

Overall Architecture and Distributed Analysis Tools

Daniella Meeker, PhD

Department of Preventive Medicine University of Southern California

Michael E. Matheny, MD, MS, MPH

Tennessee Valley Healthcare System VA Departments of Biomedical Informatics Vanderbilt University

User Experience Goal

- **PSCANER**
- The overall goal of the development of the analytic tools and user interfaces in pScanner is to:
 - Conduct analyses in a distributed national network that achieves a user experience comparable to execution on a single local node
 - > Provide menu and graphical user interface to
 - » Manage study meta-data and provenance
 - » user role drive permissions at each node
 - » Study design
 - » Analytic cohort development
 - » Support common comparative effectiveness statistical methods
 - » Provide result tables and graphs typically useful for reports and manuscripts

PCORNet Coverage

With coverage in every state — PCORnet represents thousands of conditions



Supported by the Patient-Centered Outcomes Research Institute (PCORI) Contract CDRN-1306-04819

PSCANER



- Privacy Preserving No patient row level data sharing
- Service Oriented Architecture allows for tools and modules to be stored in a different location from the end user
- Extensible established guidance and interface to allow anyone to contribute methods / user interfaces within the overall framework
- Open Source promotes transparency, independent validation of functionality, and larger user community of use
- Leveraging large initiatives / established user community PopMedNet data mart client (version proximity to Mini-Sentinel)



pSCANNER Process Model Separation

PSCANER





Process we need to represent	Standard	Rationale
Data processing rules	HQMF	CMS, ONC, HL7 endorsed Part of EHR certification process
Cohort definition rules	HQMF	100's of established data sets 1000s of cohort criteria
Data set description	QRDA PMML	QRDA – Quality Measurement EHR Certification & CMS PMML – Data analysis
Data Analysis Methods	PMML	UCSD Data Mining Group Extensible to support model specifications
Data Analysis Results	PMML	Developed to represent results
Process Workflow	BPML?	Ad-Hoc at present, in development

General to Site Specific Translator

Supported by the Patient-Centered Outcomes Research Institute (PCORI) Contract CDRN-1306-04819

- PopMedNet pScanner API
 - > Communication Layer Between:
 - » PopMedNet Client
 - » pScanner Portal
- PopMedNet DataMart Adapters
 - > Site-Specific Analytic Cluster (DAN/OCEANS)
- Contract with Lincoln Peak to extend existing API / DataMart adapters for use with pScanner





PopMedNet \rightarrow DAN

- DataMart Client
 - > Requests Analysis from Portal



- Sends Standardized Analytic Request to DAN via DataAdapter
- > On completion returns result to Portal
- DataAdapter
 - Pulls standardized requests from PopMedNet
 - Calls DAN (Site specific analytics coordination) with request to queue for processing

Distributed Analysis Network

- Three-tier Distributed Network
 - > Broker accept External Requests (e.g. DataMartClient)
 - > Cluster accepts Broker Requests based on
 - » Analysis Type
 - » Analysis Engine
 - » Availability
 - > Processing Node accepts Cluster Requests
- Design
 - > Simple Deployment as resource need increases
 - > Fault-Tolerance A node fails, and the analysis continues
 - > High-availability Distribution of load via routing algorithm
 - Analysis Engine Agnostic Workflow supports integration into widely used engines (e.g., R, Spark, SAS, OCEANS, etc.)
 - Asynchronous Queues requests for processing as resources become available



DAN – Components

PSCANER



Local Analytic Engines

- R/RServe (active)
- Revolution Analytics (in exploration)
- Potential Analytic Engines
 - > Oracle In-Database R
 - > Microsoft SQL 2016 In-Database R
 - > Apache Spark (future)
 - > SAS / SAS Grid (future)





Development Team

- Bill Clarke (Lincoln Peak)
- Claudio Farcas (UCSD)
- Josh Gieringer (VA/VU)
- Michael Matheny (VA/VU)
- Daniella Meeker (USC)
- Laura Perlman (USC)
- Dax Westerman (VA/VU)
- Bruce Swan (Lincoln Peak)

PSCA