

# Distributed Health Data Networks: Extending the PopMedNet™ Query Infrastructure to New Data Sources

Kimberly Barrett, MPH; Jessica M. Malenfant, MPH; Kyle Erickson; Adam Paczuski; Zachary Wyner, MPH; Chayim Herzig-Marx, PhD; Jeffrey S. Brown, PhD  
Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA

## OBJECTIVE

Demonstrate a new architecture and framework for an extensible point-and-click query interface in PopMedNet™ (PMN). These tools:

- Address 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: PCORnet and FDA's Sentinel Initiative

## METHODS

Tools developed include:

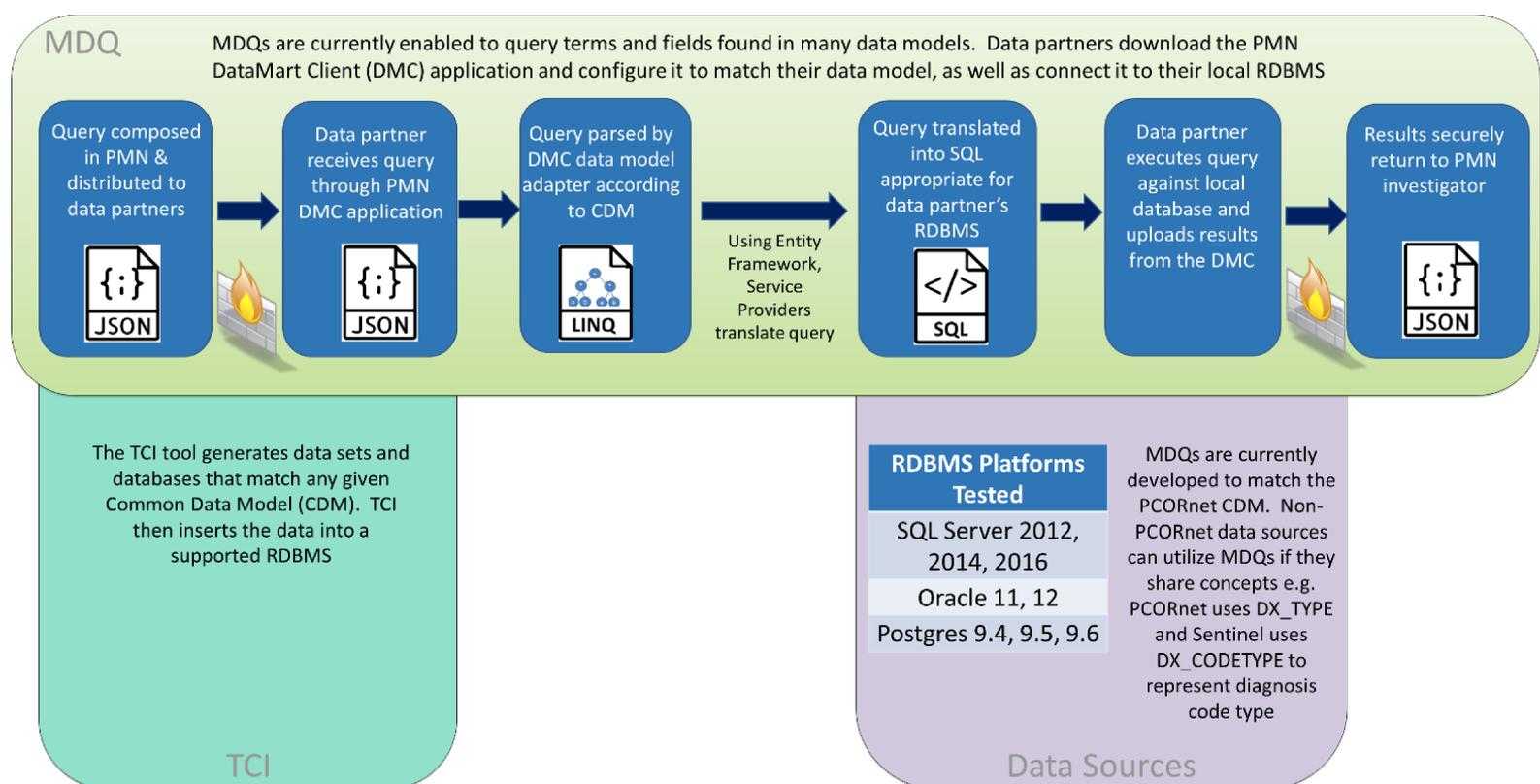
### Menu-Driven Queries (MDQs):

- 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

## IMPLEMENTATION & RESULTS



## CONCLUSION

TCI expedites the PopMedNet™ development and validation process as it accurately creates synthetic databases that adhere to any common data model, such as PCORnet, Sentinel, and ad hoc databases. This addresses the usual time-intensive process of generating test data in multiple RDBMS. Additionally, this allows us to validate that the new PMN MDQs are accurate and consistent across the different database flavors.

The tools described lay the foundation of analytic capabilities in PMN. Each tool is adaptable and can be used by both current PMN data sources and new non-PMN data sources. Data sources that do not yet have a model adapter built, can still use all of these tools if they share terms or make minor modifications to the data model

### Ongoing and future work:

- Expand the queryable terms and stratification options for existing networks
- Explore how to continue to extend the MDQ functionality for non-PMN data sources

