Distributed Health Data Networks: Implementing a Scalable Query Interface within PopMedNet for Use in Large-Scale Diverse Networks

Jessica Malenfant, MPH

Department of Population Medicine of Harvard Medical School & Harvard Pilgrim Health Care Institute

Health Care Systems Research Network Annual Conference

April 12, 2018

pcornet℠ The National Patient-Centered Clinical Research Network

DEPARTMENT OF POPULATION MEDICINE

HARVARD MEDICAL SCHOOL | Harvard Pilgrim Health Care Institute

PopMedNet™

Objective

- Demonstrate a new architecture and framework for an extensible point-and-click query interface in PopMedNetTM (PMN) that:
  - Addresses challenges in platform and software heterogeneity in PCORnet, the largest PMN network
  - Are modularized and can successfully target multiple data models and various technical ecosystems
  - Utilize widely adopted standard data exchange formats e.g. JSON, LINQ, Microsoft Entity Framework, and SQL
  - Produce consistent and valid results

Background

- PMN powers clinical and observational research through efficient and privacy-preserving methods and technologies
- PMN infrastructure permits investigators to compose and distribute custom queries through a variety of tools
- PMN is a mature platform that is used by 100s of organizations
- PMN is used in several large-scale distributed data networks including: PCORI's PCORnet and FDA's Sentinel Initiative

Problems Identified with the Initial MDQ Tool

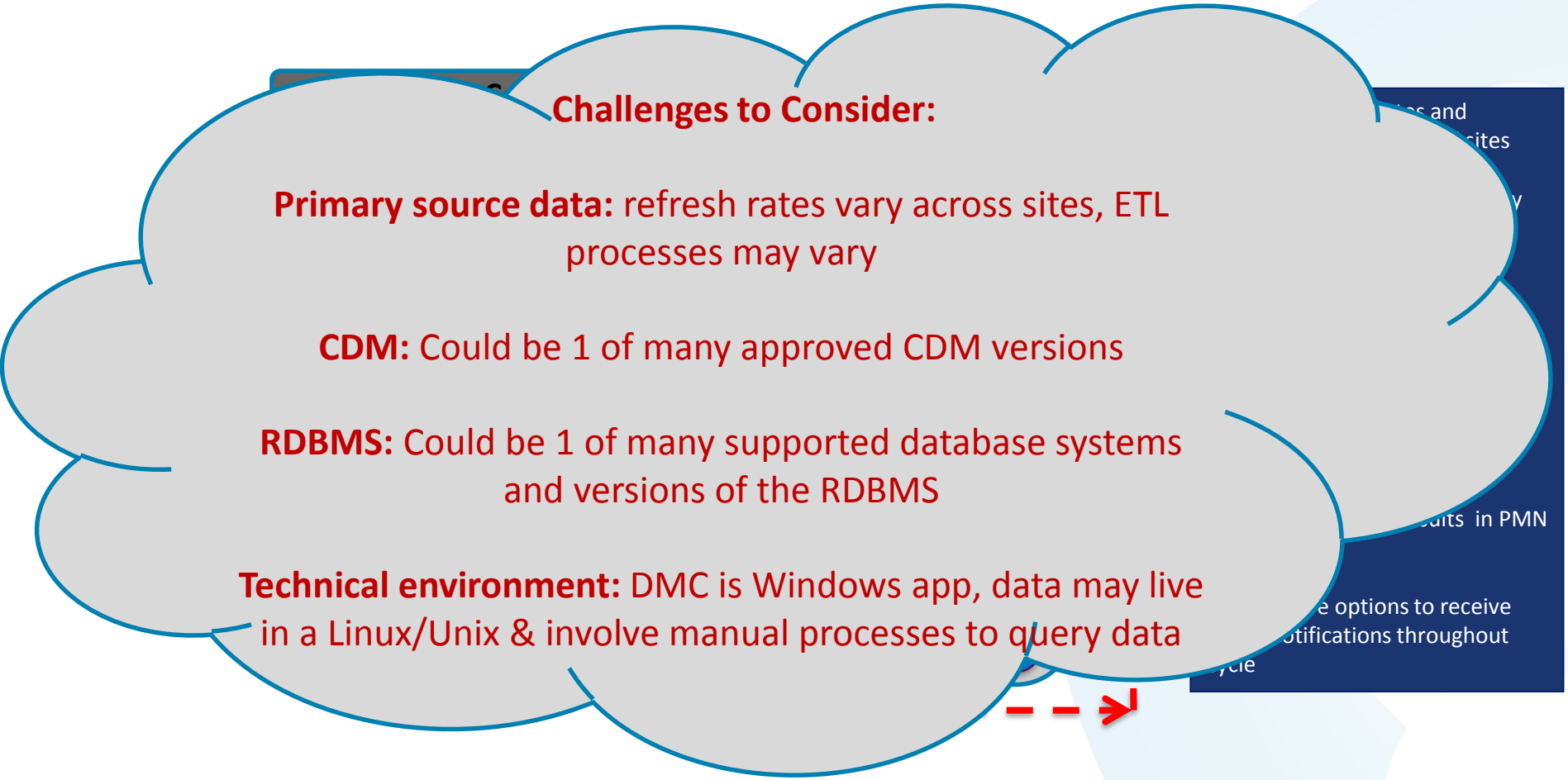
Legacy Query Composer: Developed for limited use resulting in scalability issues:

- Each query tool was hardcoded for use against a single CDM and database platform
  - The MDPHnet network's ESP data model and PostgreSQL
  - FDA's Sentinel System Summary Table data model and MS Access
- Changes required manual and redundant hard-coding
- Queryable terms could not be shared across networks (e.g. if 2 networks wanted to query race data, each query tool needed to be developed separately, even if the field names and value sets were the same)
- Changes required the sites to download a new version of the PMN DataMart Client software in order to respond to a query

Challenge: Develop a One Size Fits All MDQ Tool

- End users want a simple query tool interface and workflow
- Infrastructure should be re-usable and easily extensible and scalable, limiting CDM-specific coding
- Address the heterogeneity of technical environments across the large-scale distributed networks PMN supports
- Consider workflows for full request lifecycle including integration points with external systems

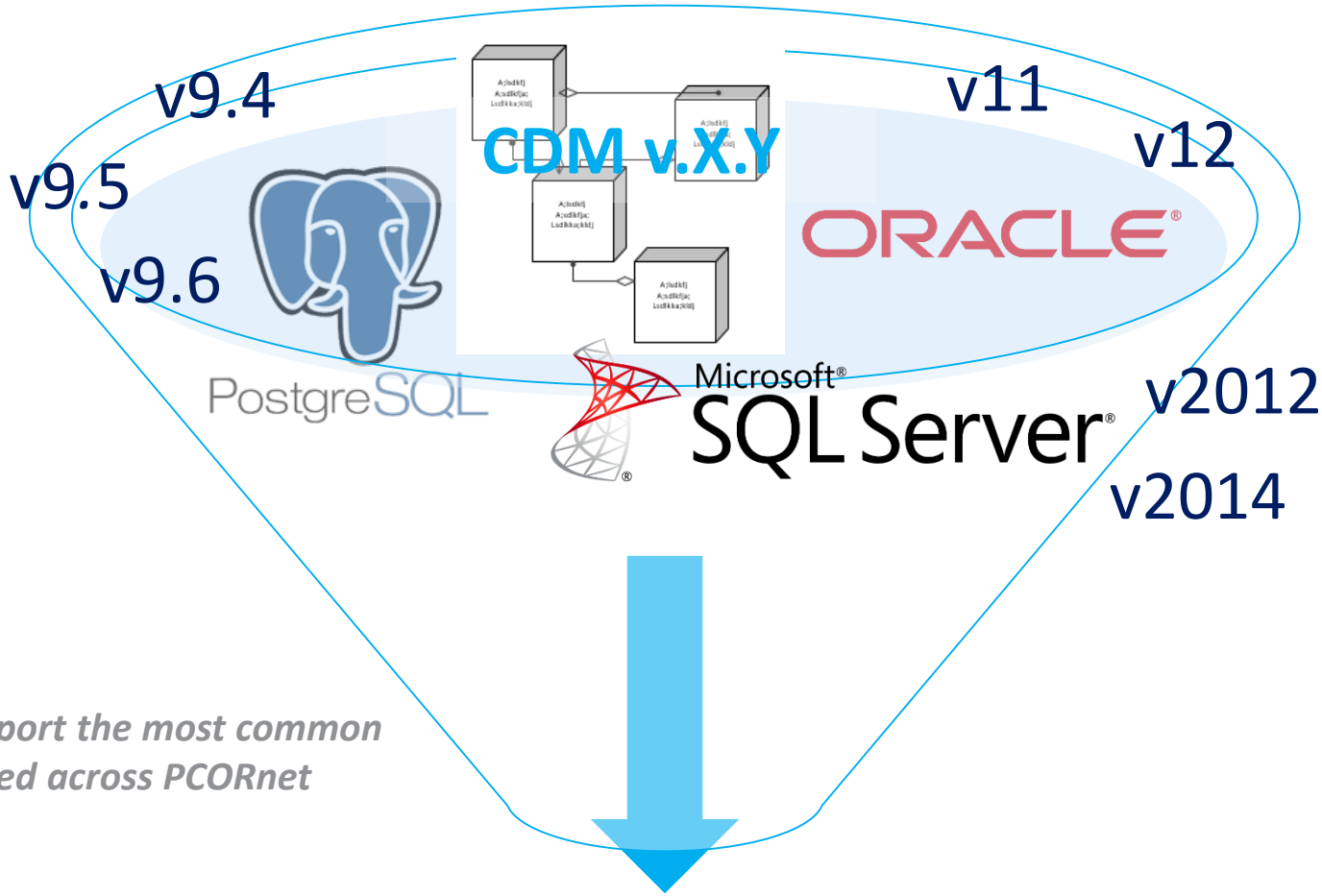
Request Cycle



Tools Developed

- Menu-Driven Queries (MDQs):
  - PMN interface supports querying terms and stratifications
  - Investigators can compose a simple or complex MDQ that includes logical operators "OR", "AND", "AND NOT" to define a cohort of interest via a user interface
  - Include software-enabled governance to determine what users can query
  - Support electronic workflows and embedded analytics
  - Include data model adapters that make the MDQs Common Data Model (CDM) aware
  - Modular design for sharing queryable terms regardless of data source
- Test Case Inserter (TCI):
  - Generates databases according to CDM specifications
  - Custom program that enables users to easily insert synthetic data into a relational database management system (RDBMS) without requiring the user to have SQL programming skills
  - Supports MDQ validation and MDQ prototypes for targeting new data sources

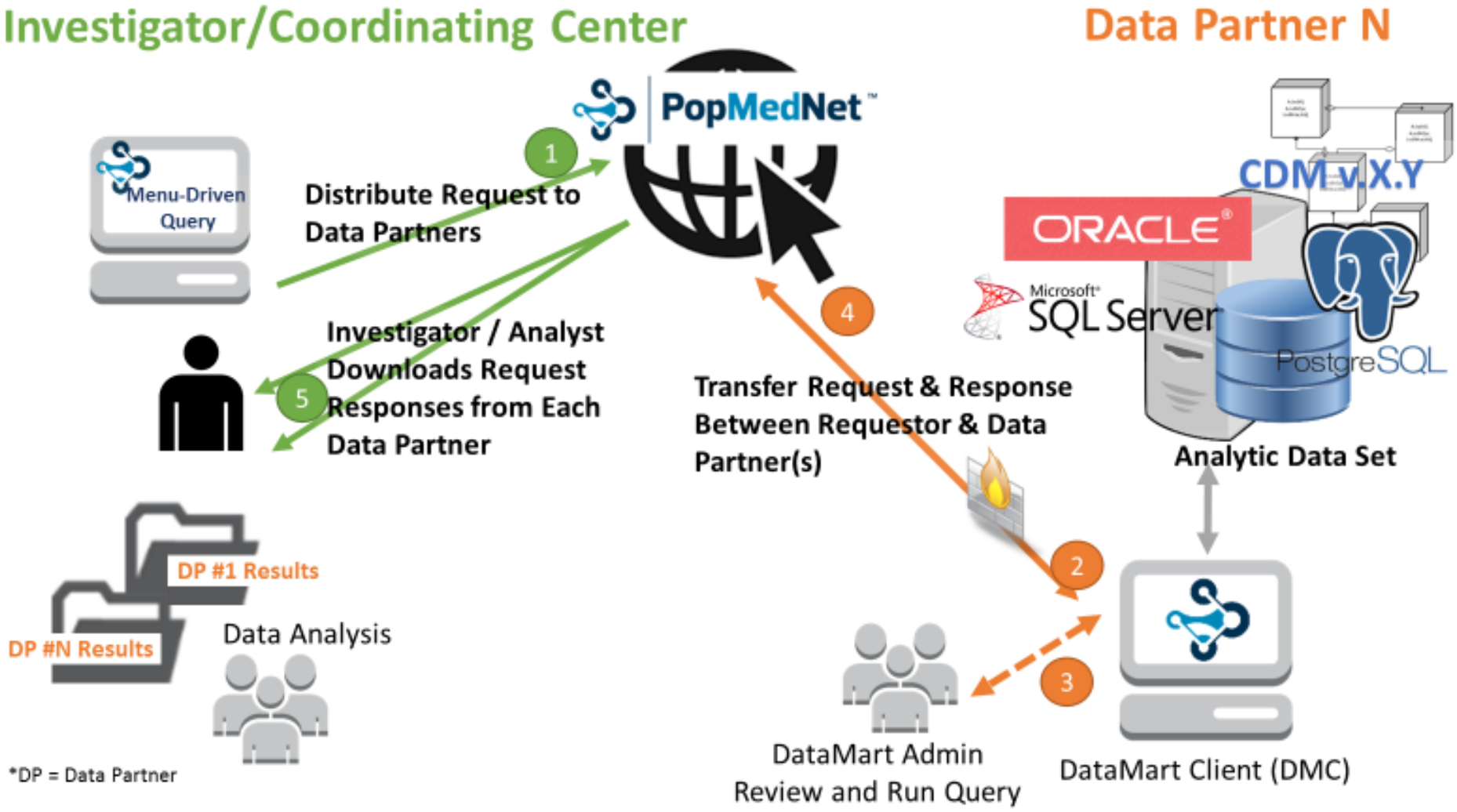
One Size fits Most* MDQ Tool



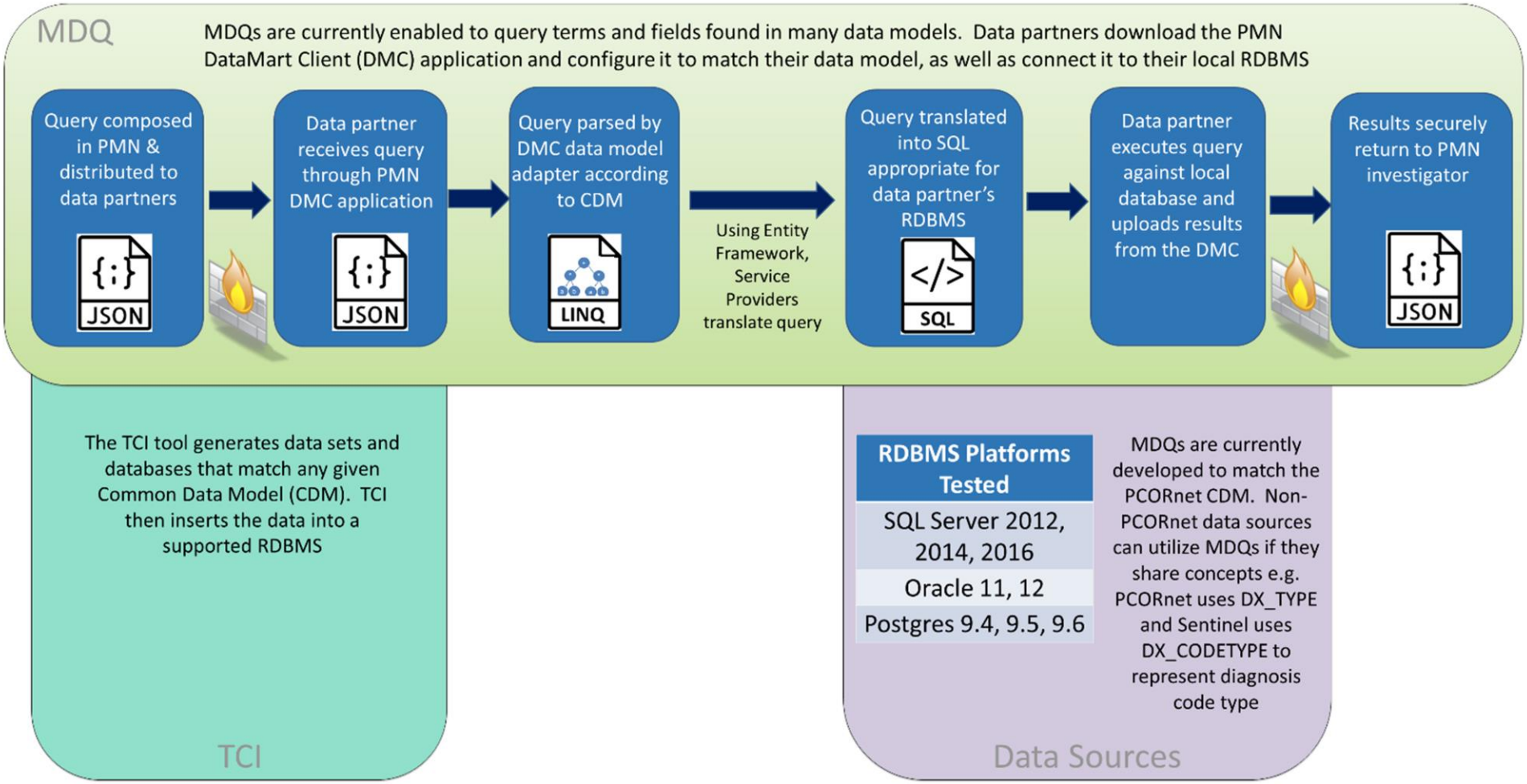
**Committed to support the most common RDMBS versions used across PCORnet*

Single MDQ Tool

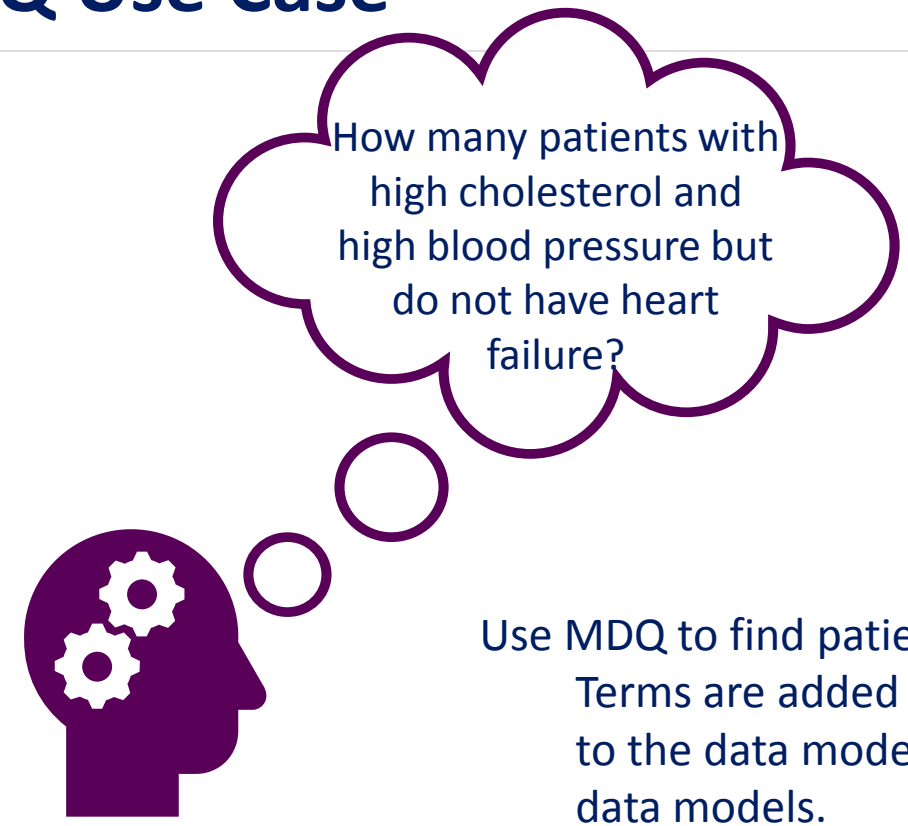
Menu-Driven Query Process



MDQ Technical Details and Validation



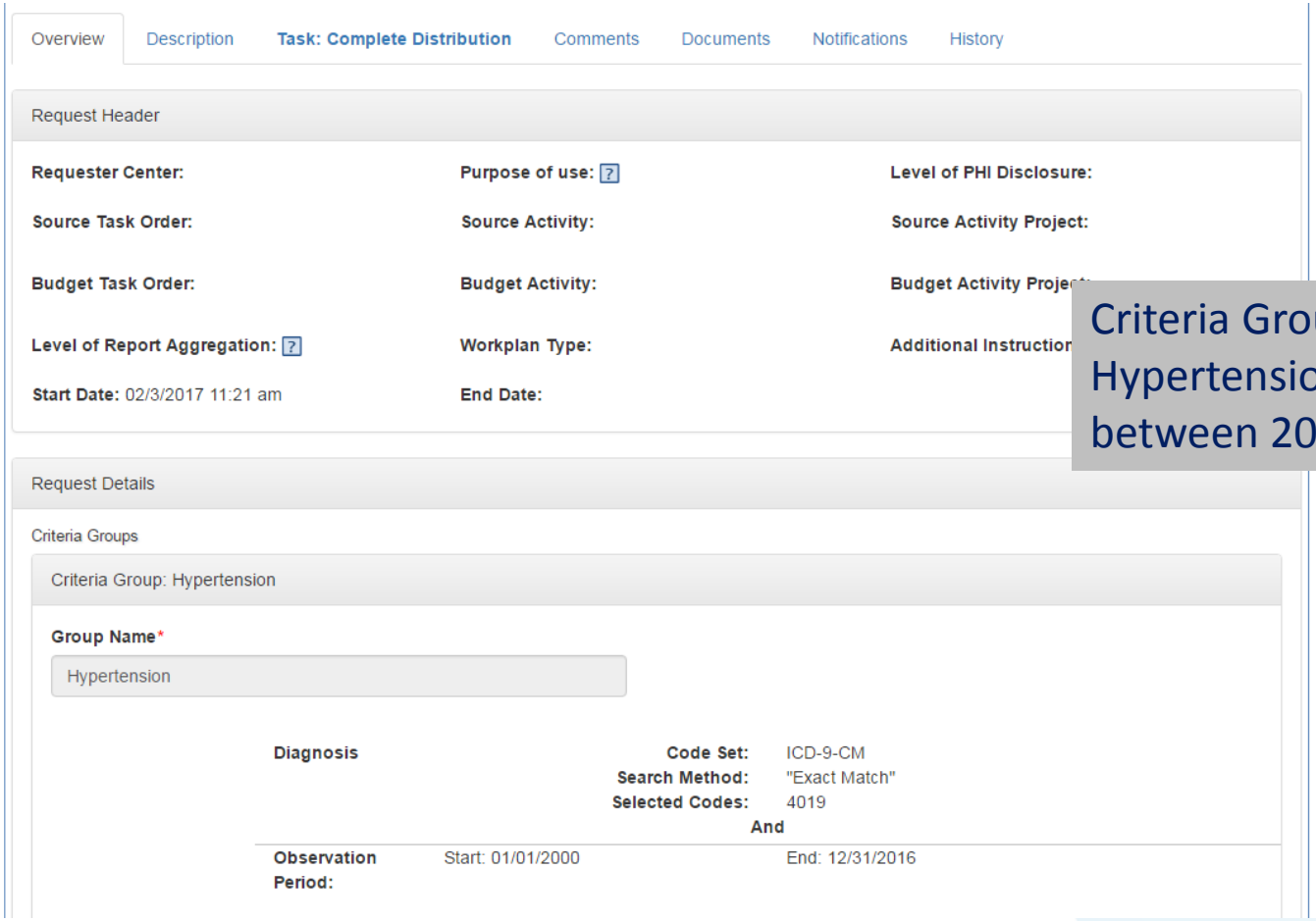
MDQ Use Case



Use MDQ to find patients of interest

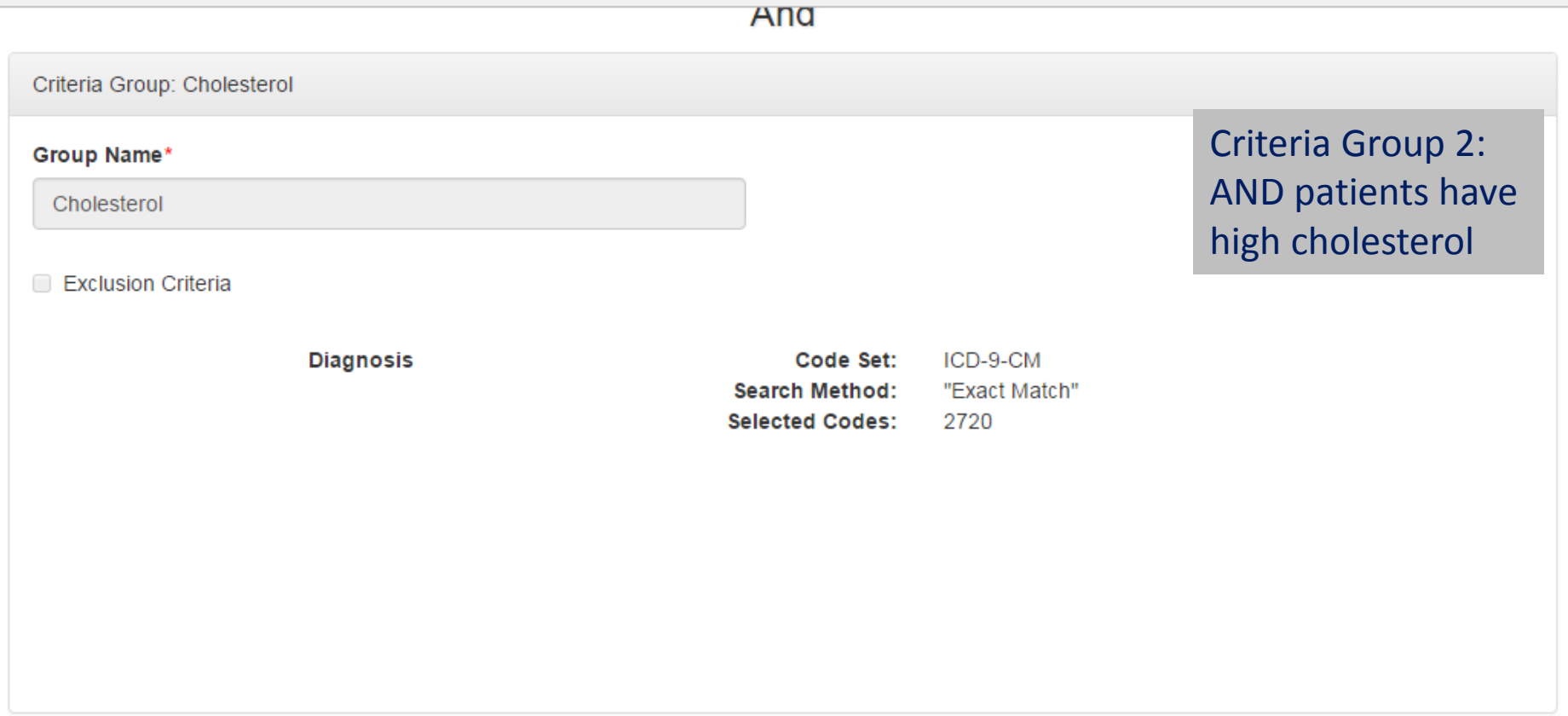
Terms are added to the PMN MDQ interface according to the data model. Terms can be re-purposed for other data models.

Query Interface to Define Cohort



Criteria Group 1: Hypertension with visits between 2000-2016

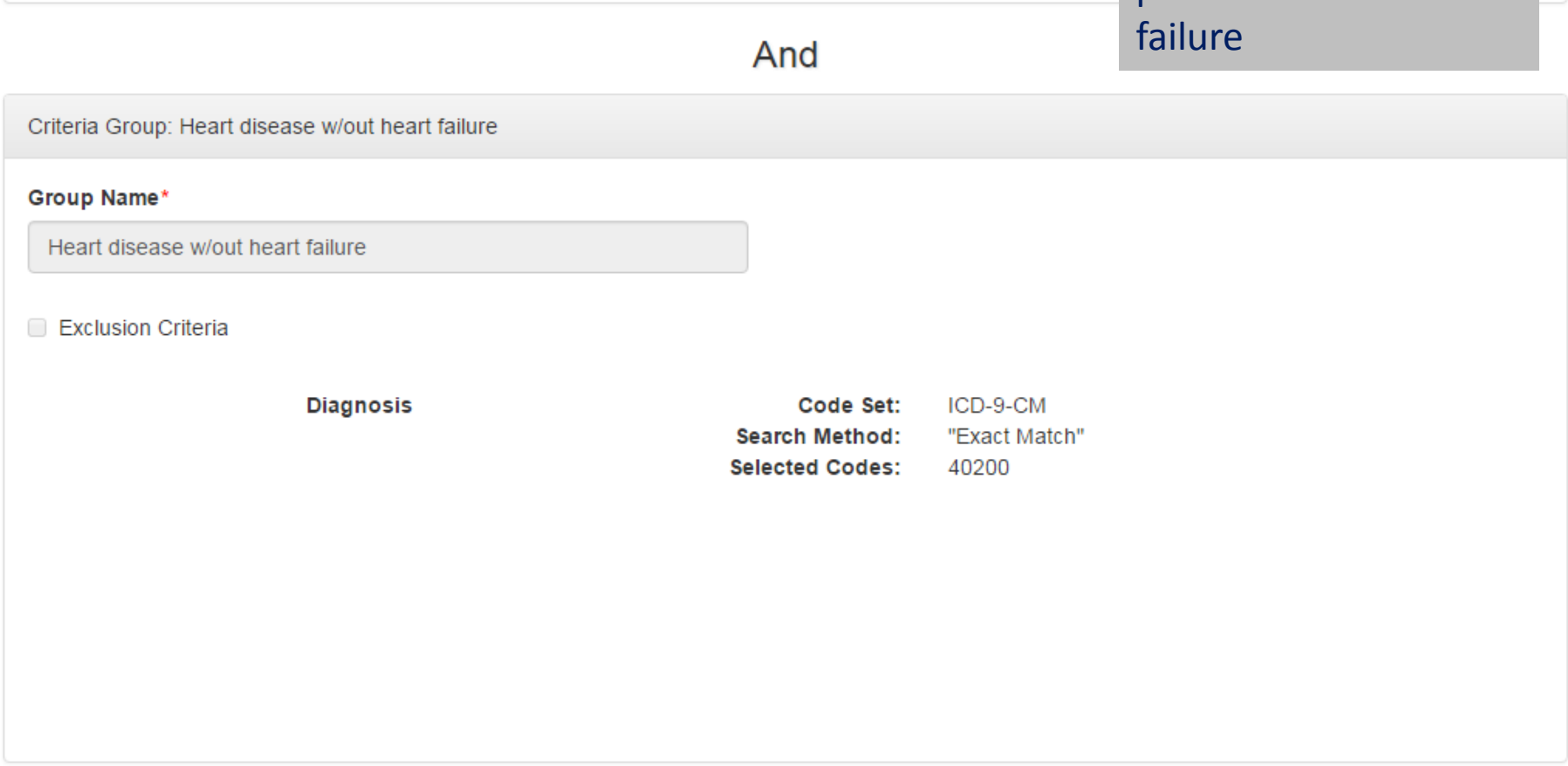
Query Interface to Define Cohort



Criteria Group 2: AND patients have high cholesterol

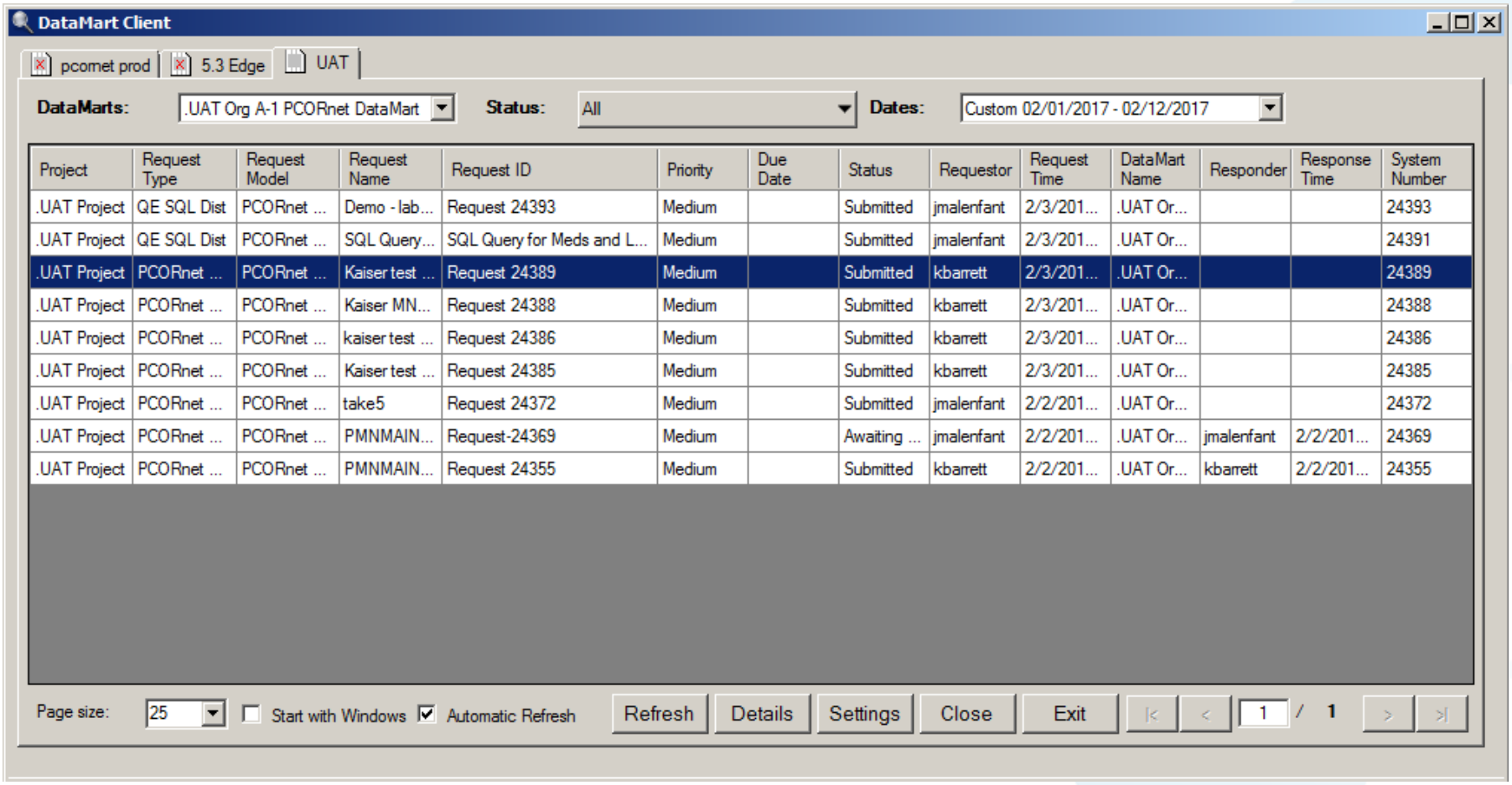
Query Interface to Define Cohort

Criteria Group 3: AND patients without heart failure



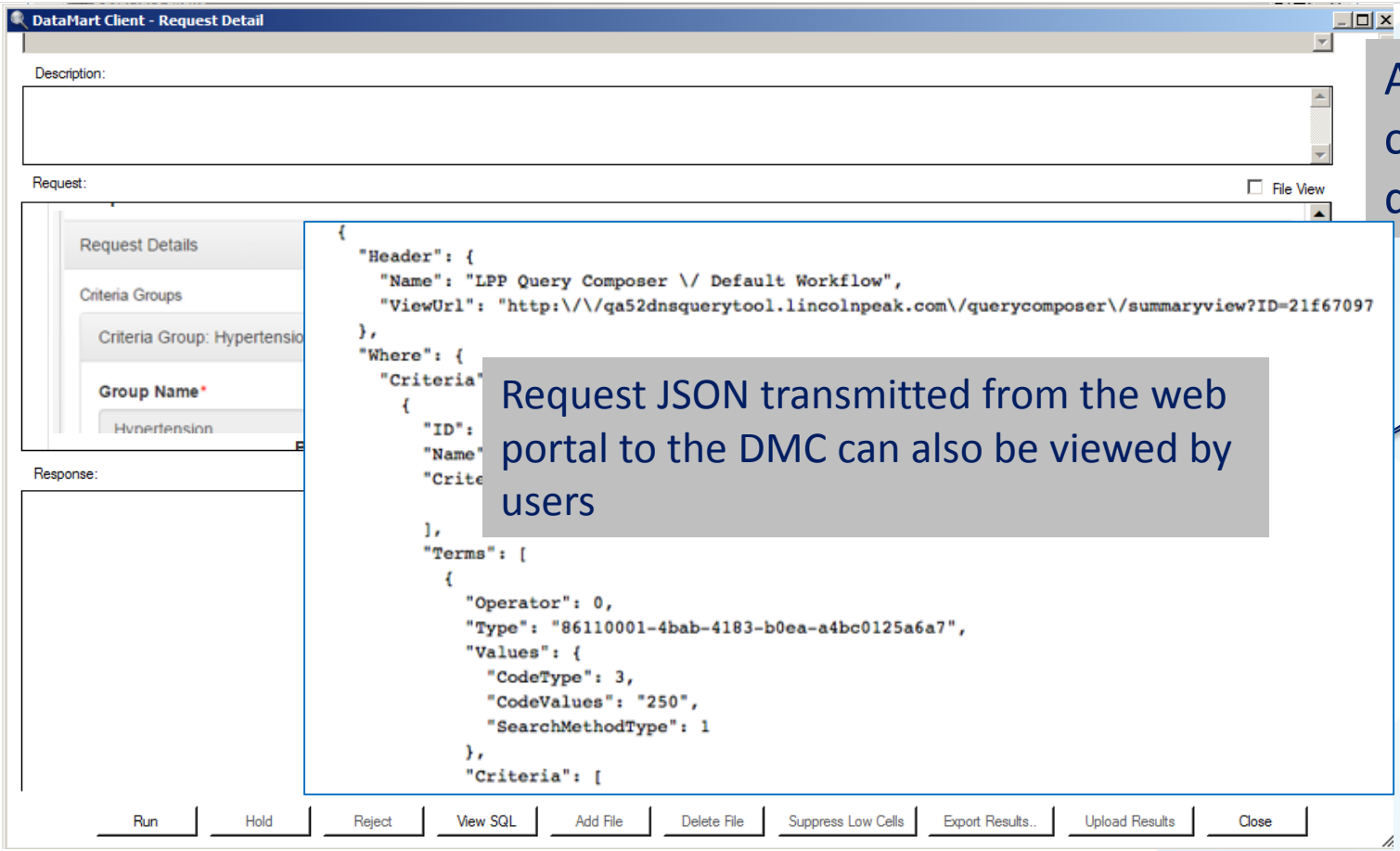
DataMart Administrator Receives the Query

DataMart Administrator Inbox – locally installed app at each site



Project	Request Type	Request Model	Request Name	Request ID	Priority	Due Date	Status	Requestor	Request Time	DataMart Name	Responder	Response Time	System Number
.UAT Project	QE SQL Dist	PCORnet ...	Demo - lab...	Request 24393	Medium		Submitted	jmalenfant	2/3/201...	.UAT Or...			24393
.UAT Project	QE SQL Dist	PCORnet ...	SQL Query...	SQL Query for Meds and L...	Medium		Submitted	jmalenfant	2/3/201...	.UAT Or...			24391
.UAT Project	PCORnet ...	PCORnet ...	Kaiser test ...	Request 24389	Medium		Submitted	kbarrett	2/3/201...	.UAT Or...			24389
.UAT Project	PCORnet ...	PCORnet ...	Kaiser MN...	Request 24388	Medium		Submitted	kbarrett	2/3/201...	.UAT Or...			24388
.UAT Project	PCORnet ...	PCORnet ...	kaiser test ...	Request 24386	Medium		Submitted	kbarrett	2/3/201...	.UAT Or...			24386
.UAT Project	PCORnet ...	PCORnet ...	Kaiser test ...	Request 24385	Medium		Submitted	kbarrett	2/3/201...	.UAT Or...			24385
.UAT Project	PCORnet ...	PCORnet ...	take5	Request 24372	Medium		Submitted	jmalenfant	2/2/201...	.UAT Or...			24372
.UAT Project	PCORnet ...	PCORnet ...	PMNMAIN...	Request-24369	Medium		Awaiting ...	jmalenfant	2/2/201...	.UAT Or...	jmalenfant	2/2/201...	24369
.UAT Project	PCORnet ...	PCORnet ...	PMNMAIN...	Request 24355	Medium		Submitted	kbarrett	2/2/201...	.UAT Or...	kbarrett	2/2/201...	24355

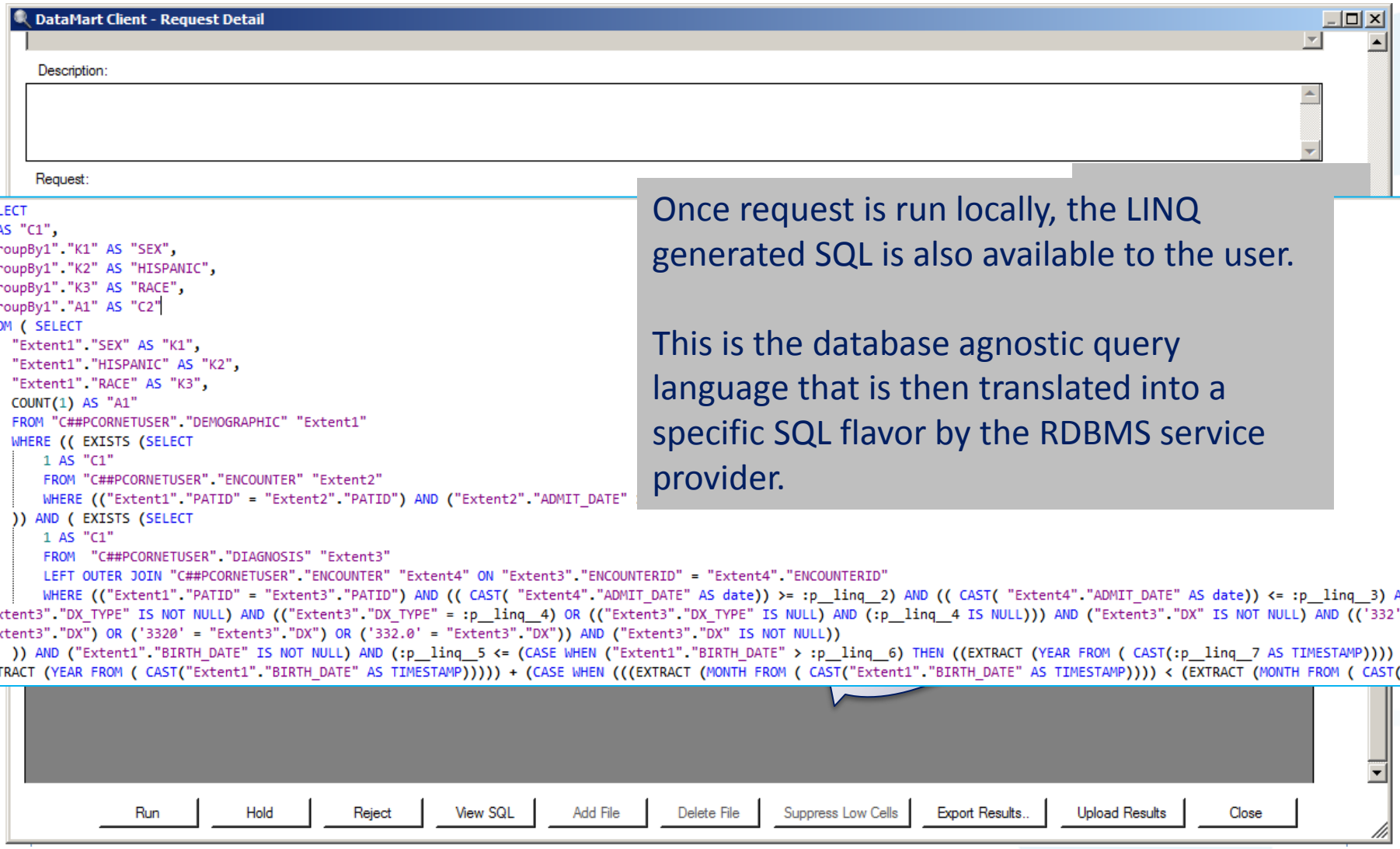
DataMart Administrator Reviews Query Details



Administrator can review query input

Request JSON transmitted from the web portal to the DMC can also be viewed by users

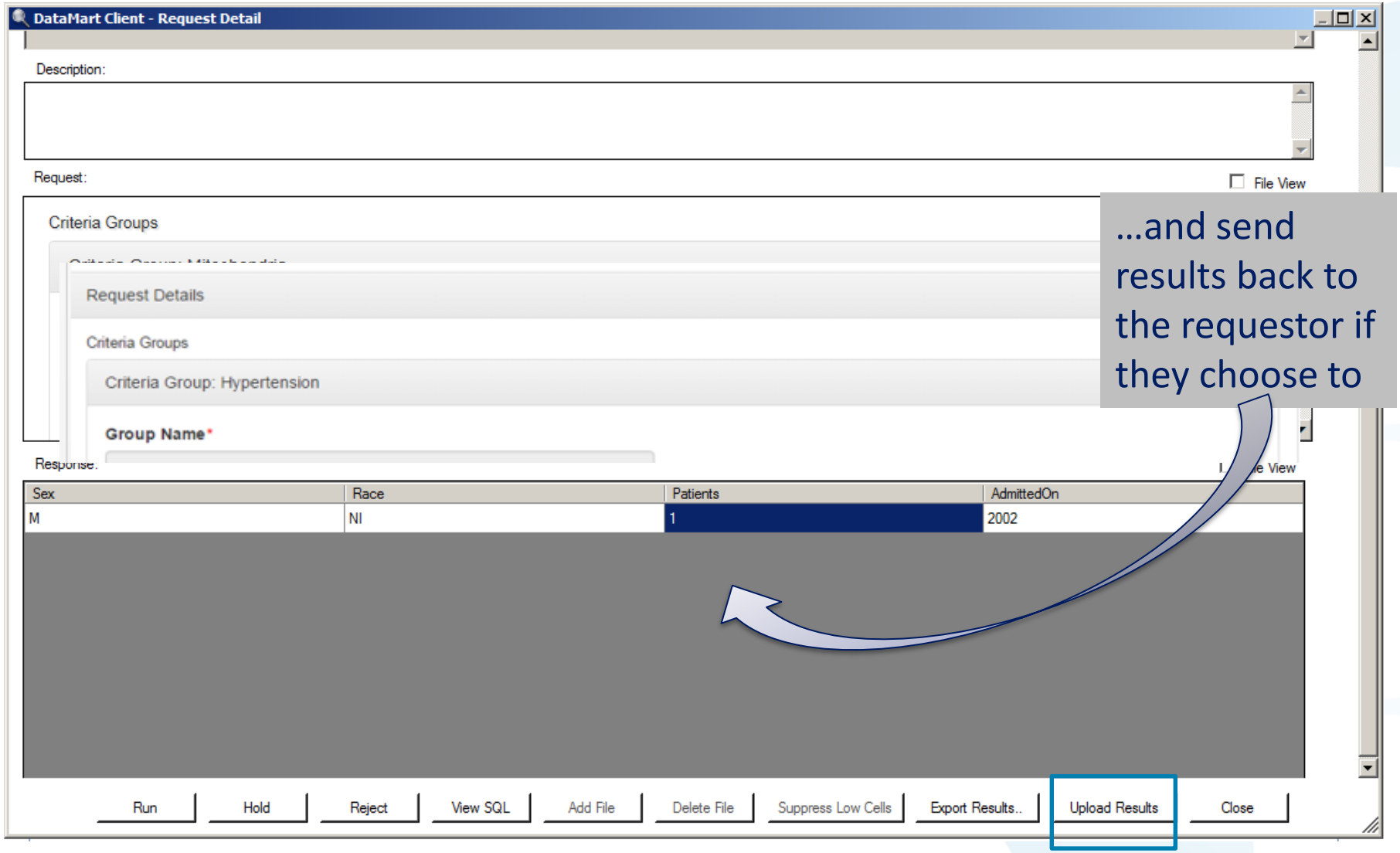
DataMart Administrator Executes the Query and Reviews Results



Once request is run locally, the LINQ generated SQL is also available to the user.

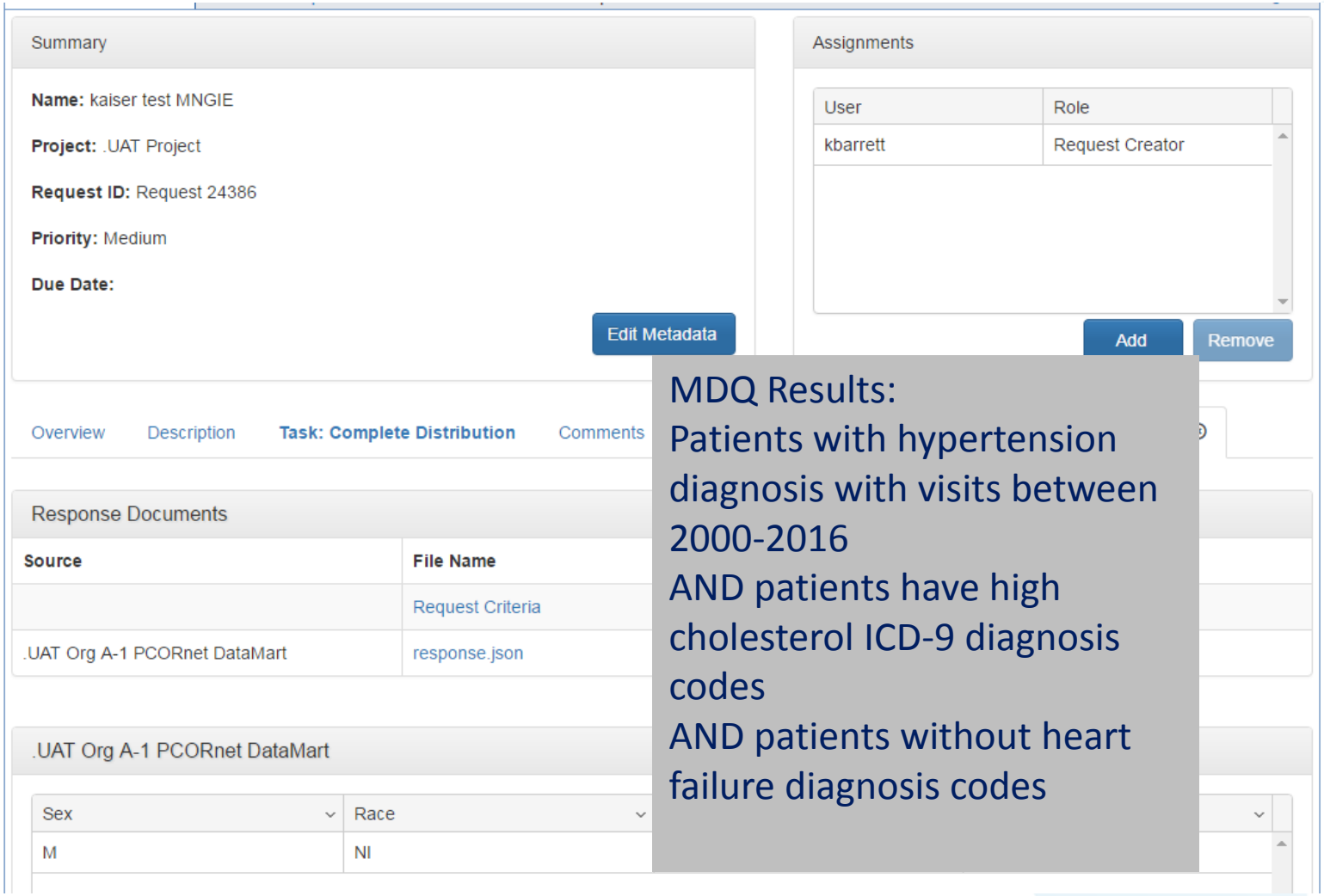
This is the database agnostic query language that is then translated into a specific SQL flavor by the RDBMS service provider.

DataMart Administrator Uploads Results



…and send results back to the requestor if they choose to

Investigator Reviews Site-Specific Results on Web Portal



MDQ Results:
Patients with hypertension diagnosis with visits between 2000-2016
AND patients have high cholesterol ICD-9 diagnosis codes
AND patients without heart failure diagnosis codes

Current Status

- Multiple filters and stratification options have been added to the MDQ tool for several fields including Race, Sex, Observation Period, Diagnosis and Procedure Codes, Height, Weight, Age, etc., more planned
- The PCORnet data adapter has been updated to process queries with the new terms and stratification options
- Testing with the TCI tool has verified that ad hoc data models that share PCORnet CDM fields can use the MDQ out-of-the box, continuing to explore how to leverage the work for other data models

Current Status

- Enhancing automation functionality
- Preparing to add functionality for users to define index events
- Validation and performance testing is in progress to evaluate how complex queries behave
- Ability to expose the actual SQL to a user prior to running a query is under investigation. The request JSON and the LINQ code are currently available to end users but require manual steps to piece the query languages together, for example:

Current Status

```
-- PrimaryObservationStart: '10/15/2013 12:00:00 AM' (Type = DateTime2, IsNullable = false)

-- PrimaryObservationEnd: '10/14/2014 12:00:00 AM' (Type = DateTime2, IsNullable = false)

-- PrimaryObservationStart: '10/15/2013 12:00:00 AM' (Type = DateTime2)

-- PrimaryObservationEnd: '10/14/2014 12:00:00 AM' (Type = DateTime2)

-- CriteriaOneCodeType: '09' (Type = String, Size = 4000)

-- CriteriaOneMinimumAge: '65' (Type = Int32, IsNullable = false)
```

```
SELECT
    1 AS [C1],
    [GroupBy1].[K1] AS [SEX],
    [GroupBy1].[K2] AS [HISPANIC],
    [GroupBy1].[K3] AS [RACE],
    [GroupBy1].[A1] AS [C2]
    FROM (

        SELECT
        [Extent1].[SEX] AS [K1],
        [Extent1].[HISPANIC] AS [K2],
        [Extent1].[RACE] AS [K3],
        COUNT(1) AS [A1]
        FROM [dbo].[DEMOGRAPHIC] AS [Extent1]
        WHERE (
            -- where the patient has an encounter between the primary criteria dates
            EXISTS (SELECT
            1 AS [C1]
            FROM [dbo].[ENCOUNTER] AS [Extent2]
            WHERE ([Extent1].[PATID] = [Extent2].[PATID]) AND ([Extent2].[ADMIT_DATE] >= @PrimaryObservationStart) AND ([Extent2].[ADMIT_DATE] <= @PrimaryObservationEnd)
        ))

        AND (
            -- from diagnosis where the codes match the primary criteria code term values and code type, and the diagnosis has an encounter associated between the primary observation dates
            EXISTS (SELECT
                1 AS [C1]
                FROM [dbo].[DIAGNOSIS] AS [Extent3]
                LEFT OUTER JOIN [dbo].[ENCOUNTER] AS [Extent4] ON [Extent3].[ENCOUNTERID] = [Extent4].[ENCOUNTERID]
                WHERE ([Extent1].[PATID] = [Extent3].[PATID]) AND ([Extent4].[ADMIT_DATE] >= @PrimaryObservationStart) AND ([Extent4].[ADMIT_DATE] <= @PrimaryObservationEnd) AND ([Extent3].[
            )

        )
```

Summary

- The MDQ tool includes a point and click, scalable query interface that supports complex logic for users to define cohorts of interest (e.g. medical codes, date ranges) for use in distributed queries
- MDQs can be executed in various technical ecosystems
- The tool is modularized, enabling functionality to be database-agnostic and can be run against multiple RDBMS platforms without custom programming as it utilizes widely adopted data exchange formats

Summary

- MDQ testing and validation process has been implemented and has shown consistent, valid results across database platforms; this process has informed tool development and continuous enhancements.
- Over 4,300 unique MDQ-to-site requests have been submitted within PCORnet, a DDN, since early 2016.
- The tool is helping to close gaps by creating more opportunities for investigators to ask research questions more easily, flexibly, and rapidly within their DDNs while adhering to their local governance and technology policies.
- Work continues to improve the MDQ tool in ways that will enable even more scalability in designing re-usable query interfaces and electronic workflows.

Acknowledgements

- PMN team at Harvard Pilgrim Health Care Institute
  - Jeffrey Brown
  - Chayim Herzig-Marx
  - Kimberly Barrett
  - Zachary Wyner
  - Adam Paczuski
- Software development partners
  - Daniel Dee
  - Dean Corriveau
- PCORnet Coordinating Center
  - Query Fulfilment Team
  - Data Characterization Team
  - Program Management Office

DEPARTMENT OF POPULATION MEDICINE
HARVARD MEDICAL SCHOOL | Harvard Pilgrim Health Care Institute

pcornet™

PopMedNet™

Jessica_malenfant@harvardpilgrim.org

popmednet.org
populationmedicine.org

@DeptPopMed