

New York State Department of Health
All Payer Database Request for Information
RFI #1205180104

Questions and Responses: Round Two

Category: General

1. Question

The RFI describes structured data queries and some proprietary products for implementing the State's vision. Would the State be interested in incorporating unstructured data mining and open source capabilities to enable cost effective extensions of the architecture?

Response

These options may be proposed. Please cost out each option separately.

2. Question

Cost Projections: The first 5 rows of the matrix are costs for tasks, tools, and services described in Sections B-F, and the last two rows are costs for Operations and Maintenance. However, Sections B-F (pgs 17-37) describe activities involved with maintenance and operations as well as implementation. Would it be correct to assume that the first 5 rows are for implementation costs only, and the last 2 rows are where all the costs for all Operations and Maintenance are to be shown?

Response

Yes, that is a correct assumption. When estimating operation and maintenance costs, please include all costs to keep system operational. For purposes of estimating maintenance, assume two releases per year impacting an average of 20% of the system functionality each release.

Category: Data Sources and Volumes

3. Question and Response

A number of questions were received regarding the amount of data that would be collected from each data source. *For purposes of responding to this RFI, please use the following estimates and approximations for the data sourced listed below:*

Commercial Health Plans:

Number of members in commercial health plans in NY: Approximately 9.3 million

Volume of claims for commercial business: Approximately 180 million claims

Medicaid:

Number of Medicaid recipients in NY: Approximately 5 million in July 2012

Volume of Medicaid claims associated with recipients: Approximately 750 million transactions in 2011; 500 million netted claims/encounters

SPARCS:

Volume: SPARCS records are added at a rate of approximately 23 million discharge records per year.

There are 234,772,6414 records estimated to be loaded into the APD; as flat files these records are estimated at 882,918,271,805 bytes.

Category: Data

4. Question

The RFI references Medicaid encounter data specifically, but we assume the State also intends for the APD to store records from capitated encounters that are privately insured. Is that correct?

Response

Yes, that is correct.

5. Question

The RFI mentions that the Department will be modifying the reporting requirements for facilities that report SPARCS data. Have those changes already been announced? When does the Department intend to implement those changes, before or during the implementation of the APD?

Response

No major changes to the SPARCS reporting specifications are anticipated for the next two years.

6. Question

On page 4, the RFI states: "The APD will build on and enhance existing DOH databases including SPARCS." Page 33 states: "Describe how you would transition SPARCS facilities to the APD." It is somewhat unclear as to what will happen to the existing SPARCS database. Is it to continue, with the State collecting the data as it does today, or shall the source data collection duty be transferred to the APD contractor and the SPARCS database as we know it today be decommissioned?

Response

The current SPARCS mainframe system will run in parallel until such time as the historic data in the MVS flat files has been fully migrated to the APD and the APD provides all the required SPARCS functionality. Please provide your best recommendations on whether the state or the APD contractor should have responsibility for the intake of SPARCS data.

7. Question

Under the headings "DOH will" and "DFS will", there is a reference to "eliminating redundancy". Are there other redundant systems besides SPARCS that the State envisions replacing with the APD?

Response

Details are to be developed during the analysis phases of the project.

8. Question

How long will data be kept live?

Response

SPARCS data is kept a rolling 35 years. It is anticipated that third party payer data will be kept live for 10 years. Details are to be developed.

9. Question

The use cases include health care transformation, comparative effectiveness, quality improvement, improving public health, assessing health care resource needs, and insurance regulation. – Can the State be more specific on the type and amount of reports/analyses that the selected vendor will be required to create/maintain?

Response

Details are to be developed during the analysis phases of the project.

10. Question

For each data source, what is the receipt frequency for each source? Daily, Weekly, Monthly, Real-time?

Response

Details are to be developed during the analysis phases of the project. The Department welcomes specific suggestions and recommendations from RFI respondents.

11. Question

“Episode grouping technology that organizes claims and discharge data into a set of clinically coherent episodes. “ Does DOH currently have software to perform these groupings? Please list specific grouping software for this criterion, and any additional licensing required above and beyond current DOH licensing.

Response

Details are to be developed during the analysis and design phases of the project.

12. Question

Is the claim data cleansed or will quality checks need to be applied?

Response

It is anticipated that claims data will need to be cleansed and quality checks applied. Please factor in this assumption in your responses.

13. Question

Will de-duplication requirements be within single claims contribution file, a single source, or multi-sourced claims data? For example, will a claim need to be checked to see if it is included in data from a separate source?

Response

Details are to be developed in the analysis phases of the project.

14. Question

For each data source, how many years of historical data needs to be converted?

Response

All historic SPARCS data will be converted. For purposes of estimation, consider at least 5 years of government and third-party payer data to be loaded. How many years of other types of historic data to be loaded into the APD will be determined as part of the analysis phases of the project. Please indicate your assumptions on sizes and years of data in your response.

15. Question

What is the expected turnaround time of data? Are feeds coming from Stakeholders on a Monthly, weekly basis? Some of the use cases might require feeds to be provided and data to be transformed in a real-time (Near-time) basis...especially if this data will serve as a means to monitor epidemic outbreaks. Do you see the need for real-time (Near-time) basis?

Response

Details are to be developed in the analysis phases of the project.

Category: Privacy and Security

16. Question

Can you please define the level of access required for NY state administrators to the APD database?

Response

Details are to be developed during the analysis phases of the project.

17. Question

Please confirm the audit trails must be HIPAA and HITECH Compliant, and that access history would be available to recipients / members whose PHI has not been de-identified, and whose PHI has been accessed. Please confirm the logs must contain the date and time of access, who accessed and the reason for access.

Response

Details are to be developed during the analysis phases of the project.

18. Question

“The APD must also adequately secure confidential data and must provide DOH the authority and capabilities needed to meet its responsibilities for data governance.” Aside from HIPAA compliant protection, what requirements does DOH security have for access to this data? Please list any specific hardware or software solutions that are required by DOH Security for the implementation of a compliant connection that are above and beyond HIPAA security measures. If additional hardware or licensing is required by DOH, please list the pricing for acquiring these solutions.

Response

Page 19 outlines additional details for security. The Department welcomes specific suggestions and recommendations from RFI respondents on this topic.

Category: Technology**19. Question**

Can DOH clarify the acronyms referenced in the diagram on page 54? (AWS RDS)

Response

AWS RDS stands for Amazon Web Service Relational Database Service.

20. Question

This potentially represents terabytes to petabytes of data. Is it the Department's intention to have all of the data continuously available online, and for this data to be accessible on a 24/7/365 basis?

Response

Please provide your best recommendations on data availability and accessibility.

21. Question

Please elaborate on when the vendor would be expected to perform edits on incorrect coding, will it be at the time it receives claims?

Response

Edits should be applied prior to accepting data into the APD warehouse.

22. Question

Please confirm the Rhapsody tool referenced is the Orion Rhapsody Integration Engine. If not, please provide the full name of the tool.

Response

The Rhapsody tool referenced is the Orion Rhapsody Integration Engine.

23. Question

Is there a Business Continuity/Disaster Recovery requirement to the APD system? If so, what are the availability requirements?

Response

Yes, there are Business Continuity/Disaster Recovery requirements. The Department welcomes specific suggestions and recommendations from RFI respondents for the details of this requirement.

24. Question

What specifically does DOH mean by the term "Oracle Appliance" in the diagram on page 58?

Response

The Oracle Database Appliance is a high availability clustered database system that includes software, servers, storage, and networking.

25. Question

"The APD system will integrate and aggregate that data in an optimum manner and create data marts with open application programming interfaces (API's) to support output data requirements." How many Datamarts are anticipated? What types of Datamarts are anticipated? Is there a current list of anticipated Datamarts and/or Data Cubes? What is meant by API in this statement? Does this refer to access to data, or does this imply the design of business objects for an existing system? Is this expected to be a database function or front end web-portal function?

Response

Details are to be developed during the analysis and design phases of the project. The Department welcomes specific suggestions and recommendations from RFI respondents.