

May 2 – 4, Philadelphia
iHealth 2017 Clinical Informatics Conference

Creating real solutions through evidence- and experience-based practice



S02: Presentations - Population Health Querying Electronic Health Data for Population Health Activities using PopMedNet™

Jessica Malenfant, MPH

Senior Health Informatics Analyst

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



DEPARTMENT OF POPULATION MEDICINE



Twitter: #iHealth17

amia.org

Disclosure

I and my partner have no relevant relationships with commercial interests to disclose.

Learning Objective

After the presentation, the participant should be better able to describe key challenges and features of implementing a query tool for distributed health data that is scalable for implementation in various settings.

Agenda

- Describe how PopMedNet (PMN) powers distributed research networks (DRN)
- Describe PMN software design & features
- Review Menu-Driven Query (MDQ) tool
 - Problem & use cases
 - Solution & Challenges
- Current status, opportunities & next steps

PopMedNet (PMN) Platform: Powering Distributed Data & Distributed Analysis

- Mature architecture using an approach shown to be **accepted by health plans, clinical sites and other data holders**
- Data partners **maintain control** over their own data
- **Distribute code** to partners for local execution
- **Sites Provide results**, not data, to the requestor
- Standardize the data using a **common data model**
- All activities **audited** and **secure**
 - Meets the **privacy, proprietary, security, and research integrity** demands of health plans and other data holders institutions' IT departments
- Especially well suited for **multi-site, multi-use networks**
- **Contribute to the Learning Health System** by providing a socio-technical platform to support the people, process, technology contributing to knowledge generation

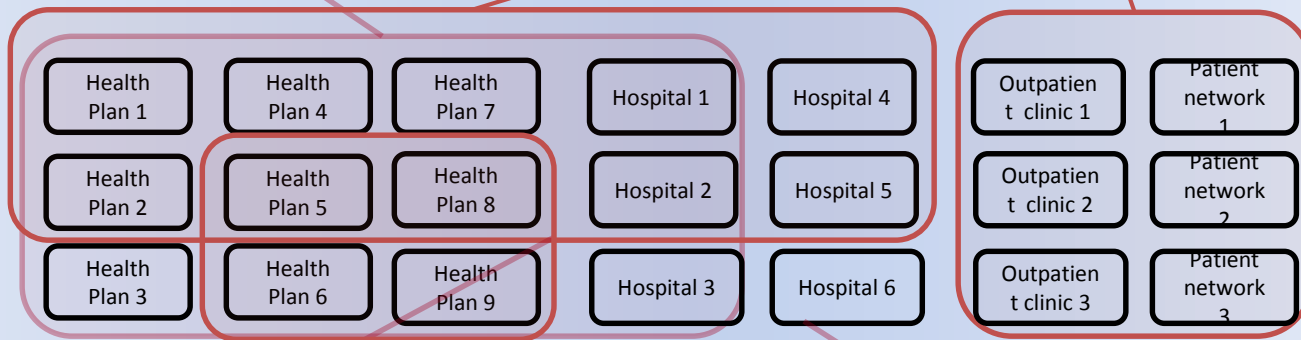
How it works: A Common Data Model

- Common Data Models (CDM) provide a **mechanism for efficient sharing of health data for secondary uses** – research and public health surveillance
- Agreed upon **structure for capturing data**
- Data owners **map their source data** (e.g. EHR, registry data, administrative claims data) into the CDM format including
 - Table names
 - Variable names
 - Value sets
 - Data formatting specifications
 - Database or data repository implementations
- Typically leverage **health IT standard coding systems and vocabularies**

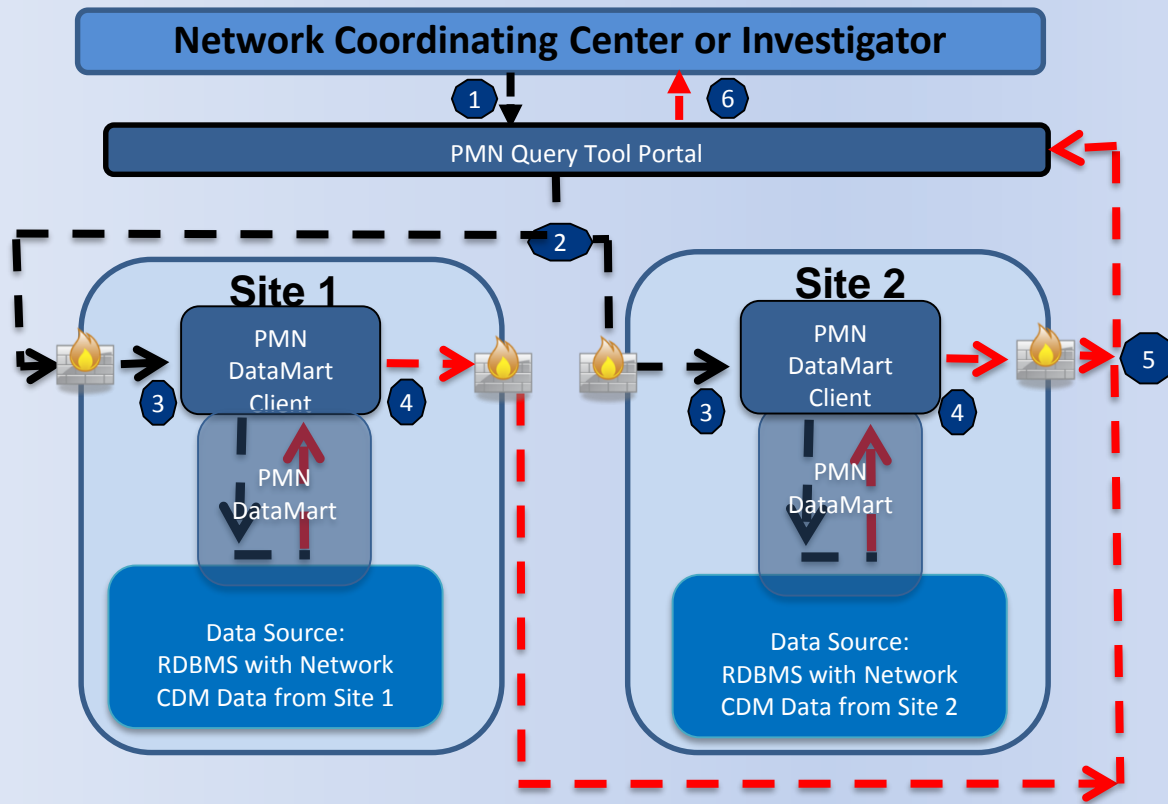
Multiple Networks Sharing PMN Infrastructure

- Each organization can **participate in multiple networks**
- Each network **benefits from architecture and security improvements** while **maintaining their unique governance and policies**
- Networks **share** analytic tools, lessons learned, and system improvements
- Each network **controls its governance** and coordination
- **Funding from each network is leveraged** across initiatives to contribute to the core PMN platform

Multiple Networks Sharing PMN Infrastructure



PMN Request Cycle: Menu Driven Query



1. Investigator creates and submits query to selected sites
 2. Individual sites retrieve query
 3. Sites review and run query directly against the CDM via the PMN DataMart Client
 4. Sites review results
 5. Individual site returns results via secure network
 6. Requestor views results in PMN Portal
- Users have options to receive notifications throughout request cycle; various automation and approval workflows available

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 RDBMS
 - The MDPHnet network's data model and PostgreSQL
 - FDA's Sentinel System Summary Table data model and MS Access database
 - All 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

Challenges to Distributed Querying

- Heterogeneity of technical environments (e.g. Windows, Linux/Unix)
- Source data systems and refresh cycles populating the CDMs vary
- Database management system (i.e. RDBMS) flavors and versions that store the CDM data vary across sites
- Data holders have local IT policies and procedures for how and where data are stored and accessed

(of course these are just a select list of challenges that need to be considered)



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

PMN Request Cycle: Menu Driven Query

Network Coordin

r In

ates and
lected sites

Challenges to Consider:

Primary source data: refresh rates vary across sites, ETL processes may vary

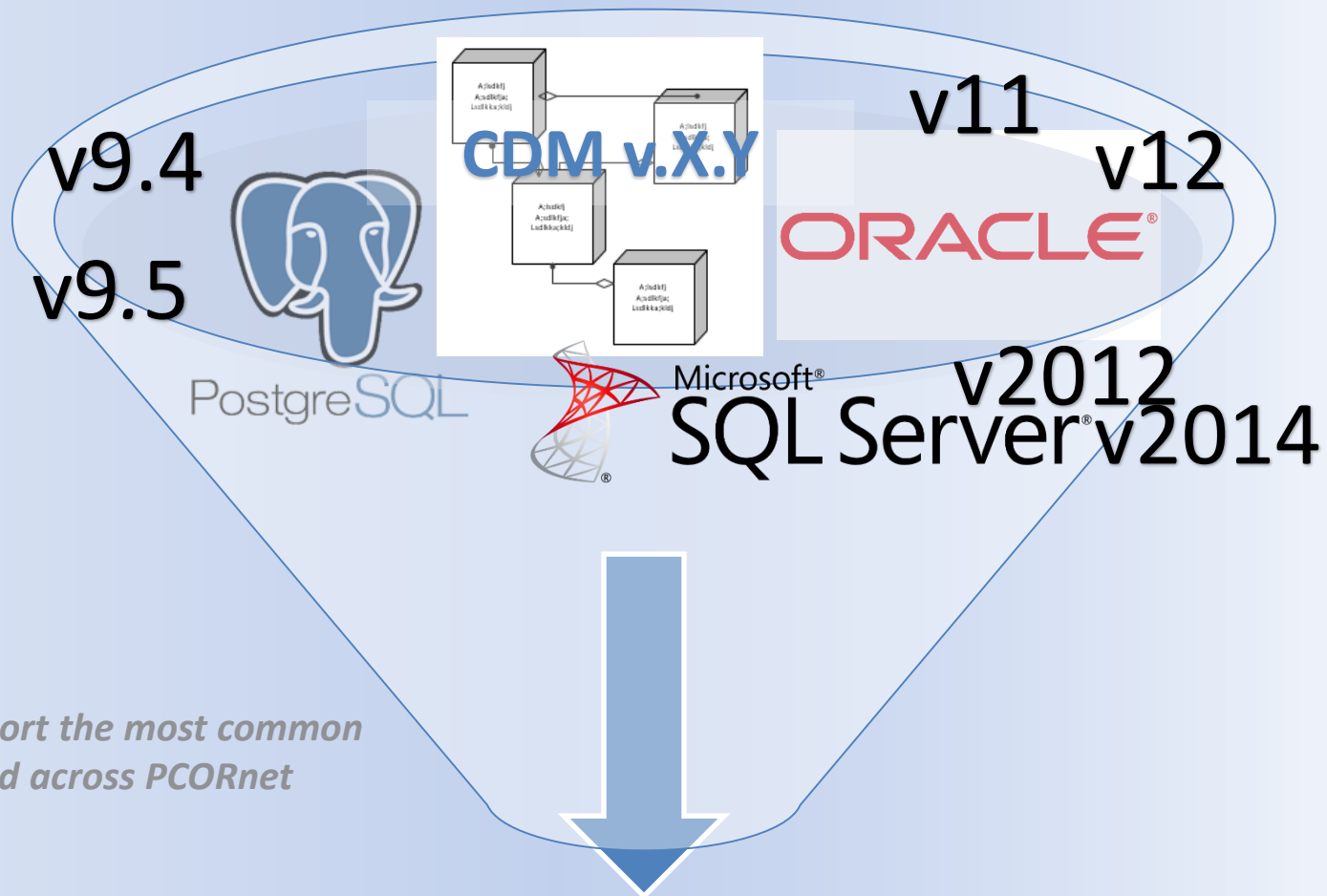
CDM: Could be 1 of many approved CDM versions

RDBMS: Could be 1 of many supported database systems and versions of the RDBMS

Technical environment: DMC is Windows app, data may live in a Linux/Unix & involve manual processes to query data

ions to receive
ions throughout

One Size fits Most* MDQ Tool

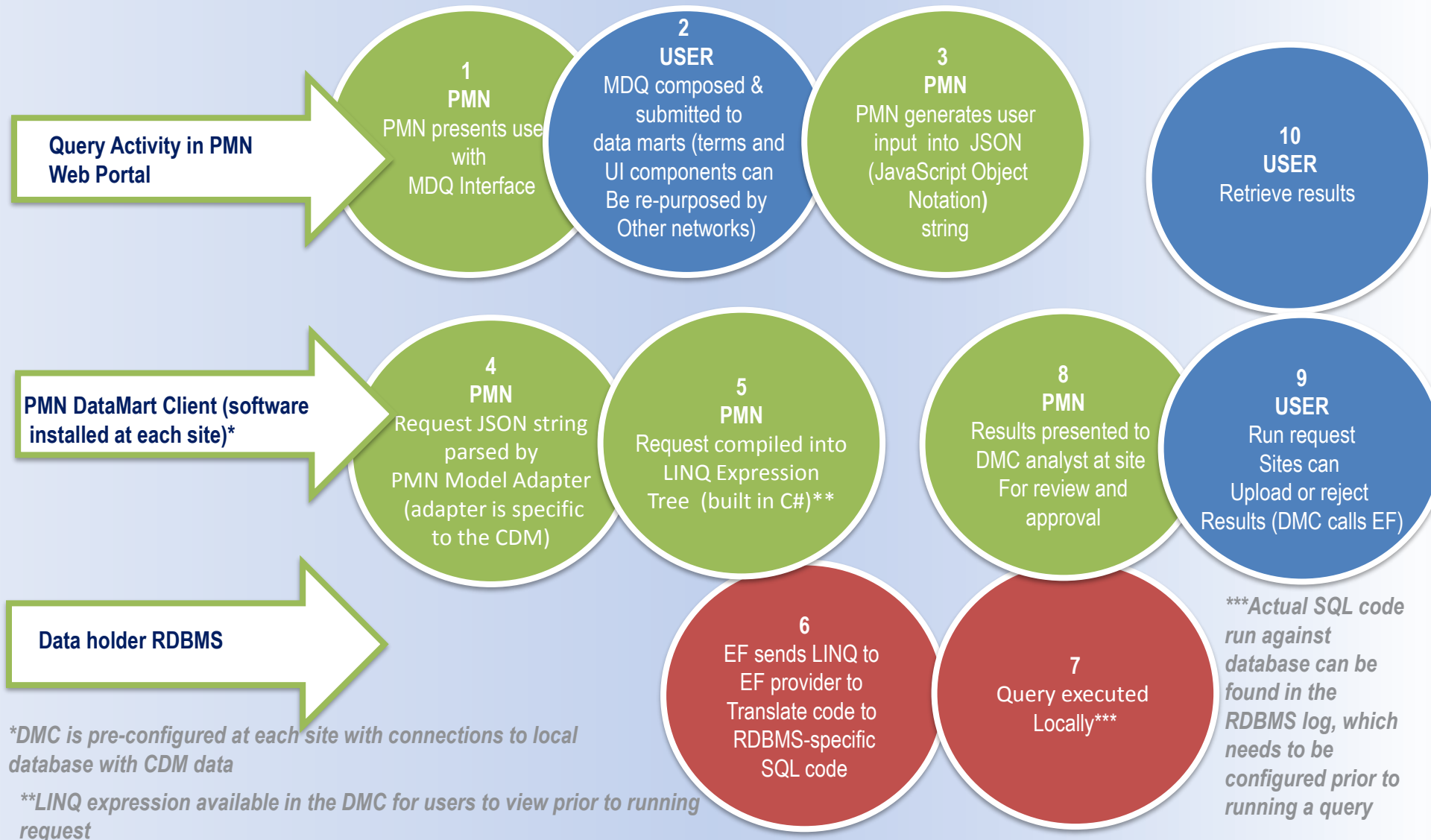


**Committed to support the most common RDBMS versions used across PCORnet*

MDQ Approach

- Leverage new, established technologies for query processing
 - .NET LINQ to Entities query
 - Microsoft Entity Framework
 - Most major database managers have providers for Entity Framework
- Develop mechanisms for queryable Terms (e.g. Race field) to be easily re-purposed for use against multiple data models and in multiple networks
- Re-design the query architecture to limit the requirements for sites to download new DataMart Client software for routine upgrades

MDQ Cycle



**DMC is pre-configured at each site with connections to local database with CDM data*

***LINQ expression available in the DMC for users to view prior to running request*



Use Case 1: Investigator Composes the MDQ Query: Why don't all people with high blood cholesterol and blood pressure get heart disease?

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.

Note that these example queries are based on the PCORnet Common Data Model

The screenshot displays the PMN MDQ interface with the following sections:

- Task: Complete Distribution** (selected)
- Request Header**:
 - Requester Center: [?]
 - Source Task Order: [?]
 - Budget Task Order: [?]
 - Level of Report Aggregation: [?]
 - Start Date: 02/3/2017 11:21 am
 - Purpose of use: [?]
 - Source Activity: [?]
 - Budget Activity: [?]
 - Workplan Type: [?]
 - End Date: [?]
 - Level of PHI Disclosure: [?]
 - Source Activity Project: [?]
 - Budget Activity Project: [?]
 - Additional Instr: [?]
- Request Details**:
 - Criteria Groups**:
 - Criteria Group: Hypertension**:
 - Group Name***: Hypertension
 - Diagnosis**: [?]
 - Code Set**: ICD-9-CM
 - Search Method**: "Exact Match"
 - Selected Codes**: 4019
 - And**: [?]
 - Observation Period**: Start: 01/01/2000, End: 12/31/2016

Criteria Group 1:
Hypertension with visits
between 2000-2016

Use Case 1: Investigator Composes the MDQ Query:

Why don't all people with high blood cholesterol and blood pressure get heart disease?

And

Criteria Group: Cholesterol

Group Name*

Cholesterol

Exclusion Criteria

Diagnosis

Code Set: ICD-9-CM
Search Method: "Exact Match"
Selected Codes: 2720

Criteria Group 2: AND patients have high cholesterol

Use Case 1: Investigator Composes the MDQ Query: Why don't all people with high blood cholesterol and blood pressure get heart disease?

And

Criteria Group: Heart disease w/out heart failure

Group Name*
Heart disease w/out heart failure

Exclusion Criteria

Diagnosis	Code Set:	ICD-9-CM
	Search Method:	"Exact Match"
	Selected Codes:	40200

Criteria Group 3: AND patients without heart failure

DataMart Administrator Receives the Query

DataMart Administrator Inbox – locally installed app at each site

The screenshot shows the DataMart Client application window. The title bar reads "DataMart Client". The interface includes a browser tab for "5.3 Edge" and a "UAT" icon. Below the browser tabs, there are filters for "DataMarts" (set to ".UAT Org A-1 PCORnet DataMart"), "Status" (set to "All"), and "Dates" (set to "Custom 02/01/2017 - 02/12/2017").

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

At the bottom of the window, there are controls for "Page size" (set to 25), checkboxes for "Start with Windows" (unchecked) and "Automatic Refresh" (checked), and buttons for "Refresh", "Details", "Settings", "Close", and "Exit". There are also navigation arrows and a page indicator showing "1 / 1".

DataMart Administrator Reviews Query Details

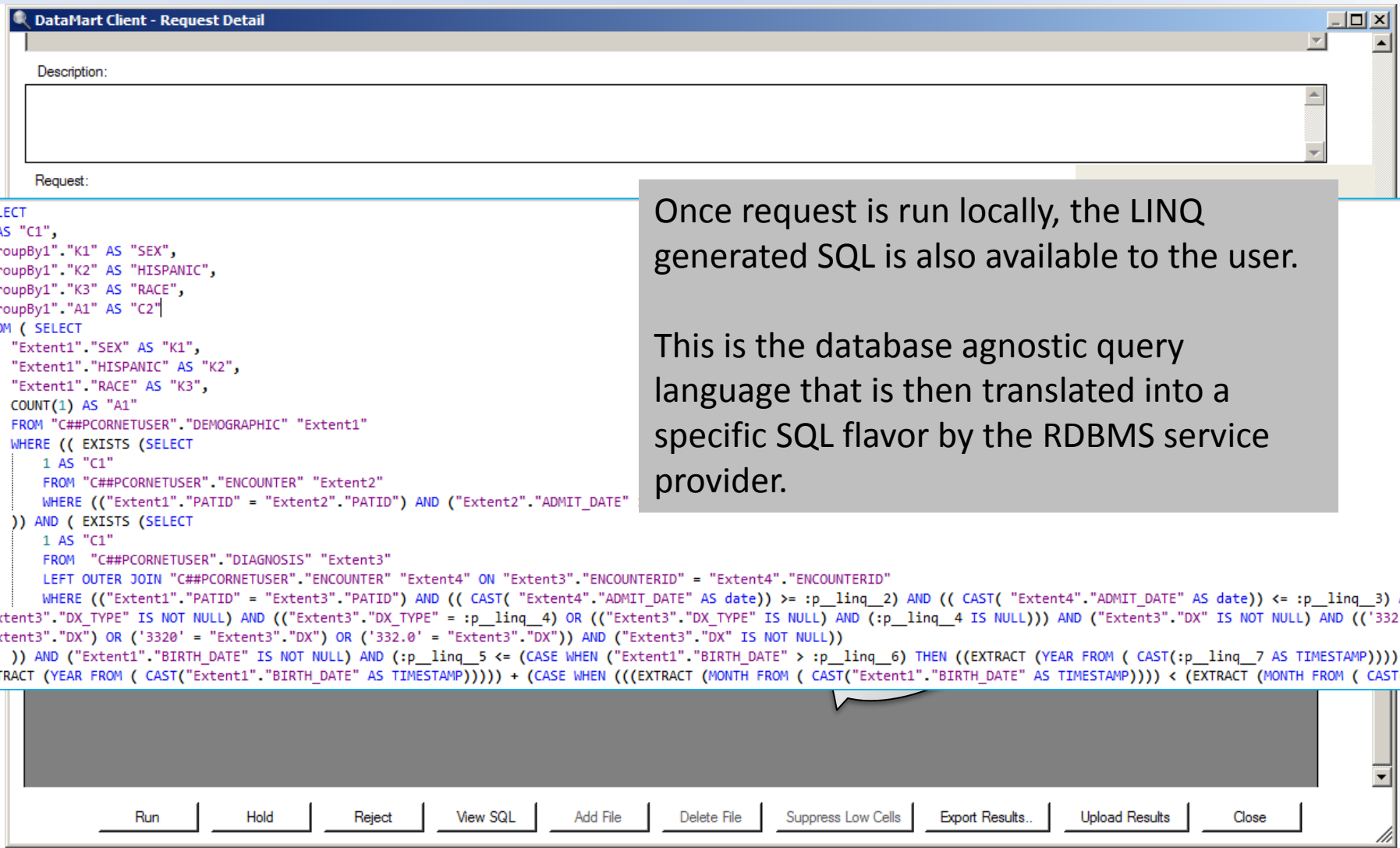
The screenshot shows the 'DataMart Client - Request Detail' window. It features a 'Request:' section with a 'File View' checkbox. A sidebar on the left contains 'Request Details', 'Criteria Groups', and 'Criteria Group: Hypertension'. The main area displays a JSON request. A grey callout box highlights the JSON content with the text: 'Request JSON transmitted from the web portal to the DMC can also be viewed by users'. The bottom of the window has a toolbar with buttons: Run, Hold, Reject, View SQL, Add File, Delete File, Suppress Low Cells, Export Results.., Upload Results, and Close.

```
{
  "Header": {
    "Name": "LPP Query Composer \\/ Default Workflow",
    "ViewUrl": "http:\\\/\\qa52dnsquerytool.lincolnpeak.com\\\/querycomposer\\\/summaryview?ID=21f67097"
  },
  "Where": {
    "Criteria": [
      {
        "ID": "86110001-4bab-4183-b0ea-a4bc0125a6a7",
        "Name": "Hypertension",
        "Criteria": [
          {
            "Operator": 0,
            "Type": "86110001-4bab-4183-b0ea-a4bc0125a6a7",
            "Values": {
              "CodeType": 3,
              "CodeValues": "250",
              "SearchMethodType": 1
            }
          }
        ]
      }
    ]
  }
}
```

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



The screenshot shows the 'DataMart Client - Request Detail' window. It has a 'Description:' field and a 'Request:' field. Below the 'Request:' field, a LINQ query is displayed. The query is a complex SELECT statement involving multiple joins and filters. At the bottom of the window, there is a control bar with buttons for 'Run', 'Hold', 'Reject', 'View SQL', 'Add File', 'Delete File', 'Suppress Low Cells', 'Export Results..', 'Upload Results', and 'Close'.

```
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 "C##PCORNETUSER"."DEMOGRAPHIC" "Extent1"
WHERE (( EXISTS (SELECT
1 AS "C1"
FROM "C##PCORNETUSER"."ENCOUNTER" "Extent2"
WHERE (("Extent1"."PATID" = "Extent2"."PATID") AND ("Extent2"."ADMIT_DATE"
)) AND ( EXISTS (SELECT
1 AS "C1"
FROM "C##PCORNETUSER"."DIAGNOSIS" "Extent3"
LEFT OUTER JOIN "C##PCORNETUSER"."ENCOUNTER" "Extent4" ON "Extent3"."ENCOUNTERID" = "Extent4"."ENCOUNTERID"
WHERE (("Extent1"."PATID" = "Extent3"."PATID") AND (( CAST( "Extent4"."ADMIT_DATE" AS date)) >= :p_linq_2) AND (( CAST( "Extent4"."ADMIT_DATE" AS date)) <= :p_linq_3) AND (
"Extent3"."DX_TYPE" IS NOT NULL) AND ((("Extent3"."DX_TYPE" = :p_linq_4) OR ((("Extent3"."DX_TYPE" IS NULL) AND (:p_linq_4 IS NULL)))) AND ("Extent3"."DX" IS NOT NULL) AND (('332' =
"Extent3"."DX") OR ('3320' = "Extent3"."DX") OR ('332.0' = "Extent3"."DX")) AND ("Extent3"."DX" IS NOT NULL)
)) AND ("Extent1"."BIRTH_DATE" IS NOT NULL) AND (:p_linq_5 <= (CASE WHEN ("Extent1"."BIRTH_DATE" > :p_linq_6) THEN ((EXTRACT (YEAR FROM ( CAST(:p_linq_7 AS TIMESTAMP)))) - (
EXTRACT (YEAR FROM ( CAST("Extent1"."BIRTH_DATE" AS TIMESTAMP)))))) + (CASE WHEN (((EXTRACT (MONTH FROM ( CAST("Extent1"."BIRTH_DATE" AS TIMESTAMP)))) < (EXTRACT (MONTH FROM ( CAST(:
```

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

Description:

Request: File View

Criteria Groups

Request Details

Criteria Groups

Criteria Group: Hypertension

Group Name*

Response: File View

Sex	Race	Patients	AdmittedOn
M	NI	1	2002

Run | Hold | Reject | View SQL | Add File | Delete File | Suppress Low Cells | Export Results.. | **Upload Results** | Close

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

Investigator Reviews Site-Specific Results on Web Portal

Summary

Name:

Project: .UAT Project

Request ID: Request 24386

Priority: Medium

Due Date:

Assignments

User	Role
	Request Creator

Response Documents

Source	File Name
	Request Criteria
.UAT Org A-1 PCORnet DataMart	response.json

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

Current Status

- Investigating issues with SQL code matching request parameters
- **Validation and performance testing** is in progress to evaluate how complex queries behave
- Enhancing automation functionality
- 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:

```

-- 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].
    )
  )
)

```

Current Status

- **Investigating issues with SQL code matching request parameters**
- **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:

Question

The PopMedNet™ (PMN) platform was enhanced to enable users to create and send custom data queries to multiple data sources using different database management systems (e.g. SQL server and Oracle). In what way does this new functionality facilitate a sharable, scalable query infrastructure?

- A. PMN was extended to allow a single query to target multiple data models at once.
- B. The tool was purpose-built to separate the front and back-end components to enable projects that use different data models to more easily leverage existing work (e.g. Race Term) to target additional data models.
- C. The system now allows for users to see which database management system is used at each site.
- D. Menu-driven queries can be used to generate patient lists that can be shared with the investigator who submitted a query.

Answer

- A. PMN was extended to allow a single query to target multiple data models at once.
- B. The tool was purpose-built to separate the front and back-end components to enable projects that use different data models to more easily leverage existing work (e.g. Race Term) to target additional data models.**
- C. The system now allows for users to see which database management system is used at each site.
- D. Menu-driven queries can be used to generate patient lists that can be shared with the investigator who submitted a query.

Explanation: This project focuses on developing a new approach to Menu-Driven-Queries (MDQ) in PMN that is scalable, extensible and enables efficient querying within a diverse health data network. The legacy query composer in PMN built for a single network was not suitable for use in the diverse ecosystem because was not scalable and much of it was hardcoded for a specific RDBMS and could not easily be repurposed. With advancements in technology, the PMN platform was substantially enhanced to introduce the Microsoft Entity Framework and custom workflow engines to produce the new MDQ tools that enable querying across RDBMS.



Jessica_malenfant@harvardpilgrim.org



[linkedin.com/in/jessicamalenfant](https://www.linkedin.com/in/jessicamalenfant)



popmednet.org
populationmedicine.org