Thompson[2], Luke V Rasmussen[3], Jennifer A Pacheco[4], Guoqian Jiang[5], Richard Kiefer[5], Qian Zhu[6], Jie Xu[7], Enid Montague[7], David S Carrell[8], Todd Lingren[9], Frank D Mentch[10], Yizhao Ni[9], Firas H Wehbe[3], Peggy L Peissig[11], Gerard Tromp[12], Eric B Larson[8], Christopher G Chute[13], Jyotishman Pathak[5,14], Joshua C Denny[1,15]
Other collaborators: Peter Speltz[1], Abel N Kho[7], Gail P Jarvik[16,17], Cosmin A Bejan[1], Marc S Williams[18], Kenneth Borthwick[19], Terrie E Kitchner[11], Dan M Roden[15,20], Paul A Harris[1]

ABSTRACT

Background Electronic health records (EHRs) are increasingly used for clinical and translational research through the creation of phenotype algorithms. Currently, phenotype algorithms are most commonly represented as noncomputable descriptive documents and knowledge artifacts that detail the protocols for querying diagnoses, symptoms, procedures, medications, and/or text-driven medical concepts, and are primarily meant for human comprehension. We present desiderata for developing a computable phenotype representation model (PheRM).

Methods A team of clinicians and informaticians reviewed common features for multisite phenotype algorithms published in PheKB.org and existing phenotype representation platforms. We also evaluated well-known diagnostic criteria and clinical decision-making guidelines to encompass a broader category of algorithms.

Results We propose 10 desired characteristics for a flexible, computable PheRM: (1) structure clinical data into queryable forms; (2) recommend use of a common data model, but also support customization for the variability and availability of EHR data among sites; (3) support both human-readable and computable representations of phenotype algorithms; (4) implement set operations and relational algebra for modeling phenotype algorithms; (5) represent phenotype criteria with structured rules; (6) support defining temporal relations between events; (7) use standardized terminologies and ontologies, and facilitate reuse of value sets; (8) define representations for text searching and natural language processing; (9) provide interfaces for external software algorithms; and (10) maintain backward compatibility.

Conclusion A computable PheRM is needed for true phenotype portability and reliability across different EHR products and healthcare systems. These desiderata are a guide to inform the establishment and evolution of EHR phenotype algorithm authoring platforms and languages.

https://www.ncbi.nlm.nih.gov/pubmed/26342218

PhEMA

Sharma et al. BMC Medical Informatics and Decision Making 2019, 19(Suppl 3):78
https://doi.org/10.1186/s12911-019-0786-z

BMC Medical Informatics and Decision Making

RESEARCH Open Access

Developing a portable natural language processing based phenotyping system

Himanshu Sharma[1†], Chengsheng Mao[2†], Yizhen Zhang[2], Haleh Vatani[1], Liang Yao[2], Yizhen Zhong[2], Luke Rasmussen[2], Guoqian Jiang[3], Jyotishman Pathak[4] and Yuan Luo[2*]

From The Sixth IEEE International Conference on Healthcare Informatics (ICHI 2018)
New York, NY, USA. 4-7 June 2018

Abstract

Background: This paper presents a portable phenotyping system that is capable of integrating both rule-based and statistical machine learning based approaches.

Methods: Our system utilizes UMLS to extract clinically relevant features from the unstructured text and then facilitates portability across different institutions and data systems by incorporating OHDSI's OMOP Common Data Model (CDM) to standardize necessary data elements. Our system can also store the key components of rule-based systems (e.g., regular expression matches) in the format of OMOP CDM, thus enabling the reuse, adaptation and extension of many existing rule-based clinical NLP systems. We experimented with our system on the corpus from i2b2's Obesity Challenge as a pilot study.

Results: Our system facilitates portable phenotyping of obesity and its 15 comorbidities based on the unstructured patient discharge summaries, while achieving a performance that often ranked among the top 10 of the challenge participants.

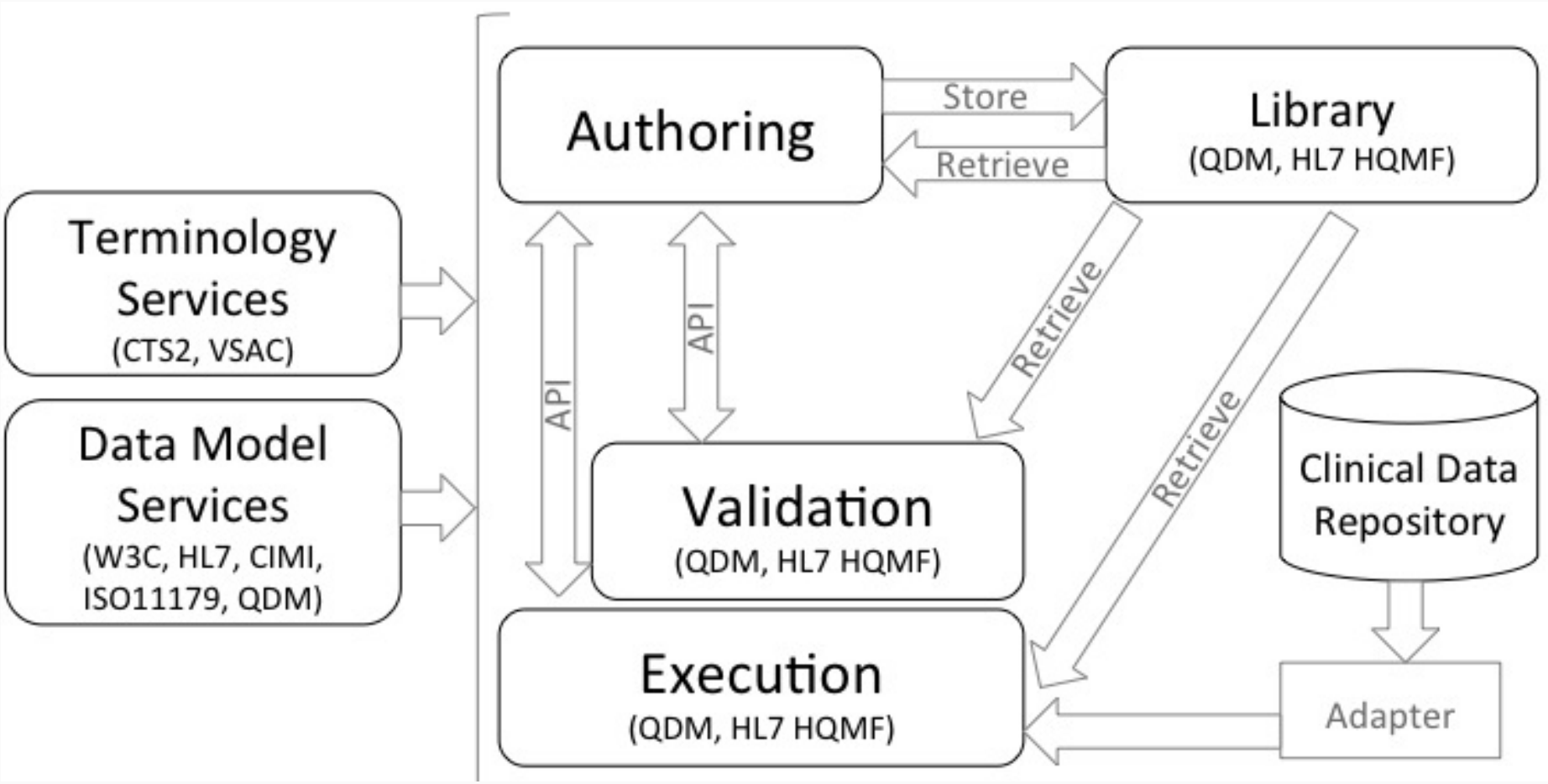
Conclusion: Our system of standardization enables a consistent application of numerous rule-based and machine learning based classification techniques downstream across disparate datasets which may originate across different institutions and data systems.

https://www.ncbi.nlm.nih.gov/pubmed/30943974

PhEMA

A Modular Architecture for Electronic Health Record-Driven Phenotyping

Luke V. Rasmussen[1]; Richard C. Kiefer[2], Huan Mo, MD, MS[3], Peter Speltz[3], William K. Thompson, PhD[4], Guoqian Jiang, MD, PhD[2], Jennifer A. Pacheco[1], Jie Xu, MS[1], Qian Zhu, PhD[5], Joshua C. Denny, MD, MS[3], Enid Montague, PhD[1], Jyotishman Pathak, PhD[2]



[Rasmussen et al. AMIA 2015] https://www.ncbi.nlm.nih.gov/pubmed/26306258

PhEMA

Step 1 of 2 – Authoring

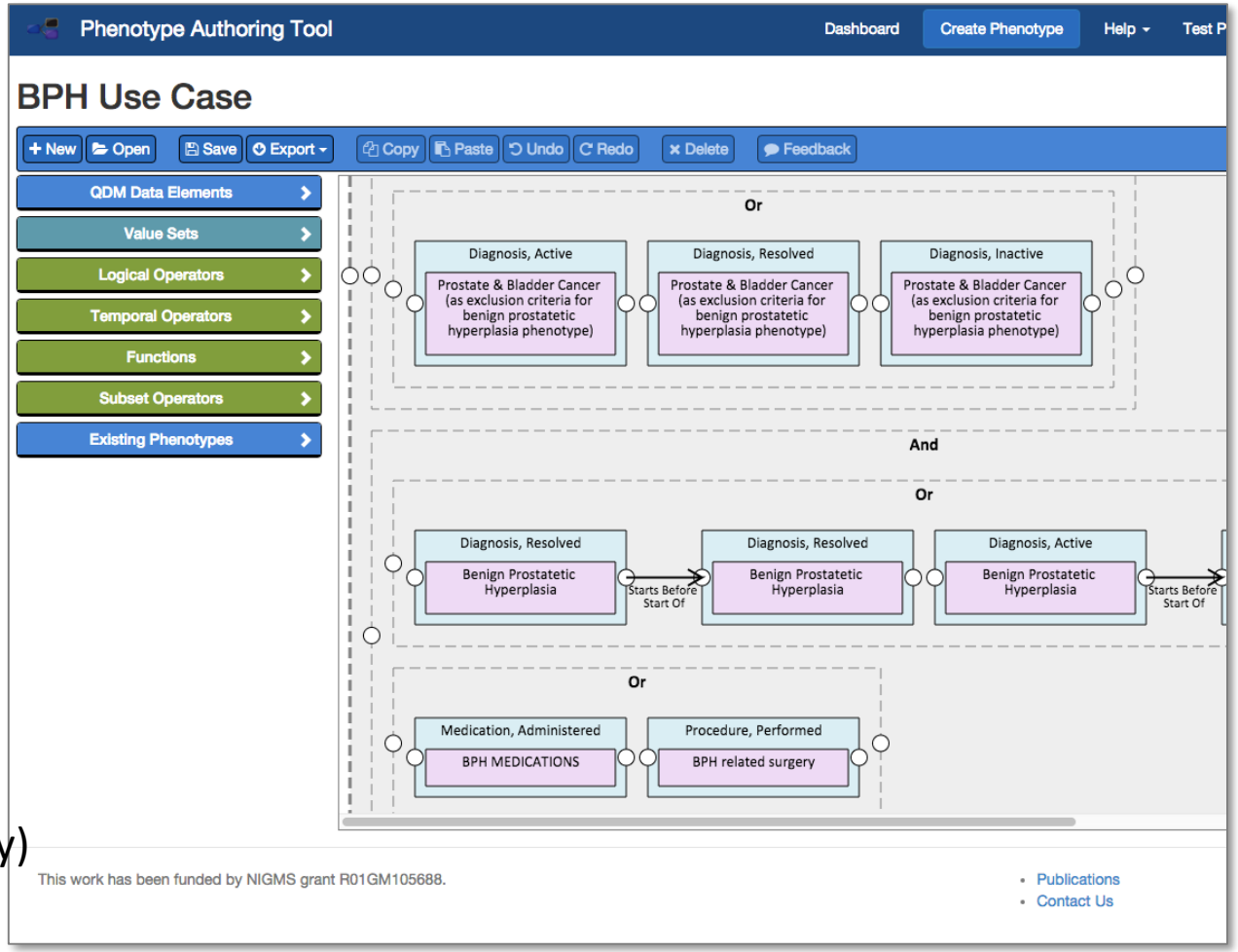
- Create using standard representation of both clinical data & phenotype algorithm
  - The entities (e.g. Diagnosis)
    - QDM (via Data Element Repository)
  - How the entities are defined
    - Value Sets
      - Value Set Authority Center
      - Custom value sets stored in local repository
  - The final artifact: both human readable & executable, & publically avail.
    - HQMF (via health-data-standards library)
    - Or CQL

PhEMA

Phenotype Authoring Tool

Search FHIR repositories

Elements loaded from DER (Data Element Repository)



Compose definition from elements, and save as HQMF XML, JSON, etc.

https://projectphema.org:8181

PhEMA

Developing a data element repository to support EHR-driven phenotype algorithm authoring and execution

Guoqian Jiang [a,*], Richard C. Kiefer [a], Luke V. Rasmussen [b], Harold R. Solbrig [a], Huan Mo [c], Jennifer A. Pacheco [d], Jie Xu [e], Enid Montague [e,f], William K. Thompson [f], Joshua C. Denny [c,g], Christopher G. Chute [h], Jyotishman Pathak [i]

G. Jiang et al. / Journal of Biomedical Informatics 62 (2016) 232–242

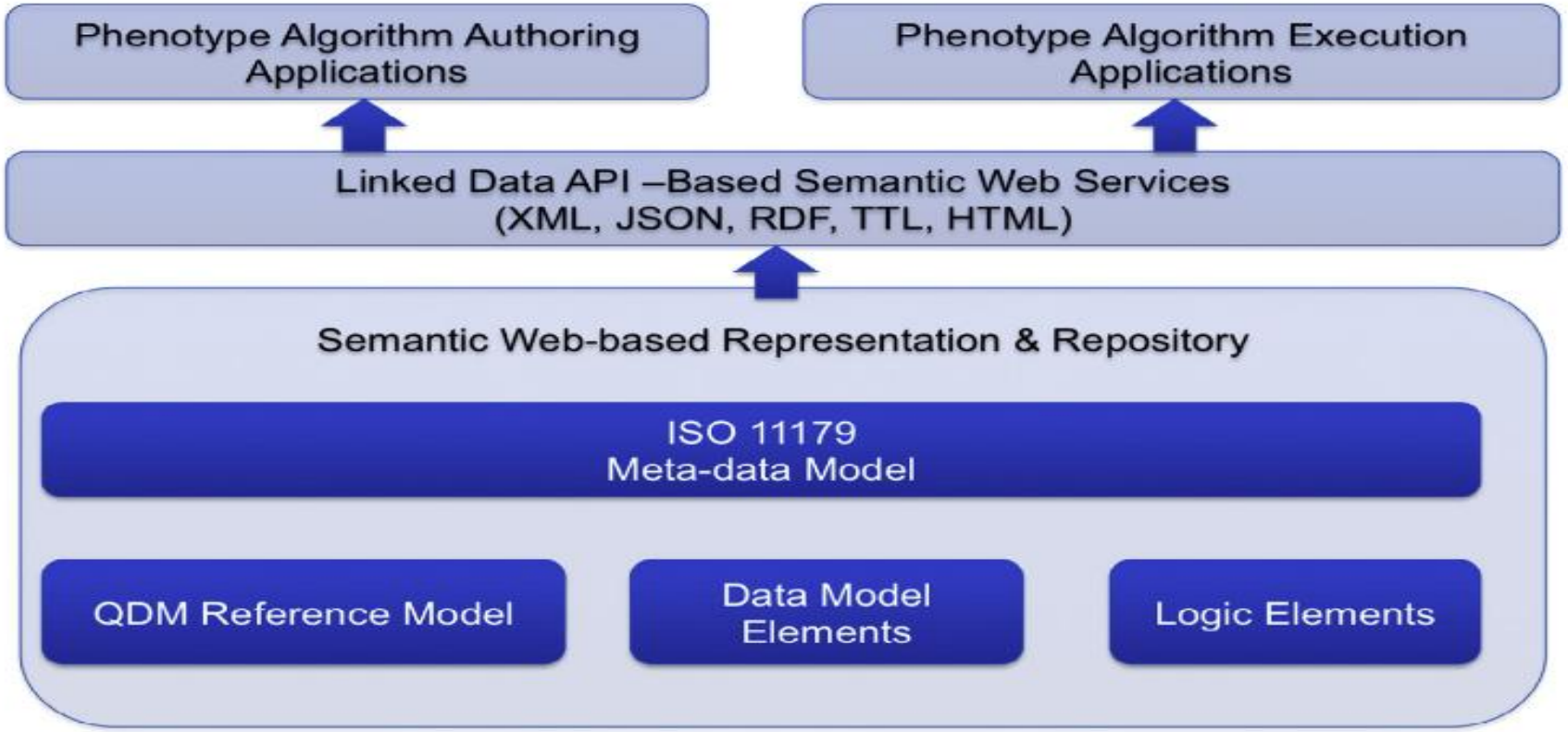
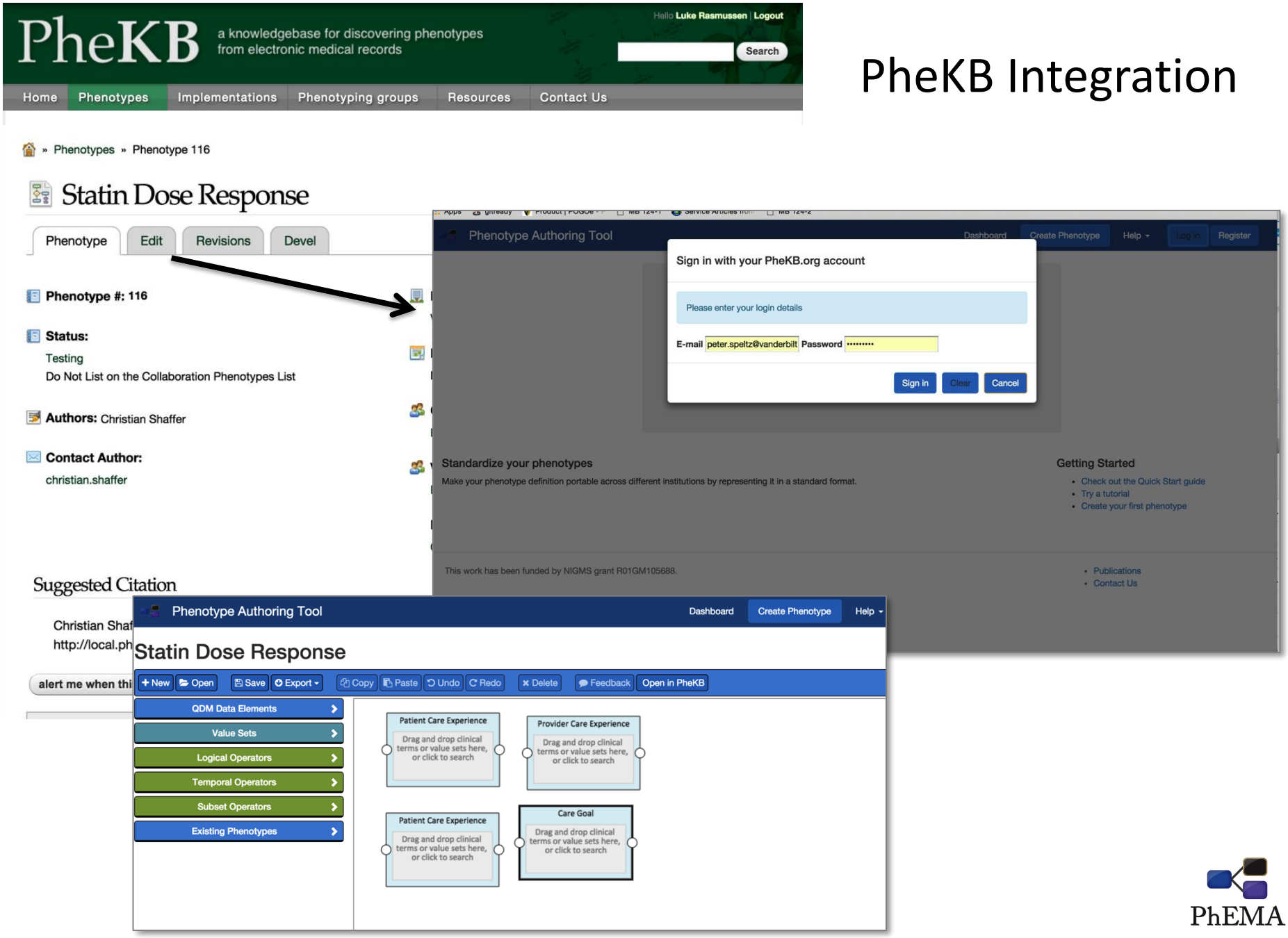


Fig. 3. System architecture.

https://www.ncbi.nlm.nih.gov/pubmed/**27392645**

PhEMA

PheKB Integration



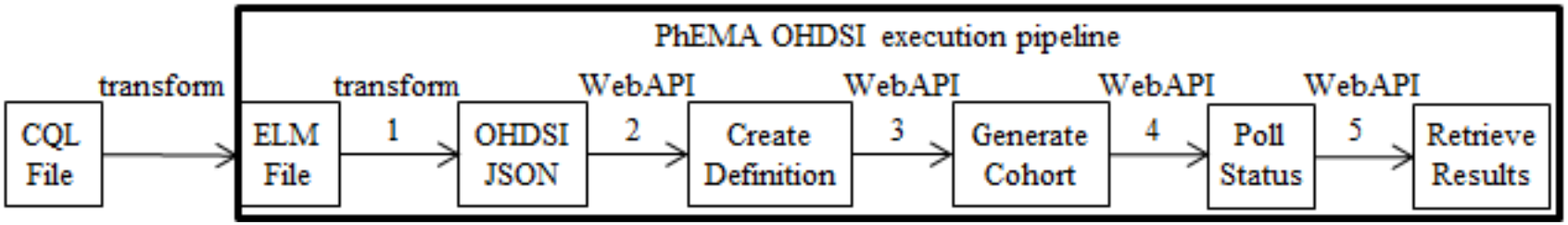
PhEMA

Step 2 of 2 - Execution

Instructions:
https://github.com/PheMA/phema-executer

Version I'm using:
https://github.com/PheMA/phema-executer/releases/tag/0.12

- Execute against i2b2 or OHDSI
  - Currently via command line
  - GUI under development
- CQL (Clinical Quality Language, via ELM (Expression Logical Model)) or HQMF translated into i2b2 & OHDSI web API calls
  - i2b2: translated to i2b2 XML
  - OHSDI: translated to OHDSI JSON:



PhEMA

Work in progress

- Ability to represent NLP & ML (machine learning) in phenotype algs.
- Supporting integration w/ CDMs:
  - Adding OHDSI OMOP export for OHDSI
  - Improved for i2b2
- Use of CQL (Clinical Quality Language)
- Focus groups and usability studies with stakeholders (eMERGE, PCORnet, others (*maybe some of you?**)

PhEMA

JAMIA
A SCHOLARLY JOURNAL OF INFORMATICS IN HEALTH AND BIOMEDICINE

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

A case study evaluating the portability of an executable computable phenotype algorithm across multiple institutions and electronic health record environments

Jennifer A Pacheco ✉, Luke V Rasmussen, Richard C Kiefer, Thomas R Campion, Peter Speltz, Robert J Carroll, Sarah C Stallings, Huan Mo, Monika Ahuja, Guoqian Jiang ...

Journal of the American Medical Informatics Association, Volume 25, Issue 11, November 2018, Pages 1540–1546, https://doi.org/10.1093/jamia/ocy101

Published: 16 August 2018

Abstract

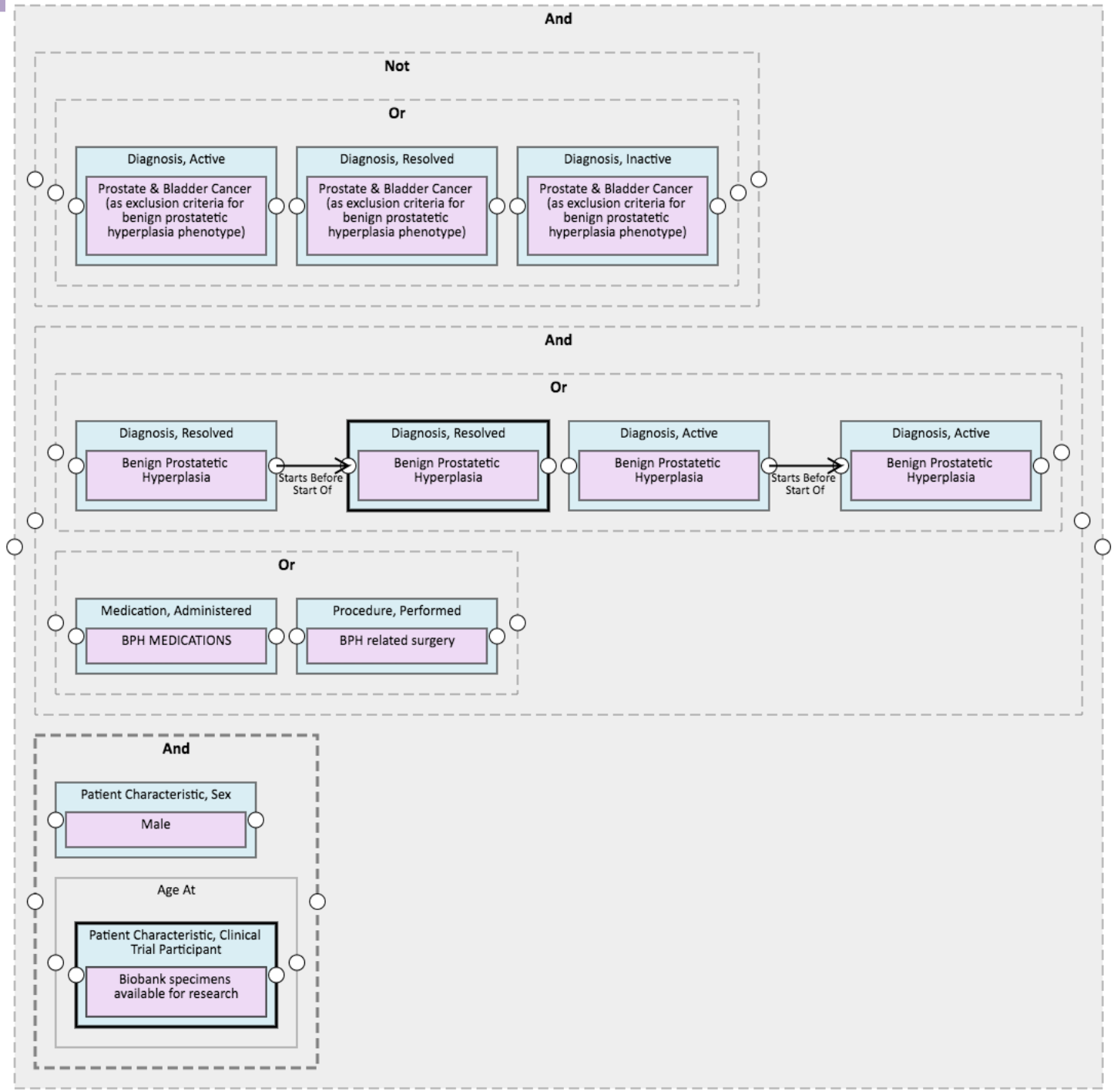
Electronic health record (EHR) algorithms for defining patient cohorts are commonly shared as free-text descriptions that require human intervention both to interpret and implement. We developed the Phenotype Execution and Modeling Architecture (PhEMA, http://projectphema.org) to author and execute standardized computable phenotype algorithms. With PhEMA, we converted an algorithm for benign prostatic hyperplasia, developed for the electronic Medical Records and Genomics network (eMERGE), into a standards-based computable format. Eight sites (7 within eMERGE) received the computable algorithm, and 6 successfully executed it against local data warehouses and/or i2b2 instances. Blinded random chart review of cases selected by the computable algorithm shows PPV ≥90%, and 3 out of 5 sites had >90% overlap of selected cases when comparing the computable algorithm to their original eMERGE implementation. This case study demonstrates potential use of PhEMA computable representations to automate phenotyping across different EHR systems, but also highlights some ongoing challenges.

https://www.ncbi.nlm.nih.gov/pubmed/**30124903**

PhEMA

BPH in PhAT

https://github.com/PheMA/bph-use-case



PhEMA

BPH CQL

```
library BenignProstateticHyperplasiaPhenotype version '1.1.000'

using QDM version '5.4'

valueset "Prostate or Bladder Cancer in ICD9CM": '2.16.840.1.113762.1.4.1053.25'
valueset "Prostate or Bladder Cancer in UMLS": '2.16.840.1.113762.1.4.1053.26'
valueset ...

parameter "Measurement Period" Interval<DateTime>

context Patient

define function "ToDate"(Value DateTime ):
      DateTime(year from Value, month from Value, day from Value, 0, 0, 0, 0, timezone from Value)

/*CalendarAgeInYearsAt calculates the calendar age (age without considering time components) in years */
define function "CalendarAgeInYearsAt"(BirthDateTime DateTime, AsOf DateTime ):
      years between ToDate(BirthDateTime)and ToDate(AsOf)

define "Patient Characteristic Birthdate Code":
      ["Patient Characteristic Birthdate": "Birth date"]

define "Patient Age 40 or Over at Start of Measurement Period":
      exists  "Patient Characteristic Birthdate Code" BirthDate
        where "CalendarAgeInYearsAt"(BirthDate.birthDatetime, start of "Measurement Period") >= 40

define "Diagnoses of BPH":
      ["Diagnosis": "BPH Diagnoses"]

define "Has BPH Diagnoses 2 or more times":
     exists ("Diagnoses of BPH" bphdFirst with "Diagnoses of BPH" bphdSecond
            such that bphdFirst.prevalencePeriod overlaps after "Measurement Period"
                  and bphdSecond.prevalencePeriod overlaps after "Measurement Period"
                  and bphdFirst.code = bphdSecond.code
                  and bphdFirst.prevalencePeriod starts before start of bphdSecond.prevalencePeriod
        )
define ...

define "BPH Case":
      "Patient Age 40 or Over at Start of Measurement Period"
            and "Patient Gender Male"
            and not "History of Prostate or Bladder Cancer"
            and "Has BPH Diagnoses 2 or more times"
            and ( "Has BPH Medications"
                        or "Has BPH-Related Surgical Procedures")
```

PhEMA

Acknowledgments: PhEMA team

- Cornell
  - Jyoti Pathak
  - Prakash Adekkanattu
  - Jessica Ancker
  - Fei Wang
  - Jie Xu
  - Zhenxing Xu
- Mayo
  - Guoqian Jiang
  - Richard Kiefer
- NU
  - Yuan Luo
  - Luke Rasmussen
  - Jennifer Pacheco
  - Chengsheng Mao
- VU: Joshua Denny
- UW: Pascal Brandt

http://projectphema.org
http://github.com/phema

Funding from NIH R01 GM105688, R01 GM103859

PhEMA

Let the show begin!

Demonstration of authoring and execution against i2b2 of a simplified BPH phenotype algorithm (no temporal relationships)

http://projectphema.org
http://github.com/phema

PhEMA