

PopMedNet: Introduction

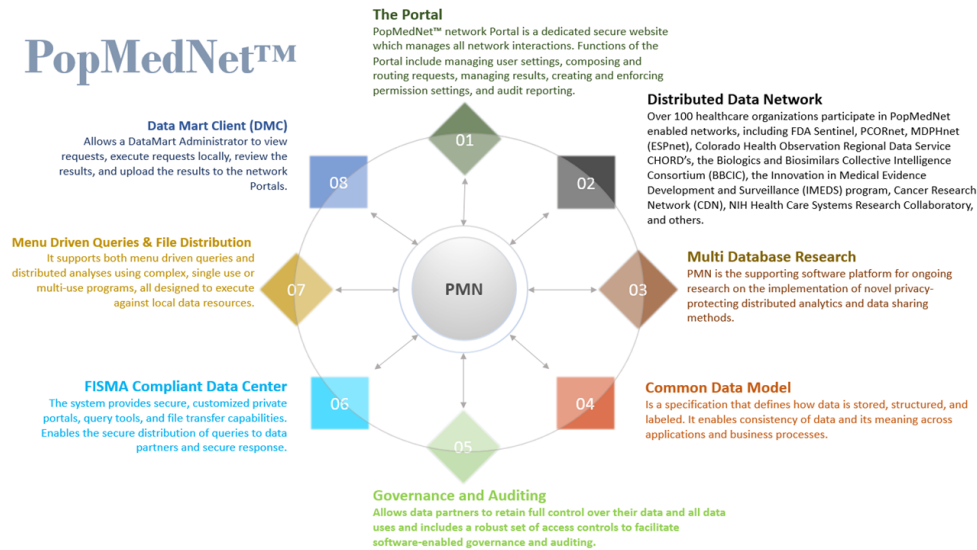


Figure 1: Overview of PopMedNet™

Overview

PopMedNet™'s development was pioneered by the Therapeutics Research and Infectious Disease Epidemiology (TIDE) group of the Department of Population Medicine (DPM) of the Harvard Pilgrim Health Care Institute (HPHCI), an appointing department of the Harvard Medical School. PopMedNet currently supports several large DDNs, including the Sentinel System, PCORnet, and the NIH Collaboratory.

PopMedNet™ is an open-source application used to facilitate multi-site health data networks. It uses a distributed network design that enables data partners to retain full control of their data. Investigators send questions to data partners¹ for review and response. PopMedNet eliminates the need for assembling patient records in a centralized repository, thus preserving patient privacy and confidentiality.

PopMedNet is made up of three primary components:

- **PopMedNet Query Tool:** The web-based application from which queries are distributed. It is typically used by the analysis center to create, distribute, and manage queries.
- **DataMart Client:** Is a locally installed application that acts as inbox for data partners to receive packages and transfer results to the analysis center.
- **Web Service API:** API used to implement all application logic that processes requests against a network. The API is also used for integrations with PMN-allied services and applications.

¹ Data partners are data owners who partners with each other in a study.

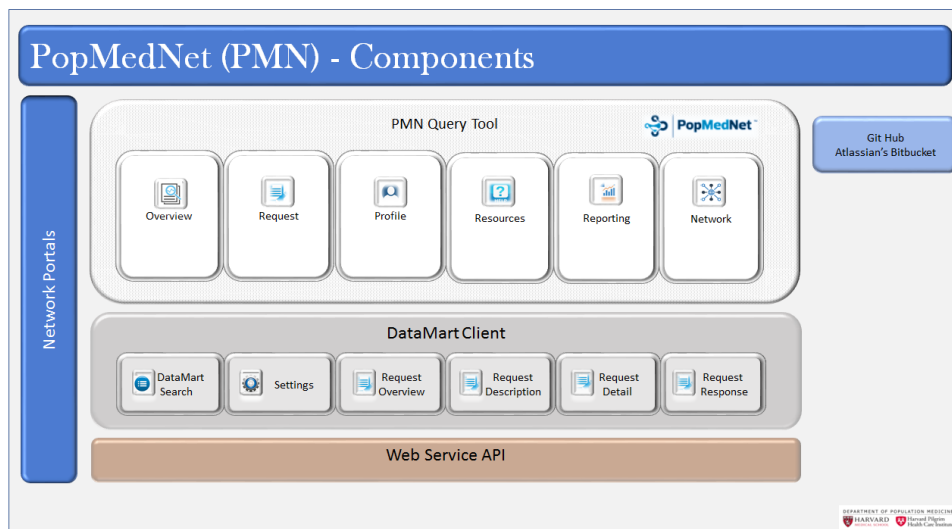


Figure 2: Illustration of PopMedNet™ Components

The following diagram shows the basic operating model of PopMedNet in a single request/response roundtrip. However, it can also operate iteratively until some conditions are met, such as that employed by the Distributed Regression data model (below). Additionally, it can be configured to run unattended.

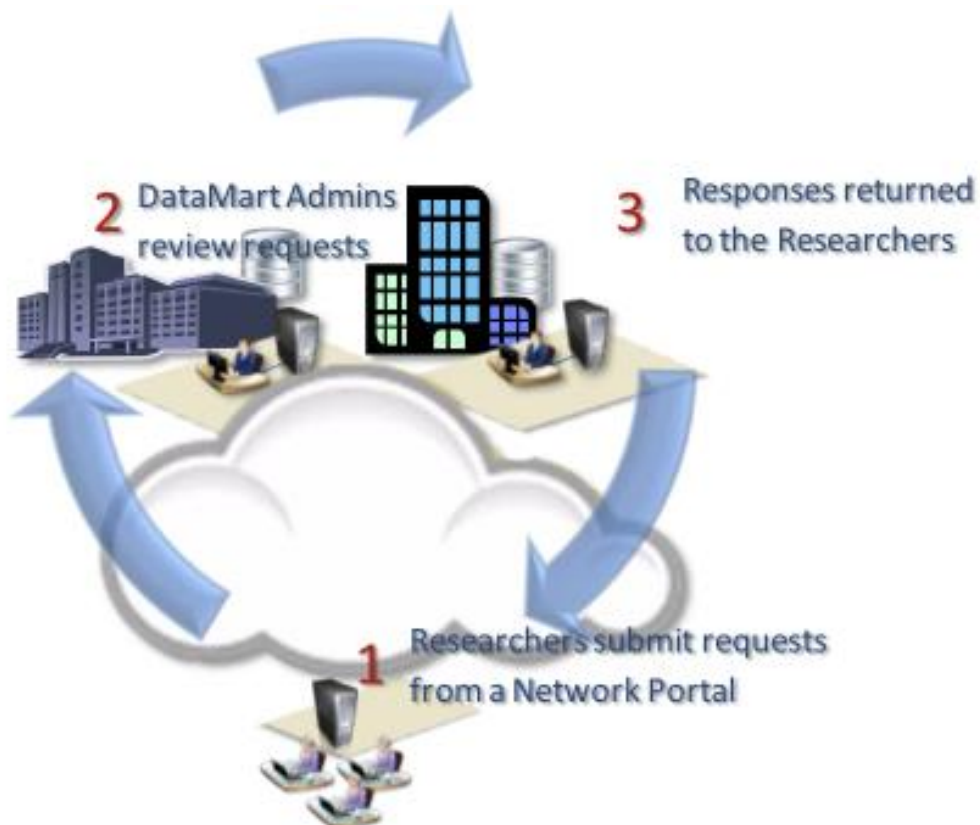


Figure 3: Basic Operating Model of PopMedNet™

The goal of PopMedNet is to facilitate distributed analysis of electronic health data in order to support medical product safety surveillance, disease surveillance, comparative effectiveness, quality, medical resource use, cost-effectiveness, and related studies. It does this by enabling the creation of health data networks that allows the data owners to keep and control uses of their data, while at the same time allows customized access and governance for each network.

The concerns of the network stakeholders are addressed in the design and governance models of the PopMedNet software platform.

Key Features

PopMedNet features:

- Multiple queries interactively composed and distributed in a single request to multiple data marts²(Multiple Menu Driven Queries – MMDQ).
- Query responses are asynchronously returned – that is, a query does not have to be responded to immediately and not all data marts have to respond.
- Customizable workflows.
- Auditing and search capabilities.
- Flexible role-based access control system enables the enforcement of varying governance policies.
- Approval processes for submitting and responding to queries
- Easy integration with project management tools such as Jira through API interfaces
- Allows for easy tracking of metadata associated to each request

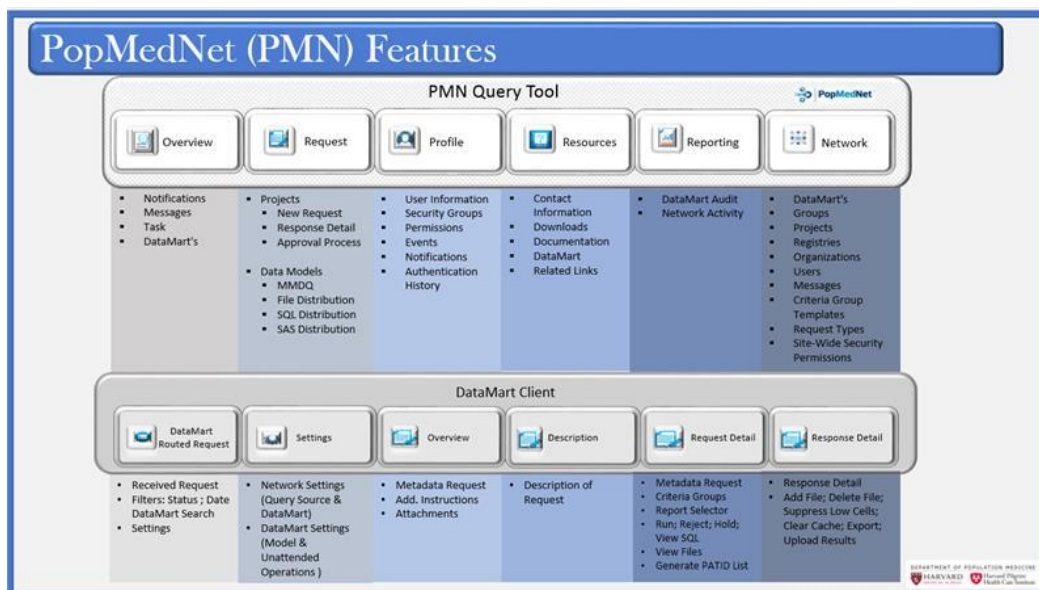


Figure 4: Diagram of PopMedNet™ Components-Feature

² Data marts are software components representing the data owners' data sources.

Architecture

PopMedNet has two components. It has a .NET web application component (Portal) and a Windows application or web-based client component (DataMart Client). The .NET web component uses a modern service-oriented architecture (SOA) with a robust API serving a dynamic UI using the MVVM pattern. The DataMart Client (DMC), uses an adapter architecture to support different data models. A variety of adapters are available, and able to support distributed regression, file transfer, ad-hoc SQL queries, and dynamically composed menu-drive queries (MMDQ) against local data sources (supporting SQL Server, Postgres and Oracle).

PopMedNet is a .NET Framework application, utilizing [ASP.NET MVC5](#) for the implementation of the website application and WinForms for the desktop application. Entity Framework 6 is used for data access, and Typescript for client-side scripting.

The DataMart Client (DMC) is a locally installed Microsoft Windows® application that acts as an inbox for data partners to receive query packages and transfer results to the analysis center. All file transfers (query requests and responses) between the data partners and the analysis center are achieved through secure HTTPS/SSL/TLS connections. There are no virtual machines, open ports, Virtual Private Networks, or any external access to data partner data, abating concerns about data security and unauthorized access. The system ensures only approved queries are submitted to and responses returned by participating data partners with several levels of software-enabled governance.

Portal (Web)		DataMart Client Application		DataMart Client Server (Web)	
View		WinForm		View	
REST API		Adapters		REST API	
.NET 4.8	EF6	.NET 4.8	EF6	.NET 4.8	EF6

Governance - Permission

Individual permissions are set on different levels within a PopMedNet instance. The permission will always apply to whatever level the permission is set at. For example, if View Request is set on a specific project, it would let that user view all requests sent in that project, but not another project. If the same permission was set on an Organization level, it would let the user view all requests sent to DataMarts that belong to that organization, regardless of which project the requests are sent in. The main levels where we set permissions in practice are:

- Site-wide: affecting everything on the PopMedNet instance
- Project: affecting everything in that project
- Project-DataMart: affecting everything for that DataMart within that project
- Project-Request Types: affecting everything for that request type within that project
- Organization: affecting everything at that organization (including DataMarts owned by that organization)

Name* Tutorial Project **Acronym*** TUTORIAL **Group*** Group Template

Start Date* 7/12/2021 **End Date** Allow Submission of Requests

Description

Project Permissions | DataMarts | Organizations | Security Groups | Events | Request Types | Request Type Permissions | Request Metadata

These permissions control access to the project and set default permissions for requests, DataMarts and other options.

Select Existing Security Group: Tutorial Project\DataMart Administrator

Permission	Inherit	Allow	Deny
Approve, Reject, and Resubmit Responses Allows the user to approve, reject, or resubmit responses that are awaiting approval	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Approve/Reject Submission Controls who can approve or reject requests that have been submitted by a user that cannot skip request approval	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Assign Request-Level Notifications	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 5: Displays the Project Permissions for a Tutorial project

S Name Organization X

TUTORIAL Modular Program DataMart Root

Overall | By Request Type | Events

These permissions control how the DataMart will respond to requests within the context of the project. These permissions override more general permissions on the DataMart and the Organization.

Select Existing Security Group: Tutorial Project\DataMart Administrator

Permission	Inherit	Allow	Deny
Approve, Reject, and Resubmit Responses Allows the user to approve, reject, or resubmit responses that are awaiting approval	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Group/Ungroup Responses Allows the user to group/ungroup responses	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hold Requests Allows the user to place a hold on submitted requests on the DataMart Client	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Modify Attachments Allows the user to add or delete Attachments.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modify Results Allows the user to upload results for File Distribution and Modular Program requests using the DataMart Client after the routing has been completed.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Override DataMart Routing Status	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 6: Displays the permissions for granting access to a specific DataMart within a Project

Other locations that may be used (this is not an exhaustive list):

- Project-Organization: affecting everything for that organization and every DataMart owned by it in that project
- Request Types: affecting that specific request type
- Registries: affecting that specific registry
- User: affecting that specific user

PopMedNet Distributed Research Network Technologies for Population Medicine Welcome, Adam Paczusi

Home Requests Profile Resources Reports Network Contact Us Log Off

Home / Users / TUTORIAL DATAMART ADMIN Active Deactivate

First Name* Middle Name Last Name* Title

TUTORIAL DATAMART ADMIN

Email Address* Phone # Fax #

popmednet_support@harvardpilgrim.org

Organization User Name*

Operations Center tutorial_datamart_admin Change Password

Security Groups Permissions Events Notifications Authentication History

These security groups control the membership for this user. By selecting a security group, you are adding this user to the security group and the user will automatically get the permissions assigned to that security group.

Name	Parent	Type	X
Tutorial Project\DataMart Administrators	(None)	DataMart Administrators	Remove

Add Security Group

Figure 7: Displays how the collections of permissions in groups are assigned to a user

Permissions set on higher, more all-encompassing levels are inherited at lower levels. This means that if an Approve/Reject Results permission is set on a project-level, that permission is effectively inherited at all Project-DataMart and Project-Organization levels within that project. You can deny permissions at a lower level, and that will overrule the inherited permission.

How It Works

In PopMedNet, the Portal handles all the governance and acts as a hub for all queries. That ensures that the coordinating center³ can expect:

- Consistency of information across data partners
- Quality of data
- Flexibility to handle different network structures with overlapping data partners
- Ability to track and report operational activities
- Managing multiple projects

The DataMart Client (DMC) allows data partners to control access to their data before sending it to the Portal. For the data partners, it will:

- Provide protection of PHI
- Prevent loss of control over proprietary information

The researcher will have no direct access to data but is shielded from the differences among networks' data models.

³ Coordinating center runs the study and issues queries from the PopMedNet Portal.

A typical PopMedNet™ request-response cycle is as follows:

1. Request: Investigators distribute data requests constructed in the point and click web interface
2. Review: Data Partners receive and review requests in their DataMart Client
3. Respond: Data partners execute the requests against their local data and upload results
4. Download: Investigators access results in the query tool

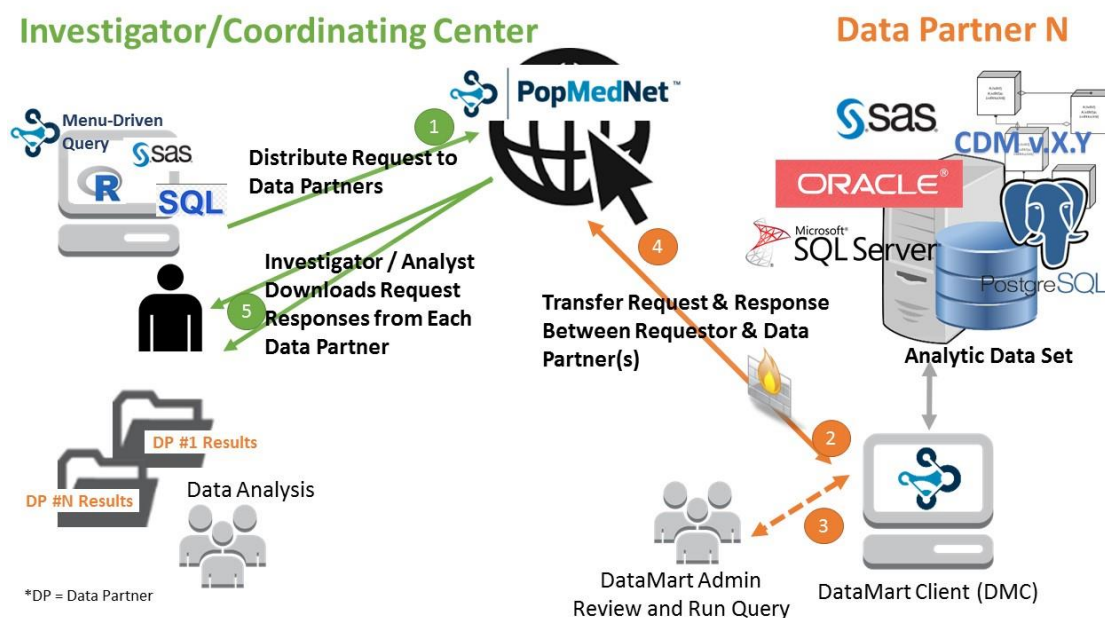


Figure 8: Diagram of PopMedNet™ Request Cycle

Key Data Models Supported

Modular Program Distribution

The Modular Program Distribution request model allows you to distribute modular program packages to the network DataMarts. The Modular Program request model provides two request types: Modular Program. The request specific criteria for these request types is the same and displayed in the figure below.

If data partners upload Modular Program results containing all-run signature files, the signature file information may be automatically read by PopMedNet and displayed on the Request Details page.

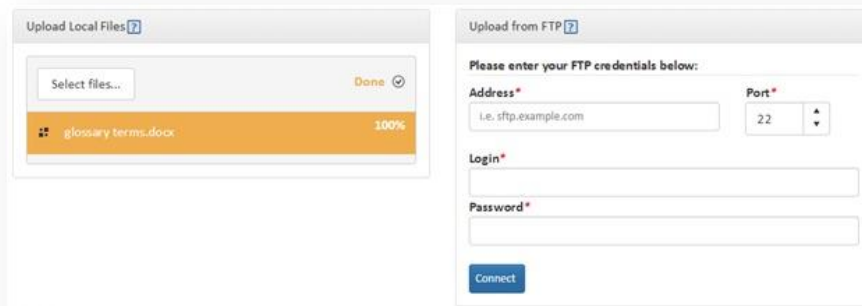


Figure 9: Illustrates how a file is attached within a Modular program request

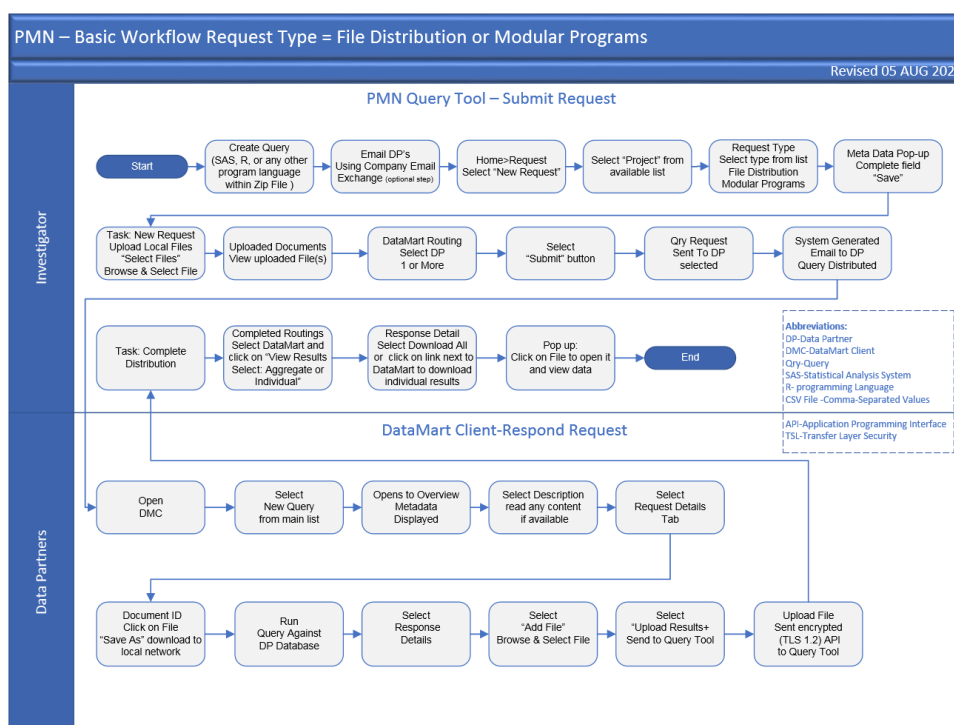


Figure 10: Modular Program workflow utilized for FDA Request

Multiple Menu Driven Query

Menu-Driven Queries (MDQs) utilize a simple point-and-click interface to add terms to the request criteria that are joined through logical operations (and/or), associations, and then grouped according to the chosen stratification(s). A major benefit is that users do not need programming expertise to create and execute MDQs. MDQs execute against data stored within a Common Data Model and output results based on the criteria entered. Additionally, files can be distributed with a query.

MDQ further supports the ability to send multiple MDQ in a single request. This would include the ability to specify distinct criteria groups and stratifications for each MDQ within the request. Data partners can run and upload the results for all MDQs simultaneously using the DataMart Client.

When an MMDQ is received by a Data Partner, the software gathers all user submitted queries, runs them simultaneously, and results are sorted into appropriate tables based on the request criteria. When this process is complete, results may be returned to the requestor. Data partners may also configure their DMC to automatically run and upload the results of MDQs and MMDQs. This auto-run and upload feature functions by queuing the queries as received by the DMC such that the first query will run and upload results prior to moving onto the next query in the queue.

MDQ adapter uses Entity Framework for conversion into SQL for SQL Server, Oracle and Postgres.

Key features of MDQ architecture include:

- Visual query composition - the MDQ architecture allows users to develop a query using a point-and-click interface using logical operations (and/or), associations (criteria group) and stratifications (group by);
- Flexibility - MDQs can be executed against a variety of database types;
- Grouped querying - the MDQ architecture allows creation and submission of multiple MDQs in a single request;
- Ability to define and share custom query templates;
- Sharing of query specifications, development of sharable phenotypes;
- Electronic workflow that guides users through query composition, governance, distribution, and response viewing;
- Ability to aggregate and export results to multiple file formats; and
- Use of PMN governance and access controls.

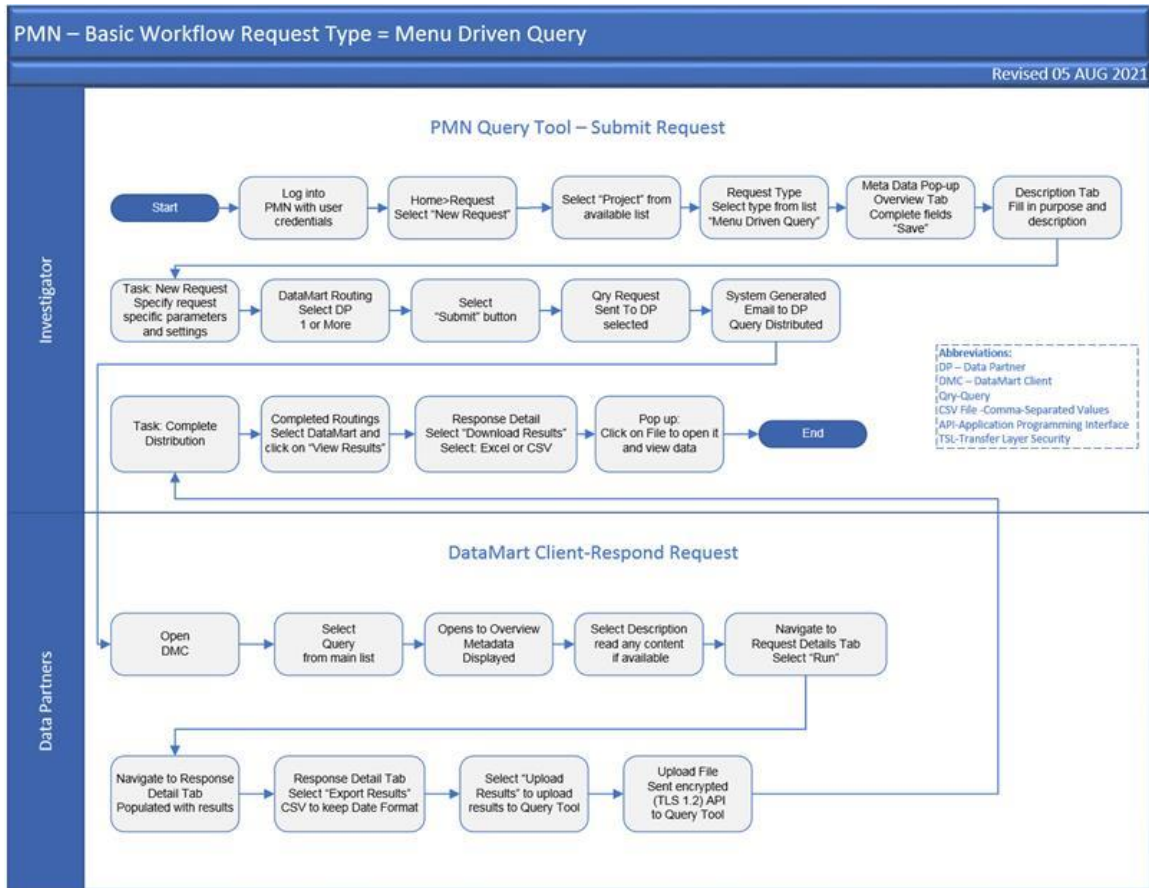


Figure 11: Menu Driven Query Workflow

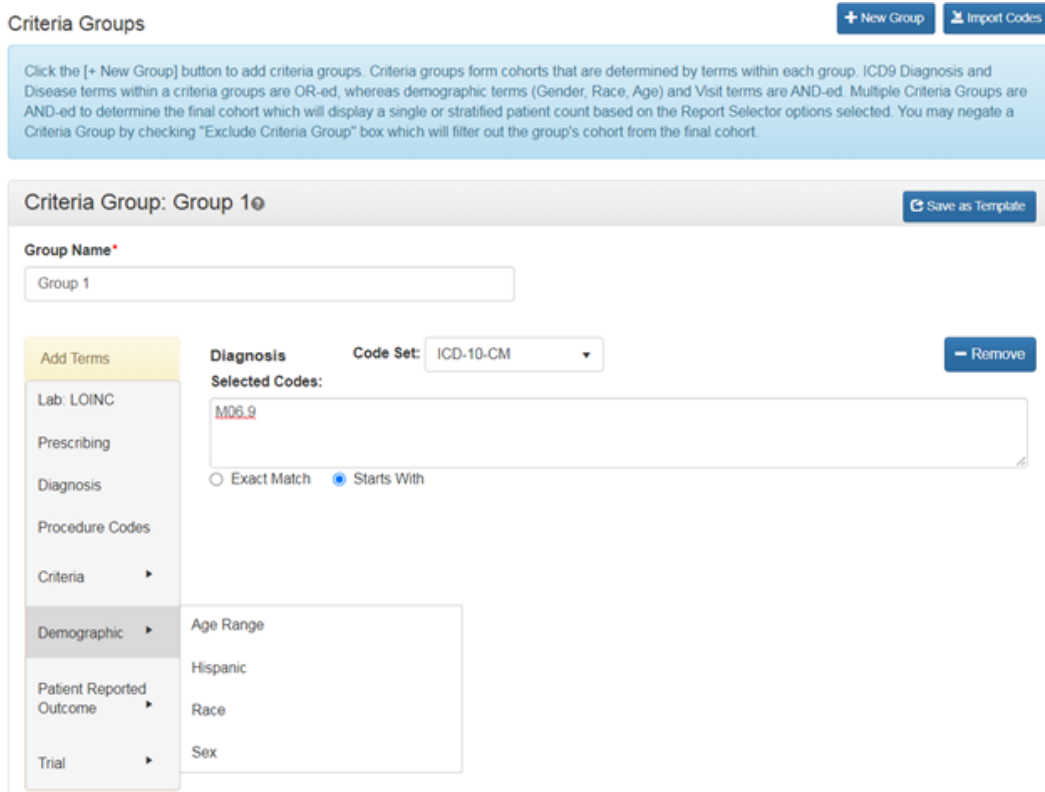


Figure 12: Menu-Driven Query User Interface.

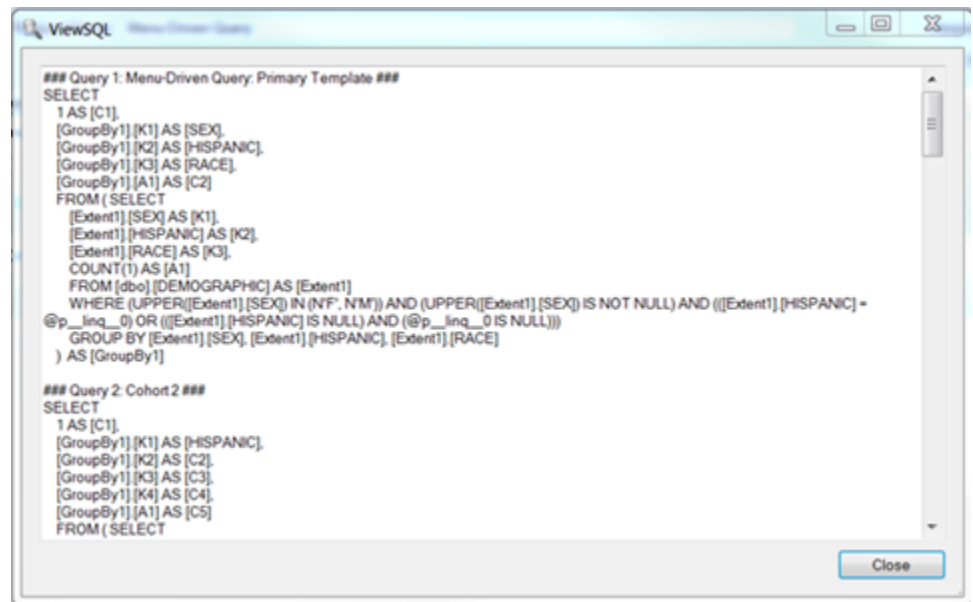


Figure 13: Structure of a Multiple Menu Driven Query (MMDQ) as seen in the DataMart Client

Distributed Regression Analysis (DRA)

Distributed Regression provides data infrastructure and methods that support robust statistical analysis for binary, survival, and continuous outcomes. It enables the ability to conduct rapid and secure distributed regression analysis in distributed data networks. DRA is mathematically equivalent to the conventional regression analysis of pooled individual-level data. It achieves the same level of statistical sophistication using only summary-level information, thereby offering better protection for individual and institutional privacy without jeopardizing the scientific rigor of the analysis.

Distributed Regression adaptor (DRA-adaptor) has the following unique characteristics:

- Supports multiple data sources with unique populations, which are often needed to answer clinical questions
- Balances analytic data requirements, patient privacy and confidentiality, local governance, and proprietary considerations

It can:

- Automatically download new files (individual and zipped) to specified location when all DataMart's responded.
- Automatically launch SAS or R interpreter and run program delivered as request payload.
- Different levels of workflow automation (completely manual, semi-automated, and fully automated).

Three regression models are supported:

1. Linear Regression
2. Logistic Regression
3. Survival Analysis (Cox Proportional Hazard Model)

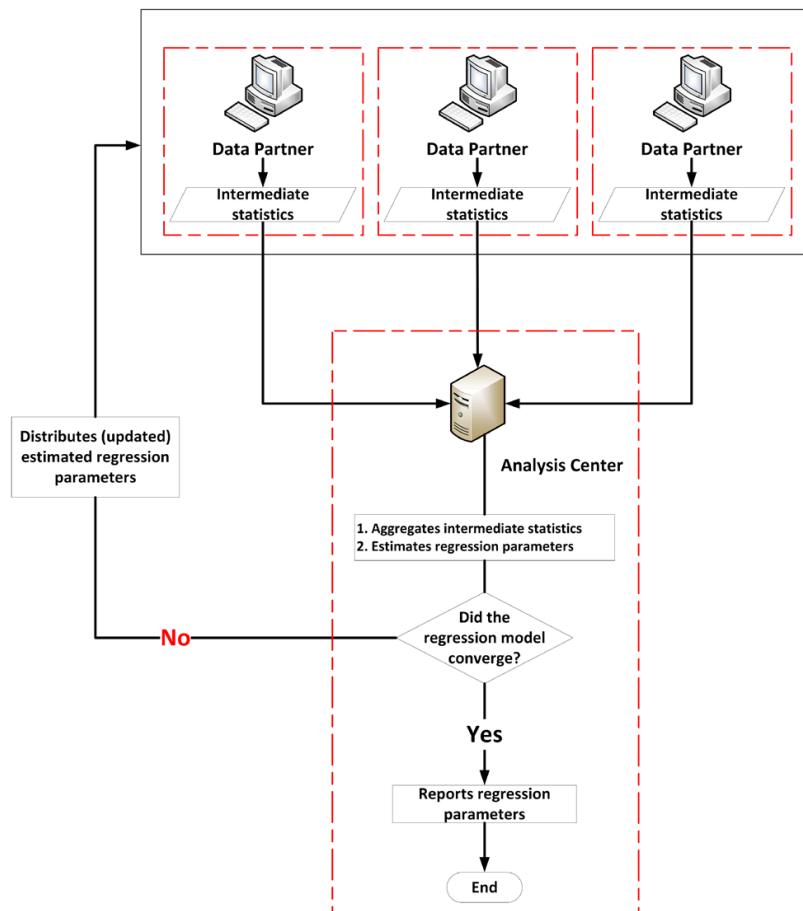


Figure 14: Iterative process to perform distributed regression analysis

A typical DRA includes 3 major steps. Step 1 involves the assembly of a harmonized individual-level analytical dataset at each data partner. In step 2, the analysis center and each data partner execute a DRA algorithm locally. Step 3 involves the iterative transfer of the DRA algorithm outputs between the data partners and the analysis center until the regression model converges or the process reaches a prespecified maximum number of iterations. PopMedNet workflow is utilized to transfer input files and computation directions interactively and automatically between data partners and the analysis center.

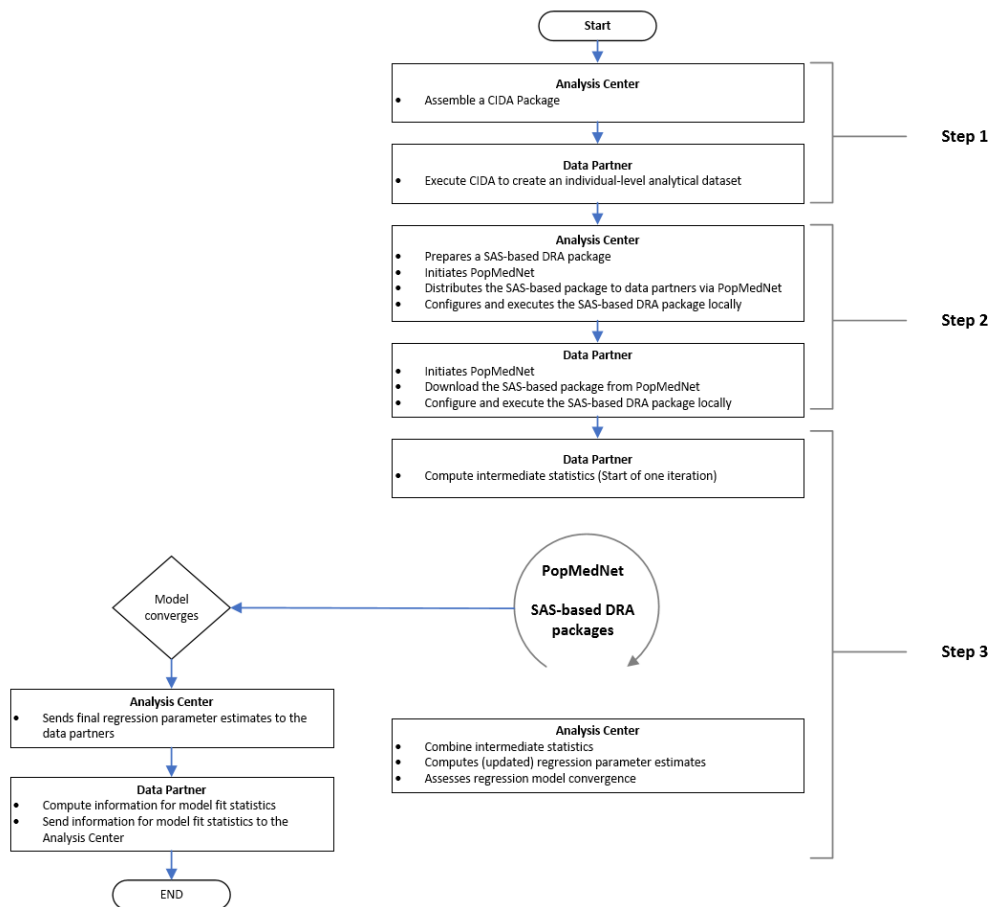


Figure 15: Three-Step process to conduct distributor regression analysis with PopMedNet. CIDA: Cohort Identification and Descriptive Analysis Tool; DRA Distributor Regression Analysis

Executing External Programs

The DataMart Client (DMC) supports data models that interoperates with external programs. For example, DataMart Client can be customized to execute an external program like a command interpreter such as R or SAS interpreters or executable programs. Alternatively, it can be customized to deposit files that are monitored by an external program. DR utilizes the latter capability of DMC to perform its iterations automatically without human intervention.

Network Settings

Query Source

Model Settings

Monitor Folder: enter a root directory of your choice

Trigger filename: Successful Initialization: job_started.ok

Trigger filename: Execution Complete: files_done.ok

Trigger filename: Execution Fail: job_fail.ok

Trigger filename: Execution Stop: job_done.ok

Output Manifest Filename: file_list.csv

Maximum Monitor Time (hours): 12

Maximum Read Attempt Time (minutes): 5

OK Cancel

Apply OK Cancel

Figure 16: DataMart Client (DMC) Model Settings

Deployment Infrastructure

The PopMedNet team recommend deploying PopMedNet is a load-balanced, clustered configuration.

PopMedNet must be installed on Windows Server running IIS and backed by SQL Server database that is not public facing.

Both the portal and the API are public facing and should require HTTPS/TLS 1.2 communication protocols.

Acknowledgements

PopMedNet™ was developed and has been extended as part of several contracts awarded by a range of federal, state, and industry stakeholders:

- FDA's Sentinel Initiative
- PCORI's National Patient-Centered Clinical Research Network (PCORnet)

In addition, Harvard Pilgrim Health Care Institute serves as the coordinating center for several other initiatives that use PMN and its associated software ecosystem including:

- Biologics and Biosimilars Collective Intelligence Consortium (BBCIC)
- Office of the Assistant Secretary for Planning and Evaluation (ASPE) & Food and Drug Administration
- Privacy-protecting distributed analysis of biomedical big data project funded by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health

References

1. Her Q, Malenfant J, Zhang Z, Vilk Y, Young J, Tabano D, Hamilton J, Johnson R, Raebel M, Boudreau D, Toh S. (2020), [A distributed regression analysis application in large distributed data networks: Analysis of precision and operational performance](#). JMIR Med Inform 2020;8(6):e15073
2. Her, QL, et al. (2018), [A Query Workflow Design to Perform Automatable Distributed Regression Analysis in Large Distributed Data Networks](#). eGEMs (Generating Evidence & Methods to improve patient outcomes), 6(1): 11, pp. 1–11. DOI: <https://doi.org/10.5334/egems.209>
3. [PopMedNet User's Guide](#)
4. [PopMedNet Standard Functionality](#)
5. [PopMedNet Modular Programming Distribution](#)
6. [PopMedNet Menu Driven Query](#)
7. Davies, M. and Erickson, K. and Wyner, Z. and Malenfant, J. (2016), M. and Rosen, R. and Brown, J., [Software-Enabled Distributed Network Governance: The PopMedNet Experience](#), The Journal of Electronic Health Data and Methods, 2016
8. Brown, J. and Holmes and J. H. and Shah, K. and Hall, K. and Lazarus, R. and Platt, R. (2010), [Distributed Health Data Networks: A Practical and Preferred Approach to Multi-Institutional Evaluations of Comparative Effectiveness, Safety, and Quality of Care](#), Medical Care, 2010/June, V48
9. Malenfant, J. and Davies, M. and Swan, B. and Brown, J. S. (2015), [PopMedNet: Collecting and Using Metadata in Distributed Research Networks](#), Journal of Patient-Centered Research and Reviews, 2015/April,